Informal Logic

Source Related Argumentation Found in Science Websites
A quantitative study

Ralph Barnes, Zoë Neumann and Samuel Draznin-Nagy

Volume 40, Number 3, 2020

URI: https://id.erudit.org/iderudit/1071502ar
DOI: https://doi.org/10.22329/il.v40i30.5984

See table of contents

Publisher(s)

Informal Logic

ISSN

0824-2577 (print)
2293-734X (digital)

Explore this journal

Cite this article

Barnes, R., Neumann, Z. & Draznin-Nagy, S. (2020). Source Related Argumentation Found in Science Websites: A quantitative study. *Informal Logic*, *40*(3), 443–473. https://doi.org/10.22329/il.v40i30.5984

Article abstract

In this paper, we consider the way that web documents seeking to persuade readers of certain science claims provide information about the sources of the arguments. Our quantitative analysis reveals that web documents in our sample include hundreds of examples in which the reader is provided information regarding the trustworthiness (or lack thereof) of sources. The web documents also contain a large number of examples in which the reader is provided with information about how many individuals hold a particular belief. We discuss ad hominem, ad verecundiam, and ad populum arguments, and the way that the examples found in our sample of documents are related to these argumentation schemes.

Source Related Argumentation Found in Science Websites: A Quantitative Study

RALPH BARNES
Psychology Department
Montana State University
Culbertson Hall, 100, Bozeman, MT, U.S.
ralph.barnes@montana.edu

ZOË NEUMANN
Psychology Department
Montana State University
Culbertson Hall, 100, Bozeman, MT, U.S.
zoenoel10@gmail.com

SAMUEL DRAZNIN-NAGY
Psychology Department
Montana State University
Culbertson Hall, 100, Bozeman, MT, U.S.
samuel.drazninnagy@g-mail.com

Abstract: In this paper, we consider the way that web documents seeking to persuade readers of certain science claims provide information about the sources of the arguments. Our quantitative analysis reveals that web documents in our sample include hundreds of examples in which the reader is provided information regarding the trustworthiness (or lack thereof) of sources. The web documents also contain a large number of examples in which the reader is provided with information about how many individuals hold a particular belief. We discuss ad hominem, ad verecundiam, and ad populum arguments, and the way that the examples found in our sample of documents are related to these argumentation schemes.

Résumé: Dans cet article, nous examinons la manière dont les documents Web qui visent à persuader les lecteurs de certaines affirmations scientifiques fournissent des informations sur les sources des arguments. Notre analyse quantitative révèle que les documents Web de notre échantillon comprennent des centaines d'exemples dans lesquels le lecteur reçoit des informations sur la fiabilité (ou l'absence de fiabilité) des sources. Les documents Web contiennent également un grand nombre d'exemples dans lesquels le lecteur est informé du nombre d'individus qui ont une croyance particulière. Nous discutons des arguments ad hominem, ad verecundiam et ad populum, et de la manière dont les exemples trouvés dans notre échantillon de documents sont liés à ces schémas d'argumentation.

Keywords: ad hominem, ad populum, ad verecundiam, fallacy, AIDS, climate change, GMO safety, vaccine safety

1. Introduction

While *ad hominem* attacks are not appropriate in some situations (e.g., claiming that a chemist cannot be trusted because she is a woman), many researchers agree that there are times when *ad hominem* attacks are relevant and reasonable (Aberdein 2014; Bondy 2015; Bowel and Kingsbury 2013; Walton 1998; Woods 2007). An example of an acceptable *ad hominem* attack might be the claim that employees of cigarette companies are likely to paint an inaccurate picture regarding the dangers of smoking. Similarly, if a politician has lied about topic X several times in the past, one could argue that the current claim that the politician has made about topic X should be viewed with a great deal of skepticism. Johnson (2009) has argued that, due to cognitive limitations, it is sometimes appropriate for individuals to appeal to character when deciding the truth of an issue.

Yap (2013) has argued that, even if an individual recognizes an *ad hominem* attack as unreasonable, the attack may still influence that individual's confidence in the claims made by the attacked source. Regardless of the issue of the reasonableness or appropriateness of such attacks, it is clear that they can be influential. Many researchers have found that trust in sources is at least as important as the scientific facts when it comes to influencing the public on issues that involve both science and public policy (Slovic 1993; Slovic, Flynn, and Layman 1991). Barnes, Johnson, MacKenzie, Tobin, and Taglang (2018) have shown that certain *ad hominem* attacks against scientists can cause people to reduce their trust in claims made by those scientists. In fact, they found that certain types of *ad hominem* attacks (i.e., accusations of scientific misconduct and conflicts of interest) can influence opinion as much as direct attacks on the empirical data presented by the researchers. Cohen's *d* values (a measure of effect size) for both type of attacks were in the moderate to moderate-large range (i.e., .39 to .68). Additionally, some researchers have found that *ad hominem* attacks in the realm of politics can be quite effective (Kaid and Boydston 2009; Sabato 1981) while others (e.g., Lau, Sigleman and Rovner, 2007) have found only a very small effect size (e.g., adjusted Cohen's *d* of .04) of negative attacks.

 Informal Logic, Vol. 40, No. 3 (2020), pp. 443–473

Pornpitakpan (2004) reviewed dozens of studies on source credibility and persuasion and found empirical evidence that qualities of the source of an argument can influence individuals. Pornpitakpan found that qualities of the source (e.g., attractiveness, reputation, etc.) were sometimes more influential and sometimes less influential than other factors (such as quality of arguments). The relationship between source characteristics and other variables was complex. Wilson and Sherrell (1993) performed a meta-analysis of the effect of sources of persuasive messages and found that source effects were consistent but small in magnitude (i.e., accounting for only 4.5% of the explained variance). It should be noted, however, that the Wilson and Sherrell study did not look specifically at *ad hominem* arguments. Many of the studies they reviewed manipulated source characteristics by, for instance, presenting some subjects with a male source and other subjects with a female source, or presenting some subjects with a physically attractive source and other subjects with an unattractive source.
Appeals to popular opinion (aka *ad populum* arguments) may not be appropriate in some situations. For instance, in the 15th century one could make the argument 'because most people believe in geocentrism, therefore geocentrism is likely to be true.' Another example would be an American politician claiming in 1800 that slavery is morally acceptable because most US citizens believed it to be so. However, appeals based on the number of people that believe a particular claim can be influential (Latané 1981; Moscovici, Mugny and van Avermaet 1985; Stangor, Sechrist, and Jost 2001; Zitek and Hebl 2007). For instance, Lewandowsky, Gignac, and Vaughan (2012) showed that perceived consensus regarding global warming had a causal role in terms of the acceptance of anthropogenic global warming. Both empirical data and statistical theory indicate that, as the number of experts who agree on some belief increases, the likelihood that the belief is true increases (Yaniv 2004). It seems that, in some cases, individuals may be well served by being influenced by popular belief.

There is an extensive literature on attack ads in politics (Gronbeck 1992; Jamieson 1992). In addition, Habernal, Wachsmuth, Gurevych, and Stein (2018) documented the amount of *ad hominem* attacks used in a social media website, and Sahlane

(2012) documented the presence of *ad hominem* attacks in newspaper editorials. There is, however, no data in the peer reviewed literature on the frequency and type of attacks on the sources of arguments in examples of science communication. The current study fills that gap in the literature. One might argue that our choice of source documents will be unlikely to provide us with many examples of personal attacks. The Mertonian norm of disinterestedness captures the idea that one value central to the institution of science is that scientists ought to act for the benefit of the scientific enterprise rather than for their own selfish interest. Since discovering the truth is an important goal of science, and insulting those making claims is not central to science, personal attacks are not consistent with disinterestedness. Therefore, to the degree that the Mertonian norm of disinterestedness captures the values of scientists, one would expect that scientists would rarely engage in personal attacks. While it is true that personal attacks are rare in the peer reviewed scientific literature, the source documents considered in the current study are not peer reviewed articles. It is also true that scientists sometimes engage in personal attacks when they are communicating outside the peer-reviewed literature (Souder and Qureshi 2012).

