

# Presence and consequences of positive words in scientific abstracts

Moritz Edlinger<sup>1</sup> · Finn Buchrieser<sup>1</sup> · Guilherme Wood<sup>1,2</sup>

Received: 13 May 2023 / Accepted: 23 October 2023 © The Author(s) 2023

# Abstract

Abstracts are the showcase of scientific studies, crafted to make an impression on the reader within a limited space and to determine the amount of attention each study receives. Systemic conditions in the sciences may change the expressive norm and incentive scientists to hype abstracts to promote their work and career. Previous studies found that terms such as "unprecedented", "novel" and "unique" have been used increasingly in recent history, to describe one's own research findings. The present study investigates the use of valence-loaded scientific jargon in the abstracts of scientific articles. Sentiment analysis with dictionaries specifically attuned to detect valence-loaded scientific jargon was employed to analyze more than 2,300,000 MEDLINE abstracts from the fields of psychology, biology, and physics. Results show that over the last four decades, abstracts have contained an increasing amount of valence-loaded scientific jargon, as previously observed in earlier studies. Moreover, our results reveal that the positive emotional content of abstracts is increasing in a way that cannot be accounted for by the increase in text length, which has also been observed in the same time period. There were small differences between scientific disciplines. A detailed analysis of the distribution of valence-loaded scientific jargon within abstracts reveals a strong concentration towards the end of the text. We discuss these results in light of psychological evidence relating positive emotions with the propensity to overestimate the value of information to inform judgment and the increase in the competition for attention due to a pressure to publish.

**Keywords** Sentiment analysis · VADER dictionaries · Hype in science · Cognitive bias · Linguistic spinning

# Introduction

Hype in science is often discussed to be a response to an incentive system that rewards publications and citations in their quantity and commercially successful research in general (Caulfield & Condit, 2012). Accordingly grants and/or tenure are being awarded to

Guilherme Wood guilherme.wood@uni-graz.at

<sup>&</sup>lt;sup>1</sup> Institute of Psychology, University of Graz, Universitaetsplatz 2, 8010 Graz, Austria

<sup>&</sup>lt;sup>2</sup> BioTechmed, Graz, Austria

individuals, institutions, nations, or journals with the greatest number of publications or citations (Meho, 2007). While an increasing number of postdoctoral researchers are looking for a limited number of permanent academic positions (Powell, 2015), the total number of global authors competing for the coveted journal space is increasing annually (Plume & van Weijen, 2014). In addition, the increasing pressure to portray one's work in terms of clinical and economic applicability is also discussed to promote hyped representations (Bubela & Caulfield, 2010). Furthermore, increasing rates of journal rejection and a bias toward non-significant results are well documented in the literature and potentially stand in the way of the ideal of scientific rigor (Fanelli, 2012; Laws, 2013; Statzner & Resh, 2010). This situation of increased competition, leads to an ever-increasing competition for the attention of the readership. In such a conflict situation coupled with a climate of "publish or perish", a non-negligible number of researchers resort to questionable research practices (QRPs) in order to be able to publish and maintain (or advance) their careers (Laws, 2013). Although there are certainly cases of deliberate attempts at deception or intentional exaggeration, hyping and other QRPs are probably the product of subliminal influences of the current incentive system that makes such practices desirable by increasing the chances of successfully publishing one's research (Caulfield, 2018). Thus, it seems natural for researchers to desire to present their work as groundbreaking, novel and successfully economically and clinically applicable.

Psychology as a discipline seems to be especially affected by systemic conditions that may promote hype (Laws, 2013). Compared to other scientific fields, psychology and psychiatry exhibit the greatest publication bias, which states that papers that are able to confirm their hypotheses in a statistically significant way are published more often, so that between 1990 and 2007 the percentage of scientific articles that were able to wholly or partially confirm the underlying hypothesis increased by over 20% (Fanelli, 2012). The social sciences, including psychology, are significantly more affected by this development than the biological sciences, which in turn are significantly more affected than physical sciences (Fanelli, 2012). In general, novel results are preferred over replication when a decision must be made about which article to publish (Neuliep & Crandall, 1993). However, journal editors in the exact sciences appear to be more receptive to non-novel results and replications than editors in the "softer" social sciences, which includes psychology (Aldhous, 2011; Madden et al., 1995). Moreover, estimates suggest that only 1% of psychological research is ever replicated (Makel et al., 2012).

