Science Communication in Online Communities

A case-study on r/science

Dan Hickey, Vetle Jahr, Jared Mantell, Julian Strietzel (Team Space)

Background

- Clear communication of scientific research is essential to help the public make informed decisions
- Good science communication is challenging! Ideally, science communication should be:
 - Free of jargon
 - Factually accurate
 - Not overly sensationalized





NBC NEWS NOW

YouTube radicalization pipeline exists, study says

Example: media will often sensationalize or misrepresent scientific findings.

Mural lymphatic endothelial cells regulate meningeal angiogenesis in the zebrafish

Neil I Bower, Katarzyna Koltowska, Cathy Pichol-Thievend, Isaac Virshup, Scott Paterson, Anne K

Lagendijk, Weili Wang, Benjamin W Lindsey, Stephen J Bent, Sungmin Baek, Maria Rondon-Galeano,

Daniel G Hurley, Naoki Mochizuki, Cas Simons, Mathias Francois, Christine A Wells, Jan Kaslin & Benjamin

M Hogan

M Hogan

Nature Neuroscience 20, 774-783 (2017) Cite this article

Scholars may also have difficulty translating their work into something that can be easily understood by non-experts

Research Questions

- What are the qualities of the science communication that is rewarded on social media?
- We use r/science as a case study, as all posts discuss peer-reviewed literature.

Namely, how do...

- Jargon
- Factual accuracy
- Sensationalism

... of Reddit posts relate to the number of comments the posts receive?

UC Berkeley



VIEW SORT O HOT " TOP POSTS PROM THE PAST HOWTH r/science Posted by u/myea 7 days ago Posted by WSteRePo 23 days ago Woman with 'mutant' gene who feels Human-raised wolves are just as 5.7k no pain and heals without scarring successful as trained dogs at working This community is a place to share and discuss new scientific research. Read about the latest advances in astronomy, biology, medicine, physics, social science, and more. Find and submit new publications and popular science coverage of current 66.2k points + 1.5k comments 100k points + 3.3k comments Join the 21.0m people in the r/science GET STARTED Caching increases read performance by making the Posted by u/Nass1m01973 10 hours ago 17.9k Control Routine vaccination of girls aged 12 or 13 years with the human papillomavirus (HPV) vaccine in Scotland has led to a dramatic reduction in cervical disease in later life, finds a new study TRIPLEBYTE 392 Comments A Share Save ... Posted by u/mives. MD-PhO-MBA | Clinical Professor/Medicine: 4 hours ago 1. Must be peer-reviewed research People who 'microdose' psychedelic drugs for a month say it increases positive emotions and 2. No summaries of summaries, rehosts. productivity, finds a study of more than a thousand reviews, or reposts volunteers from 59 countries. Increased energy, improved work effectiveness, and improved health 3. No editorialized, sensationalized, or habits were observed in clinical/ non-clinical The title and content of submissions populations. should not be editorialized. sensationalized, or biased. All titles must adhere to our headline rules. ■ 135 Comments 🖈 Share 🗓 Save ···

Research design

Use metadata such as the category of research to account for potential confounders

Measure **jargon**, **sensationalism** in the title

Intelligent men exhibit stronger commitment and lower hostility in romantic relationships | There is also evidence that intelligence supports self-regulation—potentially reducing harmful impulses in relationships.

Psychology

https://www.psypost.org/intelligent-men-exhibit-stronger-commitment-a...

u/chrisdh79 • 5 hr. ago

Q Q

Share

UC Berkeley and per

Use number of comments and votes as a measure of performance

Scrape paper abstract from URL, measure **factual consistency** between abstract and title of post

Research design

- The subreddit r/science provides us with a wealth of social media posts that directly reference science.
- We measure the jargon, sensationalism, and factual consistency of each post.



