

Is There a Role for Narrative Attributes in Scientific Literature?

Ann Hillier

A thesis

submitted in partial fulfillment of the
requirements for the degree of

Master of Marine Affairs

University of Washington

2016

Committee:

Terrie Klinger

Ryan Kelly

Program Authorized to Offer Degree:

School of Marine and Environmental Affairs

©Copyright 2016
Ann Hillier

University of Washington

Abstract

Is There a Role for Narrative Attributes in Scientific Literature?

Ann Hillier

Chair of the Supervisory Committee:

Professor, Terrie Klinger
School of Marine and Environmental Affairs

Scientists strive to produce relevant work that is both well-received within their fields and influential on the world around them. Effectively communicating science is an essential step in achieving this. Given the widespread acknowledgment that narratives are powerful tools of communication, I sought to identify a role for narrative attributes in scientific literature. I distilled six measureable indicators of narrativity from the literature, which reflect specific attributes described by narrative theory. I probed the abstracts of scientific peer-reviewed articles for these indicators by crowdsourcing, and measured their effect on article influence, as proxied by citation frequency. My results support the hypothesis that attributes associated with narrativity are positively associated with article citation frequency, with two exceptions: narrative perspective (first-person narrators are negatively associated with citations, $p < 0.01$), and conjunctions (unimodally associated). Furthermore, the regression tree analysis predicts that among articles published in lower-rank journals, abstracts with explicitly related events and that offer an evaluative standpoint are more likely to achieve higher citation outcomes. I speculate that by incorporating narrative attributes into their writing, scientists mirror the way we as humans experience the world, and in turn, captivate their readers and extend their

influence. Ultimately, this research helps form a more comprehensive understanding of how scientists can effectively communicate their work.

*“Since the fading of the original Enlightenment during the late eighteenth and early nineteenth centuries, stubborn impasse has existed in the consilience of the humanities and natural sciences. One way to break it is to collate the creative process and writing styles of literature and scientific research. This might not prove so difficult as it first seems. Innovators in both of two domains are basically dreamers and storytellers.”—E.O. Wilson, *The Social Conquest of Earth* [1].*

1. Introduction

Scientists strive to produce influential work that earns approval via the peer-review process, is published in an accredited journal, and influences subsequent studies and the larger field of inquiry. Effectively communicating one’s science is an essential step along the path from data collection and analysis to any type of influence, both within the academy and beyond. Furthermore, successfully communicating emerging science is necessary to solve the complex environmental challenges that are the stated focus of many scientific investigations. Consequently, I sought to understand the ways in which scientists—through peer-reviewed publications—may more effectively influence their fields.

Narratives are widely recognized as powerful tools of communication [2,3]. Within the environmental sciences, for example, scholars maintain that narratives help convey knowledge, complement activism, and motivate political action [4,5]. Empirical studies of narrative discourse on human comprehension illustrate these claims: narrative texts are recalled better and read faster than expository texts [3]. While the exact mechanisms are elusive, narrative theory underscores an important duality: narratives are unique to the individual yet molded by shared cultural conventions [6]. Put simply,

narratives can be highly personal, but are communicated in such a way that everyone understands them—certain ubiquitous attributes make narratives effective forms of communication.

I determine the effect of these attributes in scientific literature by measuring narrativity, defined as the extent to which an abstract features narrative attributes. While there is no single list of attributes that describes all narratives, the literature suggests that at the very least, narratives employ “causality, temporality, and character”, and are oftentimes accompanied by other optional nuanced attributes [2]. Therefore, narrativity is a “matter of degree” [7], and varies based on the presence of different combinations of attributes [7,8].

My research seeks to identify a role for narrative attributes in scientific literature. Specifically, I ask whether more narrative science writing is more influential, consistent with the theory that humans understand and remember narrative texts better. In order to determine narrativity, I distilled six measureable indicators from the literature, which reflect both the core and optional attributes described by narrative theory. I probed the abstracts of scientific peer-reviewed articles for these indicators, and measured their effect on article influence, as proxied by citation frequency. Ultimately, this research helps form a more comprehensive understanding of how scientists can effectively communicate their work.

2. Methods

2.1 Abstract Selection

I evaluated the abstracts of scientific papers based on the assumption that the abstract provides the point of entry to the paper; if a paper is cited in other publications, then at the very least, its abstract succeeded in garnering the attention of its readers. I used the database PubMed to determine the journals that published the greatest number of articles featuring the phrase “climate change” in the abstract or title between 2009 and 2012. I chose “climate change” as the overarching topic because it is a timely and rapidly growing field of inquiry, represented across a range of journals and citations. I identified the 20 journals with the greatest number of articles meeting these criteria (Table 1), and then retrieved the abstracts, citation counts, and other relevant information for articles in that set of journals through the database Web of Science (S1 Table; N = 1600 abstracts). To minimize the effect of publication year on citation counts, I included only articles published between 2009 and 2010. After removing outliers and accounting for missing data, the final dataset consisted of 675 abstracts. These abstracts had between 24 and 670 citations, with a median of 59.

Table 1: Source journals represented in dataset

Journal
1. Ambio
2. Ecological Applications: a Publication of the Ecological Society of America
3. Ecology
4. Ecology Letters
5. Environmental Management
6. Environmental Monitoring and Assessment
7. Environmental Science and Technology
8. Global Change Biology
9. International Journal of Biometeorology
10. Journal of Environmental Management
11. Molecular Ecology
12. Nature
13. Oecologia
14. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences
15. PLoS One
16. Proceedings of the National Academy of Science of the United States of America
17. Proceedings. Biological Sciences/The Royal Society
18. Science (New York, N.Y.)
19. The New Phytologist
20. The Science of the Total Environment

2.2 Crowdsourcing

I used the crowdsourcing site, CrowdFlower, to collect information regarding the narrativity of each abstract. Crowdsourcing is an efficient research tool for work that requires a degree of human assessment spread over a large number of data points, with access to a diverse, skilled workforce, and the production of reliable data in comparison to alternative methods [9,10]. This platform allowed me to: 1) collect information for a large number of abstracts; 2) collect multiple (n= 7) independent assessments (“judgments”) about the narrativity of each abstract; and simultaneously 3) include human interpretation and discretion in the quantification of narrativity. I collected multiple judgments for each abstract because achieving exact measurements is impossible; a variety of factors can influence results [11]. Rather than attempting to produce agreed-upon measures of narrativity, my methods capture and analyze the range of perceptions regarding narrativity.

Online contributors evaluated abstracts by first reading the instructions (S2 Text) and an example question, and then answering a series of six questions (S3 Text) for each abstract. These questions were intended to evaluate each abstract with respect to the indicators of narrativity (described in the next section). Contributors were paid per submitted page, each of which included five abstracts and the corresponding questions.

I used the following measures to ensure high quality responses: 1) gave access to this job only to CrowdFlower’s highest ranked contributors; 2) set a minimum completion time for each page of work; and 3) restricted contributor location to a number of countries in which English is the primary language (Australia, Canada, New Zealand,

United Kingdom, and United States). A total of 4805 contributors evaluated the abstracts used in this study.

