Hot topics in science communication: Aggressive language decreases trustworthiness and credibility in scientific debates

Lars König and Regina Jucks
University of Münster, Germany

Abstract
Current scientific debates, such as on climate change, often involve emotional, hostile, and aggressive rhetorical styles. Those who read or listen to these kinds of scientific arguments have to decide whom they can trust and which information is credible. This study investigates how the language style (neutral vs aggressive) and the professional affiliation (scientist vs lobbyist) of a person arguing in a scientific debate influence his trustworthiness and the credibility of his information. In a 2 X 2 between-subject online experiment, participants watched a scientific debate. The results show that if the person was introduced as a lobbyist, he was perceived as less trustworthy. However, the person’s professional affiliation did not affect the credibility of his information. If the person used an aggressive language style, he was perceived as less trustworthy. Furthermore, his information was perceived as less credible, and participants had the impression that they learned less from the scientific debate.

Keywords
aggressive language, credibility, professional affiliation, science communication, scientific debates, trustworthiness

1. Introduction
The evaluation of scientific claims and the role of language style

The scientific method (e.g. Gauch, 2002) can generate knowledge that is widely accepted within scientific communities and the general public; however, it cannot produce knowledge that is 100% accurate or true (e.g. Popper, 1935). Researchers who are trained in the scientific method know that their scientific findings are inherently preliminary. When confronted with new scientific claims, they know that these claims will be debated and that controversies and disagreements may arise in the process. Laypeople, on the other hand, typically do not turn to science to engage in debates about new scientific
findings. Instead, laypeople turn to science to make everyday decisions and get answers to personally relevant questions, like “Should I vaccinate my child?” or “Does processed meat cause cancer?” Since knowledge is unevenly distributed in modern societies, laypeople often lack the necessary knowledge to answer such personally relevant questions and, therefore, regularly have to rely on experts’ scientific claims (e.g. Bromme et al., 2010; Bromme and Jucks, 2017; Keil et al., 2008).

But how do laypeople decide whether they can rely on scientific claims? According to the Content-Source Integration Model (Stadtler and Bromme, 2014), laypeople decide if they should rely on scientific claims by applying two strategies. First, they can make credibility judgments (also known as firsthand evaluations) by asking themselves whether a scientific claim is compatible with their own prior knowledge and by evaluating the logical coherence of the claim. Second, they can make trustworthiness judgments (also known as secondhand evaluations) by asking themselves whether there are any reasons to suggest that the source of the scientific claim is not trustworthy. Because laypeople often have only a bounded understanding of science and because scientific claims can be highly complex and therefore hard to evaluate, trustworthiness judgments are becoming increasingly important in modern societies (Bromme and Goldman, 2014; Bromme and Thomm, 2016).

Previous research has found that laypeople use certain factors upon which to judge credibility and trustworthiness (for literature reviews, see Choi and Stevilia, 2015; Metzger and Flanagin, 2015; Pornpitakpan, 2004). When it comes to evaluating scientific claims, one factor that seems especially important is the language style that an information source uses. Thon and Jucks (2016), for example, varied whether forum posts containing scientific claims were written in a technical language style (e.g. “myocardial infarction”) or in an everyday language style (e.g. “heart attack”). The results showed that participants rated the forum posts’ authors as more trustworthy and their information as more credible when they used the everyday language style instead of the technical language style. In a more recent study, König and Jucks (2019a) varied whether a forum post describing a new scientific study was written in a positive language style (e.g. “Due to the outstanding study of Mr. Weber and its excellent execution”) or a neutral language style (e.g. “Due to the study of Mr. Weber and its execution”). The results showed that the positive language style, in comparison to the neutral language style, negatively affected trustworthiness and credibility ratings of the post and its author. More specifically, when the forum post was written in the positive language style, the participants rated the provided information as less credible, and they rated the information source as more manipulative, less benevolent, and less sincere.

The prevalence and effects of aggressive language

Over the past decades, aggressive language has emerged as a rhetoric style in public debates and it has become more prevalent in various contexts. For example, it has become more prevalent in mass media coverage and in the political arena (Mutz and Reeves, 2005). Furthermore, aggressive language frequently appears in online comment sections of newspapers, social networking sites, and online platforms like YouTube (Moor et al., 2010; Rowe, 2015). Depending on the discipline and research setting, various names have been proposed for constructs that convey the use of aggressive language; some descriptive names include verbal venting (Rösner and Krämer, 2016), flaming (Moor et al., 2010), incivility (Mutz and Reeves, 2005), negative word-of-mouth (Pfeffer et al., 2014), and attack discourse (Anderson and Huntington, 2017). Although definitions of aggressive language differ, they frequently include personal attacks and uncivil language (Blom et al., 2014; Rösner and Krämer, 2016), such as “insulting, sarcastic, teasing, negative, or cynical comments” (Lapidot-Lefler and Barak, 2012).

