Attacking science on social media: How user comments affect perceived trustworthiness and credibility

Lukas Gierth and Rainer Bromme
University of Münster, Germany

Abstract

The science on controversial topics is often heatedly discussed on social media, a potential problem for social-media-based science communicators. Therefore, two exploratory studies were performed to investigate the effects of science-critical user comments attacking Facebook posts containing scientific claims. The claims were about one of four controversial topics (homeopathy, genetically modified organisms, refugee crime, and childhood vaccinations). The user comments attacked the claims based on the thematic complexity, the employed research methods, the expertise, or the motivations of the researchers. The results reveal that prior attitudes determine judgments about the user comments, the attacked claims, and the source of the claim. After controlling for attitude, people agree most with thematic complexity comments, but the comments differ in their effect on perceived claim credibility only when the comments are made by experts. In addition, comments attacking researchers' motivations were more effective in lowering perceived integrity while scientists' perceived expertise remained unaffected.

Keywords

epistemic trust, online credibility, science communication, science conflicts, user comments

Social media are increasingly relevant as a platform for science communication and the discussion of scientific topics both because scientists use them to collaborate and distribute research findings (Van Noorden, 2014) and because laypeople turn to the Internet to search for information (Brossard and Scheufele, 2013; Su et al., 2015). Platforms like Facebook allow for a far-reaching distribution of scientific findings and for comments by the general audience. Such contributions are often short, pointed, and invite oversimplification.

Social-media-based science communicators potentially have to interact with such critical, pointed attacks in the comment section. Particularly in regard to topics that are controversial in the

Corresponding author:
Lukas Gierth, Graduiertenkolleg "Vertrauen & Kommunikation in einer digitalisierten Welt," University of Münster, Domplatz 6, Münster 48143, Germany.
Email: gierth@uni-muenster.de

public eye, such as vaccination, concern has been raised about how science deniers use social media to discredit and undermine scientific consensus (Evrony and Caplan, 2017).

To soberly assess these dangers, it is helpful to empirically study the possible negative effects of attacks on scientific findings presented on social media in order to understand which kinds of comments affect the perceived credibility of claims and trustworthiness of scientists the most. Such research requires a differentiation of types of comments as well as a map of factors, which might modify their impact. We will provide a categorical classification, which is inspired by laypersons' explanations for conflicts among scientists and will then present two studies testing its feasibility for the study of comment effects in social-media-based science communication. This will add to the understanding of how the content of the user comments relate to the audience's reasoning about the communicated information and, for social-media-based science communicators specifically, could thus facilitate a critical evaluation of the comment section to identify potentially harmful attacks.

1. Effects of user comments on the perceived credibility of online information

In general, credibility assessments of online information are determined by characteristics of the user, the source, the content, and the context, for example, website functionality or design (Choi and Stvilia, 2015; Metzger and Flanagin, 2015). Both Metzger and Flanagin (2015) and Choi and Stvilia (2015) see user comments being used as a cue for the opinions of others, in the form of social endorsement or the ratio of positive and negative comments. However, user comments do not only vary in regard to valence. For instance, negative user comments can include different types of attacks, justifications, and reasons for their negative standpoint (Lörcher and Taddicken, 2017). Such differences in regard to attacks need to be understood to understand the effect of attacks of user comments on scientific claims on social media. This is particularly important, because in contrast to aspects of the user, the source, the message, and the style or design, user comments are not present in offline publications. Since they offer a window into the real-time reactions of others to the information piece, they are one of the aspects that make social media distinctly *social*. Therefore, to understand credibility evaluations of online information, particularly on social media, it is important to understand the effect of user comments.

Thus far, studies have systematically compared user comments that were either civil or uncivil (Anderson et al., 2014; Jennings and Russell, 2019) and argumentative or subjective (Winter and Krämer, 2016). Somewhat related, studies have investigated experts commenting on each others' work using either aggressive or polite language in video or blog posts (König and Jucks, 2019a; Yuan et al., 2019). Furthermore, some studies analyzed user comments pointing out ethical and content-related flaws while manipulating whether the commenter was commenting on their *own* blog post or whether *another* critic did so (Hendriks et al., 2016a, 2016b). However, no study has systematically compared different kinds of user comments in regard to the content of the comment. To investigate this question, one needs a framework by which different types of critical user comments can be categorized.

2. Categorizing attacks on science

A recent study by Barnes et al. (2018) provided the most elaborated categorization of the content of attacks on scientific claims we know of. A scientific claim was paired with one of five kinds of ad hominem attacks (targeting the source of the scientific claim) and one content-oriented attack:

scientists' misconduct in relation to the claim, scientists' past misconducts that were not relevant to the claims, bad education of the scientist, sloppy reputation of the scientist, and conflict of interest of the scientist. The content-oriented attack was the failure to include a control group, but even in this case the critical comment made clear that it was the attacked researcher who was responsible for this flaw. Interestingly, a sloppy reputation and bad education had no effects on agreement with the study claim, while all others negatively affected agreement with the study claim with no significant differences in severity. While this study delivers an inspiring list of ad hominem attacks, it does not provide a complementary list of attacks that are more focused on the content of the study under attack. In the following, we will provide another rationale for a categorization of possible attacks on scientific claims.

Reading a scientific claim and then reading a critical comment denying its validity is a case of reading about a scientific conflict. Such conflicts among scientists are a part of the everyday business in doing research. Nevertheless, the public sometimes understands these controversies as indicative of a weakness of the discussed claims. Therefore, laypersons' subjective explanations for controversies among scientists have been studied. Based on interview studies (Bromme et al., 2015; Kajanne and Pirttilä-Backman, 1999) about laypersons explanations' for disagreements among health experts, Thomm et al. (2015) found four explanations laypeople generate for scientists' disagreements: differences in competence, differences in motivations, differences in the research process, and the overall thematic complexity of the topic of contention. Studies of two German (Thomm and Bromme, 2016; Thomm et al., 2015), an Israeli (Thomm et al., 2017), and three American (Dieckmann et al., 2015; Dieckmann and Johnson, 2019; Johnson, 2017) samples paint a largely robust picture of these four conceptually distinct causal explanations of scientific conflicts by laypeople.

Which explanation is favored depends on person factors (e.g. cognitive ability; see Dieckmann et al., 2015) as well as on information about the scientists who are in dispute. For instance, people are more likely to attribute a scientific conflict to differences in motivations if it is between a university and an industry-funded researcher (Thomm et al., 2015). Furthermore, when presented with a scientific conflict, people rate the less competent source as less credible—but this was partially mediated by whether they viewed the conflict as being due to the different levels of competence (Thomm and Bromme, 2016). Thus, conflict explanations can also affect source credibility.