2.3 Independent Variables: Narrative Indicators

I adapted methods and indicators based on comparable studies [8,12,13] and supported by relevant literature from various fields, including narrative theory [e.g. 7,14,15], psychology [3,16,17], communications [18], philosophy [19], and history [11]. I selected indicators as described below.

Indicator 1: Setting

One of the fundamental components of narratives is a description of where and when the events take place [14]. These spatial and temporal dimensions help create a mental image that distinguishes narratives from other forms of discourse [14]. I assessed this by asking contributors if there is a specific mention of place or time in the abstract [12].

Indicator 2: Narrative Perspective

Lejano et al. [8] acknowledge that at minimum, the presence of a narrator distinguishes narratives from other forms of communication—narrators tell narratives. The narrator is responsible for eliciting emotions in the reader [20], and first-person narrators have a stronger narrative presence than other narrative perspectives, such as third-person or no narrator [3,12]. I assessed narrative perspective by asking contributors whether or not the narrator referred to himself in the text (e.g., through use of pronouns such as we, I, and our).

Indicator 3: Sensory Language

Lejano et al. [8] argue that narrativity is in part determined by the existence of a personal identity. This personal identity can be established through the narrator expressing “emotions, attitudes, beliefs, and interpretations” [14]. Bilandzic et al. [12] identify this attribute through the use of language that appeals to the senses and emotions. Similarly, I assessed this attribute by asking contributors to count the number of times that sensory or emotional language appeared in the abstract.

Indicator 4: Conjunctions

A temporal or causal ordering of events is an essential, and distinguishing, characteristic of narratives [8,19,21,22]. This type of structuring guides the events in a way that makes rational and logical sense to the reader [14,15]. I used the presence of conjunctions to determine the extent to which an abstract is logically ordered, defining it according to Sydserff and Weetman [13]: “Conjunction...is concerned with the specific words or phrases...which function as links and bind narratives together...The presence of such devices distinguishes a narrative from what would otherwise be a collection of unconnected units”. I assessed this attribute by asking contributors to count the number of times that conjunctions appeared in the text. Specifically, I asked contributors to identify words or phrases that signify cause and effect, contrast, or temporal ordering [13].

Indicator 5: Connectivity

Meaningful connections in narratives can also be identified through explicitly related events or topics. I defined connectivity according to Sydserff and Weetman [13]: “...the degree to which each succeeding part of a narrative answers a question that could

have been generated out of an earlier part”. I assessed connectivity by asking contributors to count the number of times that words or phrases from one sentence were used to create an explicit link to the sentence immediately before it. I provided contributors the additional instruction to look for logical linkage between ideas. Words or phrases that create explicit links include either a specific reference back to the same thing or repetition of a word from the previous sentence, provided it carries the same meaning [13].

Indicator 6: Appeal

Appeal refers to the moral, evaluative orientation of a narrative. For example, Andrews et al. [22] suggest that narratives must turn to the reader and offer commentary or reflection about its events, in order for the reader to understand the crucial point of the story. Daiute and Lightfoot [15] describe a similar concept that they refer to as the “landscape of consciousness”. This indicator answers the question of *why* the story is being told. I assessed this attribute by asking contributors if the text makes an explicit appeal to the reader or a clear recommendation of action [12].

2.4 Dependent Variable: Citation Frequency

I used citation frequency as a measurement of article influence. There is a wide body of literature supporting the use of citation analyses as frameworks for evaluating science communication [23–26]. Key to understanding the practical uses of citation analyses is what Lawani and Bayer [23] describe as “...the cumulative nature of science, with later work generally building upon earlier scholarship”. Citations reflect the extent to which a piece of work is represented in a body of literature, and can therefore be used as to evaluate its impact. I also collected information on publication year, source journal,

number of authors, and abstract length (number of words), all factors that are known to influence the citation rate of peer-reviewed scientific publications [24,26,27].

2.5 Quality Control

I treated question 2, “Does the narrator refer to himself in the text?” as a test question, or secondary quality-control mechanism, due to its objectivity (i.e., unlike some of the other narrative indicators, the existence of a first-person narrator has a “true” answer). After considering all seven responses for this question, contributors who responded in the majority were included in the analysis, whereas responses in the minority were assumed to be in error and entirely omitted from the analysis. This helped ensure that the analysis include only responses from trustworthy contributors. I took the mean of the remaining trusted responses to generate a dataset with one value per indicator for each abstract.

I detected outlier abstracts by fitting a normal probability distribution to the independent variables “conjunctions” and “connectivity”, and excluding responses with a < 0.01 likelihood. Consequently, abstracts with very large (or small) numbers of conjunctions or connective phrases were removed from the dataset. I similarly removed abstracts with extreme values of word count. Outliers for the independent variable “sensory language” were manually removed because this variable was not normally distributed; the majority of abstracts were evenly distributed across a specific range of values, so abstracts on the extreme ends were removed. A total of 50 outliers were removed from the dataset. Finally, I removed any incomplete rows. This left 675 abstracts in the dataset.

Questions with “yes/no” categorical responses (i.e., the indicators “setting”, “narrative perspective”, and “appeal”) were assigned the value 0 for “no” and 1 for “yes”, and the means were analyzed on a 0 to 1 scale. The means for the questions that required contributors to count (i.e., the indicators “conjunctions”, “connectivity”, and “sensory”) were analyzed on a continuous scale, starting at 0. This turned an otherwise discrete variable into a continuous variable, creating an index that captured variations in perceptions of narrativity. For example, contributors might count different numbers of connective phrases and links in a piece of text. Taking the mean, and thereby including the disagreement among responses, produced an overall measure of perceived connectivity for that piece of text. These methods incorporated the subjective nature of narrativity into the results.

3. Analyses

I sought to identify relationships between individual indicators of narrativity and citation frequency, as well as the unique combinations of indicators that led to higher citation outcomes. The primary analyses included in this study are described in the next sections. Supplementary analyses are included in the Supporting Information section (S4 Figure-S5 Figure). Ordination techniques did not succeed in separating groups according to citation frequency, and therefore are not included in the analysis.

3.1 Linear Regression

I used the linear model function in R [28] to evaluate the relationship between individual independent variables, or indicators, and the log of the dependent variable, citation frequency. I used a log transformation in order to see potential linear relationships in the data [29]. Significant linear relationships ($p < 0.05$) were expressed with an R-squared value and a p-value, to characterize the strength of the linear association between the independent and dependent variable and the level of statistical significance.

3.2 Nonlinear Model

I fit a second-order nonlinear model to data with apparent nonlinear trends, and compared this to the fit of a linear regression for the same data; I used the log likelihood ratio test in R [28] to compare the goodness of fit. The log likelihood ratio is used to compute a test statistic (Chi-square distribution with one degree of freedom), with an associated p-value. From this, I determined whether or not to reject the null model (linear) in favor of the alternative model (nonlinear).