Ad hominem and *ad populum* attacks are of interest to those working in the fields of argumentation and informal logic. Some researchers in those fields concern themselves with questions such as ‘when is an *ad hominem* or *ad populum* argument reasonable?’ and ‘what is the best taxonomy of *ad hominem* argumentation schemes?’ We hope that some researchers wrestling with these questions will find information on the frequency of different forms of argument as they appear in the real world to be useful. For example, researchers may want to dedicate more time and resources to *ad hominem* arguments that are extremely common and dedicate relatively little time and resources to types of *ad hominem* arguments that are quite rare. Barnes and Church (2013) and Barnes, Church, and Draznin-Nagy (2017) have found that, in certain ways, the frequency and type of arguments vary as a function of whether-or-not the communicator advocates a position that is consistent with mainstream scientific thought. Those two publications, however, considered only one scientific topic (evolution), and neither of those publications tackled the issues of *ad hominem* or *ad populum* arguments. It is an open question whether the rhetorical differences described in those two studies are indicative of a larger pattern or are limited to the debate regarding origins.

While other researchers have referred to the *ad hominem*, *ad verecundiam*, and *ad populum* argumentation schemes, we prefer to frame the issue somewhat differently. From our perspective, one can make a statement about the source of data or claim (using Toulmin's 1958 definitions of data and claim) that might indicate that the source is perfectly credible/trustworthy, not at all credible/trustworthy, or having any degree of credibility/trustworthiness between those two extremes. We use the term 'appeal to source quality' to refer to any attempt to describe a person/people/thing responsible for data or a claim as credible/trustworthy or not credible/untrustworthy. So rather than treating *ad verecundiam* and *ad hominem* as two distinct argumentation schemes, we treat them as two examples that are part of a single dimension. On one end of the dimension, appeal to source quality could take the form, 'claim X is likely false because the person making the claim has bad qualities.' On the other end of the dimension, appeal to source quality could take the form, 'claim X is likely true because the person making the claim has good qualities.' Additionally, one can make a statement about how many individuals endorse a particular data or claim. One could claim that everyone believes X, no one believes X, or any value in between. We refer to statements of this type as appeals to source quantity. Like appeals to source quality, we feel that appeals to source quantity should also be seen as a dimension that involves a continuum (e.g., few/most/all people agree that X is true).

While many taxonomies of argumentation schemes have been proposed over the years, we feel that our simple taxonomy is best suited to our current purposes. The taxonomy we propose is partly informed by research in the field of psychology. For instance, Chaiken's heuristic-systematic model (Chaiken 1980; 1987) distinguishes between two broad classes of influence on reasoning: message characteristics and source characteristics (which can be either positive or negative). In psychological jargon, the term 'source' refers to an individual or individuals who are responsible for the origin of the claim and/or data of a persuasive argument. There is a wealth of psychological research that indicates that individuals may be influenced by thoughts and feelings about the source(s) of an argument. For instance, characteristics of a source (e.g., moral character, bias, trustworthiness, attractiveness, etc.) may impact attitudes towards a claim (Briñol and Petty 2009).

We may deviate from some researchers in our use of the term source. For us, a source answers the question, 'where did the data or claim come from?' or 'who/what do we attribute the data or claim to?' A source could be a person (e.g., Dr. Smith) or a group of people (e.g., the Intergovernmental Panel on Climate Change (IPCC), the American Medical Association (AMA), etc.). Groups such as the IPCC and AMA sometimes release policy statements (e.g., 'human forced climate change is real,' 'vaccines do not cause autism,' etc.), and in that capacity they are sources. We hope that that labeling such groups as sources will not be controversial. In some cases, authors describe entities such as journals to be sources, in the sense that journals produce data and claims, and communicators sometimes comment on the credibility/trustworthiness of specific journals. The credibility/trustworthiness of a given journal is due to the people employed by the journal (not the office furniture owned by the journal). For instance, a communicator might claim that a particular predatory journal is less trustworthy than a non-predatory journal with a high impact factor. Finally, the most controversial things that we will consider as sources are tests and measurements. In the peer reviewed scientific literature, as well in less formal examples of science communication, individuals have been known to comment on the trustworthiness and credibility of various tests and measurements. In many scientific fields, the accuracy, precision, error rate, etc. of various tests and measures are of central importance and are a common topic of discussion. For instance, psychologists might discuss the credibility/trustworthiness of various measures of intelligence (Neisser, et al. 1996), while epidemiologists might discuss the trustworthiness and validity of memory-based dietary assessment methods (M-BMs) (Barnes 2019). A patient suffering from a knee injury might get both an X-ray and an MRI. The doctor or patient might feel that one of those sources of data is more trustworthy than the other. If a thing can be an answer to the question, 'who/what do we attribute the data or claim to?' then we consider that thing to be a source. For that reason, we consider tests and measures to be sources.

Barnes and Church (2013) were interested in the rhetoric found in science-related web documents that were A) argumentative in nature and B) focused on the topic of origins (e.g., creationism and evolution). To that end they used a series of search terms to locate the most popular web documents that argued for the truth of either creationism or evolution. We are also interested in the rhetoric employed in science-related web documents. However, our concern is not with origins, but with a particular set of four controversial science topics (i.e., AIDS, climate change, GMO safety, vaccine safety). Because we wanted to find web documents that promoted positions both consistent with and opposed to the scientific mainstream, we were limited in the topics we could choose. For instance, if we choose the scientific claim that lead in drinking water is harmful, it would be unlikely that we could find many websites promoting the non-mainstream view (i.e., that lead in drinking water is not harmful). The authors of the current study settled on AIDS denialism, climate change, GMO safety, and vaccine safety as the four topics most likely to have large numbers of web documents making arguments for and against the scientific mainstream. The primary goal of this descriptive study is to measure and report the frequency of source related arguments appearing in web documents that deal with any of those four controversial science topics. The secondary goal of this study is to report the effect size for comparisons of each source related argument as they appear in our two different categories of documents (i.e., those consistent with and those inconsistent with the scientific mainstream). This study is exploratory, and therefore we do not conduct any hypothesis testing nor report any *p* values.

Method

In order to locate potential source documents, we searched the internet using search terms similar to those used by Barnes and Church (2013). Our search terms were tailored to the four topics we wished to explore: AIDS, climate change, GMO safety, and vaccine safety. For each search term tailored to locate a website promoting a position consistent with the mainstream scientific position on a topic, we also include a matching search term tailored to locate a website promoting a position inconsistent with the mainstream scientific position on that topic. A complete list of the search terms we employed have been uploaded to an online depository hosted by openICPSR (Barnes, Neumann, and Draznin-Nagy 2020). The total number of documents as a function of topic and position (i.e., consistent or inconsistent with mainstream science) can be found in Table 1. Because we wished to obtain web documents that were likely to be frequently accessed by those browsing the web, we limited our search to the top 100 hits for each search phrase. In addition, we excluded audio and video clips, documents over 20,000 words in length, peer-reviewed journal articles, books, book chapters, and peer reviewed journal articles. The 69 web sites we ultimately selected included 12 related to AIDS, 28 related to climate change, 15 related to GMOs, and 15 related to vaccines. To guarantee that the text would not change over the course of the coding process, the content of each of the web documents was copied and pasted into word processing documents. The complete text of all the web documents in our sample may be obtained by emailing the first author.