However, this relationship could also be reversed. To create a conflict in people's minds, these four explanations of scientific conflict could be turned into arguments attacking scientists. For instance, a science denier might state that scientists lack the necessary expertise, or are too biased to make valid claims in regard to an issue of contention. Similarly, they might state that the methods used are insufficient or too unreliable and that the inherent complexity and randomness of the topic forbids making concrete claims about it. In fact, science deniers do this; some of the most used arguments of global warming skeptics are that temperature records and climate models are too unreliable, and that climate scientists are a part of a conspiracy (Elsasser and Dunlap, 2013; Skeptical Science, 2019). Hence, the four dimensions of laypersons' explanations for conflicts among scientists provide a heuristic for categorical distinctions between four types of attacks against the credibility of a scientific claim.

Since the suspected cause of disagreement can lead to further inferences about the source and the credibility of claims put forward by that source, which types of attacks affect the credibility of an attacked claim and which types of attacks affect the trustworthiness of the source of an attacked claim is an open empirical question. In particular, no study has evaluated the effect of arguments stating that controversial topics are too complex to make any concrete claims about them.

3. Study rationale and research questions

In this study, we aim to investigate how the content of critical user comments attacking scientific claims and sources on social media affect the perceived trustworthiness of the attacked source and the credibility of the attacked claim. The content of the user comments is adapted from the four explanations for scientific conflict: accusations of incompetence, accusations of conflict of interest, pointing out the dependence of the results on the research methods, and pointing out the thematic complexity of the topic. Since previous research has found that preference for the different conflict explanations is varied, it is important to also understand how much people agree with such counterarguments based on these explanations.

Apart from people's agreement with the critical user comments, there are two levels to consider in regard to judgments about the attacked claim: past research found that laypeople seem to differentiate between what they see as credible, and what they personally agree with (Bromme et al., 2015; Scharrer et al., 2013; Thomm and Bromme, 2012). Therefore, alongside the perceived credibility of scientific claims we also assessed participants' agreement with these claims.

A strong predictor of how plausible, or credible, a scientific claim is deemed to be are preconceived notions and attitudes about the topic of the claim (Sinatra et al., 2014). When certain outcomes of reasoning are more desirable compared with others, reasoning will be skewed in the direction of these outcomes, which is a process subsumed under the term of motivated reasoning (Kraft et al., 2015; Kunda, 1990; Nauroth et al., 2014).

To account for the effects of a priori attitudes, we assessed attitude as a covariate. Furthermore, we used four different controversial topics in the Facebook posts (potential dangers of homeopathy, health dangers of genetically engineered food, crime rates of refugees in Germany, and potential dangers of early childhood vaccinations) to ensure a certain heterogeneity of opinions to be present in our sample.

What can we expect in regard to how the different kinds of attacks on a scientific claim affect agreement with the user comment, agreement with and credibility of the attacked claim, and perceived trustworthiness of the source of the claim? The expertise and motivations of scientists have been found to affect trustworthiness and credibility (Critchley, 2008; Hendriks et al., 2015; König and Jucks, 2019b; Lombardi et al., 2013; Thon and Jucks, 2017). Furthermore, people use information about research methods to evaluate scientific claims (Sadler, 2004; Wolters et al., 2016) and the certainty or uncertainty of a scientific claim predicts preference, plausibility, and trustworthiness (Jensen, 2008; Lombardi et al., 2013).

However, it is an open question whether claims about the expertise, motivations, research methods, or the thematic complexity would also affect trustworthiness and credibility when they are presented as a user comment, without any further justifications. Therefore, this is an exploratory study with the following research questions:

RQ1. Does the level of agreement with user comments attacking a scientific claim depend on the type of critical user comment?

RQ2. Is the level of agreement with a scientific claim that is attacked by a user comment affected by the type of critical user comment?

RQ3. Is the perceived credibility of a scientific claim that is attacked by a user comment affected by the type of critical user comment?

RQ4. Is the perceived trustworthiness of a source of a scientific claim that is attacked by a user comment affected by the type of critical user comment?

We also want to investigate whether certain types of attacks are more effective in regard to some topics but not others. Domain-specific preferences for explanations of scientific disagreement have already been found with people favoring complexity and research process explanations in regard to a dispute in the domain of biology, but motivations and competence-based explanations in regard to a dispute in the domain of history (Thomm and Bromme, 2016). In this sense, our study also differs from Barnes et al. (2018), since we aim to uncover such possible interactions of topics and different kinds of attacks:

RQ5. Are the effects outlined in RQ1–RQ4 moderated by the topic of the scientific claim?

4. Study 1

Methods

Participants. Of the study's 144 participants, 98 were female. The median age was 22 (M = 23.01 years, standard deviation (SD) = 3.94 years). In all, 84.03% of the participants were students. Participants were recruited through Facebook groups for first-semester students of specific German universities and compensated with a €5 Amazon gift card. We used www.unipark.de as the online software environment for conducting the study. All participants gave informed consent to participate in the study at the beginning of the study, and, after a debriefing at the end of the study, again consented that their responses could be analyzed in the context of a scientific study.

Materials. All materials were in German. Participants saw screenshots of Facebook pages. Each screenshot included one Facebook post with several comments, of which all but the top comments were blurred. There were four different posts made by the fictitious science journalism page "Wissen//Online" (German for Knowledge//Online). Each post was about a scientific study on one of the four controversial topics and included as the headline a scientific claim made by that study. The topics were homeopathy ("Study: Homeopathy holds health risks"), genetic engineering of food produce ("Study: Genetically modified produce not dangerous for humans"), refugee crime statistics ("Study: More crime due to refugees from North Africa"), and childhood vaccinations ("Study: Early childhood vaccinations not dangerous"). The study claims were, while oversimplified, approximately in line with real scientific findings (Ernst, 2002; Pfeiffer et al., 2018; Taylor et al., 2014; Tsatsakis et al., 2017).

User comments. We manipulated the topmost user comment under every post. All user comments were critical of the credibility of the study claim, but included different types of critiques. There were four types of comments: claiming that the researchers who conducted the study were incompetent (incompetence comment, e.g. "A lot of these researchers are not familiar with the teachings of homeopathy. So this study can't be taken seriously."), claiming that the researchers who conducted the study had conflicts of interest (motivations comment, e.g. "A lot of these researchers get research grants from the pharmaceutical industry. So this study can't be taken seriously."), claiming that the methods of the study were not appropriate or that the result of the study was too heavily dependent on the type of methods used (research process comment, e.g. "Homeopathy uses completely different methods and theories compared to classic medicine. So this study can't be taken seriously."), and finally, claiming that the field of study was simply too complex to allow straightforward claims to be made about it (thematic complexity comment, e.g. "Homeopathy is a very broad and complex field. So this study can't be taken seriously.").