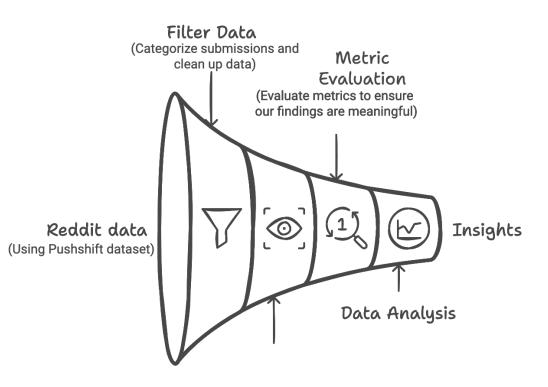
Research questions

Our measures of jargon, sensationalism, and factual consistency in titles indicate the quality of science communication

- RQ1: How does science communication quality differ by the type of source used? (e.g., news article, research paper)
- RQ2: Is high-quality science communication favored by users?
- **RQ3:** How has the quality of science communication changed over time?

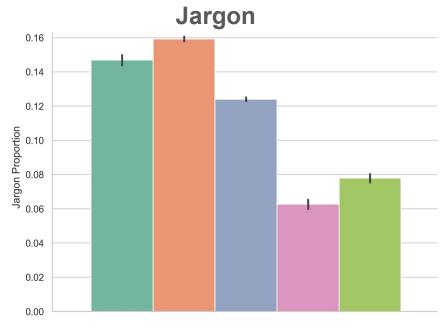
Progress

- Reddit data
- 🔽 Filter and clean up data
- Extract abstracts
- Metric evaluation
- Data Analysis
- Insights

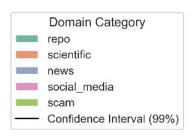


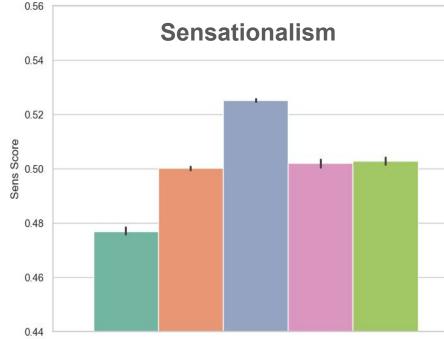
Extract
Abstracts
(Scrape for identifier and use Semantic

Scholar for data)

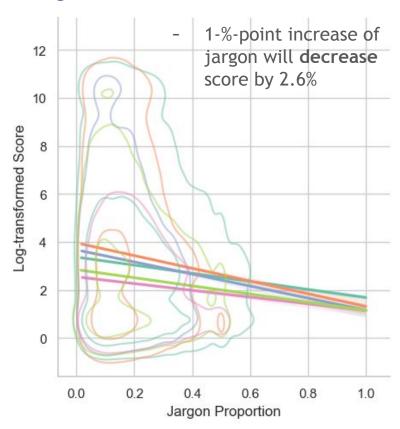


RQ1: How does the quality of science communication differ by the type of source used? (e.g., news article, research paper)



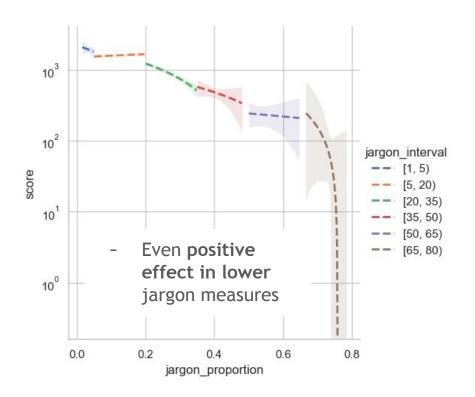


Jargon

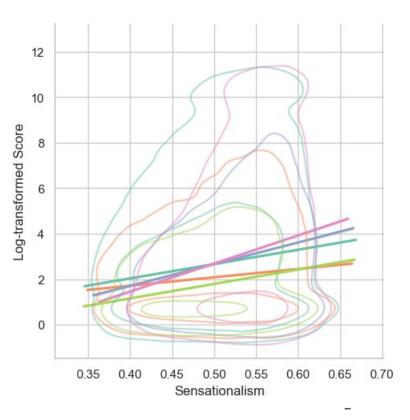


UC Berkeley

RQ2: Is high-quality science communication favored by users?