In some scientific fields, a common practice is to take a random sample and then perform a statistical test of the sample to see if it is likely that the sample came from a specific population. We have not used that practice, however, for to get a random sample of all websites related to the four topics of interest would first require that we evaluate 100,000+ websites and then categorize them as either meeting, or failing to meet, our inclusion criterion. The Google search algorithm we used was PageRank (Wikipedia 2020), which counts the number and quality of links to a page as a means to determine the page's importance. Seventy-five percent of the users of internet search engines will not look beyond the first page of results (Agrawal 2017). However, for each of our searches, we combed through the first ten pages of results. For these reasons, while our 69 source documents were not randomly selected, the documents we selected were A) judged by PageRank to be of high importance and B) those that likely received the overwhelming majority of views (among those searching for the four topics included in the present study). Because we were concerned with the types of arguments that people are commonly exposed to, the sample we used is more appropriate than a random sample.

Table 1. Frequency and word count of all source web documents

Topic	Document position	Number of documents	Word count
AIDS	Consistent	5	6530
	Inconsistent	6	9881
Climate change	Consistent	16	16754
	Inconsistent	12	10647
GMOs	Consistent	9	11359
	Inconsistent	6	7420
Vaccines	Consistent	11	26494
	Inconsistent	4	5589
Total Consistent		41	61137
Total Inconsistent		28	33537
Total		**69**	**94674**

Two individuals coded each document in terms of its position (i.e., arguing for a claim that is either consistent with or inconsistent with the mainstream science position). The four possible main claims for the inconsistent positions were 1) HIV does not causes AIDS, 2) global climate change is not occurring and/or it is not due to human actions, 3) GMOs are not safe, 4) the health risks of vaccines (e.g. autism) outweigh the benefits. The four possible main claims of the documents consistent with mainstream science were the opposite of the claims of the inconsistent documents. Two individuals independently coded each document as either consistent or inconsistent with the mainstream science position. Two individuals then highlighted and numbered each example of text that praised or derogated a source. By source, we mean a person, people or thing that is considered to be the origin of either data or claim. This would include *ad hominem* attacks (e.g., researcher X is a shill for Monsanto) and appeals to expertise (e.g., X is one of the most respected researchers in her field). Additionally, we also highlighted what we refer to as appeals to source quantity. The authors developed three rubrics and practiced coding using sources other than the 69 web documents used in the current study.

The initial rubric was used to classify each section of highlighted text into one of twelve different categories. The rubric included definitions and examples for ten different types of source-negative statements as well as source-positive statements and source quantity statements. The twelve category names and a brief description for each can be found in Table 2. The full text of all three rubrics can be found in an online depository hosted by openICPSR (Barnes, Neumann, and Draznin-Nagy 2020). Once all coding was complete, the raw frequencies, and frequency per 10,000 words were calculated as a function of topic (AIDS, climate change, GMOs, and vaccines) and document type (consistent, inconsistent).

Table 2. Categories and brief descriptions of the initial rubric

General category	**Specific category**	**Brief description**
Appeals to source quality	Expertise/credentials	Attack on data/claim because the source lacks expertise or has been formally disavowed by some institution
	Emotions/passion	Attack on data/claim because the source is emotional or blinded by passion
	Ethics/character	Attack on data/claim because the source is unethical or of bad character
	Intelligence/bad thinking processes	Attack on data/claim because of the way the source thinks, or doesn't think (e.g., stupid, close-minded, ignorant)
	Financial interest	Attack on data/claim because source has a vested financial interest
	Appeal to negative non-financial motivations	Attack on data/claim because motives of the source are bad for non-financial reasons
	Association with bad	Attack on data/claim because

	people/groups	the source is a member of an organization that is known to be bad or because the source associates with people known to be bad
	Bad action	Attack on data/claim because the source engaged in a bad action
	Generic abusive	Attack on data/claim because the source is bad in a vague/generic sense (e.g., is called a jerk)
	Other negative qualities	An attack on data/claim because of something negative about the source, but it does not fit into any other categories above
	Appeals to positive qualities	Support for data/claim because source has good qualities or has committed good actions
Appeal to source quantity	Appeal to source quantity	Arguing for or against data/claim because of the number of people that agree/disagree with the data/claim

A summary of the action rubric can be found in Table 3 and was only used to identify the items coded as 'bad action' using the initial rubric. Finally, the positive rubric is the inverse of the rubric presented in Table 2 and was only used to identify the items coded as 'appeals to positive qualities' using the initial rubric.

Coding for all three rubrics was conducted by two individuals. Coders practiced coding science websites that did not meet the criteria for inclusion in the present data set. During this period, the coders periodically compared their results, calculated the inter-rater reliability, and discussed reasons for their disagreements. When inter-rater reliability had reached an acceptable level, each coder began to independently code the content. In order to determine inter-rater reliability, we compared the coding results of each pair of coders for each rubric and calculated a kappa value for each comparison (Cohen 1960). The inter-rater reliability for the coding of document position (consistent, inconsistent) was perfect (κ = 1). The inter-rater reliability for the initial rubric reached acceptable levels (κ = .92) as did the inter-rater reliability for the action rubric (κ = .84) and the positive rubric (κ = .9). However, we must note that the identification of relevant fragments of text was done by a single coder. While there was potential for interrater disagreement in terms of selecting the appropriate code for each fragment, there was no potential for interrater disagreement in terms of selecting fragments. For this reason, it is likely that kappa values may be inflated.

Table 3. Categories and brief descriptions of the action rubric

Category	**Brief description**
Harm	Action causes physical or psychological harm to people, animals, or environment
Information control	An action related to information control (e.g., lying, withholding information)
Money	Controlling money in some way (e.g., funding/defunding research)
Other	An action that doesn't fit into the above three categories

Results

All raw data for the study have been uploaded to an online depository hosted by openICPSR (Barnes, Neumann, and Draznin-Nagy 2020). The effect sizes (in the form of Cohen's *d*), raw frequencies and frequencies per 10,000 words for the initial rubric are summarized in Table 4. Cohen's *d* (Cohen 1988) is a standardized measure of effect size. In particular, Cohen's *d* captures the difference in condition means divided by the relevant measure of variability. For the effect size calculations we present, variability was measured by the pooled standard deviation of the data. An effect size of 1 indicates that the difference between the means of two different conditions is equal to the pooled standard deviation, while an effect size of .5 would indicate that the difference between means is half as large as the standard deviation. Cohen (1988) suggests that effect sizes of small, medium, and large correspond to Cohen's *d* values of .2, .5, and .8, respectively. Frequency, and frequency per 10,000 words, will be helpful to researchers interested in how commonly various argumentative strategies appear in internet-based science writing. Effect size information will provide researchers with a sense of how large the differences are between consistent and inconsistent conditions.