Given the high prevalence of aggressive language in various contexts, it is essential to understand how this language style might affect laypeople’s credibility and trustworthiness judgments. In some contexts, researchers have already begun to investigate the effects of aggressive language.
In the context of political debates, for example, research has shown that using an aggressive language style to attack political opponents and engaging in negative campaigning can backfire and harm the politicians who use such strategies (e.g. Carraro et al., 2010; Cavazza, 2017; Nau and Stewart, 2014). Nau and Stewart (2014), for example, let their participants read a series of excerpts from congressional debates and committee hearings that were either written in their original neutral language style (e.g. “It just strikes me as unbelievable…”) or were manipulated to represent a more aggressive language style (e.g. “It just strikes me as unbelievable and hypocritical…”). Results showed that politicians who allegedly used the aggressive language style, in comparison to politicians who used the neutral language style, were rated as less trustworthy, less knowledgeable, less competent, and less likable. Furthermore, researchers have found that political incivility increases interest but at the same time lowers political trust (Mutz and Reeves, 2005).

**Aggressive language and hostility in scientific debates**

So far, no study has specifically investigated how often scientists use aggressive language in the context of scientific debates. However, recent events suggest that scientific debates have also been becoming more hostile and aggressive. Furthermore, this development gained wide public attention after the *New York Times Magazine* published an article about the change of tone in scientific debates (Dominus, 2017). The article links the increasing aggressiveness in scientific debates with the emergence of a methodological reform movement that has challenged the methodological understanding and malpractices in the scientific community of psychology research (see Nelson et al., 2018). The rise of this methodological reform movement has sparked various necessary scientific debates. The *New York Times Magazine* article, however, does not specifically focus on the scientific debates that were caused by the reform movement itself. Instead, the article focuses on the aggressiveness that is inherent in these current scientific debates.

In recent years, there have been numerous prominent examples of aggressive language use in the context of scientific debates. For example, during a presentation at the 13th Annual Meeting of the Society for Personality and Social Psychology, a prominent psychologist “interrupted loudly from the front row,” a behavior that the *New York Times Magazine* article described as “violating standard academic etiquette.” Another prominent incident happened after some researchers (Doyen et al., 2012) could not replicate the effects of a well-known psychological study (Bargh et al., 1996). As a response to the failed replication, one of the authors of the original study suggested in a *Psychology Today* blog that the researchers of the replication study were “incompetent or ill informed” (see Bartlett, 2013; Dominus, 2017). Another noteworthy incident happened after two statisticians wrote an article for the popular online magazine *Slate* (Gelman and Fung, 2016), in which they claimed that some scientific claims “are nothing but tabloid fodder.” One scientist who was attacked in the *Slate* article turned to Facebook and wrote that she had been “quiet and polite long enough” (see Gelman, 2016). She went on to suggest that she had been “attacked in a destructive way” and that she was “tired of being bullied.” Although these are just a few prominent examples of aggressive attacks in the context of scientific debates, they demonstrate that scientists make use of aggressive language to stress their arguments and communicate their scientific claims and positions. Therefore, it is essential to elaborate on the potential consequences of using an aggressive language style in scientific debates.

**This study: How does aggressive language influence the credibility of scientific claims and the trustworthiness of science communicators?**

One highly relevant question is whether the increasing use of aggressive language in scientific debates influences the credibility of scientific claims and the trustworthiness of scientists. Based
on the earlier mentioned findings from the political domain, it can be hypothesized that using an aggressive language style in scientific debates will reduce the credibility of the provided information and reduce the trustworthiness of the information source. Besides investigating the overall effect of an aggressive language style in scientific debates, it is important to understand whether this language style effect might be intensified or weakened according to some additional characteristics of the debater. Other personal characteristics such as credibility might be at play, as Burgoon et al. (2002) argued in the context of language expectancy theory that “highly credible communicators have the freedom (wide bandwidth) to select varied language strategies and compliance-gaining techniques in developing persuasive messages, while low-credible communicators must conform to more limited language options if they wish to be effective.” They go on by proposing that “highly credible sources can be more successful using low-intensity appeals and more aggressive compliance-gaining messages than can low-credible communicators using either strong or mild language.” Following this argumentation, it can be hypothesized that the aggressive language style effect will be weaker (less harmful to credibility and trustworthiness) for highly credible sources like scientists (e.g. Fiske and Dupree, 2014) than for less credible sources like lobbyists (e.g. Gallup, 2017). Furthermore, it can be hypothesized that being a low-credible information source (being a lobbyist instead of being a scientist) will by itself somewhat reduce credibility and trustworthiness judgments. To test these hypotheses, we staged a scientific debate and varied the professional affiliation of the person arguing in the debate (whether the person was introduced as a scientist or a lobbyist) and language style (whether a neutral language style or an aggressive language style was used). In a between-subject online experiment, research participants watched a video of the staged scientific debate and subsequently evaluated the trustworthiness of one of the debaters and the credibility of his information. The goal was to answer the following research questions:

Research question 1: Does an aggressive language style, in comparison to a neutral language style, decrease the trustworthiness of a person arguing in a scientific debate and the credibility of his information?