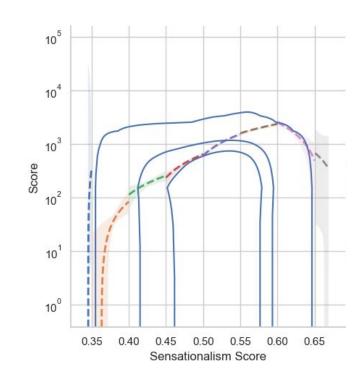


Sensationalism



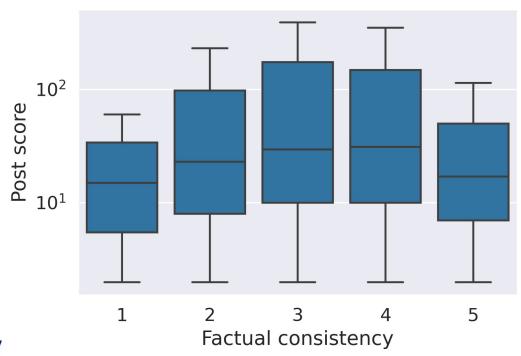
RQ2: Is high-quality science communication favored by users?

- ← 1-%-point increase of sensationalism will increase score by 10%
- → Decreasing or steady effect
- → Extreme sensationalism will decrease performance

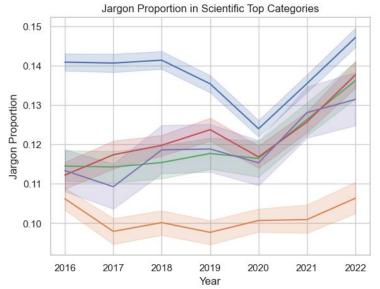


Factual consistency

RQ2: Is high-quality science communication favored by users?



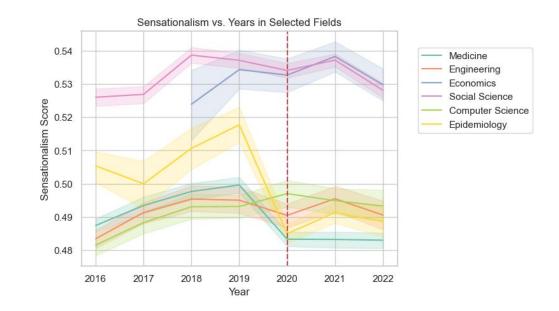
RQ3: How has the quality of science communication changed over time?



 Noticeable jargon drop associated with the coronavirus pandemic



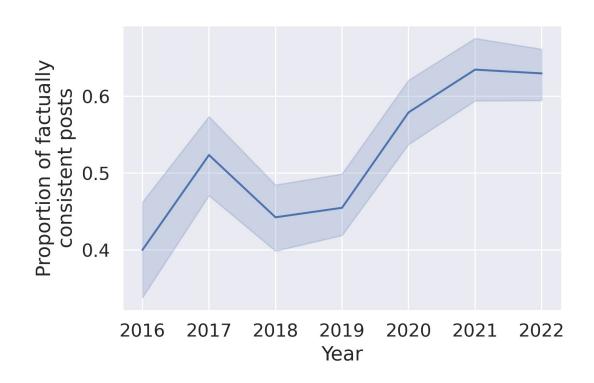
 Additionally decreased sensationalism in Epidemiology





RQ3: How has the quality of science communication changed over time?

Factual consistency



Conclusions

- Users prefer moderately sensational, moderately factual, low-jargon posts
 - There seems to be a tradeoff between low jargon and low sensationalism or high factual consistency
- Results tend to be more distorted when sharing from news articles compared to the primary literature
- Events like COVID can have an impact on the interest in and quality of science communication

Questions?

Thanks!