For both types of document position (consistent, inconsistent), appeals to negative qualities of sources were quite common in documents of all four topics. Appeals to negative qualities of source were more common in the inconsistent than the consistent documents for three of the four topics (AIDS, climate change, vaccines). However, in the GMO documents, the use of appeals to negative qualities of sources was nearly evenly matched between the consistent and inconsistent conditions. For both consistent and inconsistent documents in all four categories, appeals to negative qualities of sources were generally more common than appeals to positive qualities of sources. One exception to this pattern can be found in the consistent AIDS condition: appeals to positive qualities of sources were more common than appeals to negative qualities of sources. The second exception to this pattern can be found in the consistent GMO condition: appeals to positive qualities of sources were nearly as common as appeals to negative qualities of sources. For all four topics, the most common appeal to negative qualities of source was the category of 'bad action.' For all four topics, 'generic abusive' attacks (which would generally correspond to abusive *ad hominem*) were used relatively infrequently.

Table 4. Standardized effect size (reported as Cohen's *d*) and frequencies of various categories (from the initial rubric) for all four topics individually (AIDS, climate change, GMO, vaccine), and the four topics in aggregate. Frequency per 10,000 words is shown in parentheses

	Category	**Effect size**	**Consistent**	**Incon-sistent**	**Total**
AIDS	Exper-tise/credentials	0	0(0)	0(0)	0(0)
	Emotions/passion	.08	4(6.1)	4(4)	8(4.9)
	Ethics/character	.58	0(0)	1(1)	1(.6)
	Intelligence/bad thinking processes	.65	1(1.5)	11(11)	12(7.3)
	Financial interest	1.15	0(0)	4(4)	4(2.4)

	Appeal to negative non-financial motivations	.78	0(0)	4(4)	4(2.4)
	Association with bad people/groups	0	0(0)	0(0)	0(0)
	Bad action	.86	4(6.1)	28(28.3)	32(19.5)
	Generic abusive	.63	1(1.5)	0(0)	1(.6)
	Other negative qualities	.93	4(6.1)	21(21.3)	25(15.2)
	Total appeals to negative qualities	**.84**	**14(21.4)**	**73(73.9)**	**87(53)**
	Appeals to positive qualities	.28	23(35.2)	19(19.2)	42(25.6)
	Appeal to source quantity	.09	12(18.4)	13(13.2)	25(15.2)
	Total	**.45**	**49(75)**	**105(106.2)**	**154(93.8)**
Climate change	Exper-tise/credentials	.08	1(.6)	1(.9)	2(.7)
	Emotions/passion	.61	0(0)	6(5.6)	6(2.2)
	Ethics/character	.55	0(0)	7(6.6)	7(2.6)
	Intelligence/bad thinking processes	.92	0(0)	12(11.3)	12(4.4)
	Financial interest	.35	2(1.2)	4(3.8)	6(2.2)
	Appeal to negative non-financial motivations	.49	0(0)	11(10.3)	11(4)
	Association with bad people/groups	.41	0(0)	1(.9)	1(.4)
	Bad action	1.3	0(0)	87(81.7)	87(31.8)
	Generic abusive	1.33	0(0)	17(16)	17(6.2)
	Other negative qualities	1.16	4(2.4)	26(24.4)	30(11)
	Total appeals to negative qualities	**1.31**	**7(4.2)**	**172(161.5)**	**179(65.3)**
	Appeals to positive qualities	.59	6(3.6)	34(31.9)	40(14.6)
	Appeal to source quantity	.81	6(3.6)	30(28.2)	36(13.1)
	Total	**1.08**	**19(11.3)**	**236(221.6)**	**255(93.1)**
GMO	Exper-tise/credentials	.67	3(2.6)	0(0)	3(1.6)
	Emotions/passion	.67	3(2.6)	0(0)	3(1.6)
	Ethics/character	.58	0(0)	1(1.3)	1(.5)
	Intelligence/bad thinking processes	.69	16(14.1)	1(1.3)	17(9.1)

	Financial interest	.89	1(.9)	4(5.4)	5(2.7)
	Appeal to negative non-financial motivations	0	3(2.6)	2(2.7)	5(2.7)
	Association with bad people/groups	.58	0(0)	1(1.3)	1(.5)
	Bad action	.45	28(24.7)	32(43.1)	60(32)
	Generic abusive	.48	6(5.3)	1(1.3)	7(3.7)
	Other negative qualities	.64	7(6.2)	0(0)	7(3.7)
	Total appeals to negative qualities	**.05**	**67(59)**	**42(56.6)**	**109(58)**
	Appeals to positive qualities	1.41	60(52.8)	12(16.2)	72(38.3)
	Appeal to source quantity	.38	54(47.5)	17(22.9)	71(37.8)
	Total	**.53**	**181(159.3)**	**71(95.7)**	**252(134.2)**
Vaccine	Expertise/credentials	1.13	10(3.8)	0(0)	10(3.1)
	Emotions/passion	.04	3(1.1)	1(1.8)	4(1.2)
	Ethics/character	.64	1(.4)	4(7.2)	5(1.6)
	Intelligence/bad thought processes	.38	23(8.7)	4(7.2)	27(8.4)
	Financial interest	.04	30(11.3)	10(17.9)	40(12.5)
	Appeal to negative non-financial motivations	.76	4(1.5)	11(19.7)	15(4.7)
	Association with bad people/groups	0	0(0)	0(0)	0(0)
	Bad action	.66	47(17.7)	42(75.1)	89(27.8)
	Generic abusive	.1	7(2.6)	2(3.6)	9(2.8)
	Other negative qualities	.39	32(12.1)	3(5.4)	35(10.9)
	Total appeals to negative qualities	**.2**	**157(59.3)**	**77(137.8)**	**234(72.9)**
	Appeals to positive qualities	.06	57(21.5)	19(34)	76(23.7)
	Appeal to source quantity	.5	20(7.5)	14(25)	34(10.6)
	Total	**.19**	**234(88.2)**	**110(196.8)**	**344(107.2)**
Four topics in aggregate	Expertise/credentials	.55	14(2.3)	1(.3)	15(1.6)
	Emotions/passion	.16	10(1.6)	11(3.3)	21(2.2)
	Ethics/character	.5	1(.2)	13(3.9)	14(1.5)
	Intelligence/bad	.01	40(6.5)	28(8.3)	68(7.2)

	thought processes				
	Financial interest	.01	33(5.4)	22(6.6)	55(5.8)
	Appeal to negative non-financial motivations	.44	10(1.6)	25(7.5)	35(3.7)
	Association with bad people/groups	.39	0(0)	2(.6)	2(.2)
	Bad action	.8	79(12.9)	189(56.4)	268(28.3)
	Generic abusive	.32	14(2.3)	20(6)	34(3.6)
	Other negative qualities	.19	47(7.7)	50(14.9)	97(10.2)
	Total appeals to negative qualities	**.46**	**248(40.6)**	**361(107.6)**	**609**(64.3)
	Appeals to positive qualities	.1	146(23.9)	84(25)	230(24.3)
	Appeal to source quantity	.09	92(15)	74(22.1)	166(17.5)
	Total	**.31**	**486(79.5)**	**519(154.8)**	**1005**(106.2)

In terms of the use of appeals to source quantity, the data revealed no clear patterns. In the climate change and vaccine source documents, the inconsistent documents relied on these appeals more than the consistent documents. However, the reverse of that pattern was found in the GMO condition. In the AIDS documents, the use of this type of argument was fairly evenly matched between the consistent and inconsistent web documents.

The effect sizes (in the form of Cohen's *d*), raw frequencies, and frequencies per 10,000 words for the action rubric are summarized in Table 5. What bad actions were the sources allegedly engaging in? Here the pattern is clear: for all four topics, information control was (by far) the most frequent action that sources were accused of. In fact, only a negligible number of the other three types of actions appeared in the source documents. The other clear pattern in the data is that sources inconsistent with the scientific mainstream were more likely to accuse sources of engaging in bad actions. Not only was the pattern consistent across all four topics, but the effect sizes associated with that effect were consistently large.