3.3 Regression Tree

I used a regression tree analysis to determine combinations of indicators that best predict article citation frequency. The regression tree model weighed all of the predictors of article citation frequency used in this study (the six independent variables and other factors, including source journal, abstract length, and number of authors), and selected the predictor and cutpoint value that minimized the tree's residual sum of squares (RSS) [30]. (RSS is a measure of discrepancy between the data and the model's predicted values, so a lower RSS indicates a closer fit). The predictor and cutpoint with the lowest RSS is represented at the top of the tree, and represent the most important factor in predicting citation frequency.

The remaining predictor regions were further subdivided through the same process, resulting in a series of cutpoints, or internal nodes, and final predictor spaces, or terminal nodes. The final tree is a simplified, "pruned" version of itself, with the number of terminal nodes determined by the minimum cross-validation error [30]. The citation predictions are expressed at terminal nodes, located at the bottom of the model. These predictions are the mean of the responses for the observations in that terminal node; the number of observations is indicated in parentheses.

4. Results

4.1 Individual Indicators of Narrativity

The relationship between individual independent variables, or indicators, and the log of the dependent variable, citation frequency, are shown in Fig.1. Regression lines are included on plots where there is a significant linear relationship between the independent and dependent variable ($p < 0.05$). In general, these results support the hypothesis that attributes associated with narrativity are positively associated with article citation frequency, with two exceptions: narrative perspective (negatively associated), and conjunctions (unimodally associated). I obtained similar results when holding the year of publication on the dataset constant, suggesting that there is not a strong effect of year on the dataset (S6 Figure); this is confirmed in the next section. The coefficient of determination (R-squared value) and p-value of statistical significance for each indicator are summarized in Table 2.

Figure 1: Linear regression for individual indicators of narrativity

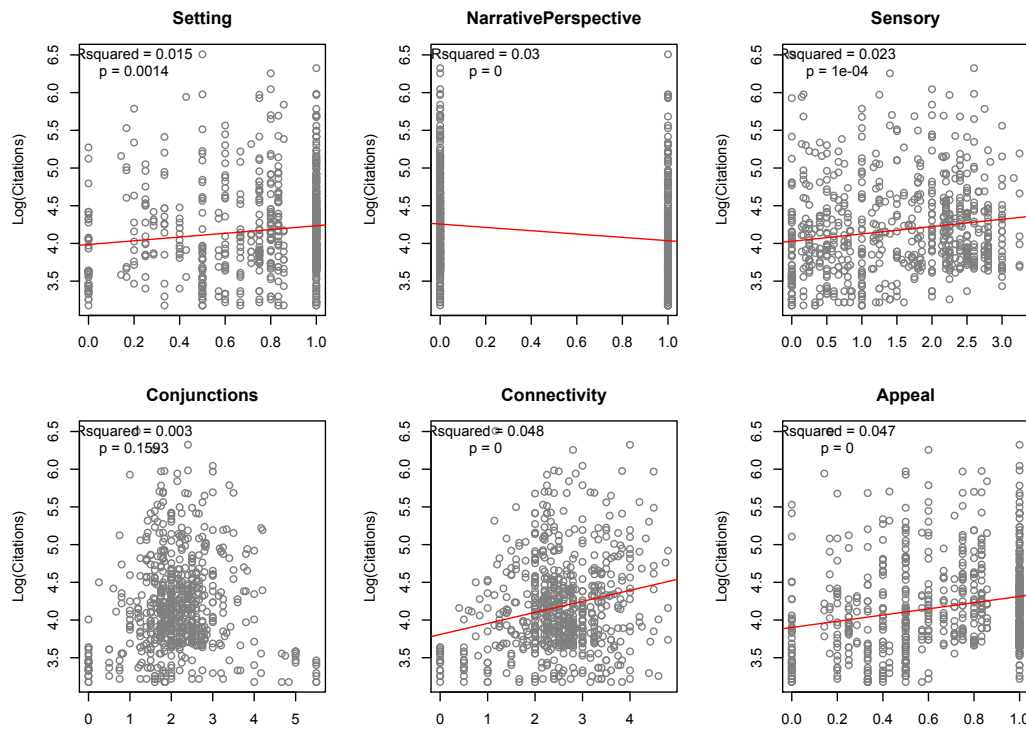


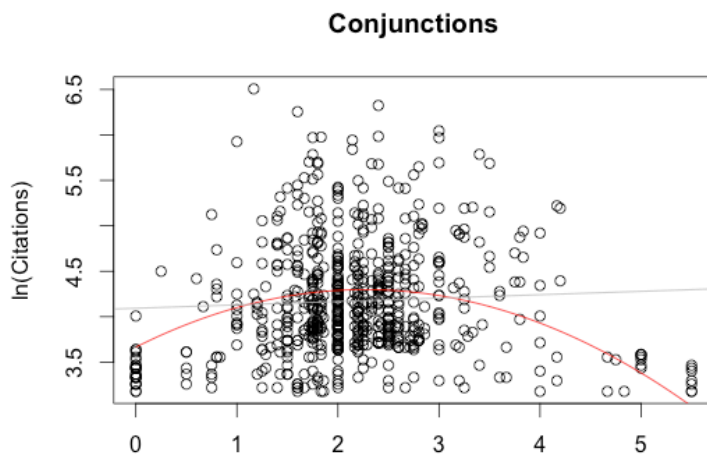
Table 2: Coefficient of determination and statistical significance for each indicator

	R-squared	p-value
Setting	0.015	0.0014
Narrative Perspective	0.03	<0.00001
Sensory	0.023	0.0001
Conjunctions	0.003	0.1593
Connectivity	0.048	<0.00001
Appeal	0.047	<0.00001

As the individual plots depict, I found positive associations ($p < 0.01$) between the indicators “setting”, “sensory”, “connectivity”, and “appeal”, and number of citations. Conversely, I found a negative association ($p < 0.01$) between the indicator “narrative perspective” and number of citations, suggesting that abstracts with first person narrators are cited less often than abstracts with third person narrators or no narrator. I found a

strong nonlinear association between the indicator “conjunctions” and number of citations (Fig. 2, with the linear regression model for comparison). The log likelihood ratio test produced statistically significant results, $p \ll 0.0001$, offering strong support that the nonlinear model is a better fit for the data than the linear model. These results suggest that as the number of conjunctions in an abstract increase, citations rise and then fall. In other words, at a certain point an abstract can feature too many conjunctions, which negatively influences citations.