Jargon

- Using existing PMI-scores for various fields [1]
- Classifying subfields from [1] into each reddit category
- Classifying every word with PMI-score >0.1 as 'jargon-word'
- Calculating jargon-proportion as the proportion of words that are jargon in that category compared to all of the words in the title
 - e.g. if 50/100 words are jargon, their score is 0.5



Jargon evaluation

- 3 annotators counting the number of jargon words in a reddit title
- Comparing the average of the annotations with the values that were calculated using the keyword dictionary

Correlations:

Spearman correlation: SignificanceResult(statistic=0.10708850960390848, pvalue=0.4453294414958735)
Pearson correlation: PearsonRResult(statistic=-0.019035486976025275, pvalue=0.8923840597923983)

Correlation: -0.01903548697602531

Kendall tau: SignificanceResult(statistic=0.08697179581478527, pvalue=0.37102213878789503)

Correlations without 0 values:

 $Spearman\ correlation:\ Significance Result (statistic=0.1478150785011751,\ pvalue=0.35637817154109164)$

Pearson correlation: PearsonRResult(statistic=0.05263337796129377, pvalue=0.7438022868391653)

Correlation: 0.05263337796129374

Kendall tau: SignificanceResult(statistic=0.1208728606727125, pvalue=0.2745181872022512)

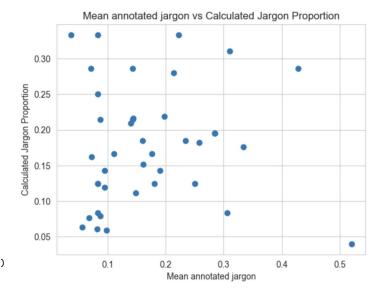
Correlations without outliers:

 $Spearman\ correlation:\ Significance Result (statistic=0.23632858445273083,\ pvalue=0.1420592455160639)$

Pearson correlation: PearsonRResult(statistic=0.2332416327658079, pvalue=0.14749497225350627)

Correlation: 0.23324163276580787

Kendall tau: SignificanceResult(statistic=0.17967309687492472, pvalue=0.10931812315719142)



- Removing 0 values and outliers improved the agreement and p-values, but the agreement values are still not ideal
- Requires future work

Factual Consistency Evaluation

- → Ask LM about consistency of reddit post towards abstract.
- Correlation between human evaluation shown in:
 - Our tests on SummEval (~.45)
 - Prior research (~.53) [1,2]

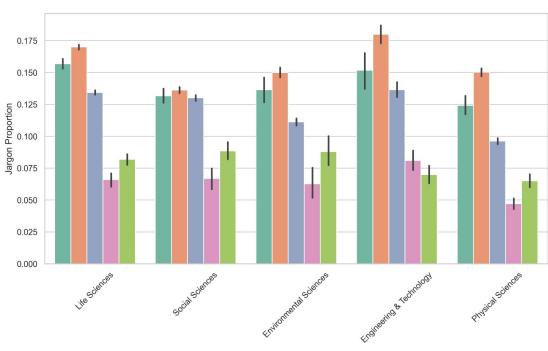
(DA Prompt)

Score the following [task-ins] with respect to [aspect] on a continuous scale from 0 to 100, where a score of zero means "[ant-aspect]" and score of one hundred means "perfect [aspect]". Note that [aspect] measures [aspect-ins].

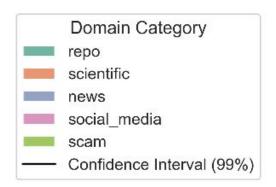
```
[Conditioned Text]
[Generated Text]
```

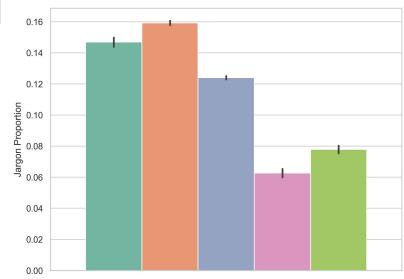
Scores:

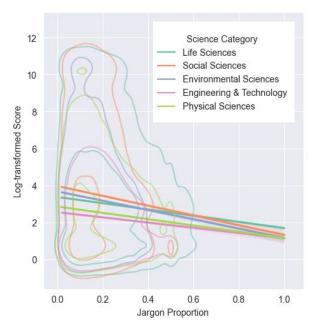




- Jargon is most prominent in Scientific News
- Relatively steady between fields
- Scam and Social Media least jargony