#### Hype and scientific language use

Since top journals want the most innovative and cutting-edge research results for their publications and researchers aim to publish in these journals, there is a tendency to overestimate the value and evidential value of one's own research, i.e., to hype one's own work (Ioannidis, 2008; Young et al., 2008). The abstracts of scientific papers play a special role in this regard, as they are the "condensed form" of an article and are often used as a first port of call when assessing whether reading the full article is a worthwhile endeavor (Islamaj Dogan et al., 2009). In consequence of this situation the language scientists employ in describing their findings has already changed, as Otte et al. (2022) describe the use of "creative linguistic solutions" (p. 2) in randomized controlled trials when statistical significance is desired but not achieved. In a type of "significance dance" by researchers, which can also be referred to as "linguistic spin," there is often an attempt to mislead the

reader's perception into a more favorable interpretation of the results, despite non-significant primary results (Otte et al., 2022, p. 2). In over 500,000 randomized controlled trials published on PubMed over the past 30 years, it has been observed that phrases such as "failed to reach statistical significance", "not quite significant", "all but significant", "did not quite reach statistical significance", "difference was apparent" and "approaches statistical significance" are used less and less in scientific papers, while phrases such as an "increasing trend", "a positive trend", "a numerical trend" and "nominally significant" are increasingly used (Otte et al., 2022, p. 1). In a similar vein, salient expressions such as "paradigm shift" and "pushing the envelope" have also been increasingly used in the titles of papers published on PubMed in recent decades (Atkin, 2002). Similar results were also obtained by Vinkers et al. (2015) examining PubMed abstracts between 1974 and 2014 for the occurrence of certain predefined positive, negative, and random/neutral terms. While there was no change in the relative frequency of neutral terms, there was a slight increase in the use of negative words and a large increase in the use of positive words. In particular, the words "robust", "novel", "innovative" and "unprecedented" contributed to this trend, increasing in relative frequency by as much as 15,000% over a 40-year time period (Vinkers et al., 2015). These authors attributed this trend to a potential deliberate overemphasis and exaggeration of the results in order to stand out in the scientific competition. A drawback of this study is that the predefined words were not examined for their contextual usage and thus the relation to scientific usage cannot always be clearly assumed. Similarly, possible changes in abstract length were not considered, as longer abstracts increase the likelihood that certain words will be used, and the increase in use of salient terms may also be in part due to an increase in abstract length. These research findings suggest that scientific communication and the vocabulary of scientists has changed over the past decades. Cancer research, for example, is characterized by public reporting that is dominated by superlatives (Abola & Prasad, 2016). Terms such as "groundbreaking", "breakthrough", "game changer" and similar expressions are regularly used in news articles to describe both approved and unapproved drugs, sometimes even for drugs that have not yet been tested in humans (Abola & Prasad, 2016). Although this circumstance is likely in part attributable to news industry incentive schemes, the research by Yavchitz et al. (2012) suggests that a great deal of spin in reporting is also due to scientists' choice of words. Therefore, the choice of words and language used by scientists matters when the focus consists in disseminating truth and facts, rather than unbecoming exaggerations. There is a line separating undue hype and the objective provision of information, and even if it may not be easily recognizable in individual cases, it runs contrary to the strict scientific ethos to seek "creative linguistic solutions" and to hype one's own research in order to gain better chances of publication.

# Information seeking behavior and its motivational and cognitive basis

Research on scientific information seeking behavior and the role of positive emotions in cognition is especially relevant in this context as it has been shown that under deadline pressure and in a context of information overload, cues and heuristics grow in relevance (Schwieder, 2016). Emotions are such a powerful cue during information processing (Savolainen, 2014). While negative emotions trigger deeper and more elaborate processing (Taylor, 1991), contents evoking positive emotions are processed faster and judged to be more similar to each other (Alves et al., 2017; Garcia et al., 2012; Unkelbach et al.,