Research question 2: Does being introduced as a lobbyist, in comparison to being introduced as a scientist, decrease the trustworthiness of a person arguing in a scientific debate and the credibility of his information?

Research question 3: Does being introduced as a lobbyist in combination with using an aggressive language style result in especially low trustworthiness and credibility ratings?

2. Method

Design and material

We used a 2 (language style: neutral language vs aggressive language) × 2 (professional affiliation: scientist vs lobbyist) between-subject experimental design, resulting in four experimental conditions. For each experimental condition, a video of a scientific debate was created that consisted of two parts. In the first part of the video, three male actors were seen on stage: a debate moderator was in the middle, and there were two panelists, one on the left and one on the right. Shortly after the video started, the moderator turned to the right panelist and summarized the conclusions of a study that was allegedly described earlier by the left panelist. Then, the moderator asked the right panelist whether he agreed with the study conclusions. The first part of the video was the same in all four experimental conditions. In the second part of the
Table 1. Manuscript of the scientific debate.

Debate Moderator:
Mr. Schneider, we have just heard that - according to Mr. Becker’s study - the use of antidepressants has almost no effect when treating patients with depression. Do you agree with this?

Right Panelist:
I have a completely different opinion. I am convinced that the use of antidepressants is perfectly justified when treating patients with depression.

The reason for Mr. Becker’s conclusion - that psychotherapy without antidepressants is just as effective as psychotherapy with antidepressants - lies in a series of really dumb methodological mistakes that Mr. Becker made in his study and in his amateurish approach.

Mr. Becker’s first daft mistake was that he didn’t control whether the participants of the three groups had varying depression levels at the beginning of the study and that is simply stupid. If you cannot tell in the beginning stages of a study whether the groups are equally depressed, you will not be able to conclude that one treatment was more effective than the others. That’s why Mr. Becker’s approach was not correct from a methodological perspective and rather incompetent and naïve.

Mr. Becker’s second idiotic mistake was that he revealed his hypotheses to his patients in the beginning of the study which was completely moronic. The problem is that the knowledge about the hypotheses can distort the responses of the patients and, therefore, you can no longer make distinct conclusions about the effectiveness of the different treatments - that’s the reason why you normally wouldn’t do such a dumb thing.

The third dimwitted mistake was that Mr. Becker used an inadequate depression scale that cannot accurately differentiate between mild and moderate depression. The use of this inadequate depression scale is the third reason why we should be skeptical about Mr. Becker’s conclusions and it shows once again how inconsiderately and idiotically he behaved.

All in all, you can say that Mr. Becker’s study features some methodological shortcomings and that his approach was dilettantish. That’s why I am still convinced that the use of antidepressants is justified and reasonable when treating patients with depression.

The aggressive language style version of the scientific debate contained the words and phrases printed in bold face and the neutral language style version did not contain these words and phrases. The table shows an English translation of the German manuscript. Therefore, the translated version may not appear as authentic to native English speakers as the original version appears to native German speakers. The original German version of the manuscript can be obtained from the authors.

video, the right panelist answered the moderator’s question by disagreeing with the study conclusions and criticizing the methodology of the study on several grounds. The topic of the scientific debate was the effectiveness of antidepressants. The study that was debated had allegedly found that antidepressants are not effective in treating depression; the topic was chosen because the effectiveness of antidepressants is a widely discussed topic in psychology and medicine. The scientific study discussed in the staged debate was fictitious, but this was not mentioned during the video presentation.

The experimental manipulations were realized in the second part of the video. Depending on the experimental condition, the right panelist described and criticized the study and its methodology either in (1) a neutral language style or (2) an aggressive language style. The aggressive language style version of the video was created by adding insulting, negative and cynical comments (see Lapidot-Leffler and Barak, 2012) to the panelist’s argumentation. Table 1 contains a manuscript of the scientific debate that shows the language style manipulation. Furthermore, during the second part of the video, a continually visible banner at the bottom of the video displayed the right panelist’s name (Johannes Schneider) and his professional affiliation. Depending on the experimental condition, the
right panelist was introduced either as (1) a scientist who worked for a pharmacological institute at a university or (2) as a lobbyist who worked at a pharmacological lobbying organization.

**Sample**

To facilitate external validity, the goal was to recruit participants with an intrinsic interest in depression-related topics, who would potentially watch online videos to acquire relevant information. Therefore, German university students enrolled in psychology programs were chosen as participants. Participants were contacted via email and social network sites and received €5 for participating in the online experiment. Participants who indicated at the end of the study that they answered the questions honestly and completed the study without interruption and technical problems were included in data analyses. A total of 14 participants were excluded from data analyses because they took much longer (two standard deviations longer) than the average participant to complete the study. The final sample contained 221 (182 female, 39 male) participants with an average age of 23 years ($M=22.87, SD=4.00$). Furthermore, the average participant had been enrolled in their study program for four semesters ($M=4.40, SD=3.29$) and took 17 minutes ($M=16.51, SD=5.20$) to complete the study.