Figure 2: Nonlinear model for conjunctions and log citations



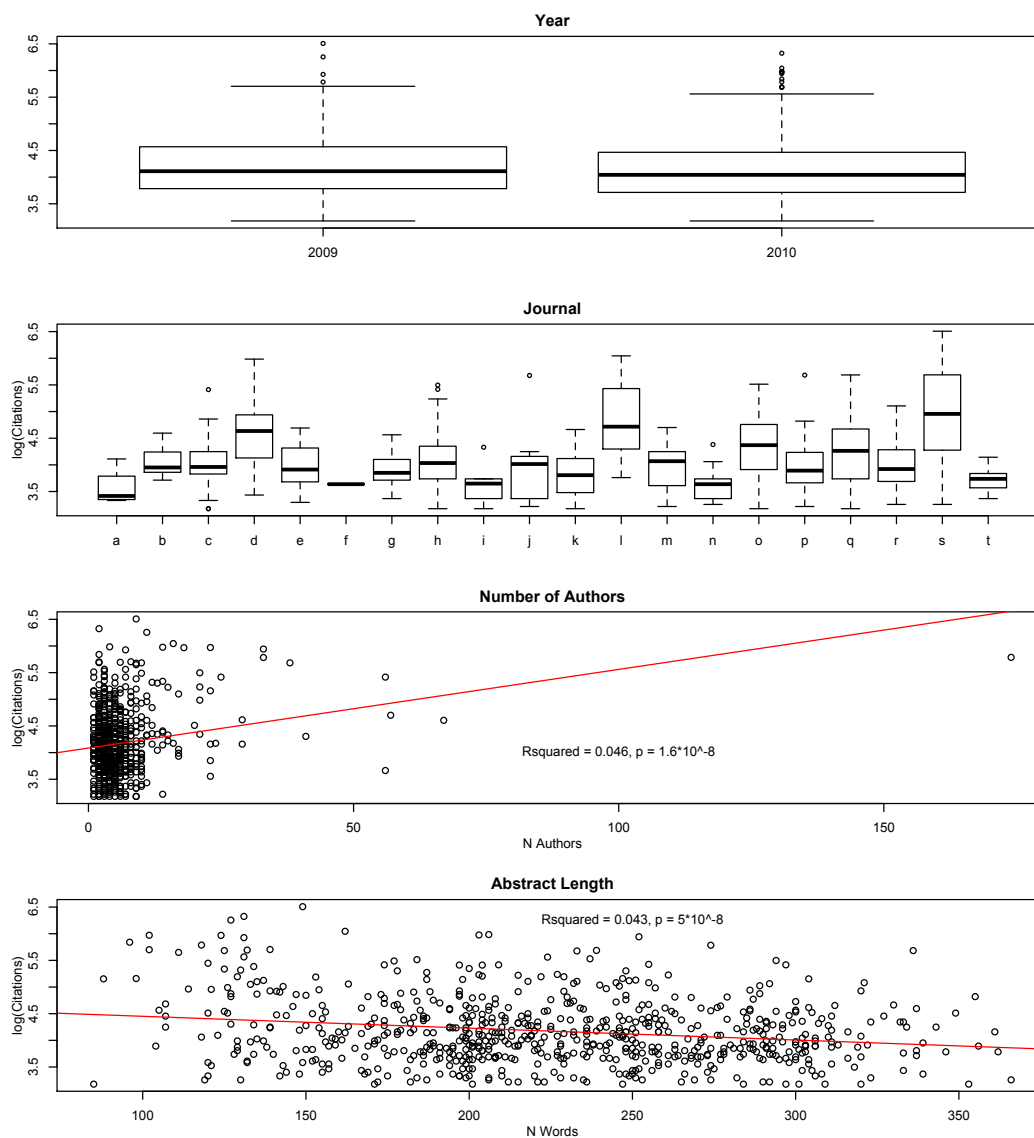
4.2 Non-narrative Independent Variables

I used box plots and linear regression to evaluate the relationship between non-narrative independent variables and citation frequency (Fig. 3). The year of publication did not have a significant effect on citations ($p > 0.05$). The source journal did have a significant effect on citations ($p < 0.01$, $R\text{-squared} = 0.25$). There was a significant positive linear association between the number of authors and citations ($p < 0.01$, $R\text{-$

squared=0.046), indicating that articles with more authors tend to receive more citations.

This trend was not driven by the outlier(s); removal of papers with N authors > 20 produced a significant positive association. There was a significant negative linear association between abstract length (word count) and citation frequency ($p < 0.01$, R -squared=0.043), where shorter abstracts tend to receive more citations.

Figure 3. Effect of non-narrative variables on log citations



4.3 Regression Tree for Predicting Citation Outcomes

The regression tree (Fig. 4) stratifies the data into seven regions of predictor space for citation frequencies. The predicted numbers of mean citations for each region are located at the terminal nodes; the number in parentheses corresponds to the number of training observations used to determine the prediction value. The tree can be interpreted as follows: the source journal is the most important factor in an article determining the number of citations, and articles published in *Nature*, *Science*, or *Ecology Letters* tend to be more highly cited. Among articles published in *Nature*, *Science*, or *Ecology Letters*, those with fewer than 13 authors are cited less than articles with 13 authors or more. For those articles not published in *Nature*, *Science*, or *Ecology Letters*, the number of authors plays little role in determining citations. Rather, the indicator “connectivity” comes into play, and articles with low connectivity (connectivity < 1.1) tend to be cited less than articles with high connectivity (connectivity ≥ 1.1). Among those articles with high connectivity, articles published in *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* or *Proceedings of the National Academy of Science of the United States of America (PNAS)* tend to be more highly cited. Among the articles not published in these journals, the indicator “appeal” plays a role in receiving citations, and articles with low appeal (appeal < 0.6) tend to be cited less than those with higher appeal (appeal ≥ 0.6). Among articles with higher appeal, articles with fewer than 19 authors tend to be cited less than articles with more than 19 authors. In sum, the independent variables “connectivity” and “appeal” influence citation frequency in articles published in lower-rank journals. The four left-most terminal nodes correspond to the list of journals in Table 3. In sum, the six node splits included in this regression tree explain

about 34% of the variance in the dataset. Additional information regarding the predictive accuracy of the regression tree is summarized in S7 Fig.