Table 5. Standardized effect size (reported as Cohen's *d*) and frequencies of various categories (from the action rubric) for all four topics individually (AIDS, climate change, GMO, vaccine), and the four topics in aggregate. Frequency per 10,000 words is shown in parentheses.

	Category	**Effect size**	**Consistent**	**Inconsistent**	**Total**
AIDS	Harm	.51	1(1.5)	5(5.1)	6(3.7)
	Information control	.84	3(4.6)	21(21.3)	24(14.6)
	Money	.58	0(0)	1(1)	1(.6)
	Other	.58	0(0)	1(1)	1(.6)
	Total	**.83**	**4(6.1)**	**28(28.3)**	**32(19.5)**
Climate change	Harm	.72	0(0)	4(3.8)	4(1.5)
	Information control	1.3	0(0)	68(63.9)	68(24.8)
	Money	0	0(0)	0(0)	0(0)
	Other	.85	0(0)	15(14.1)	15(5.5)
	Total	**1.28**	**0(0)**	**87(81.7)**	**87(31.8)**
GMO	Harm	.91	0(0)	6(8.1)	6(3.2)
	Information control	.32	22(19.4)	23(31)	45(24)
	Money	.58	0(0)	1(1.3)	1(.5)
	Other	0	3(2.6)	2(2.7)	5(2.7)
	Total	**.52**	**25(22)**	**32(43.1)**	**57(30.4)**
Vaccine	Harm	.79	4(1.5)	10(17.9)	14(4.4)
	Information control	.62	37(14)	27(48.3)	64(19.9)
	Money	.43	2(.8)	0(0)	2(.6)
	Other	.6	4(1.5)	5(8.9)	9(2.8)
	Total	**.66**	**47(17.7)**	**42(75.1)**	**89(27.7)**
Four topics in aggregate	Harm	.58	5(.8)	24(7.1)	29(3.1)
	Information control	.78	62(10.1)	139(41.4)	201(21.2)
	Money	.08	2(.3)	2(.6)	4(.4)
	Other	.61	7(1.14)	22(6.6)	29(3.1)
	Total	**.8**	**76(12.4)**	**187(55.8)**	**263(27.8)**

The effect sizes (in the form of Cohen's *d*), raw frequencies and frequencies per 10,000 words for the positive rubric are summarized in Table 6. There is no clear pattern of results for appeals to positive source qualities. For two topics (AIDS, GMO), consistent documents employed more of these appeals than inconsistent documents, however, for the other two topics (climate change, vaccines) the opposite was true. The most frequent type of appeal to positive qualities of sources varied as a function of document topic. 'Other' was the most frequent category in the AIDS and vaccines documents, 'expertise/credentials' was the most frequent in the climate change documents, and 'financial interest' was the most frequent in the GMO documents. While 'bad action' was the most frequent category from among the appeals to negative qualities of sources, its opposite (good action) was not very common relative to the other appeals to positive qualities.

Table 6. Standardized effect size (reported as Cohen's *d*) and frequencies of various categories (from the positive rubric) for all four topics individually (AIDS, climate change, GMO, vaccine), and the four topics in aggregate. Frequency per 10,000 words is shown in parentheses.

	Category	**Effect size**	**Consistent**	**Inconsistent**	**Total**
AIDS	Expertise/credentials	.95	1(1.5)	8(8.1)	9(5.5)
	Emotions/passion	0	0(0)	0(0)	0(0)
	Ethics/character	0	0(0)	0(0)	0(0)
	Intelligence/good thinking processes	0	0(0)	0(0)	0(0)
	Financial interest	0	0(0)	0(0)	0(0)
	Appeal to positive non-financial motivations	.63	2(3.1)	0(0)	2(1.2)
	Association with good people/groups	0	0(0)	0(0)	0(0)
	Good action	.63	1(1.5)	0(0)	1(.6)
	Generic praise	.42	4(6.1)	10(10.1)	14(8.5)
	Other	.97	15(23)	1(1)	16(9.7)
	Total	**.28**	**23(35.2)**	**19(19.2)**	**42(25.6)**
Climate change	Expertise/credentials	.37	4(2.4)	9(8.5)	13(4.7)
	Emotions/passion	0	0(0)	0(0)	0(0)
	Ethics/character	0	0(0)	0(0)	0(0)

	Intelligence/good thinking processes	.41	0(0)	1(.9)	1(.4)
	Financial interest	.41	0(0)	1(.9)	1(.4)
	Appeal to positive non-financial motivations	0	0(0)	0(0)	0(0)
	Association with good people/groups	.41	0(0)	2(1.9)	2(.7)
	Good action	.61	0(0)	2(1.9)	2(.7)
	Generic praise	.51	0(0)	10(9.4)	10(3.6)
	Other	.63	2(1.2)	9(8.5)	11(4)
	Total	**.59**	**6(3.6)**	**34(31.9)**	**40(14.6)**
GMO	Exper-tise/credentials	.57	13(11.4)	4(5.4)	17(9.1)
	Emotions/passion	.47	1(.9)	0(0)	1(.5)
	Ethics/character	0	0(0)	0(0)	0(0)
	Intelligence/good thinking processes	.1	2(1.8)	1(1.3)	3(1.6)
	Financial interest	.93	27(23.8)	4(5.4)	31(16.5)
	Appeal to positive non-financial motivations	.15	1(.9)	1(1.3)	2(1.1)
	Association with good people/groups	.47	3(2.6)	0(0)	3(1.6)
	Good action	.13	2(1.8)	1(1.3)	3(1.6)
	Generic praise	1.19	4(3.5)	0(0)	4(2.1)
	Other	.82	7(6.2)	1(1.3)	8(4.3)
	Total	**1.41**	**60(52.8)**	**12(16.2)**	**72(38.3)**
Vaccine	Exper-tise/credentials	.37	3(1.1)	2(3.6)	5(1.6)
	Emotions/passion	0	0(0)	0(0)	0(0)
	Ethics/character	0	0(0)	0(0)	0(0)
	Intelligence/good thinking processes	.6	3(1.1)	0(0)	3(.9)
	Financial interest	.4	13(4.9)	1(1.8)	14(4.4)
	Appeal to positive non-financial motivations	.41	4(1.5)	3(5.4)	7(2.2)
	Association with good people/groups	0	0(0)	0(0)	0(0)
	Good action	.7	3(1.1)	4(7.2)	7(2.2)
	Generic praise	.33	5(1.9)	3(5.4)	8(2.5)
	Other	.18	26(9.8)	6(10.7)	32(10)
	Total	**.06**	**57(21.5)**	**19(34)**	**76(23.7)**

Four topics in aggregate	Expertise/credentials	.24	21(3.4)	23(6.9)	44(4.6)
	Emotions/passion	.22	1(.2)	0(0)	1(.1)
	Ethics/character	0	0(0)	0(0)	0(0)
	Intelligence/good thinking processes	.14	5(.8)	2(.6)	7(.7)
	Financial interest	.42	40(6.5)	6(1.8)	46(4.9)
	Appeal to positive non-financial motivations	.05	7(1.1)	4(1.2)	11(1.2)
	Association with good people/groups	0	3(.5)	2(.6)	5(.5)
	Good action	.19	6(1)	7(2.1)	13(1.4)
	Generic praise	.35	13(2.1)	23(6.9)	36(3.8)
	Other	.22	50(8.2)	17(5.1)	67(7.1)
	Total	**.1**	**146(23.9)**	**84(25)**	**230(24.3)**