2008). When in a better mood, information seekers are prone to process general and less specific information (Zhang & Jansen, 2009). For instance, in an investigation on planned risk information seeking and a vaccine booster for COVID, Li et al. (2023) observed that citizens may overlook potential risks or knowledge gaps when in a positive mood. Moreover, even if triggered shortly, positive emotions are able to influence decisions (Alves et al., 2017; Deutsch & Strack, 2008). For instance, Topolinski and Strack (2009a) showed that positive emotions increase the probability of perceiving semantic, visual, and grammatical coherence even when in reality there is none. In another study, Topolinski and Strack (2009b) also showed that positive affect is used as an internal cue in *intuitive* judgments of semantic coherence and increases the probability of attributing coherence to contents beyond their semantic properties. Moreover, positive emotion increases both cognitive flexibility and distractibility (Dreisbach & Goschke, 2004) and is a cue to refrain from socalled proactive cognitive control (Fröber & Dreisbach, 2014), e.g. refraining from reacting to *irrelevant* information. Along these lines, Verde et al. (2010) showed that positive words exaggerate the perception of familiarity and decrease accuracy in memory retrieval tasks. Similarly, judgements about the clarity (Whittlesea, 1993), likeability (Topolinski & Strack, 2009a) or truth (Reber & Schwarz, 1999) of stimuli can also be made on the basis of a subtle, positive affect. In this context, and also for all purposes in the present study, positive words are not necessarily those describing positive emotions but rather those able to elicit a positive emotional reaction. One way to determine whether words elicit positive emotions is to ask people to rate their content, such as the developers of the sentiment analysis tool VADER did (Hutto & Gilbert, 2014). Words selected in this way have been shown to elicit neural correlates of positive emotions (Citron, 2012; Mahrukh et al., 2023; Scott et al., 2009). In summary, when positive emotions can be triggered quickly, using simple stimuli such as positive words, they work as a strong cue during information processing, lower the critical threshold required to establish semantic coherence, and generate an artificial feeling of familiarity with novel task contents, lowering the proactive cognitive control required to focus on the information seeking behavior.

Considered in the context of this study, it becomes apparent that the excessive use of terms with positive connotations, triggering a low-threshold positive affect, may have an impact on the judgments of the readership of the respective abstracts. Thus, judgements about the familiarity, likeability and truth of previously unread article abstracts may be influenced towards a more beneficial interpretation. This perspective is especially relevant in current times, in which scientists are confronted with an exploding number of research articles, in which it is increasingly difficult to stay up-to-date with the literature and fast decisions about an article's value must be made (Fraser & Dunstan, 2010).

The information seeking behavior among scientists is experiencing dramatic changes in the last decades (Niu & Hemminger, 2012) because of the large availability of studies, time economy is an essential commodity. When working under time pressure, information seeking behavior is under strong influence of cues, such as emotions to decide about the relevance of a given study. This leads to the phenomenon of citing without reading. By analyzing how authors inadvertently copy misprints from references, Simkin and Roychowdhury (2003) estimated that mere 20% of the citers also read the original papers. Inaccurate citations are a common problem, for recent studies show that at least one mistake in the literature list of papers can often be found in the most diverse research fields as biomedicine (15%; Pavlovic et al., 2021), surgery (15%; Sauder et al., 2022), tourism (37%, Moyle et al., 2022). In ecology, incorrect attribution of original findings and ideas can be found in 22% and 15% of review papers respectively (Teixeira et al., 2013). Even in the field of meta-research, which emerged from efforts to mitigate problems with the quality of scientific output, an apparent lack of critical engagement with the cited literature is observed, leading to incorrect reproductions of claims and overgeneralization of supposed research findings (Horbach et al., 2021). Even if the estimates by Simkin and Roychowdhury (2003) were exaggerated: with the number of new publications growing exponentially, attention becomes an increasingly expensive commodity and concerns about "reading before citing" are expected to become only more serious in the near future. Advances in artificial intelligence such as chatGPT only complicate the picture for they accelerate further paper production (Chen, 2023; Hosseini & Horbach, 2023) and are the basis for tools to summarize scientific papers. Even if none of these pieces of evidence is conclusive regarding our interpretation of the role of positive emotions in abstracts, they are numerous and consistent regarding inaccuracies produced by citing without reading (a.k.a. reading only the abstract). Figure 1 depicts the process of deciding on citing a manuscript with or without careful reading. A positive emotional context generated by words with a positive valence may work as a cue to stop information seeking early and conclude for the citability of a study based only on its abstract. As recently shown by Liu and Zhu (2023), positive words increase the citation of scientific studies. In view of the high prevalence of citation errors, the number of wrong citations is expected to increase too.