Figure 4: Regression tree for predicting citation outcomes

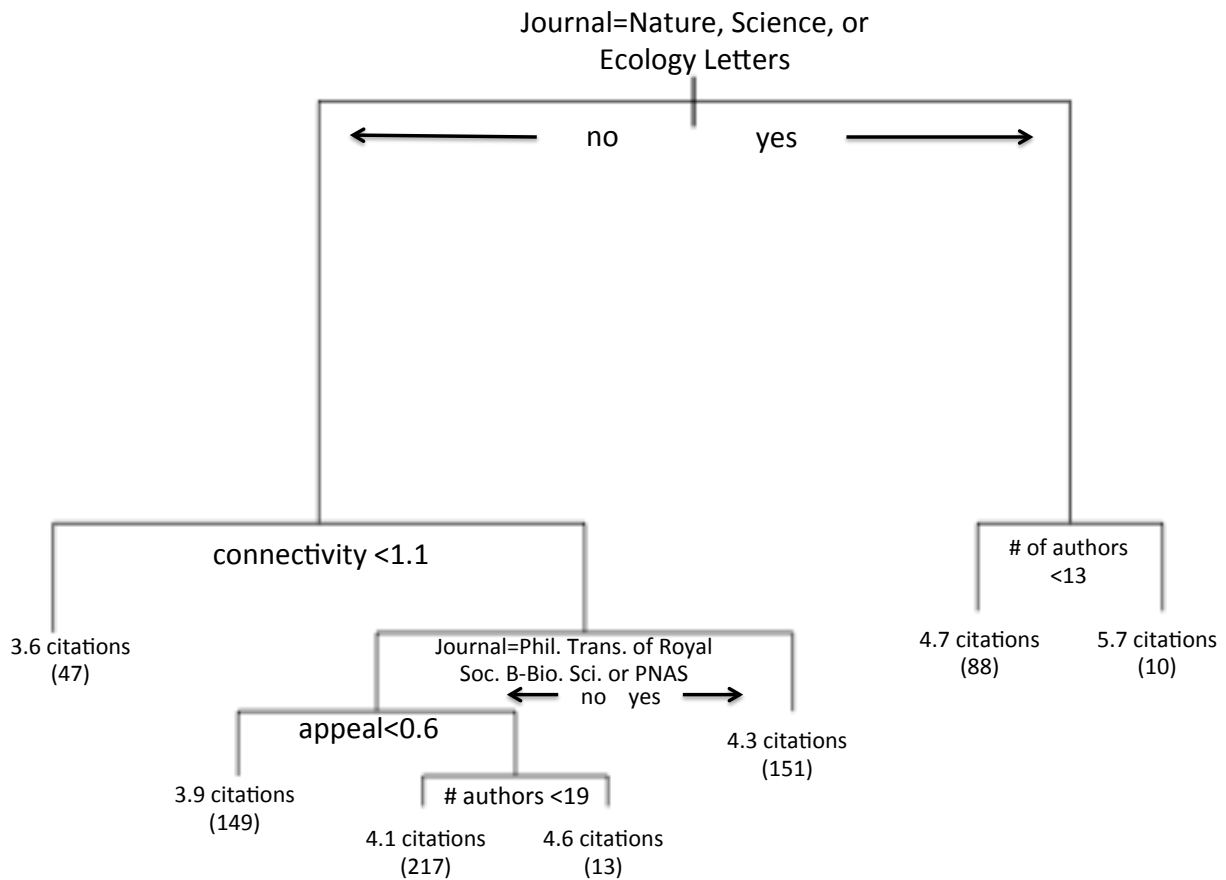


Table 3: Journals represented in left-most terminal nodes (lower citation predictions)

Journal
1. Ambio
2. Ecological Applications: a Publication of the Ecological Society of America
3. Ecology
4. Environmental Management
5. Environmental Monitoring and Assessment
6. Environmental Science and Technology
7. Global Change Biology
8. International Journal of Biometeorology
9. Journal of Environmental Management
10. Molecular Ecology
11. Oecologia
12. PLoS One
13. Proceedings. Biological Sciences/The Royal Society
14. The New Phytologist
15. The Science of the Total Environment

5. Discussion

5.1 Individual Indicators of Narrativity

I found that all six indicators of narrativity have significant effects on citation outcomes. The linear regression model suggests that abstracts that rank higher for “setting”, “sensory”, “connectivity”, and “appeal” are more highly cited, consistent with the hypothesis that attributes associated with narrativity are positively correlated with article citation frequency. I speculate that this relationship is facilitated by the ability of these indicators to provide an experience that is relatable, and therefore more understandable, to the reader. For example, a mention of place or time in an abstract provides a situational context that the reader can conceptualize and more readily derive meaning from. Similarly, the use of language to describe physical and emotional observations establishes a sense of personal identity in the abstract; the use of words or phrases to create explicit links between sentences more accurately reflects the way humans experience the world, as a continuous story rather than a series of disconnected events; an appeal to the reader about the significance of the work or a recommendation of action resonates with a common morality in humans. I believe by reflecting a universally understood human experience, abstracts succeed at both drawing the reader in and effectively conveying significance. To the reader, the relevance of the work is clear, and therefore more citable.

Conversely, the linear regression model suggests that abstracts with first person narrators are less often cited than abstracts with no narrator or third person narrators. These results are surprising because first person narration is associated with a stronger narrative presence, and therefore provide a perspective that is presumably easier for the

reader to take on [17]. It is possible that these results reflect differences in disciplinary preferences; the abstracts evaluated in this study cover multiple disciplines, and a similar study on the effect of common scientific writing advice found that citation outcomes can vary depending on the field [27]. Another study found that in order for readers to embody the perspective of the narrator, they must be readily introduced as the subject of the text [16]. Thus, the amount of contextual information provided in an abstract may affect the ability of the reader to adopt the internal perspective that is typically induced through the use of first person pronouns.

The linear regression model failed to explain the trend in the data for conjunctions. Instead, I found a unimodal association between conjunctions and citation outcomes, with the maximum number of citations achieved when an abstract features between two and three conjunctions. These results suggest that after that point, an abstract can feature too many conjunctions. This makes intuitive sense; if conjunctions help logically order events through cause and effect, contrast, and temporal ordering, a high number of conjunctions indicates a high number of events, which become difficult to keep track of. My results reveal a sweet-spot for comprehending the events in an abstract, after which an abstract can simply have too much going on.

5.2 Non-narrative Variables

I found that an abstract's source journal has a strong influence on citation outcomes, explaining about 25% of the variance in citations. Furthermore, the source journal has a significant interaction with abstract length, indicating that these two variables are not independent of each other. These findings suggest that an abstract's

length is often determined by the journal and its publication requirements, and the most highly cited journals tend to have shorter abstracts. The number of authors is not influenced by the source journal, indicating that abstracts with more authors tend to be more highly cited regardless of the journal of publication.

5.3 Unique Combinations for Predicting Citation Outcomes

Consistent with the previous findings, I found the journal of publication has the greatest influence on citation frequency. That is, for articles published in *Nature*, *Science*, or *Ecology Letters*, narrativity has little impact. Instead, the number of authors has a positive influence on citation outcomes for articles published in these journals, consistent with findings of bibliometric studies [24]. However, for articles published in lower rank journals, specific attributes associated with narrativity do have a significant effect on citation outcomes.

After journal of publication, “connectivity” is the greatest predictor of citation outcomes, with a higher degree of connectivity leading to more citations. This is a significant and interesting finding because it suggests that in many journals, narrativity is an important factor; the extent to which an abstract features explicitly related events or findings is even more important than publication in many top-tier journals, or other factors such as abstract word count. Furthermore, the presentation of scientific information typically reflects deductive reasoning, where accepted truths retain their meaning regardless of surrounding content [2]. This finding provides support for a more narrative-based, inductive reasoning approach to science communication, where individual units must be connected to each other in order to coherently convey a message.

The regression tree analysis indicates that after connectivity, publication in the journals *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* or *Proceedings of the National Academy of Science of the United States of America (PNAS)* influences citation outcomes. Among articles not published in either of these top-tier journals, the indicator “appeal” is an important predictor of citation frequency, where articles that more clearly convey the significance of the work or make a recommendation of action are cited more often. This is also an interesting finding because scientists receive conflicting advice about signaling importance [27], and this finding suggests that doing so is one of the top predictors of achieving citations in journals others than those in the top tier.

In sum, two indicators of narrativity lead to higher citation outcomes: “connectivity” and “appeal”. These indicators represent attributes that may not be top priorities for scientists when performing research or reporting findings. However, these attributes are particularly important for articles published in lower rank journals, which account for a large proportion of published science. In other words, although narrativity may not matter at a certain caliber of recognition by the academy, as demonstrated by publication in *Science*, *Nature*, or *Ecology Letters*, it does improve citation outcomes for articles published in numerous other journals. Specifically, my results suggest that scientists benefit from incorporating well-connected events and an appeal to the reader into their abstracts.