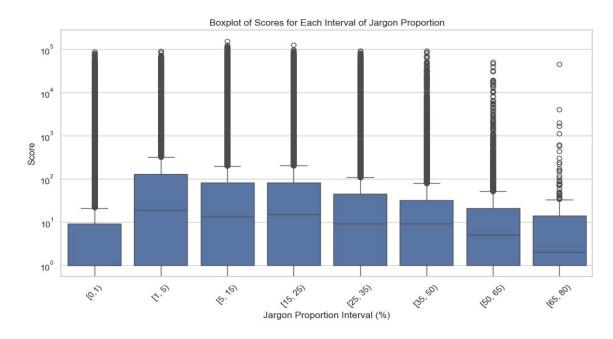


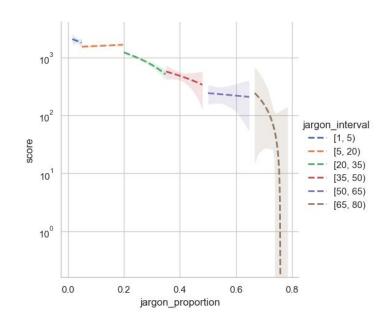
1-%-point increase of jargon will **decrease** score by 2.6%

U	CB	erk	elev
		CITT	CIC y

	$\log(\text{score})$	
Jargon Proportion (%)	-0.026^{***}	
	(0.001)	
Constant	3.474***	
	(0.047)	
Observations	97,672	
Adjusted R ²	0.069	

Note: $\begin{array}{c} {}^*p{<}0.1; \ ^{**}p{<}0.05; \ ^{***}p{<}0.01 \\ \\ \text{Filtered for score} > 1 \ \text{and jargon proportion} > 0. \\ \\ \text{Corrected for year, month, and top category.} \end{array}$





Even **positive effect in lower** jargon measures

	$Dependent\ variable:$		
	$\log(\text{score})$	score	
	(1)	(2)	
Jargon x (0,5]	-0.018*	-40.574	
	(0.011)	(37.351)	
Jargon x (5,15]	0.003	35.550**	
	(0.004)	(14.044)	
Jargon x (15,25]	-0.045***	-105.258***	
	(0.004)	(14.242)	
Jargon x (25,35]	-0.079***	-135.598^{***}	
5 (, ,	(0.008)	(28.072)	
Jargon x (35,50]	-0.018***	-19.362	
5 (,]	(0.006)	(19.875)	
Jargon x (50,65]	-0.002	-38.898	
5 (, ,	(0.027)	(90.170)	
Jargon x (65,80]	0.026	-70.473	
0 (,]	(0.053)	(179.551)	
Constant	, ,	, ,	
Observations	97,672	97,672	
\mathbb{R}^2	0.073	0.018	
Adjusted R ²	0.072	0.018	
Residual Std. Error (df = 97634)	2.272	7,685.550	
F Statistic (df = 37; 97634)	207.314***	49.005***	

Note:

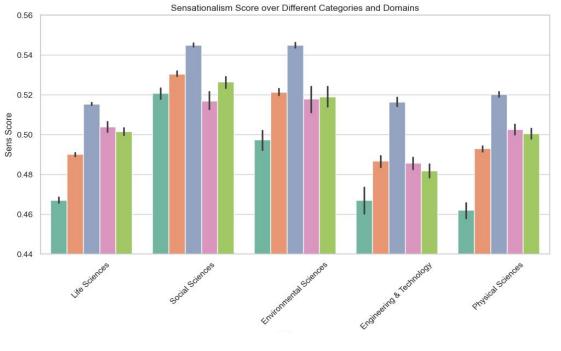
 $^*\mathrm{p}{<}0.1;\,^{**}\mathrm{p}{<}0.05;\,^{***}\mathrm{p}{<}0.01$

Filtered for score > 1.