#### Aims and research questions

We approach the question of hype in science from the specific point of view of the emotional context created by words with a non-zero emotion valence. Evidence from the neuropsychology of emotions shows how specific vocabulary evokes emotions, even if the words themselves do not describe emotions or emotional states (Citron, 2012; Mahrukh et al., 2023; Scott et al., 2009). As discussed above, emotions are used as cues to guide and speed up decision processes (Topolinski & Strack, 2009a, 2009b) during



Fig. 1 Flowchart of the influence of + jargon words on information seeking behavior and citation probability

information seeking. With this focus in mind, we ask the following research questions: Did the language employed by scientists to present their findings become more positive in recent decades, and do the fields of psychology, biology, and physics differ in this development and in terms of the positive charge of their abstracts? To answer this question, this paper examines the emotional content of over 2 million PubMed scientific abstracts in the aforementioned disciplines using sentiment analysis. Previous studies used dictionaries unfiltered for technical terms (Baes et al., 2022; Liu & Zhu, 2023; Vinkers et al., 2015; Yuan & Yao, 2022). This can lead to biased estimates of the emotion valence and its change over the years, as the emergence of new technical terms in later years may inflate the sentiment estimates. In the present study, we adapted the sentiment analysis dictionary VADER (Valence Aware Dictionary for sEntiment Reasoning, Hutto & Gilbert, 2014), an empirically validated, gold standard sentiment dictionary. As these words have been scored by hundreds of participants as sentiment words, one can expect that they produce the expected positive or negative emotional reactions as described in the previous sections. Our adaptation of VADER produced dictionaries containing only the so-called valence-based scientific jargon. Valence-based scientific jargon was defined in the present study as (i) words that belong to the dictionary of VADER and individually or in groups of two or three yield a positive valence. These words also (ii) are frequent in the vocabulary of psychology, biology and physics and allows comparisons between these disciplines and (iii) according to our methods of dictionary tuning are not technical terms such as "support" in "support vector machines" or "care" in "intensive care unit". Hereafter we refer to these terms that are characteristic of valence-loaded scientific jargon either as + jargon or—jargon according to their positive or negative emotional valence.

The aforementioned findings of Vinkers et al. (2015), Otte et al. (2022), and Atkin (2002) suggest that language use has changed and certain positive terms, in describing one's research, are used more frequently. Research findings such as those of Fanelli (2009, 2010, 2012), Madden et al. (1995), and Laws (2013) suggest that the psychological discipline is exposed to those forces that create hype in science to a greater extent than "harder" sciences. Furthermore, we expect abstracts to present a higher concentration of + jargon close to the end of the ms, because at that position the positive emotion context is more available during decision-making. Following this line of thought we hypothesized that:

**Hypothesis 1** The ratio of neutral abstracts, defined as abstracts without a positive sentiment value, has decreased over time.

**Hypothesis 2** The ratio of neutral abstracts, defined as abstracts without a positive sentiment value, is lower in psychology than in biology and physics.

Hypothesis 3 The positivity of the abstracts of scientific articles has increased over time.

**Hypothesis 4** The positivity of the abstracts in psychological articles is greater than in biological or physical articles.

**Hypothesis 5** The different parts of the abstract text, will be dissimilarly embedded in a positive context, in particular the last part of the abstract exhibiting higher levels of positivity.

# Methods

The analyses were performed using the statistical program R version 4.1.3. In Fig. 2, the study flow chart is depicted.

# Data collection

Data collection took place from November 2021 to December 2021 at the Karl-Franzens-University in Graz. The package *rentrez* was used for these purposes, as it provides a simple and consistent interface to the NCBI (National Center for Biotechnology Information) databases (Winter, 2017). Using the search terms depicted in Table 4 of the Supplementary Materials, this package was used to extract the IDs, publication dates and abstract texts of the articles from PubMed. Because the MeSH terms were used for the search, the PubMed search was limited to results from the MEDLINE database only. In this way, a total of 2,318,946 million abstracts were extracted for further analysis, 373,856 in physics, 752,339 in biology, and 1,192,751 for psychology.

The database used in this study is MEDLINE, which is an English-language bibliographic database maintained by the National Library of Medicine (NLM) in the United



Fig. 2 Flowchart of the study procedure

States. It contains more than 28 million references to scientific articles dating back to 1946, with a focus on research related to the field of biomedicine. More than 5200 scientific journals are included in the MEDLINE database (National Library of Medicine, n.d., Section: MEDLINE: Overview). MEDLINE is the largest component of the well-known literature database PubMed and can be queried through PubMed. It applies the hierarchically organized controlled vocabulary of the NLM called MeSH (medical subject headings) to facilitate a literature search process by means of a common vocabulary so that research on similar fields is labeled with a common term. The MeSH vocabulary is updated annually in a time- and resource-intensive process in which NLM staff manually assign terms from the MesH structure to new publications and expand it as necessary should new research areas open up. The hierarchical structure of MeSH terms includes an outline of specific terms under broad concepts, and a search of higher-level concepts (e.g. psychology) includes an automatic search of lower-level terms. To avoid subjective selection of search terms and to represent the respective disciplines as thoroughly as possible, the literature search was conducted using all of the more specific MeSH terms of the three scientific areas of interest, namely psychology, biology and physics. Table 4 in the Supplementary Materials shows the three science areas along with the subordinate search terms and the number of available papers per search term.