6. Conclusion

The abstract is a scientist's lure. It is the only section of the paper immediately available on databases such as PubMed [31], and therefore has the potential to hook the reader, enticing them to pursue, and cite, the article itself. Abstracts are also succinct, and typically must conform to a structure specific to the journal of publication [31]. The challenge is to write an abstract that meets those content requirements, while also alluring the reader with its relevance and style.

In terms of style, peer-reviewed, scientific discourse is often viewed as its own form of communication, exempt from the qualities of narratives that humans inherently relate to. However, my findings support an alternative viewpoint: scientists can draw in readers and achieve more citations by incorporating narrative attributes into their abstracts. The most important attribute to consider is the abstracts' connectedness, or the extent to which the events in an abstract are related. By referencing aspects of previous sentences, scientists can tie their work together into a coherent story that the reader easily follows. Scientists can also feature evaluative commentary in order to communicate significance and to offer a sense of moral grounding that readers inherently identify with. Scientists might also consider conjuring up a sense of time or place in their abstracts, and using more descriptive, sensory language to establish a personal identity. By incorporating these attributes into their writing, scientists mirror the way we as humans experience the world, and in turn, captivate their audience and extend their reach.

7. Supporting Information

S1 Table. Information collected for each abstract from Web of Science

PMID
Title
Abstract
Author(s)
Journal
Year
Abstract word count
Times cited

S2 Text. CrowdFlower job instructions

Identify Attributes of Scientific Writing

Overview

In this job, you will be presented with a piece of text from a scientific article. After reading the text, you will be asked to identify various attributes of scientific writing. Your input will be used to help form a better understanding of science communication.

Process

Read the text and answer each question. You may wish to review the questions before reading the text for the first time.

Do:

Read each question carefully. For each question, read the text at least once. Refer to the text as often as necessary. For questions that require you to count, you may wish to have a pencil and paper handy to keep running tallies.

Do not:

Do not read the text only once. Please do not randomly guess answers!

Summary:

You will read the text and identify the attributes that the questions describe. Pay close attention to the instructions for each question, and refer to the text as often as necessary.

Thank You!

Your input is valuable to us. Thank you for taking the time to complete this job!

S3 Text. CrowdFlower job questions

Question 1: Is there mention of a specific place or time?

(Yes/No)

Question 2: Does the narrator refer to himself in the text? Look for pronouns such as we, our, I, and my.

(Yes/No)

Question 3: Count the number of times that sensory or emotional language is used. Look for words that represent sight, sound, smell, taste, touch, or feelings.

(Select a number)

Question 4: Count the number of times that conjunctions are used in the text. Look for words or phrases that signify cause and effect (e.g. consequently; therefore; as a result; so; for this reason), contrast (e.g. however; although; surprisingly), or temporal ordering (e.g. then; next; first; second).

(Select a number)

Question 5: Count the number of times that words or phrases from one sentence are used to create an explicit link to the sentence immediately before it (e.g. these benefits; this finding; or 'costs...the costs are...'; 'trend...this trend indicates...'; 'profit...a significant contribution to profit...').

(Select a number)

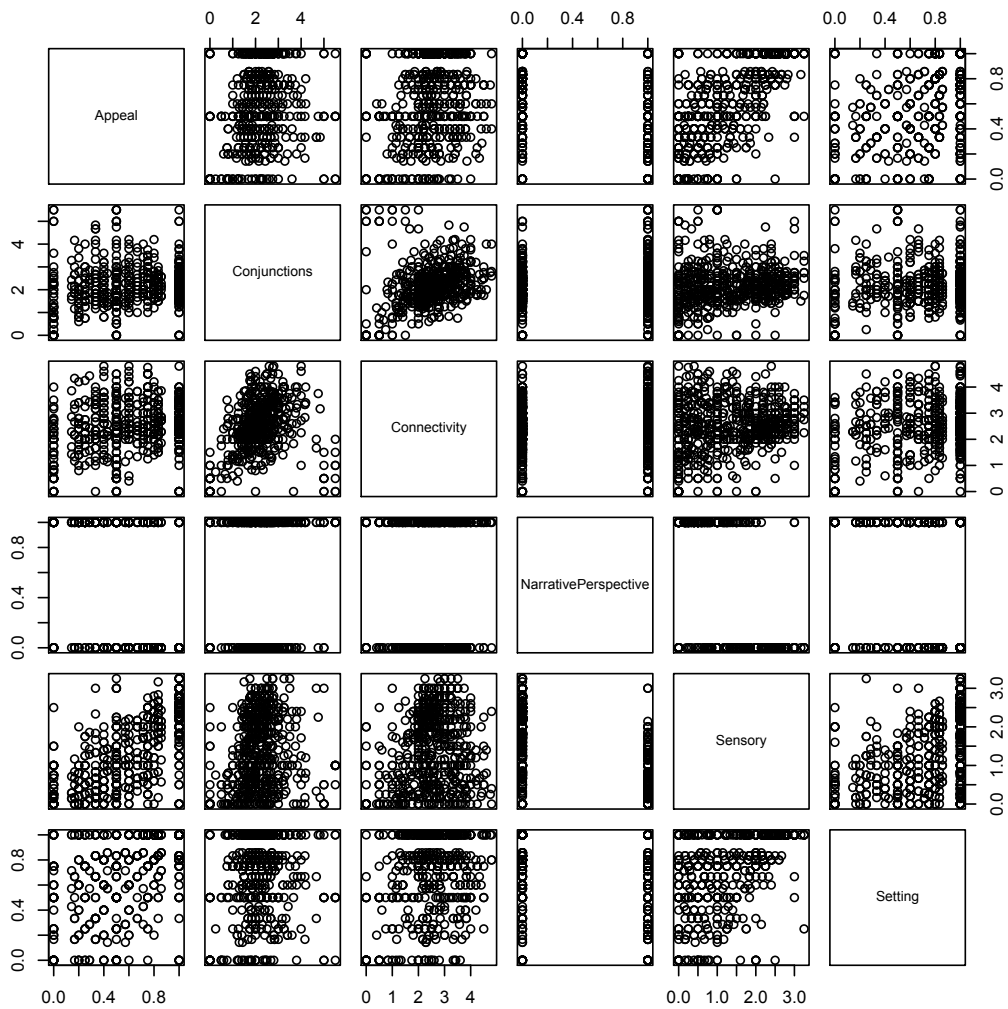
Tip: Look for logical linkage between ideas. Words or phrases that create explicit links include either a specific reference back to the same thing or repetition of a word from the previous sentence, provided it carries the same meaning.

Question 6: Does the text make an explicit appeal to the reader or a clear recommendation of action (e.g. We hope conservation professionals value the potential of...)?