**Procedure**

The experiment was conducted online using the Questback EFS Survey© platform for data collection. Before the experiment started, participants were told that the experiment would address the communication of scientific information in online videos. Furthermore, they were informed about the general procedure of the upcoming experiment and that they could end the experiment at any time. To start the experiment, participants had to indicate that they had read all provided information and that they agreed to take part in the experiment. On the remaining pages, participants indicated their age, gender, whether they studied psychology at the bachelor’s or master’s level, the university where they studied psychology, and the semester they were currently in. Furthermore, they answered the control measures (see section “Control measures”). Following this, participants were randomly assigned to one of the four experimental conditions and watched the corresponding online video (see section “Design and material”). After watching the online video, participants answered the dependent measures (see sections “Credibility measures” and “Trustworthiness measures”). After answering the dependent measures, participants answered the manipulation check questions (see section “Manipulation check”) and were debriefed. They were told about the manipulations of the experiment, that the study discussed in the debate and its results were fictions and that they could contact the leading scientist if they had any further questions or comments. Furthermore, they could choose to leave their information to get reimbursed for their participation. The experiment was designed to comply with the ethical guidelines developed by the American Psychological Association (APA) and the German Psychological Society (DGPs).

**Control measures**

Four control measures were included to assess whether the experimental groups differed in regard to characteristics that could affect the study results. It is possible that people who frequently watch online videos are better able to identify the quality of online videos and their content than people who just occasionally watch online videos. Therefore, participants answered the questions “How often do you watch videos online?” (General Use) and “How often do you watch videos
online to learn something new or acquire new skills?” (Educational Use) on a scale ranging from 1 (very rarely) to 7 (very often). Furthermore, it is possible that people who are well informed about a topic make different credibility and trustworthiness judgments than people who are less informed about the same topic. Therefore, participants answered the question “How much do you know about depression and antidepressants?” (Prior Knowledge) on a scale ranging from 1 (very little) to 7 (very much) and indicated their agreement with the statement “Antidepressants are effective drugs for the treatment of depression” (Prior Attitude) on a scale ranging from 1 (totally disagree) to 7 (totally agree).

**Credibility measures**

The credibility of information can be assessed in diverse ways. For the purpose of this study, three measurements were used: a measure that assessed the general credibility of the provided information (Message credibility), a specific credibility measure that assessed how much participants agreed with the main statement of the online video (Antidepressants attitude), and a measure that assessed the subjective learning gain of the participants (Subjective comprehension). Participants indicated how much they agreed with the provided statements on a scale ranging from 1 (totally disagree) to 7 (totally agree). For the Message credibility and Subjective comprehension measures, a total score was generated by calculating the mean.

**Message credibility.** As a general credibility measure of the provided information, the Message Credibility Scale (Appelman and Sundar, 2016) was translated and adapted. Participants indicated their agreement with three statements, for example, “The provided information was believable.”

**Antidepressants attitude.** As a more specific credibility measure, participants were asked how much they agreed with the main statement of the online video (“Antidepressants are effective drugs for the treatment of depression”).

**Subjective comprehension.** As a more indirect credibility measure, the Subjective Comprehension Subscale from the Recipient Orientation Scale (Bromme et al., 2005; Zimmermann and Jucks, 2018) was adapted to assess the participants’ subjective learning gain. Participants indicated their agreement with five statements, for example, “I have the feeling that I have learned something new by watching the online video.”

**Trustworthiness measures**

Researchers have used various measures to assess trustworthiness, depending on the research question and setting. For the purpose of this study, three measurements were used: a measure that tapped into the manipulative character of a person (Machiavellianism), a measure developed for the assessment of experts’ trustworthiness in online contexts (Expertise, Integrity, Benevolence), and a more general likability measure (Likability). For each measure, a total score was generated by calculating the mean.

**Machiavellianism.** The German version of the Machiavellianism Subscale from the Dirty Dozen Scale (Jonason and Webster, 2010; Küfner et al., 2014) was used to assess how manipulative the person arguing in the scientific debate was perceived to be. Participants indicated their agreement with four statements, for example, “Johannes Schneider has used deceit or lied to get his way,” on a scale ranging from 1 (totally disagree) to 7 (totally agree).
**Expertise, integrity, benevolence.** The Muenster Epistemic Trustworthiness Inventory (Hendriks et al., 2015) was used to assess how trustworthy the person arguing in the scientific debate was perceived to be. Participants rated 15 items on a scale ranging from 1 (not trustworthy at all) to 7 (very trustworthy). The items measured expertise (six items, e.g. “unqualified—qualified”), benevolence (four items, e.g. “immoral—moral”) and integrity (four items, “insincere—sincere”).