#### Sentiment-analysis

In this study, sentiment analysis is used to assess the emotional content of scientific abstracts over time. The sentiment tool VADER implements an empirically validated, gold standard sentiment dictionary and is well suited for fast and accurate sentiment analysis, especially with large datasets such as those encountered in this study (Hutto & Gilbert, 2014). VADER not only distinguishes individual words in terms of their polarity, unlike some other instruments, VADER also takes differences in the intensity of the words into account, so that emotionally stronger words, such as "love," are scored higher than "like." The sentiment values of the individual words of the entire VADER lexicon were generated by human raters in the construction process of this instrument (Hutto & Gilbert, 2014). Another advantage of VADER is that it incorporates relationships between words using five rules that fundamentally improve the accuracy of the instrument, regardless of the underlying lexicon (Hutto & Gilbert, 2014). These rules include: considering punctuation, as well as capitalization, considering words that change the degree of strength (e.g. "very"), considering contrastive conjunctions using "but," and analyzing the three words preceding each sentiment word to detect negations that change the polarity of the text (Hutto & Gilbert, 2014). Consequently, VADER analyzes texts based on the individual words it is composed of, along with the five heuristics using grammatical and syntactic cues just described, and thus produces a sentiment value that indicates whether a text is predominantly positive, negative and neutral. Furthermore, VADER incorporates abbreviations for social media contexts (e.g. "lol", "nh", etc.) which were excluded for this study.

# Data preprocessing

Before the abstracts were analyzed, they were filtered using a preprocessing pipeline. During preprocessing, articles that appeared repeatedly, articles for which the abstract was missing, as well as all abstracts that were not exclusively written in English were excluded. Additionally few abstracts produced an error when being analyzed by VADER and had to be excluded. Afterwards abstracts that contained less than 50 words and more than 500 words were also excluded. Liberal criteria for abstract length were deliberately chosen to incorporate a wide range of abstracts, as different journals and scientific disciplines impose different guidelines. Additionally due to sparse data, abstracts prior to 1980 were excluded. Finally, truncated abstracts, i.e. abstracts that were not fully available on PubMed, were excluded from subsequent analyses. In total 1,124,356 abstracts were removed this way, leaving 1,194,590 abstracts for further analyses.

#### The creation of science-specific dictionaries

Since VADER was not designed specifically for usage in scientific contexts, the initial analysis of scientific texts was always flawless, since everyday language is often fundamentally different from scientific language. For instance, considering the word "cancer", which has negative connotations in everyday language, is used almost exclusively as a technical term in science, which can lead to an artificial increase in sentiment in some scientific fields in which cancer research is commonly conducted. This problem is illustrated in Fig. 6 in the Supplementary Materials, which illustrates which words contributed most to the sentiment score in each field when the dictionary of VADER was not adapted. Words such as "care", "growth" and "energy" are prominent examples of words that are used neutrally in science but are generally regarded as positive sentiment words. To circumvent this problem, VADER's lexicon was adapted to specifically address the study of scientific language. This approach offers the advantage that, on the one hand, only relevant terms are examined while on the other hand VADER's structure is applied to detect *intensifications*, attenuations, and negation changes, thus increasing the accuracy of the analysis. For the creation of the revised lexicons, the assessment of the context was crucial, since VADER cannot differentiate between different contexts. For example, the word "support" contributes to the sentiment score in the same way regardless of whether it is applied as technical vocabulary ("support vector machine") or to positively represent results ("these findings support"). To examine contextual usage, network graphs were created visualizing the most frequent connections between words, and the most frequent bi- and tri-grams of the relevant terms were examined to decide whether a term primarily represents a neutral content word or + jargon putting the work in a good light. An n-gram is a word combination of n consecutive words, meaning that the most frequent n-gram indicates which n number of words mostly appear in conjunction. For the context analyses, ten to fifteen percent of the abstracts were randomly selected, depending on the scope of the scientific field, since the analyses would otherwise have been too resource-intensive. Stopwords (common words such as "the", "of", "it", etc.) were removed for the analyses. We defined valence-based scientific jargon as (i) words that belong to the dictionary of VADER and individually or in bigram or trigram yield a positive valence. Moreover, these words also (ii) are frequent in the vocabulary of all three fields of science investigated in the present study and (iii) according to our methods of dictionary tuning are not technical terms such as "support" in "support vector machines" or "care" in "intensive care unit".