In most cases in which a source was either a test or measurement, the appeal to source quality was coded as 'other negative qualities' or 'other positive qualities.' It was rare that an appeal to source quality (either positive or negative) directed toward a human was coded as 'other.' For this reason, the frequency of the 'other' responses in Tables 4 and 6 may serve as a rough approximation of how often the authors of our sample of web documents criticized or praised non-human sources, such as tests and measures. The final rows of Tables 4 through 6 present the data collapsed across the four topics and can therefore provide an idea of the relationship between consistent and inconsistent documents. Table 4 revealed that, when collapsed across all four topics, inconsistent documents tend to rely more heavily on appeals to negative qualities (Cohen's d = .46). Table 5 revealed that inconsistent documents rely much more heavily on description of the bad behavior of scientists and other sources they disagree with (Cohen's d = .8). Finally, Table 6 failed to reveal any meaningful difference between consistent and inconsistent documents in terms of appeals to positive qualities (Cohen's d = .1).[i]

Discussion

We provided frequency data for several types of appeals to source quality and quantity across four different topics. The frequency data were broken down by topic (AIDS, climate change, GMO safety, vaccine safety) and perspective (consistent with mainstream science, inconsistent with mainstream science). The low number of items coded as 'other,' and our high kappa values indicate that the rubric we created was well suited for coding the source documents we had selected. We found hundreds of examples of appeals to source quality and quantity, both in documents arguing for claims that are inconsistent with mainstream science and in documents that argue for claims that are inconsistent with mainstream science.

All of our source documents had an agenda (i.e., to persuade the reader of a particular claim about a science-related issue). Because laypeople can detect invalid argument schemes (van Eemeren, Garssen and Meuffels 2009), we might expect that the authors of our 69 source documents would employ reasonable arguments more often than unreasonable arguments. Our results are consistent with that notion. For instance, generic abusive attacks are considered to be generally unreasonable by a number of authors (van Eemeren, Garrsen and Meufells 2009; van Eemeren, Garssen and Meufells 2012; Walton 1998), and we found in our sample that the generic abusive attacks were used relatively infrequently. Additionally, the most common type of action that sources were accused of was control of information. Control of information includes things like lying, hiding information, manipulating data, and committing fraud. Rather than being an irrelevant or unreasonable attack, accusations of engaging in information control seem to be both relevant and reasonable. Individuals *should* distrust a source if that source has a record of deception.

In web documents dealing with all four topics, we found that appeals to negative source qualities were more common than appeals to positive source qualities and that authors were more likely to describe the characteristics of the sources than to describe their actions. This fact might indicate that the authors of the documents feel that appeals to negative source qualities are more effective than appeals to positive source qualities. Unfortunately, the data collected in the current study is unable to address the reasons why authors might prefer to use one type argument more than another. The final rows of Tables 4 and 5 revealed that source documents promoting ideas inconsistent with the scientific mainstream are more likely to mention the bad qualities and deeds of those they disagree with. The current data cannot tell us why this pattern emerges, but we can put forth a hypothesis. It may be that those arguing for a claim that does not have widespread acceptance may be willing to use a wider range of argumentative strategies (even those that are often frowned upon) in order to increase their market share of public opinion. Since those who argue for a claim that does have widespread acceptance are 'winning,' they may be more conservative in the argumentative strategies they rely on. It is possible that increased reliance on *ad hominem* attacks may be accompanied by a reduction of discussion of empirical data. However, further research would be required to explore this possibility.

Our results are consistent with the literature. Habernal, Wachsmuth, Gurevych, and Stein (2018) identified and coded *ad hominem* attacks in the social media website, Reddit. Like the present study, Habernal et al. found many instances of appeals to negative source qualities. Additionally, the taxonomy of *ad hominem* attacks used by Habernal et al. has some similarities with that used in the current study. Sahlane (2012) documented extensive use of appeals to negative source qualities in newspaper opinion/editorial items focused on the conflict between the US and Iraq. However, neither Habernal et al. nor Sahlane identified or counted instances of appeals to positive source qualities or appeals to source quantity. We know of no study that recorded the relative frequency of the three types of appeals considered in the current study (i.e., appeals to negative source qualities, positive source qualities, and source quantities). Additional work will be needed in order to determine if the relative frequencies for the three types of appeals considered in the current study is similar or dissimilar to that found in other document types and other forms of media.

In the present study, the final rows of Tables 4 and 5 revealed that sources that promote claims that are inconsistent with mainstream science appear to use *ad hominem* attacks with greater frequency than sources that promote claims consistent with scientific orthodoxy. This finding is consistent with the results presented by Barnes and Church (2013) and Barnes, Church, and Draznin-Nagy (2017). Data reported in both papers revealed that those with messages inconsistent with the scientific mainstream (i.e., creationism) rely on very different argumentative and rhetorical strategies than those arguing for messages consistent with the scientific mainstream (i.e., evolution). It must be noted, however, that when the current data are looked at in disaggregate form, the pattern of data is less clear.

There are a number of limitations of the current study. First, there is no reason to think that the absolute and relative frequency data presented here will generalize to other forms of media. For instance, due to norms, conventions, and explicit rules, we know that *ad hominem* attacks are extremely rare in the peer-reviewed scientific literature. Additionally, journalistic norms also discourage the use of appeals to negative source qualities. An additional limitation of the current study is that the content we explored was limited to only four science topics (AIDS, climate change, GMO safety, vaccine safety). Authors of web documents dealing with other science issues may use appeals to source quality and quantity in a very different manner than that revealed in the current results. Finally, it should be noted that the sample of sources used in the current study were all sources with an agenda: the primary goal of each document was to convince the reader of a particular science claim. Science communication that is not intended to push an agenda will likely have different characteristics than the documents included in the current study.

The data that we have presented may be useful to those working in the area of informal fallacies. The research on informal fallacies may benefit from an increased awareness of the relative frequency of potentially fallacious arguments as they appear in the wild. Some personal attacks are used more often than others, so it may make sense for researchers to spend more time addressing the types of attacks that appear most often. Additionally, researchers may also use the current data as a check to determine if attacks that are perceived to be most reasonable are also the types of attacks that occur most often.

Research indicates that trust in sources is at least as important as the scientific facts when it comes to influencing the public. The current study reveals how frequently science communication presented in websites veers from 'just the facts' in order to focus on the sources of those facts. What is the answer to the question, 'why is trust in sources so influential when it comes to science claims?' In light of the current results, the answer to that question might be that the public has been provided a great deal of information about the sources of the claims and may therefore be aware of many reasons to trust/distrust certain sources.

Acknowledgments

We would like to thank Glenn Branch (National Center for Science Education) for his valuable feedback on an early draft of this manuscript.