**Likability.** The Reysen Likability Scale (Reysen, 2005) was translated and adapted to assess how likable the person arguing in the scientific debate was perceived to be. Participants indicated their agreement with 11 statements, for example, “I would ask Johannes Schneider for advice,” on a scale ranging from 1 (totally disagree) to 7 (totally agree).

**Manipulation check**

Two measures were included to assess whether the participants remembered the language style and the professional affiliation correctly.

**Language style.** To assess whether the participants remembered the language style, they were asked whether certain aggressive expressions were used during the scientific debate. Participants could choose between “Yes,” “No,” and “I do not know.”

**Professional affiliation.** To assess whether the participants remembered the professional affiliation of the debater, they were asked “For whom did Johannes Schneider work?” Participants could choose between “A university,” “A lobbying organization,” and “I do not know.”

**3. Results**

**General procedure and data availability**

For all analyses, the statistical software IBM© SPSS© Statistics Version 25 was used. For the main analyses of the dependent measures, two-way between-subject analyses of variance were conducted with language style (neutral language vs aggressive language) and professional affiliation (scientist vs lobbyist) as independent variables. Type three sums of squares were used. For all analyses, the alpha level was set at $\alpha=0.05$. The research dataset and the used syntax code are available from the authors on request. The dataset contains further variables that have not been described in this article and have not been analyzed yet because they exceed the scope of this article.

**Manipulation check**

Of the 221 participants, 195 (88%) correctly remembered the professional affiliation of the person arguing in the scientific debate and 205 (93%) correctly remembered whether he used a neutral or an aggressive language style; 179 (81%) participants remembered both correctly. Since it was of interest whether these two factors interacted with each other, the following data analyses were based on the 179 participants who remembered both factors correctly.

**Control measures**

Before analyzing the dependent measure, four one-way between-subject analyses of variance were conducted with experimental condition as the independent variable and the control measures as
dependent variables to analyze whether the participants in the four experimental groups differed in aspects that could influence the study results. The results showed that the participants in the four experimental groups did not significantly differ in regard of their general online video use (General Use: \( F(3, 175) = 2.202, p = .090 \)), online video use for educational purposes (Educational Use: \( F(3, 175) = 0.370, p = .775 \)), prior knowledge (Prior Knowledge: \( F(3, 175) = 0.296, p = .829 \)), and their attitude toward antidepressants (Prior Attitude: \( F(3, 175) = 1.176, p = .320 \)). Therefore, the four control measures were not included in further analyses.

**Credibility measures**

*Message credibility.* There was a significant main effect of language style (\( F(1, 175) = 21.130, p < .001, \eta_p^2 = .108 \)) on message credibility, indicating that aggressive language led to lower message credibility ratings than neutral language. There was no main effect of professional affiliation (\( F(1, 175) = 2.081, p = .151, \eta_p^2 = .012 \)) and no interaction effect (\( F(1, 175) = 1.826, p = .178, \eta_p^2 = .010 \)).

*Antidepressants attitude.* There were no main effects of language style (\( F(1, 175) = 0.009, p = .924, \eta_p^2 < .001 \)) and professional affiliation (\( F(1, 175) = 0.196, p = .659, \eta_p^2 = .001 \)) on antidepressants attitude. Furthermore, there was no interaction effect (\( F(1, 175) = 0.547, p = .460, \eta_p^2 = .003 \)).

*Subjective comprehension.* There was a significant main effect of language style (\( F(1, 175) = 14.034, p < .001, \eta_p^2 = .074 \)) on subjective comprehension, indicating that aggressive language led to lower subjective comprehension ratings than neutral language. Furthermore, there was no main effect of professional affiliation (\( F(1, 175) = 0.332, p = .565, \eta_p^2 = .002 \)) and no interaction effect (\( F(1, 175) = 0.050, p = .824, \eta_p^2 < .001 \)). Table 2 shows the means and standard deviations of the credibility measures.

**Trustworthiness measures**

*Machiavellianism.* There was a significant main effect of language style (\( F(1, 175) = 9.645, p = .002, \eta_p^2 = .052 \)) on Machiavellianism, indicating that aggressive language led to higher Machiavellianism ratings than neutral language. Furthermore, there was a significant main effect of professional affiliation (\( F(1, 175) = 8.714, p = .004, \eta_p^2 = .047 \)) on Machiavellianism, indicating that being a lobbyist led to higher Machiavellianism ratings than being a scientist. There was no interaction effect (\( F(1, 175) = 0.001, p = .976, \eta_p^2 < .001 \)).

*Expertise.* There was a significant main effect of language style (\( F(1, 175) = 59.100, p < .001, \eta_p^2 = .252 \)) on expertise, indicating that aggressive language led to lower expertise ratings than neutral language. There was no main effect of professional affiliation (\( F(1, 175) = 0.394, p = .531, \eta_p^2 = .002 \)) and no interaction effect (\( F(1, 175) = 0.545, p = .461, \eta_p^2 = .003 \)).