Covariates. We measured participants' attitudes toward the topics via questionnaires consisting of five 5-point Likert-type scale items for each topic (for reliability and factor analysis statistics of all multi-item scales, see S1–S3 in the Supplemental Materials). We also measured self-reported knowledge and involvement for each topic with single items. These covariates are not analyzed here.

Dependent variables. For each topic, we measured agreement with the study claim, perceived credibility of the study, and agreement with the user comment with one item each. We measured trustworthiness via the Muenster Epistemic Trustworthiness Inventory (Hendriks et al., 2015). It is designed to measure three dimensions of trustworthiness of a scientist: expertise, benevolence, and integrity. It includes fourteen 7-point semantic differentials in total, six of which comprise the expertise scale and four of which comprise the benevolence and integrity scales, respectively.

Design. The experiment was a mixture of a within-subjects and between-subjects design. Since there were four topics and four different types of critical comments, there were 16 possible topic–argument combinations. Participants were randomly allocated to four topic–argument combinations with the following restrictions: every participant had to see each topic and each argument exactly once throughout the experiment. This led to 26 possible sets of four argument–topic combinations.

Procedure. After filling out an informed consent form, participants filled out questionnaires containing attitude scales, knowledge, and involvement items for each of the four topics. These questionnaires were presented in random order, as were the items within each attitude scale. Then, participants viewed the stimulus screenshots and after viewing each screenshot, they completed a questionnaire including the dependent variables. The blocks of post screenshot and accompanying questionnaires were presented in random order. Following this part of the experiment, participants completed a short demographics questionnaire and a debriefing.

Statistical analyses. For further computations, attitude scores were averages of the five single-item scores per topic. To assess the effect of prior attitudes and the user comment manipulations, we computed mixed-effects models for all dependent variables and performed significance testing on the incremental chi-square scores between models from the least saturated to the most saturated. Our fixed effects were entered in blocks: first attitude, then post topic, then user comment condition, and finally the interaction between post topic and user comment condition. To account for the fact that multiple measurements per person were used, we included a random effect to control for participant ID. Topic and comment type, respectively, were represented by three dummy-coded variables each. All analyses were performed in R (R Core Team, 2018).

Results

Prior attitudes. Including attitude as a predictor (see Table 1) significantly improved the prediction of agreement with the user comment (higher attitude scores were associated with lower agreement with the user comment), agreement with the claim (higher attitude scores were associated with higher agreement with the claim), claim credibility (higher attitude scores were associated with higher credibility of the claim) and all three dimensions of trustworthiness (higher attitude scores were associated with higher expertise, integrity, and benevolence judgments).

Table 1. Model fit statistics for mixed-effects models.

Fixed effects predictor blocks	Study 1				Study 2			
	AIC	BIC	χ^2	p	AIC	BIC	χ^2	p
Comment agreement								
Attitude	1627.01	1644.44	82.22	<.001	1737.83	1755.68	84.19	<.001
Topic	1625.16	1655.66	7.85	.049	1739.77	1771.00	4.07	.254
Comment	1559.62	1603.18	71.55	<.001	1702.58	1747.20	43.18	<.001
Topic × comment	1532.66	1615.42	44.96	<.001	1693.67	1778.44	26.91	.001
Claim agreement								
Attitude	1359.69	1377.11	287.54	<.001	1544.01	1561.85	281.99	<.001
Topic	1338.91	1369.41	26.77	<.001	1544.92	1576.15	5.09	.165
Comment	1342.98	1386.54	1.94	.585	1548.97	1593.59	1.95	.584
Topic × comment	1351.79	1434.56	9.18	.421	1563.70	1648.47	3.27	.953
Credibility								
Attitude	1483.49	1500.92	164.80	<.001	1669.39	1687.23	163.20	<.001
Topic	1454.52	1485.01	34.97	<.001	1666.25	1697.48	9.13	.028
Comment	1455.97	1499.54	4.55	.208	1662.88	1707.49	9.38	.025
Topic × comment	1467.09	1549.86	6.88	.649	1675.06	1759.83	5.82	.758
Expertise								
Attitude	1529.59	1547.02	94.48	<.001	1752.51	1770.36	78.06	<.001
Topic	1525.81	1556.30	9.79	.020	1748.67	1779.90	9.84	.020
Comment	1526.10	1569.67	5.70	.127	1753.71	1798.32	.96	.810
Topic × comment	1535.19	1617.96	8.91	.446	1765.88	1850.65	5.83	.757
Integrity								
Attitude	1583.81	1601.24	140.41	<.001	1722.17	1740.02	91.96	<.001
Topic	1588.21	1618.71	1.60	.659	1720.60	1751.83	7.57	.056
Comment	1582.86	1626.43	11.35	.010	1718.78	1763.39	7.83	.050
Topic × comment	1581.93	1664.69	18.94	.026	1732.53	1817.29	4.25	.894
Benevolence								
Attitude	1620.15	1637.58	140.65	<.001	1821.61	1839.46	108.07	<.001
Topic	1613.85	1644.34	12.31	.006	1806.95	1838.18	20.66	<.001
Comment	1611.44	1655.00	8.41	.038	1800.18	1844.79	12.77	.005
Topic × comment	1615.06	1697.83	14.38	.110	1815.57	1900.34	2.60	.978

AIC: Akaike information criterion; BIC: Bayesian information criterion.
The independent variables topic, comment, and the interaction term topic × comment were entered as blocks into the mixed-effects regression via dummy-coded variables.

Comment effects. We entered the comment manipulation into the model after controlling for attitude and topic. That is, we isolated the effect of the comments on the dependent variables while keeping attitude and other aspects of the topic as covariates. We then tested the successive models (see Table 1) for significant increases in predictive performance. If the model fit the data significantly better than the previous model, we computed post hoc Tukey tests for pairwise comparisons of the comment conditions. The marginal means, which these pairwise comparisons were computed upon, are shown in Figure 1.