(Yes/No)

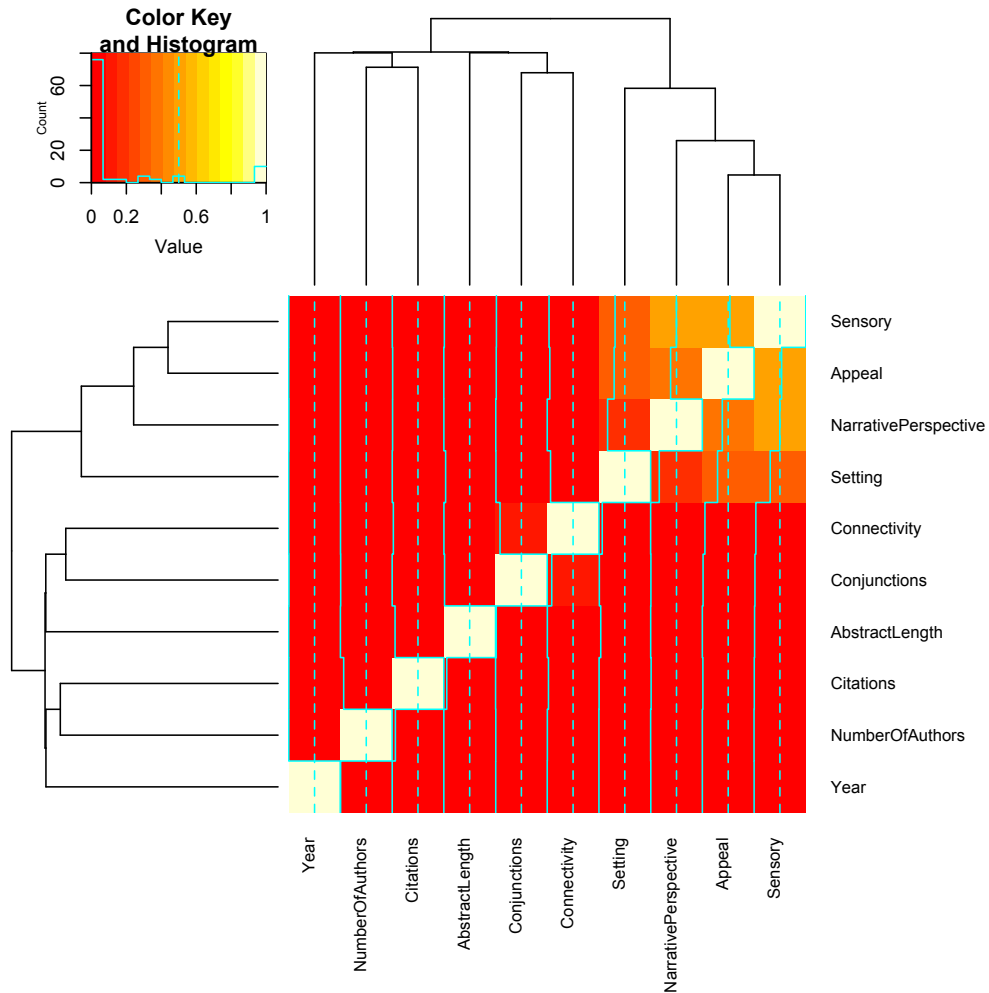
S4 Figure. Results for correlations between narrative attributes

I used a pairwise correlation matrix to depict the correlations between narrative indicators. Each indicator is perfectly correlated with itself, therefore those plots are replaced with the indicator's name; the plot is symmetrical along this diagonal line of names.



S5 Figure. Heat map

The heat map visualizes the correlations between indicators with colors instead of numbers. The indicators “sensory”, “appeal”, “narrative perspective”, and “setting” are more highly correlated with each other than with the indicators “connectivity”, “conjunctions”, and other factors, including abstract length, citations, number of authors, and year of publication. However, there are similarities within these groupings, which the dendrogram depicts.

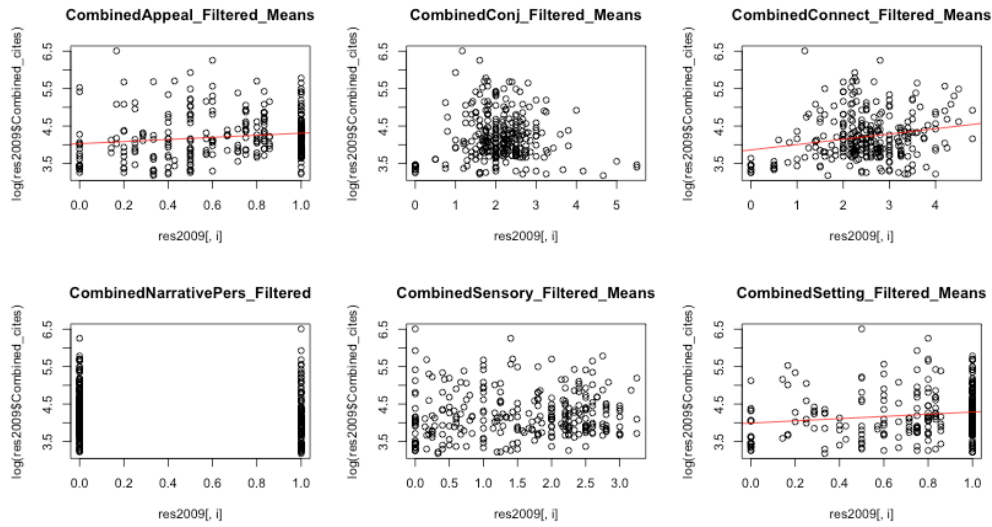


S6 Figure. Results for individual indicators of narrativity, holding year constant.

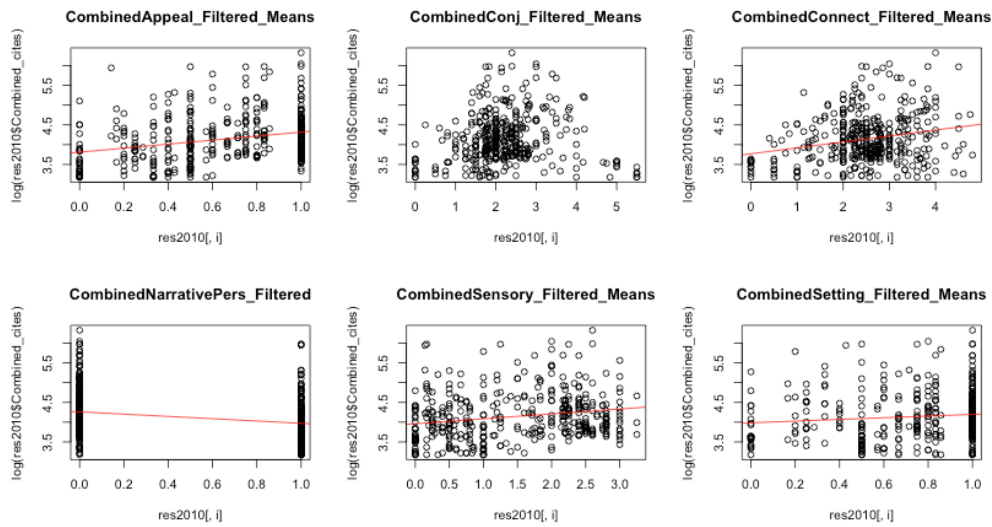
The year of publication did not have a significant effect on citations ($p>0.05$).