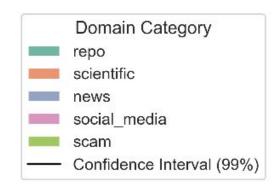
Repo, scientific, news, social media and scam domains.

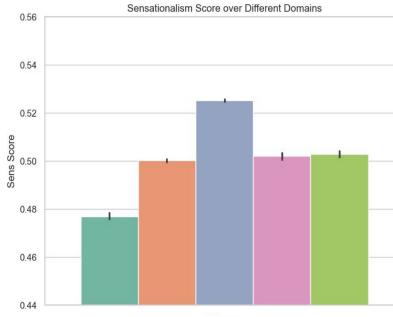
Corrected for year, month, and domain category.

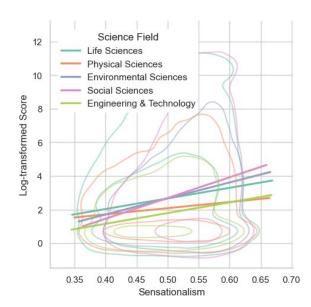




- More serious **domains** have less sensationalism
- Differs between scientific fields

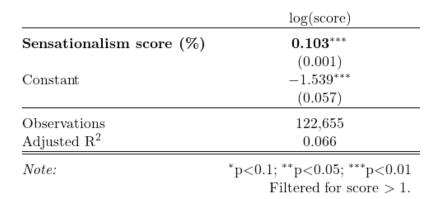


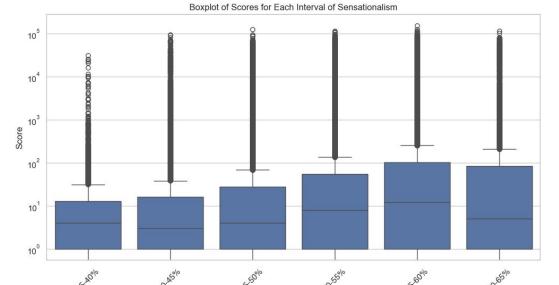




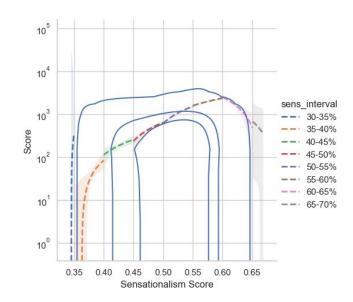
1-%-point increase of sensationalism will increase score by **10**%

U	C	B	er	<i>ke</i>	lev
	<u> </u>		CI		LUY





Sensationalism Interval (%)



- Decreasing or **steady effect**
- Extreme sensationalism will decrease performance

U	C	Be	rk	el	ev	
			111		-y	

	$Dependent\ variable:$		
	log(score)		
Sensationalism x (30,35]	7.794		
	(7.376)		
Sensationalism x (35,40]	0.106***		
	(0.031)		
Sensationalism x (40,45]	0.094***		
, , , ,	(0.012)		
Sensationalism x (45,50]	0.102***		
, , , ,	(0.009)		
Sensationalism x (50,55]	0.093***		
· / 1	(0.008)		
Sensationalism x (55,60]	0.104***		
, , <u>1</u>	(0.009)		
Sensationalism x (60,65]	$-0.105^{'***}$		
(, ,	(0.036)		
Sensationalism x (65,70]	-1.754		
, , <u>1</u>	(1.503)		
Constant	-268.312		
	(256.966)		
Observations	122,655		
\mathbb{R}^2	0.122		
Adjusted R ²	0.122		
Residual Std. Error	2.154 (df = 122615)		
F Statistic	438.636^{***} (df = 39; 122615)		

Note:

 $\label{eq:problem} p{<}0.1;\ ^{**}p{<}0.05;\ ^{***}p{<}0.01$ Filtered for score >1. Corrected for year, month, and domain category.

Conclusions/future directions

 Moving forward, we will incorporate sensationalism and factual consistency measures into the regression model

 We are making progress on scraping research papers

 We will also incorporate more potential confounders such as the popularity of research papers outside of Reddit