Comment agreement significantly depended on the type of comment (RQ1). There were significant differences between the comments for complexity and incompetence ($p < .001$), complexity and motivations ($p = .01$), motivations and incompetence ($p < .001$), and between incompetence
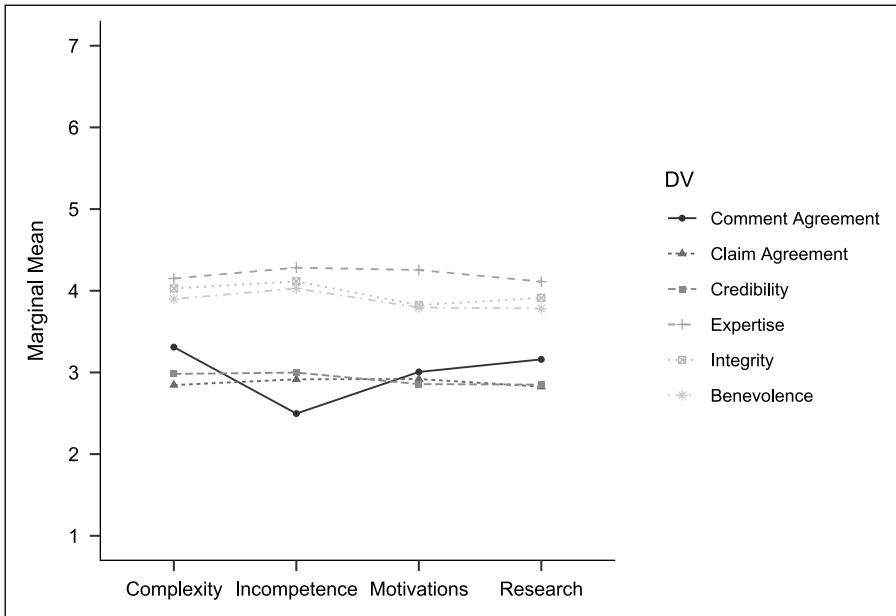
Figure 1. Marginal means for study 1.

and research process comments ($p < .001$). The other two content evaluations, claim agreement (RQ2) and claim credibility (RQ3), did not depend on the comment manipulation.

For source evaluations (RQ4), expertise was not dependent on the type of comment present, but integrity and benevolence were (see Table 1). Integrity judgments were the lowest in the motivations condition (see Figure 1). However, the only significant difference was between the motivations comment and the incompetence comment ($p = .009$). Benevolence judgments were the lowest in the research process condition, the second lowest in the motivations condition, the third lowest in the complexity condition, and the highest in the incompetence condition (see Figure 1). There were no significant differences between any two comments.

Comment and topic interactions. The model fit for comment agreement improved significantly after inclusion of the interaction terms (RQ5, see Table 1). Looking at the comment agreement marginal means for each topic (see Figure 2), it is clear that agreement with some types of comment, particularly complexity and incompetence, is quite similar for all topics, while there is more variance in regard to the motivations and research process. Tukey post hoc tests reveal that for the topics refugee crime and genetically modified organisms (GMOs) the complexity comments are rated significantly higher than the motivations comments (refugee crime: $p < .001$; GMO: $p < .001$), but not for the topics homeopathy and vaccinations. Similarly, the research process comment was rated significantly lower than the complexity comment for the GMOs topic ($p = .009$), but not for the other topics.

The model fit for integrity judgments also improved significantly after inclusion of the interaction terms. As with comment agreement, the topic matters particularly for motivations comments. This is mainly due to how motivations comments affect integrity judgments in regard to the vaccinations topic. While for the three other topics there are no significant differences between any of the comments, for the vaccination topic integrity judgments were significantly lower when a
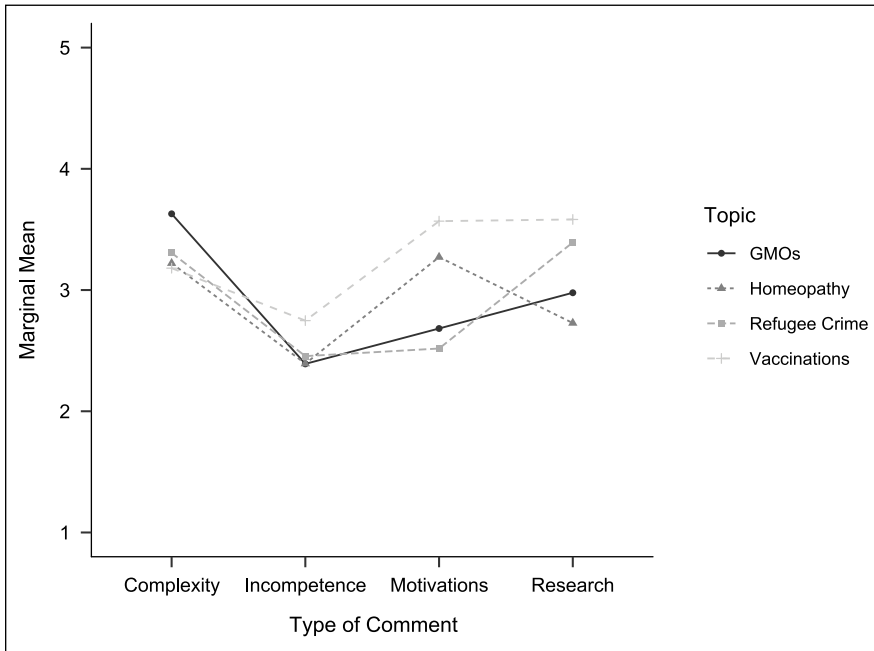
Figure 2. Comment agreement marginal means per topic for study 1.

motivations comment was displayed compared with when a complexity ($p=.008$) or an incompetence ($p<.001$) comment was displayed.

Discussion

The data show that people generally agree with user comments alluding to thematic complexity compared with other types of science-critical user comments. However, complexity did not lower trustworthiness compared with the other comments. While expertise remained stable, regardless of the type of critical comment, integrity and benevolence judgments were dependent on the type of comment present. In particular, comments about the motivation of the scientist lowered perceived integrity compared with comments about the scientists' incompetence. This shows that—after controlling for effects of attitude—the content of user comments determines people's evaluations of those comments and of the source that was attacked by the comments.

Apart from these "topic-agnostic" effects, we also found that participants agreed more with the motivations comment in regard to the health industry-related topics vaccinations and homeopathy. Similarly, accusing scientists of bias affected integrity judgments more strongly than the complexity and incompetence comments in regard to the vaccination topic.

However, even though people did reason about the content of the comments to adjust their evaluations of the comments and the attacked source, they did not adjust their evaluations of credibility accordingly. One reason why this might be the case is that the user comments were perceived as weak evidence for evaluating the credibility of a scientific claim, potentially because people have no information about the source of the comments. People are potentially aware that, due to the interactivity of social media, any person, even non-experts, could comment. To analyze whether the perceived expertise of the source mattered for the effect of the attacks, we conducted a second study in which the attacks were still presented as user comments, but by users who were also scientists.