*Integrity.* There was a significant main effect of language style (\( F(1, 175) = 14.187, p < .001, \eta_p^2 = .075 \)) on integrity, indicating that aggressive language led to lower integrity ratings than neutral language. Furthermore, there was a significant main effect of professional affiliation (\( F(1, 175) = 8.825, p = .003, \eta_p^2 = .048 \)) on integrity, indicating that being a lobbyist led to lower integrity ratings than being a scientist. There was no interaction effect (\( F(1, 175) = 0.038, p = .845, \eta_p^2 < .001 \)).

*Benevolence.* There was a significant main effect of language style (\( F(1, 175) = 36.752, p < .001, \eta_p^2 = .174 \)) on benevolence, indicating that aggressive language led to lower benevolence ratings
than neutral language. Furthermore, there was a significant main effect of professional affiliation ($F(1, 175) = 4.328, p = .039, \eta^2_p = .024$) on benevolence, indicating that being a lobbyist led to lower benevolence ratings than being a scientist. There was no interaction effect ($F(1, 175) = 0.466, p = .496, \eta^2_p = .003$).

Likability. There was a significant main effect of language style ($F(1, 175) = 57.495, p < .001, \eta^2_p = .247$) on likability, indicating that aggressive language led to lower likability ratings than neutral language. Furthermore, there was a significant main effect of professional affiliation ($F(1, 175) = 9.378, p = .003, \eta^2_p = .051$) on likability, indicating that being a lobbyist led to lower likability ratings than being a scientist. There was no interaction effect ($F(1, 175) = 0.085, p = .771, \eta^2_p < .001$). Table 2 shows the means and standard deviations of the trustworthiness measures.

### 4. Discussion

Three research questions addressed the effects of language style (neutral language vs aggressive language), professional affiliation (scientist vs lobbyist), as well as their interaction in a scientific debate. Research question 1 asked whether language style (neutral language vs aggressive language) influences trustworthiness judgments and the credibility of information. We hypothesized that an aggressive language style, in comparison to a neutral language style, would negatively affect the trustworthiness of the debater and the credibility of the provided information. The results support this hypothesis. If the debater used an aggressive language style, in comparison to a neutral language style, he was perceived as less trustworthy and his information as less credible. More specifically, if the debater used an aggressive language style, he was perceived as more manipulative (Machiavellianism), less competent (Expertise), less sincere (Integrity), less benevolent (Benevolence), and less likable (Likability). Furthermore, the information he provided was
perceived as less credible (Message credibility), and participants had the impression that they learned less from the scientific debate (Subjective comprehension). The only credibility measure not affected by language style was the participants’ attitude toward the topic of the scientific debate (Antidepressants attitude).

Research question 2 asked whether the professional affiliation (scientist vs lobbyist) of a person arguing in a scientific debate influences his trustworthiness and the credibility of the provided information. We hypothesized that being introduced as a lobbyist, in comparison to being introduced as a scientist, would negatively affect the trustworthiness of the debater and the credibility of the information he provided. The results partly support this hypothesis. If the debater was introduced as a lobbyist, in comparison to being introduced as a scientist, he was perceived as less trustworthy. More specifically, the debater was perceived as more manipulative (Machiavellianism), less sincere (Integrity), less benevolent (Benevolence), and less likable (Likability). However, expertise (Expertise) was not affected. Furthermore, the professional affiliation of the scientific debater did not affect the credibility of the provided information (Message credibility, Antidepressants attitude, Subjective comprehension).

Finally, research question 3 asked whether the language style (neutral language vs aggressive language) and the professional affiliation (scientist vs lobbyist) of a person arguing in a scientific debate would interact with each other to influence trustworthiness and the credibility of the provided information. We hypothesized that being a lobbyist would intensify the negative effects of the aggressive language style, but the results showed that the two factors did not interact with each other.

The results demonstrate that an aggressive language style negatively affects the trustworthiness of a person arguing in a scientific debate and the credibility of his information. The effect was found on every measure except the credibility measure of attitude. Another interesting finding is that the professional affiliation did not influence the trustworthiness of the arguing person and the credibility of the provided information in the same direction. The results show that the lobbyist was seen as less trustworthy in comparison to the scientist, but professional affiliation did not affect the credibility of the provided information. A reason for this finding might be that lobbyists are generally perceived as less trustworthy, but people do not automatically infer that therefore their information is also less credible. Instead, they might just conclude that a lobbyists’ information is less credible if he had the chance to actively manipulate the provided information to his advantage. This interpretation would be in line with findings from König and Jucks (2019b), who showed that scientific claims from a lobbyist are less credible if they are supported by studies that were conducted by the lobbyist himself, in comparison to studies that were conducted by other scientists. In this study, the lobbyist did not have the opportunity to manipulate any information; instead, he criticized an opposing position by delivering scientifically sound and strong arguments, and, therefore, the credibility of his information was not affected. Unexpectedly, the language style and the professional affiliation did not interact with each other. This finding is especially surprising because our participants perceived the scientist as more trustworthy than the lobbyist, and therefore the scientist, in accordance with language expectancy theory, should have had the freedom to choose a more aggressive language style.