References

Aberdein, Andrew. 2014. In defence of virtue: The legitimacy of agent-based argument appraisal. *Informal Logic* 34(1): 77-93. DOI: https://doi.org/10.22329/il.v34i1.3938

Agrawal, A. J. 2017. How to optimize your SEO results through content creation. *Forbes.* https://www.forbes.com/sites/ajagrawal/2017/08/30/how-to-optimize-your-seo-results-through-content-creation/#150da86d2aa3

Barnes, Ralph Michael. 2019. Identifying and clarifying arguments in a recent debate regarding measures based on memory-based methods. *Journal of Clinical Epidemiology* 111: 115-123. https://doi.org/10.1016/j.jclinepi.2019.03.013

Barnes, Ralph Michael and Rebecca A. Church. 2013. Proponents of creationism but not proponents of evolution frame the origins debate in terms of proof. *Science & Education* 22(3): 577-603. DOI: 10.1007/s11191-012-9451-y

Barnes, Ralph Michael, Church, Rebecca A. and Samuel Draznin-Nagy. 2017. The Nature of the arguments for creationism, intelligent design, and evolution. *Science & Education* 26(1-2): 27-47. DOI: 10.1007/s11191-017-9875-5

Barnes, Ralph Michael, Johnson, Heather M., MacKenzie, Noah, Tobin, Stephanie J. and Chelsea M. Taglang. 2018. The effect of ad hominem attacks on the evaluation of claims promoted by scientists. *PLOS ONE.* https://doi.org/10.1371/journal.pone.0192025

Barnes, Ralph Michael, Neuman, Zoë and Samuel Draznin-Nagy. 2020. Data and supplementary information for 'Source related argumentation found in science websites: a quantitative study.' https://www.openicpsr.org/openicpsr/project/120216/version/V2/view/ https://doi.org/10.3886/E120216V2

Benjamin, Daniel and James Berger. 2019. Three recommendations for improving the use of *p*-values. *The American Statistician* 73(S1): 186-191. https://doi.org/10.1080/00031305.2018.1543135

Bondy, Patrick. 2015. Virtues, evidence, and ad hominem arguments. *Informal Logic* 35(4): 450-466. DOI: https://doi.org/10.22329/il.v35i4.4330

Bowel, Tracy and Justine Kingsbury. 2013. Virtue and argument: Taking character into account. *Informal Logic* 33(1): 22-32. DOI: https://doi.org/10.22329/il.v33i1.3608

Briñol, Pablo. and Richard E. Petty. 2009. Source factors in persuasion: A self-validation approach. *European Review of Social Psychology* 20(1): 49-96. https://doi.org/10.1080/10463280802643640

Calin-Jageman, Robert and Goeff Cumming. 2019. The new statistics for better science: Ask how much, how uncertain, and what else is known. *The American Statistician* 73(S1): 271-280. https://doi.org/10.1080/00031305.2018.1518266

Camerer, Colin and Anna Dreber, et al. 2016. Evaluating replicability of laboratory experiments in economics. *Science* 351: 1433-1436. DOI: 10.1126/science.aaf0918

Chaiken, Shelly. 1980. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology* 39(5): 752-766. http://dx.doi.org/10.1037/0022-3514.39.5.752

Chaiken, Shelly. 1987. The heuristic model of persuasion. In *Social Influence: The Ontario Symposium* (Vol. 5, pp.3-39), eds. M.P. Zanna, J.M. Olson, and C.P. Herman. Hillsdale, NJ: Earlbaum.

Cohen, Jacob. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1): 37–46. https://doi.org/10.1177/001316446002000104

Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Routledge.

Colquhoun, David. 2019. The false positive risk: A proposal concerning what to do about *p*-values. *The American Statistician* 73(S1): 192-201.
https://doi.org/10.1080/00031305.2018.1529622

Dienes, Z. 2011. Bayesian versus orthodox statistics: Which side are you on? Perspectives on *Psychological Science* 6(3): 274-290. DOI: 10.1177/1745691611406920

van Eemeren, Frans H., Bart Garssen, and Bert Meuffels. 2009. *Fallacies and judgements of reasonableness: Empirical research concerning the pragma-dialectical discussion of rules.* Dordrecht, Netherlands: Springer.

van Eemeren, Frans H., Bart Garssen, and Bert Meufells. 2012. The disguised abusive ad hominem empirically investigated: Strategic manoeuvring with direct personal attacks. *Thinking & Reasoning* 18(3): 344-364.
https://doi.org/10.1080/13546783.2012.678666

Fidler, Fiona, Thomason, Neil, Cumming, Goeff, Finch, Sue and Joanna Leeman. 2004. Editors can lead researchers to confidence intervals, but can't make them think. *Psychological Science* 15(2): 119-126.
https://doi.org/10.1111/j.0963-7214.2004.01502008.x

Gronbeck, Bruce E. 1992. Negatives narratives in 1988 presidential campaign ads. *Quarterly Journal of Speech* 78: 333-346. https://doi.org/10.1080/00335639209384000

Habernal, Ivan, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. 2018. Before name-calling: Dynamics and triggers of ad hominem fallacies in web argumentation. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* Volume 1, 386-396. DOI: 10.18653/v1/N18-1036

Hubbard, Raymond. 2016. Will the ASA's efforts to improve statistical practice be successful? Some evidence to the contrary. *The American Statistician* 73(S1): 31-35. https://doi.org/10.1080/00031305.2018.1497540

Ioannidis, John. 2005. Why most published research findings are false. *PLOS Medicine* 2(8): e124. DOI:10.1371/journal.pmed.0020124

Jamieson, Kathleen Hall. 1992. *Dirty politics: Deception, distraction, and democracy*. New York: Oxford University Press.

Johnson, Christopher M. 2009. Reconsidering the ad hominem. *Philosophy* 84(2): 251-266. doi: 10.1017/S0031819109000217

Kaid, Lynda Lee and John Boydston. 2009. An experimental study of the effectiveness of negative political advertisements. *Communication Quarterly* 35(2): 193-201. https://doi.org/10.1080/01463378709369680

Latané, Bibb. 1981. The psychology of social impact. *American Psychologist* 36(4): 345-356. DOI:10.1037/0003-066X.36.4.343

Lau, Richard R., Sigelman, Lee and Ivy Brown Rovner. 2007. The effects of negative political campaigns: A meta-analytic reassessment. *Journal of Politics* 69(4): 1176-1209. https://doi.org/10.1111/j.1468-2508.2007.00618.x

Lewandowsky, Stephan, Gignac, Gilles E. and Samuel Vaughan. 2012. The pivotal role of perceived scientific consensus in acceptance of science. *Nature Climate Change* 3: 399-404. DOI: 10.1038/NCLIMATE1720

McShane, Blakeley, Tackett, Jennifer, Böckenholt, Ulf and Andrew Gelman. 2019. Large-scale replication projects in contemporary psychological research. *The American Statistician* 73(S1): https://doi.org/10.1080/00031305.2018.1505655

Mogil, Jeffrey and Malcolm Macleod. 2017. No publication without confirmation. *Nature* 542: 409-411. DOI: 10.1038/542409a

Moscovici, Serge, Mugny, Gabriel and Eddy van Avermaet (eds.). 1985. *Perspectives on minority influence.* Cambridge: Cambridge University Press.

Mulaik, Stanley, Raju, Nambury and Richard Harshman. 1997. In *What if there were no significance tests?* eds. Lisa Harlow, Stanley Mulaik and James Steiger, 65-115. Mahwah, NJ: Lawrence Erlbaum Associates.

Neisser, Ulric, Boodoo, Gwyneth, Bouchard, Thomas J., Boykin, A. Wade, Brody, Nathan, Ceci, Stephen J., Halpern, Diane F., Loehlin, John C., Perloff, Robert, Sternberg, Robert J. and Susana Urbina. 1996. Intelligence: Knowns and unknowns. *American Psychologist* 51(2): 77-101. DOI:10.1037/0003-066X.51.2.77

Open Science Collaboration. 2015. *Science* 349. DOI: 10.1126/science.aac4716

Pornpitakpan, Chanthika. 2004. The persuasiveness of source credibility: A critical review of five decades' Evidence. *Journal of Applied Social Psychology* 34(2): 243-281. DOI: 10.1111/j.1559-1816.2004.tb02547.x

Sabato, Larry J. 1981. *The rise of political consultants: New ways of winning elections.* New York: Basic Books, Inc., Publishers.