5. Study 2

Study 1 showed that people agree differently with different kinds of user comments. One other way in which science deniers aim to undermine scientific consensus is by sponsored or fake scientists who argue with their cause in mind (Cho et al., 2011). Do the effects of different types of user comments attacking science depend on the expertise of the commenter? This might be the case since as we said before, compared with the study by Barnes et al. (2018), in our study the attacks were presented as comments and not as facts and, therefore, the trustworthiness of the comment source might be a cue for the veracity of the comment. To explore this question, we manipulated the expertise of the commenter, introducing them as a scientist, in a second study. In large parts, study 2 was a replication of study 1. For information on study 2's methods not covered in the following section, please refer to the "Methods" section of study 1.

Method

Participants. The sample consisted of 160 participants, of which 114 were female. The median age was 23 years ($M = 24.05$ years, $SD = 5.54$ years) and 80% of the participants were students. As in study 1, the participants were recruited through various Facebook groups for first-semester students of specific universities, but different groups were chosen to reduce the possibility of overlap between the two samples. Since all Facebook groups were targeted at first-semester students for a specific year and a specific university, and it is difficult for German students to be enrolled at two different universities concurrently, it is highly unlikely that participants of study 1 also participated in study 2.

Materials. Compared to study 1, we changed two things to make our participants view the user comments as having been posted by experts. First, we introduce the fictitious Facebook page "Wissen//Online" as a page run by scientists for scientists, where only other scientists may comment. Second, we added a "Dr." title to the name of the commenter.

Results

Prior attitude. Including the prior attitude scores into a mixed-effects model significantly improved the prediction quality compared with the null model for comment agreement, claim agreement, claim credibility, expertise, integrity, and benevolence (see Table 1). All coefficients were in the same direction as in study 1. This further confirms the findings from study 1.

Comment effects. The fit statistics of the mixed-effects models predicting the dependent variables in study 2 are given in Table 1. Figure 3 plots the marginal means.

As in study 1, comment agreement significantly depended on the type of comment. Again, the marginal means for complexity were the highest, followed by research process, motivations, and incompetence, in that order. There were significant differences in comment agreement between the comments for complexity and incompetence ($p < .001$), complexity and motivations ($p = .001$), motivations and incompetence ($p = .032$) and between incompetence and research process comments ($p < .001$). As in study 1, claim agreement was not affected by the comment manipulation. However, in study 2 the prediction of credibility scores improved significantly when the comment manipulations were entered into the mixed-effects model. The marginal credibility means were the lowest in the complexity condition, followed by motivations, research process, and incompetence,
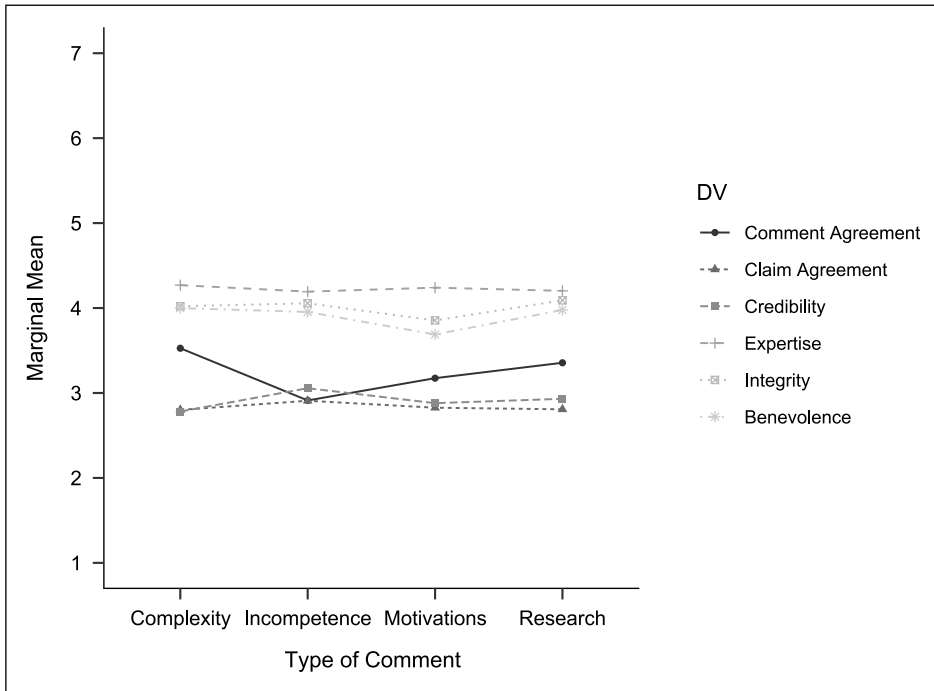
Figure 3. Marginal means for study 2.

in that order. However, only the difference between the complexity and incompetence conditions was statistically significant ($p = .014$).

As in study 1, expertise was not dependent on the type of comment present, but integrity and benevolence were (see Table 1). Integrity judgments were the lowest in the motivations condition (see Figure 3). The only significant difference was between the motivations comment and the research process comment condition ($p = .047$). Benevolence judgments were the lowest in the motivations condition, second lowest in the incompetence condition, third lowest in the research process condition, and highest in the complexity condition (see Figure 3). There were significant differences between the motivations condition and the research process condition ($p = .018$) and the complexity condition ($p = .001$).