Further, similarities between the two datasets, depicted below, support the finding that the year of publication did not confound the results in this study.

2009



2010



S7 Figure: R output for predictive accuracy of regression tree (“printcp” function)

This table summarizes the predictive accuracy of the regression tree [28]. The column labeled “CP” refers to “complexity parameter”, and controls the number of nodes in the tree. When the cost of adding a node split to the tree is greater than CP, tree building discontinues in R. The column labeled “nsplit” refers to “node split”. As the figure indicates, the regression tree has six node splits, each with two corresponding error rates. “Rel error” refers to relative error, or the error rate for predictions with the data used to generate the tree. “X error” refers to the cross-validated error, or the error rate for predictions with new data. “Xstd” refers to the cross-validated standard deviation, another measurement used to determine tree growth; R chooses the lowest level where $xstd + rel\ error < xerror$. The error rates in row 7 indicate that the six nodes included in this regression tree have a relative error of 66.24%, and a cross-validated error rate of 77.73%. In sum, the six node splits included in this regression tree explain about 33.76% of the variance in the dataset.

	CP	nsplit	rel error	xerror	xstd
1	0.179772	0	1.00000	1.00045	0.062138
2	0.046483	1	0.82023	0.82447	0.045270
3	0.037341	2	0.77375	0.81946	0.046762
4	0.033518	3	0.73640	0.80685	0.046833
5	0.027479	4	0.70289	0.78253	0.046617
6	0.012986	5	0.67541	0.79089	0.051439
7	0.012713	6	0.66242	0.77729	0.050927

8. References

1. Edward O. Wilson. *The social conquest of earth*. 1st ed. New York: Liveright PubCorp; 2012.
2. Dahlstrom MF. Using narratives and storytelling to communicate science with nonexpert audiences. *Proc Natl Acad Sci U S A*. 2014;111: 13614–13620. doi:10.1073/pnas.1320645111
3. Graesser AC, Olde B, Klettke B, Green M, Strange J, Brock T. How does the mind construct and represent stories? *Narrative Impact: Social and cognitive foundations*. Hoboken: Taylor & Francis; 2002. pp. 229–262.
4. Ingram H. The Power of Narratives in Environmental Action: New Approaches. *Environ Sci Policy Sustain Dev*. 2014;56: 4–5. doi:10.1080/00139157.2014.881698
5. Kelly RP, Cooley SR, Klinger T. Narratives can motivate environmental action: the Whiskey Creek ocean acidification story. *Ambio*. 2014;43: 592–9. doi:10.1007/s13280-013-0442-2
6. Atkinson P, Delamont S. Rescuing narrative from qualitative research. *Narrat Inq*. 2006;16: 164–172. doi:10.1075/ni.16.1.21atk
7. Herman D, Jahn M, Ryan M-L. *Routledge encyclopedia of narrative theory*. London: New York: Routledge; 2005.
8. Lejano R, Ingram M, Ingram H. *The Power of Narrative in Env. Networks*, ch. 3. 2013.
9. De Clercq O, Hoste V, Desmet B, Oosten PV, De Cock M, Macken L. Using the Crowd for Readability Prediction. *Nat Lang Eng*. 2014;1: 59. doi:10.1017/S1351324912000344
10. Behrend TS, Sharek DJ, Meade AW, Wiebe EN. The viability of crowdsourcing for survey research. *Behav Res Methods*. 2011;43: 800–813. doi:10.3758/s13428-011-0081-0
11. Gross AG, Harmon JE, Reidy M. *Communicating Science: The Scientific Article from the 17th Century to the Present*. New York, NY: Oxford University Press; 2002.
12. Bilandzic H, Dahlstrom M, Busselle R, Wagner A. Exemplars, anecdotes or narratives? A meta-analysis of narrativity in exemplification research and narrative persuasion. 2014.

13. Sydserff R, Weetman P. A texture index for evaluating accounting narratives: An alternative to readability formulas. *Account Audit Account J.* 1999;12: 459–488.
14. Holstein J, Gubrium JF. *Varieties of narrative analysis.* Los Angeles [Calif.]: Sage; 2012.
15. Daiute C, Lightfoot C. *Narrative analysis studying the development of individuals in society.* Thousand Oaks, Calif: Sage; 2004.
16. Brunyé TT, Ditman T, Mahoney CR, Augustyn JS, Taylor HA. When you and I share perspectives pronouns modulate perspective taking during narrative comprehension. *Psychol Sci.* 2009;20: 27–32.
17. Mano Y, Harada T, Sugiura M, Saito DN, Sadato N. Perspective-taking as part of narrative comprehension: A functional MRI study. *Neuropsychologia.* 2009;47: 813–824. doi:10.1016/j.neuropsychologia.2008.12.011
18. Dahlstrom MF. The moderating influence of narrative causality as an untapped pool of variance for narrative persuasion. *Commun Res.* 2015;42: 779–795.
19. Fisher WR. Narrative rationality and the logic of scientific discourse. *Argumentation.* 1994;8: 21–32. doi:10.1007/BF00710701
20. Peterson EE, Langellier KM. The performance turn in narrative studies. *Narrat Inq.* 2006;16: 173–180.
21. Norris SP, Guilbert SM, Smith ML, Hakimelahi S, Phillips LM. A theoretical framework for narrative explanation in science. *Sci Educ.* 2005;89: 535–563. doi:10.1002/sce.20063
22. Andrews, M. Andrews, M., Squire, Corinne, & Tamboukou, Maria. (2008). *Doing narrative research.* Los Angeles [Calif.] ; London: SAGE.
23. Lawani SM, Bayer a E. Validity of citation criteria for assessing the influence of scientific publications: new evidence with peer assessment. *J Am Soc Inf Sci.* 1983;34: 59–66. doi:x
24. Hurley L a., Ogier AL, Torvik VI. Deconstructing the collaborative impact: Article and author characteristics that influence citation count. *Proc ASIST Annu Meet.* 2013;50. doi:10.1002/meet.14505001070
25. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A.* 2005;102: 16569–16572. doi:10.1073/pnas.0507655102
26. Tijssen RJW, Raan a FJV, Netherlands T. Net Citation Balances: A Measure of Influence between Scientific Journals. *Science.* 1990;41: 298–304.

27. Weinberger CJ, Evans J a., Allesina S. Ten Simple (Empirical) Rules for Writing Science. PLOS Comput Biol. 2015;11: e1004205. doi:10.1371/journal.pcbi.1004205
28. R Core Team. R: A language and environment for statistical computing. [Internet]. R Foundation for Statistical Computing, Vienna, Austria.; 2014. Available: <http://www.R-project.org/>
29. McDonald JH. Data transformations - Handbook of Biological Statistics [Internet]. [cited 16 May 2016]. Available: <http://www.biostathandbook.com/transformation.html>
30. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning [Internet]. New York, NY: Springer New York; 2013. Available: <http://link.springer.com/10.1007/978-1-4614-7138-7>
31. Andrade C. How to write a good abstract for a scientific paper or conference presentation. Indian J Psychiatry. 2011;53: 172. doi:10.4103/0019-5545.82558