Besides considering the significance levels and descriptive statistics, it is important to consider the effect sizes. The largest effect of the language style manipulation was found on the Expertise scale ($\eta^2_p = .252$), and the largest effect of the professional affiliation manipulation was found on the Likability scale ($\eta^2_p = .051$). As small effect sizes are the rule in various disciplines (Valkenburg and Peter, 2013), the relatively large effect sizes found in this study additionally stress the importance of the language style and the professional affiliation of a person arguing in a scientific debate.
Limitations

Although the findings of this study highlight the importance of the language style and the professional affiliation of people arguing in a scientific debate, there are limitations to the generalizability of the results. Some limitations result from the chosen study sample. For example, previous research has identified age differences in the suggestibility to misinformation and source monitoring (e.g. Mitchell et al., 2003). This might be a problem because the study sample consisted of university students who were, on average, relatively young. Furthermore, younger generations are probably more used to aggressive language than older generations and therefore they might react differently to aggressive language styles. Hence, future research should replicate this study with different age groups to see whether the results are generalizable in this regard. Furthermore, previous research has found that information literacy influences credibility and trustworthiness judgments (e.g. Choi and Stevilia, 2015). This might be a problem because the study sample consisted of undergraduate and graduate students who might have had higher information literacy than the average person. Therefore, future research should replicate this study with participants who differ in information literacy to analyze whether this might moderate the effects found in this study.

Further limitations to the generalizability of the results might arise from the chosen topic of the scientific debate. For example, this study found that participants’ attitudes toward the topic of the scientific debate (Antidepressants attitude) was not affected by language style. A possible reason for this finding might be that the study sample consisted of psychology students who are repeatedly confronted with information about depression and antidepressants during their study programs. Therefore, the study participants might already have had strong attitudes about the topic, and one scientific debate was not enough to change their minds significantly. Furthermore, psychology students might be especially interested in a debate about antidepressants and therefore might react differently than the general public to the presented information. Therefore, future research should replicate this study with participants that do not have a background in psychology or medicine.

Another factor that might limit the generalizability of the results is the setting in which the scientific debate took place. In this study, the debater directed his criticism and aggressive language directly at the other panelist in a face-to-face discussion. The study participants might have perceived this behavior as particularly inadequate because it contradicts discussion norms in university classrooms. Therefore, it would be interesting to replicate this study in other communication settings, where other discussion norms might exist. For example, previous research has found that aggressive language is used in Twitter discussions (Anderson and Huntington, 2017). Since Twitter and other social media outlets make it possible to use aggressive language to attack people in discussions without facing them in person, it would be interesting to test whether this circumstance would affect the magnitude of the effects found in this study. Besides changing the discussion setting, future research should also explore how other professional affiliations might alter the effects. As mentioned earlier, aggressive language is quite common in political debates. Therefore, it might not be as detrimental for a politician to use aggressive language in a scientific debate as it is for a scientist.

Further limitations to the generalizability of the results arise from the geographical location of the experiment. According to Jasanoff (2011), countries have developed different civic epistemologies throughout history. Civic epistemologies refer to the ways in which societies evaluate and discuss the importance of (scientific) knowledge claims. In the United States, for example, “information is typically generated by interested parties and tested in public through
overt confrontation between opposing, interest-laden points of view” (Jasanoff, 2011). Germany, in contrast, possesses a more consensual civic epistemology which focuses on “building communally crafted expert rationales, capable of supporting a policy consensus.” Due to this consensual approach, discussions in Germany might be generally less aggressive than in other countries. Consequently, the results of this study are probably not generalizable across different countries and therefore no normative recommendations (e.g. “Science communicators around the world should use neutral language in scientific debates.”) should be drawn from it. However, it would be highly interesting to replicate this study in countries with different civic epistemologies to see whether the results would differ.

5. Conclusion

This study has shown that when people are confronted with aggressive language by science communicators, they will judge these communicators to be less trustworthy and deem the communicated information as less credible. Furthermore, when the science communicators are introduced as lobbyists, they will be perceived as less trustworthy. However, their information will not be deemed less credible. These findings are important for various disciplines and professions, because they demonstrate that information seekers do not just react to scientific information itself. Instead, they are sensitive to the way scientific information is presented and who presents it.

Acknowledgements

We thank Dr Robin Junker, Dr Jens Riehemann, and Dr Jonathan Barenberg for their great performance in the scientific debate. Furthermore, we thank Dr Celeste Brennecka for the language editing and our research assistant Hannah Meinert for her support. Lastly, we thank Dr Susan Howard and the anonymous reviewers for their insightful comments and valuable suggestions.

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) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: this work was supported by the German Research Foundation (DFG) within the framework of the Research Training Group GRK 1712 Trust and Communication in a Digitized World. The funding was granted to the second author.