Comment and topic interactions. In regard to comment agreement, introducing the interaction terms into the mixed-effects model again significantly increased the model fit (see Table 1). Looking at the comment agreement marginal means per comment and topic (see Figure 4), we again see that agreement with the complexity comment is rather high and agreement with the incompetence comment is rather low. Again, agreement with the motivations comment is more dependent on the topic. For the topics GMOs and refugee crime, agreement with the motivations comment is significantly lower than with the complexity comment (GMOs: $p < .001$; refugee crime: $p < .001$), while for the topics vaccination and homeopathy agreement with the complexity and the motivations comments do not differ significantly. In contrast to study 1, agreement with research process comments is less dependent on the topic and rather high, not being significantly different from agreement with the complexity comment for any topic. Further contrasting study 1, there were no
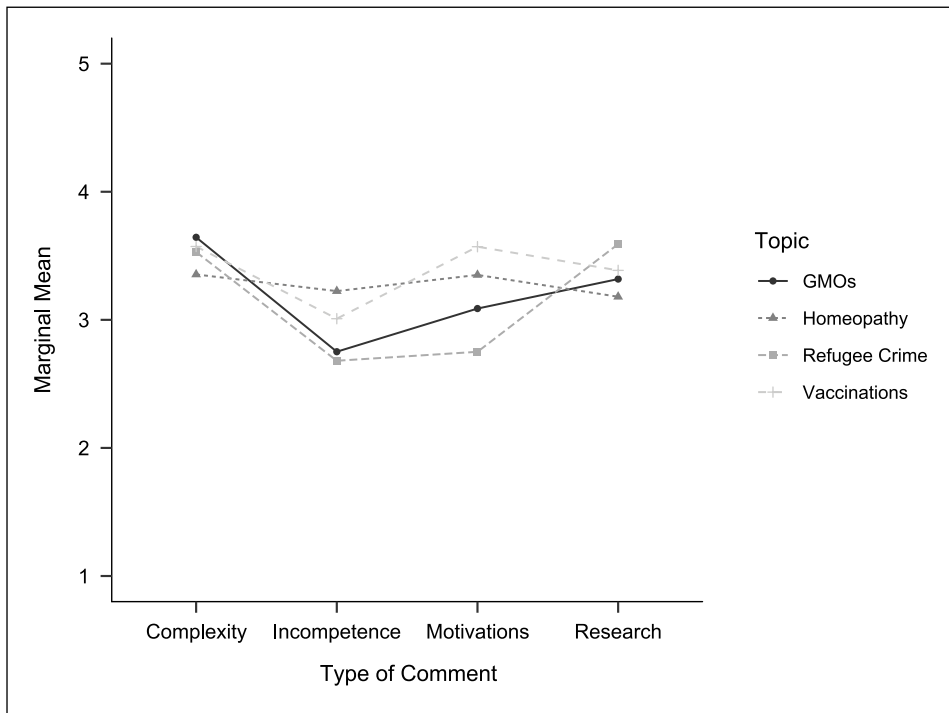
Figure 4. Comment agreement marginal means per topic for study 2.

significant improvements in model fit for integrity judgments after inclusion of the comment and topic interaction terms.

Discussion

We replicated two findings from study 1: first, participants agree more with critical user comments targeting the inherent complexity of scientific findings compared with other types of critical user comments, and, second, integrity judgments are most affected by critical user comments targeting the motivations of scientists. However, in contrast to study 1, the credibility of a scientific claim on social media also depended on the type of user comment present. Specifically, user comments targeting thematic complexity have a stronger negative impact on credibility judgments compared to user comments claiming scientists' incompetence.

6. General discussion

Prior attitudes

Prior attitude was the strongest predictor of comment agreement, claim agreement, claim credibility, and source trustworthiness. This is in line with the vast previous research on motivated reasoning (Kunda, 1990; Nauroth et al., 2014; Sinatra et al., 2014). This was also not surprising for another reason: since there was only limited information presented to the participants, they had to rely more on prior attitudes to make up their mind. This, however, also reflects the actual situation on social media: people sometimes only read the headlines and sites are often set up in a way that

only the most relevant or even just the most recent comment is shown. We wanted to know how reasoning would be affected by user comment content in precisely this context.

Effects of comments

First, we found that people favor some kinds of attacks over others (RQ1); across both experiments they agreed most with comments using thematic complexity as an anti-science argument, second most with research process arguments, third most with motivations arguments, and least with comments attacking the expertise of researchers. This finding resonates with previous findings where people generally preferred thematic complexity as an explanation of scientific conflict (Dieckmann et al., 2015; Thomm and Bromme, 2016; Thomm et al., 2015).

In study 1, credibility did not depend on the type of comment present (RQ3). Since there was no condition with a neutral or even positive comment, we cannot say that the comments did not affect credibility at all, just that there was no significant variation between comment types. In study 2, where the commenter was introduced as an expert, there were significant differences between the comment types. In particular, we find that the complexity comment significantly lowers credibility compared with the expertise comment.

Furthermore, trustworthiness judgments, particularly in regard to the integrity and benevolence of the source of the scientific claim were also dependent on the content of the user comment (RQ4). Particularly, attacks on the motivations of the scientists led to lower integrity judgments in both studies and lower benevolence judgments in study 2.

It therefore seems likely that people attend to the content of user comments and thus, models of online credibility evaluations (e.g. Choi and Stvilia, 2015; Metzger and Flanagin, 2015) should account for the content of user comments instead of just the ratio of positive and negative comments. However, this effect also depends on the source of the comment as the different comments did not differ in their effect on credibility judgments when coming from laypeople, but did when they were coming from experts. If people were to simply agree *more* with expert comments regardless of content and thus uniformly rate the credibility of the claim lower, we would have not observed this effect. Therefore, one could interpret this finding as the expertise of the commenter moderating the effect of different types of negative user comments on credibility and making the differences between the comments in regard to credibility more pronounced.

Furthermore, it is important to note the relative effectiveness of comments addressing the research process and the thematic complexity, as both of these are inherent to science. Science, as it becomes more developed, is inherently complex and the research process is highly institutionalized, ideally being so far removed from the person of the researcher that other researchers could reproduce a given study with enough information. These inherent qualities of science can also be seen as a reason why science is generally trusted.

These concepts are even taught in the context of science education, particularly in regard to socio-scientific issues. However, addressing these specifically to attack a scientific claim could also act as a cue that the commenting source has some knowledge about science, and is therefore trustworthy in the matter. Indeed, we found that the expertise of the commenter determines whether the comment's content is used as a cue to form a credibility judgment. Problematically, evoking the illusion of trustworthiness by using science-based attacks and corporate-sponsored experts and think tanks are a part of strategic attempts at spreading misinformation, particularly in regard to controversial topics (Cho et al., 2011; Garrett, 2017).

Such strategies have been linked to the creation of a post-truth epistemic climate in which feelings matter more than facts (Lewandowsky, Cook and Ecker, 2017; McIntyre, 2018). More aptly though, in such a climate facts still matter but there is room to be selective about which facts do, since even those arguing against the scientific consensus have become adept at signaling epistemic authority.

It is worrying that they seemingly do not have to go to great lengths to lessen the credibility of scientific claims. A supposed expert's user comment stating that the complexity of a topic is too great to make any concrete claims about it was effective in this study, and, even more worryingly, this attack was uniformly agreeable across four very different controversial topics. Potentially, scientists and science communicators should be wary when such attacks are present in their own user comment sections. Since the manual moderation of user comment sections, which are typically open to anyone, is laborious, it could be helpful to focus on such critical user comments, which signal expertise either by claiming to have certain credentials or by using science-based attacks.