Sahlane, Ahmed. 2012. Argumentation and fallacy in the justification of the 2003 war on Iraq. *Argumentation* 26(4): 459-488. https://doi.org/10.1007/s10503-012-9265-8

Simmons, Joseph, Nelson, Leif and Uri Simonsohn. 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22(11): 1359-1366. DOI: 10.1177/0956797611417632

Slovic, Paul. 1993. Perceived risk, trust, and democracy. *Risk Analysis* 13(6): 675-682. https://doi.org/10.1111/j.1539-6924.1993.tb01329.x

Slovic, Paul, Flynn, James H. and Mark Layman. 1991. Perceived risk, trust, and the politics of nuclear waste. *Science* 254(5038): 1603-1607. DOI: 10.1126/science.254.5038.1603

Souder, Lawrence and Furrah Qureshi. 2012. Ad hominem arguments in the service of boundary work among climate scientists. *Journal of Science Communication* 11(1): 1-11. https://doi.org/10.22323/2.11010201

Stangor, Charles, Sechrist, Gretchen B. and John T. Jost. 2001. Changing racial beliefs by providing consensus information. *Personality and Social Psychology Bulletin* 27(4): 486-496. DOI: 10.1177/0146167201274009

Toulmin, Stephen E. 1958. *The uses of argument.* Cambridge: Cambridge University Press.

Walton, Douglas. 1998. *Ad hominem arguments.* Tuscaloosa, AL: University of Alabama Press.

Wasserstein, Ronald and Nicole Lazar. 2016. The ASA's statement on p-values: context, process, and purpose. *The American Statistician* 70(2): 129-133. DOI: 10.1080/00031305.2016.1154108

Wasserstein, Ronald, Schirm, Allen and Nicole Lazar. 2019. Moving to a world beyond "p < .05". *The American Statistician* 73(S1): 1-19. https://doi.org/10.1080/00031305.2019.1583913

PageRank, Wikipedia, 2020, https://en.wikipedia.org/wiki/PageRank

Wilson, Elizabeth J., and Daniel L. Sherrel. 1993. Source effects in communication and persuasion research: A meta-analysis of effect size. *Journal of the Academy of Marketing Science* 21(2): 101-112. DOI: 10.1007/BF02894421

Woods, John. 2007. Lightening up on the ad hominem. *Informal Logic* 27(1): 109-134. DOI: https://doi.org/10.22329/il.v27i1.467

Yaniv, Ilan. 2004. The benefit of additional opinions. *Current Directions in Psychological Science* 13(2): 75-7*8. https://doi.org/10.1111/j.0963-7214.2004.00278.x*

Yap, Audrey. 2013. Ad hominem fallacies, bias, and testimony. *Argumentation* 27: 97-109. DOI 10.1007/s10503-011-9260-5

Ziliak, Steve and Deirdre McCloskey. 2008. *The cult of statistical significance: How the standard error costs us jobs, justice, and lives*. Ann Arbor, MI: University of Michigan Press.

Zitek, Emily M. and Michelle R. Hebl. 2007. The role of social norm clarity in the influenced expression of prejudice over time. *Journal of Experimental Social Psychology* 43(6): 867-876. DOI: 10.1016/j.jesp.2006.10.010

---

[i] In the social and biological sciences quantitative results are frequently presented along with inferential statistics. Null-hypothesis significance tests (NHST) are currently the most common inferential tests presented in the social and biological sciences, although Bayesian statistics are slowly gaining in popularity.

Over the past 20 years, it has become evident that several sciences that employ NHST (a form of inferential testing that relies on the reporting of *p* values) have developed a replication crisis (Open Science Collaboration 2015; Camerer and Dreber et al. 2016; Ioannidis 2005; McShane Tackett Böckenholt and Gelman 2019). What this means is that many findings reported using NHST cannot be replicated. This makes it very difficult for researchers to determine which results can be trusted.

Why might reporting statistical significance and *p* values lead to misunderstanding and replication problems? One problem is that though *p* values do not provide information about either the likelihood of either the alternative or the null hypothesis (Wasserstein and Lazar 2016), many researchers falsely believe that *p* values do provide this type of information. For a study with a *p* value of .99, the probability that the alternative hypothesis is true could take any value between 1 and 0. Similarly, for a study with a *p* value of .001 the probability that the alternative hypothesis is true could take any value between 1 and 0. Because of widespread misunderstanding of the meaning of *p* values and the term ‘statistical significance,’ reporting either may lead to misunderstandings about the data.

A second problem is that calculations of *p* are appropriate only when all hypotheses and all details of the statistical analysis have been determined prior to looking at the data. Unfortunately, HARKing (hypothesizing after the results are known) has been a common feature of published research in fields that report *p* values and statistical significance. The problem is that researchers conduct exploratory research (which is valuable to science and not problematic in itself) and then subject those results to inferential statistics and report *p* values and statistical significance. Reporting *p* values for a study in which all hypotheses and details of the statistical analysis were agreed upon prior to conducting the study is the equivalent of shooting an arrow at a bullseye painted on the side of a barn. Reporting *p* values for a study in which the hypotheses and specific analyses were chosen after the results were known, HARKing, is the equivalent of shooting an arrow at the side of a barn, and *then* painting a target around the arrow. Simmons, Nelson, and Simonsohn (2011) claimed that nearly any result can be shown to be statistically significant (e.g., a *p* value equal to or below .05) by HARKing. They demonstrated their claim by choosing the details of their statistical analysis after the results were collected in order to show that listening to the song “While I’m sixty-four” caused people to become 18 months younger ($p = .04$).

For a number of reasons (including the two reasons listed above), many statisticians feel that one of the critical causes of this replication crisis is NHST (Hubbard 2016; Wasserstein Schirm and Lazar 2019; Ziliak and McCloskey 2008). Some experts feel that the best way to reduce untrustworthy research results is to require researchers to be more sophisticated in their use of NHST (Benjamin and Berger 2019; Colquhoun 2019; Mulaik Raju and Harshman 1997). However, others feel that *p* values should be abandoned and replaced with confidence intervals (Calin-Jageman and Cumming 2019; Fidler and

Thomason et al. 2004) or Bayesian statistics (Dienes 2011). The American Statistical Association has recently suggested that reporting statistical significance (a status that is based on *p* values of a certain value) should be abandoned (Wasserstein and Lazar 2016; Wasserstein Schirm and Lazar 2019).

Mogil and Macleod (2017) have suggested that a solution to the problem is to first carry out exploratory studies that are not subjected to any formal statistical inference. If researchers believe that these findings are worthy of additional study, then they should then conduct pre-registered hypothesis testing studies subject to statistical inference in order to test hypotheses that were inspired by the earlier exploratory research. We agree with Mogil and Macleod in that drawing a clear distinction between exploratory research and hypothesis testing research will help mitigate the replication crisis and improve the reporting of results. For that reason, we have chosen not to conduct or report inferential statistics on the results of the current exploratory study. We do feel, however, that the means, effect sizes, and sample sizes reported in our Results section will convince many readers that a number of our results are worthy of follow-up in the form of hypothesis testing studies. We have provided readers with the raw data from our study, so that researchers who wish to conduct inferential statistics (e.g., Poisson regression, Bayesian analysis, etc.) can do so. Additionally, by making public our raw data, any research team that wishes to calculate Bayes factors that take into account data from both a follow-up study and our original exploratory study are able to do so.