ORCID iD

Lars König https://orcid.org/0000-0003-1450-8449

References

Anderson AA and Huntington HE (2017) Social media, science, and attack discourse: How Twitter discussions of climate change use sarcasm and incivility. Science Communication 39(5): 598–620.

Appelman A and Sundar SS (2016) Measuring message credibility: Construction and validation of an exclusive scale. Journalism & Mass Communication Quarterly 93(1): 59–79.

Bargh JA, Chen M and Burrows L (1996) Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. Journal of Personality and Social Psychology 71(2): 230–244.
Keil FC, Stein C, Webb L, Billings VD and Rozenbilt L (2008) Discerning the division of cognitive labor: An emerging understanding of how knowledge is clustered in other minds. *Cognitive Science* 32(2): 259–300.

König L and Jucks R (2019a) *It’s an amazing drug!* Positive language decreases the credibility of health information and the trustworthiness of health communicators in online health forums. Manuscript Submitted for Publication.

König L and Jucks R (2019b) When do information seekers trust scientific information? Insights from recipients’ evaluations of online video lectures. *International Journal of Educational Technology in Higher Education* 16(1).

Küfner ACP, Dufner M and Back MD (2014) Das Dreckige Dutzend und die Niederträchtigen Neun: Kurzskalen zur Erfassung von Narzissmus, Machiavellismus und Psychopathie [The Dirty Dozen and the Nasty Nine: Short scales for measuring narcissism, Machiavellianism and psychopathy]. *Diagnostica* 61(2): 76–91.

Lapidot-Leffler N and Barak A (2012) Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior* 28(2): 434–443.

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

Mitchell KJ, Johnson MK and Mather M (2003) Source monitoring and suggestibility to misinformation: Adult age-related differences. *Applied Cognitive Psychology* 17(1): 107–119.

Moor PJ, Heuvelman A and Verleur R (2010) Flaming on YouTube. *Computers in Human Behavior* 26(6): 1536–1546.

Mutz DC and Reeves B (2005) The new videomalaise: Effects of televised incivility on political trust. *American Political Science Review* 99(1): 1–15.

Nau C and Stewart CO (2014) Effects of verbal aggression and party identification bias on perceptions of political speakers. *Journal of Language and Social Psychology* 33(5): 526–536.

Nelson LD, Simmons J and Simonsohn U (2018) Psychology’s renaissance. *Annual Review of Psychology* 69: 171–1724.

Pfeffer J, Zorbach T and Carley KM (2014) Understanding online firestorms: Negative word-of-mouth dynamics in social media networks. *Journal of Marketing Communications* 20: 117–128.

Popper K (1935) *Logik der Forschung: Zur Erkenntnistheorie der Modernen Naturwissenschaft* [The Logic of Research: The Epistemology of Modern Natural Science]. Berlin, BE: Springer-Verlag.

Pornpitakpan C (2004) The persuasiveness of source credibility: A critical review of five decades’ evidence. *Journal of Applied Social Psychology* 34(2): 243–281.

Reysen S (2005) Construction of a new scale: The Reysen Likability Scale. *Social Behavior and Personality* 33(2): 201–208.

Rösner L and Krümer NC (2016) Verbal venting in the social web: Effects of anonymity and group norms on aggressive language use in online comments. *Social Media and Society*. Epub ahead of print 16 August. DOI: 10.1177/2056305116664220.

Rowe J (2015) Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, Communication & Society* 18(2): 121–138.

Stadler M and Bromme R (2014) The content–source integration model: A taxonomic description of how readers comprehend conflicting scientific information. In: Rapp DN and Braasch JLG (eds) *Processing Inaccurate Information: Theoretical and Applied Perspectives from Cognitive Science and the Educational Sciences*. Cambridge, MA: The MIT Press, pp. 379–402.

Thon FM and Jucks R (2016) Believing in expertise: How authors’ credentials and language use influence the credibility of online health information. *Health Communication* 32(7): 828–836.

Valkenburg PM and Peter J (2013) Five challenges for the future of media-effects research. *International Journal of Communication* 7: 197–215.

Zimmermann M and Jucks R (2018) How experts’ use of medical technical jargon in different types of online health forums affects perceived information credibility: Randomized experiment with laypersons. *Journal of Medical Internet Research* 20(1): 1–13.
Author biographies

Lars König (MSc in Psychology, University of Würzburg, Germany) works at the University of Münster, Germany. He is a member of the Research Training Group “Trust and Communication in a Digitized World.” Currently, his research focuses on science/health communication, persuasion in online environments, and the strategic design of digital learning environments.

Regina Jucks (Dr Phil, University of Münster, Germany) is a Full Professor at the University of Münster, Germany. Her research focuses on various instructional settings, ranging from doctor-patient communication to higher education. Jucks is a Principal Investigator in the Research Training Group “Trust and Communication in a Digitized World” and a Consortium Partner in the EU Horizon 2020 project “Dialogue and Argumentation for Cultural Literacy Learning in Schools (DIALLS).”