Furthermore, it is important to consider the topic of a post in relation to the user comment's content. We found that comment preference depended on the topic: while people view complexity as a strong argument and accusations of incompetence as a weak argument regardless of topic, their relative preference for research process and motivations comments depend more on the topic. They agreed much more strongly when the commenter claimed that the researchers were biased in regard to the topics homeopathy and vaccinations compared with the topics genetically engineered foods and refugee crime statistics.

For the vaccinations topic specifically, attacking the motivations was also more effective at lowering perceived integrity in study 1. In comparison, accusations of political motivations in regard to the refugee crime topic were less effective. This may be due to recent public debates highlighting the impact of "economically big players" like the pharmaceutical industry on the German healthcare system. These stakeholders have become visible "gorillas in the mist" (Lewandowsky et al., 2017), even compared with the stakeholders of GMO activities. Science communicators operating on social media should therefore also be aware of this when sharing information relating to such topics. Potentially, information about funding sources should be communicated proactively for topics where the impact of funding sources on the scientific process has been discussed critically. Particularly on social media, where many people first see a post linking to an external article, alongside potentially critical user comments, it could be considered to move funding information to the short description that is contained in the post and immediately visible.

Categories of scientific attacks

This research was also conducted to investigate how feasible the explanations for scientific conflict are as a framework for categorizing attacks on science, and it is therefore important to integrate our results with other research surrounding anti-science arguments. Barnes et al. (2018) found that arguments based on misconduct by a scientist affected the agreement with claims made by that scientist compared with, for instance, claims about that scientist's expertise. In our study, we did not find any effect on agreement with a study claim (RQ2). This could be because we presented our arguments as the opinion of another source, the commenter, and not as fact. In addition, our user comments were vague and imprecise: they were just barely stating the argument and without any further details. Considering this, it is remarkable that there were any differences in how agreeable the comments were, which indicates that such a classification of science-critical user comments is helpful for further empirical research.

Limitations

Our study has some limitations. First, the sample consisted mainly of students. Potentially, students differ in meaningful ways from the rest of the population in regard to science reception. To avoid including true experts into our sample, we posted the call for participants in Facebook groups targeted at first-semester students. Nevertheless, the relatively young age of our sample and their

interest in pursuing academic studies could have an effect on their science attitudes. This could especially be the case in regard to their preference for the complexity argument, since Dieckmann et al. (2015) found that people with higher cognitive ability prefer complexity as an explanation for scientific conflict. Future research on this subject would therefore benefit from a more diverse sample.

As said before, we included no condition in which user comments were absent or irrelevant. However, we did not attempt to assess whether critical user comments would affect credibility and trustworthiness at all, but whether differences in content between different kinds of critical user comments would be noticed and influence the reasoning of participants. Another limitation is that there might be a difference in how people regard science-related content on their own Facebook feed compared with the screenshots given to participants in the experiment. However, the presentation of anti-science arguments within even just fabricated Facebook screenshots should hold more external validity than a comparable experiment presenting the argument manipulation without any real-world context.

Future directions and implications

This study had two main areas of interest: the effect of different kinds of science-critical user comments on social media and how explanations of scientific conflict could be turned into anti-science arguments. It has implications for both fields. First, we found that the content of critical user comments has an effect on the credibility of scientific claims if the commenters were introduced as experts. This raises obvious questions about when and how people come to reason about the sources of user comments when they encounter information about science online. In one way, the readers share a common view with the commenters as they are both part of the audience. For controversial topics, however, audience members might (if they are aware of the controversy) become vigilant when commenters engender extreme praise or vitriol. In our study, information about the commenters was controlled. However, many people have accessible profiles on social media. Future research could therefore investigate the conditions under which people gather information about a commenting source, particularly in regard to online debates about controversial scientific issues.

In addition, one interesting finding in regard to the four categories of scientific attacks was that some were more agreeable or more effective in regard to certain topics. We hypothesized that attacks on scientists' motivations would be more familiar in regard to topics where big economic stakeholders are widely known. To follow on from these thoughts, a classification of different types of *topics* (e.g. in regard to political or economic influences) would be needed, which could then be cross-matched with the categories of scientific attacks established in this article.

In sum, even though our work was exploratory in nature, we think it is justified to conclude that the content of user comments can affect credibility evaluations of science information on social media, and that science communicators should be wary of the comparative effectiveness of the complexity argument as an attack on science claims.

ORCID iDs

Lukas Gierth [iD] https://orcid.org/0000-0001-5550-9304
Rainer Bromme [iD] https://orcid.org/0000-0002-6123-7269

Supplemental material

Supplemental material for this article is available online.

References

Anderson AA, Brossard D, Scheufele DA, Xenos MA and Ladwig P (2014) The "nasty effect": Online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication* 19(3): 373–387.

Barnes RM, Johnston HM, MacKenzie N, Tobin SJ and Taglang CM (2018) The effect of ad hominem attacks on the evaluation of claims promoted by scientists. *PLoS ONE* 13(1): e0192025.

Bromme R, Scharrer L, Stadtler M, Hömberg J and Torspecken R (2015) Is it believable when it's scientific? How scientific discourse style influences laypeople's resolution of conflicts. *Journal of Research in Science Teaching* 52(1): 36–57.

Brossard D and Scheufele DA (2013) Science, new media, and the public. *Science* 339(6115): 40–41.

Cho CH, Martens ML, Kim H and Rodrigue M (2011) Astroturfing global warming: It isn't always greener on the other side of the fence. *Journal of Business Ethics* 104(4): 571–587.

Choi W and Stvilia B (2015) Web credibility assessment: Conceptualization, operationalization, variability, and models. *Journal of the Association for Information Science and Technology* 66(12): 2399–2414.

Critchley CR (2008) Public opinion and trust in scientists: The role of the research context, and the perceived motivation of stem cell researchers. *Public Understanding of Science* 17(3): 309–327.

Dieckmann NF and Johnson BB (2019) Why do scientists disagree? Explaining and improving measures of the perceived causes of scientific disputes. *PLoS ONE* 14(2): e0211269.

Dieckmann NF, Johnson BB, Gregory R, Mayorga M, Han PKJ and Slovic P (2015) Public perceptions of expert disagreement: Bias and incompetence or a complex and random world? *Public Understanding of Science* 26(3): 325–338.

Elsasser SW and Dunlap RE (2013) Leading voices in the denier choir: Conservative columnists' dismissal of global warming and denigration of climate science. *American Behavioral Scientist* 57(6): 754–776.

Ernst E (2002) A systematic review of systematic reviews of homeopathy. *British Journal of Clinical Pharmacology* 54(6): 577–582.

Evrony A and Caplan A (2017) The overlooked dangers of anti-vaccination groups' social media presence. *Human Vaccines & Immunotherapeutics* 13(6): 1–2.

Garrett RK (2017) The "echo chamber" distraction: Disinformation campaigns are the problem, not audience fragmentation. *Journal of Applied Research in Memory and Cognition* 6(4): 370–376.

Hendriks F, Kienhues D and Bromme R (2016a) Disclose your flaws! *Studies in Communication Sciences* 16(2): 124–131.

Hendriks F, Kienhues D and Bromme R (2016b) Evoking vigilance: Would you (dis)Trust a scientist who discusses ethical implications of research in a science blog? *Public Understanding of Science* 25(8): 992–1008.

Hendriks F, Kienhues D and Bromme R (2015) Measuring laypeople's trust in experts in a digital age. *PLoS ONE* 10(10): e0139309.

Jennings FJ and Russell FM (2019) Civility, credibility, and health information: The impact of uncivil comments and source credibility on attitudes about vaccines. *Public Understanding of Science* 28(4): 417–432.

Jensen JD (2008) Scientific uncertainty in news coverage of cancer research: Effects of hedging on scientists' and journalists' credibility. *Human Communication Research* 34(3): 347–369.

Johnson BB (2017) "Counting votes" in public responses to scientific disputes. *Public Understanding of Science* 27: 594–610.

Kajanne A and Pirttilä-Backman A-M (1999) Laypeople's viewpoints about the reasons for expert controversy regarding food additives. *Public Understanding of Science* 8(4): 303–315.

König L and Jucks R (2019a) Hot topics in science communication: Aggressive language decreases trustworthiness and credibility in scientific debates. *Public Understanding of Science* 28(4): 401–416.

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): 1.

Kraft PW, Lodge M and Taber CS (2015) Why people "don't trust the evidence." *The ANNALS of the American Academy of Political and Social Science* 658(1): 121–133.

Kunda Z (1990) The case for motivated reasoning. *Psychological Bulletin* 108(3): 480–498.

Lewandowsky S, Cook J and Ecker UKH (2017) Letting the gorilla emerge from the mist: Getting past post-truth. *Journal of Applied Research in Memory and Cognition* 6(4): 418–424. Available at: https://DOI.org/10.1016/j.jarmac.2017.11.002

Lombardi D, Seyranian V and Sinatra GM (2013) Source effects and plausibility judgments when reading about climate change. *Discourse Processes* 51(1–2): 75–92.

Lörcher I and Taddicken M (2017) Discussing climate change online. Topics and perceptions in online climate change communication in different online public arenas. *JCOM* 16(2): 1–21.

McIntyre LC (2018) *Post-truth.the MIT Press Essential Knowledge Series*. Cambridge, MA: MIT Press.

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

Nauroth P, Gollwitzer M, Bender J and Rothmund T (2014) Gamers against science: The case of the violent video games debate. *European Journal of Social Psychology* 44(2): 104–116.

Pfeiffer C, Baier D and Kliem S (2018) On the development of violence in Germany. Available at: https://www.zhaw.ch/storage/shared/sozialearbeit/News/gutachten-entwicklung-gewalt-deutschland.pdf

R Core Team (2018) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Sadler TD (2004) Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching* 41(5): 513–536.

Scharrer L, Britt MA, Stadtler M and Bromme R (2013) Easy to understand but difficult to decide: Information comprehensibility and controversiality affect laypeople's science-based decisions. *Discourse Processes* 50(6): 361–387.

Sinatra GM, Kienhues D and Hofer BK (2014) Addressing challenges to public understanding of science: Epistemic cognition, motivated reasoning, and conceptual change. *Educational Psychologist* 49(2): 123–138.

Skeptical Science (2019) Climate myths sorted by taxonomy. Available at: https://skepticalscience.com/argument.php?f=taxonomy (accessed 20 May 2019).

Su LY-F, Akin H, Brossard D, Scheufele DA and Xenos MA (2015) Science news consumption patterns and their implications for public understanding of science. *Journalism & Mass Communication Quarterly* 92(3): 597–616.

Taylor LE, Swerdfeger AL and Eslick GD (2014) Vaccines are not associated with autism: An evidence-based meta-analysis of case-control and cohort studies. *Vaccine* 32(29): 3623–3629.

Thomm E and Bromme R (2012) It should at least seem scientific! Textual features of "scientificness" and their impact on lay assessments of online information. *Science Education* 96(2): 187–211.

Thomm E and Bromme R (2016) How source information shapes lay interpretations of science conflicts. *Reading and Writing* 29(8): 1629–1652.

Thomm E, Barzilai S and Bromme R (2017) Why do experts disagree? *Learning and Instruction* 52: 15–26.

Thomm E, Hentschke J and Bromme R (2015) The Explaining Conflicting Scientific Claims (ECSC) Questionnaire. *Learning and Individual Differences* 37: 139–152.

Thon FM and Jucks R (2017) Believing in expertise: How authors' credentials and language use influence the credibility of online health information. *Health Communication* 32: 828–836.

Tsatsakis AM, Nawaz MA, Tutelyan VA, Golokhvast KS, Kalantzi O, Chung DH, et al. (2017) Impact on environment, ecosystem, diversity and health from culturing and using GMOs as feed and food. *Food and Chemical Toxicology: An International Journal Published for the British Industrial Biological Research Association* 107(Part A): 108–121.

Van Noorden R (2014) Online collaboration: Scientists and the social network. *Nature* 512(7513): 126–129.

Winter S and Krämer NC (2016) Who's right: The author or the audience? *Communications* 41(3): 339–360.

Wolters EA, Steel BS, Lach D and Kloepfer D (2016) What is the best available science? A comparison of marine scientists, managers, and interest groups in the United States. *Ocean & Coastal Management* 122: 95–102.

Yuan S, Ma W and Besley JC (2019) Should scientists talk about GMOs nicely? Exploring the effects of communication styles, source expertise, and preexisting attitude. *Science Communication* 41: 267–290.

Author biographies

Lukas Gierth is a cognitive psychologist working at the University of Münster. His research focuses on epistemic trust, epistemic vigilance, and informal science communication on social media.

Rainer Bromme is a Senior Professor of Educational Psychology at the University of Münster. His research focuses on formal and informal learning contexts, science communication, and the public's trust in science.