Technical terms were excluded in several steps. In a first step we identified the 200 most influential sentiment words of each of the three scientific fields using the standard VADER lexicon. The influence of a sentiment word results from its average sentiment value, which may vary slightly due to the presence of intensifications and attenuations in a specific context, and the frequency of each term. Then an intersection was performed to determine which of these 200 words appear in all three fields of science. The

rationale behind using the intersection of these terms is that valence-loaded scientific jargon is used in all fields, regardless of content. This procedure allows us to exclude technical terms from specific fields that are classified as sentiment words by VADER, such as "attracted" in physics, "vitamin" in biology or "optimism" in psychology. This intersection yielded 79 positive terms which in a second step were examined contextually (via bi- and trigrams and network graphs) to determine whether they represent valence-loaded jargon in the context of scientific communication, as defined above. This approach allowed us to identify and exclude terms that are initially classified as sentiment words by VADER and occur in all disciplines to varying degrees, but have a purely technical connotation.

To minimize dependencies on a single set of words, three differently sized dictionaries were created. The short dictionary consists of 15 words, the medium dictionary consists of 25 words and the long dictionary consists of 35 words. The medium dictionary is an extension of the short dictionary, and the long dictionary is in turn an extension of the medium dictionary. Each word was assigned to the corresponding dictionary/dictionaries depending on the proportion of the specific word being used as a + jargon word and not being used in a technical context. This was done by investigating the frequencies of the corresponding bi- and trigrams of each word. The smaller the dictionary, the stricter this inclusion criterion was. All terms however in all of the three dictionaries were primarily used in a manner consistent with our definition of valence-loaded scientific jargon. Hence, for example the word "strong" was assigned to the longer dictionary because, although it is primarily used, in the context of "strong correlation" or "strong evidence," as a word that enhances the perceived importance of a text, it also appeared to a lesser extent as a technical term in physics in connection with "strong magnetic field." The word "important" however was assigned to the short dictionary because it is virtually exclusively used in a manner consistent with our definition of + jargon words while its usage as a technical term is negligible, as can be identified through the corresponding bi- and trigrams. Following this procedure, words that fit our definition of valence-loaded scientific jargon can be distinguished from technical terms. Of course this separation is gradual rather than qualitative in nature and the decision about the classification of the words may be delicate and has to be judged individually for each term. Table 10 in the Supplementary Materials of this paper lists frequency of the bi- and trigrams, broken down by scientific field, which formed the basis for the decision to include words in the respective revised dictionaries, in detail. Words outside the intersection of the 200 most influential words were also included, insofar as they stood out in the analyses as sentiment words; for example, the words "progress" and "opportunity" were included. The word "significantly" was also included in the analyses, while the word "significant" was not included because it was commonly associated with negations in the bigrams and tri-grams. The inclusion of this word seems justified and desirable as statistical significance is often regarded as a key element when deciding whether to publish or cite a paper (Vinkers et al., 2021), whereby it becomes advisable to talk in terms of statistical significance to make a positive impression with one's work. However, since "significantly" was not represented in the standard VADER lexicon, it was assigned the same value that VADER uses for the word "significant." The exact word lists including the sentiment value of each word can be seen in Table 1. The analyses with the adapted dictionaries took place with Python in version 3.10.4, which also enables analyzing using VADER while modifying the underlying dictionary. Figure 3 illustrates the most common bi-grams of the words implemented in the long dictionary, to depict the contextual usage of these terms in this sample of scientific abstracts. Once the custom dictionaries were complete, each abstract was split into three different parts of the same length using Python. Then sentiment values

Table 1         The terms of the three           science-specific dictionaries         Image: Specific dictionaries	Short	Medium	Long	Sentiment value
		'beneficial	'beneficial	1.9
		'benefit	'benefit	2.0
		'benefits	'benefits	1.6
	'better	'better	'better	1.9
			'effectively	1.9
		'efficient	'efficient	1.8
	'excellent	'excellent	'excellent	2.7
		'greater	'greater	1.5
			'help	1.7
			'importance	1.5
	'important	'important	'important	0.8
	'improve	'improve	'improve	1.9
	'improved	'improved	'improved	2.1
		'improvement	'improvement	2.0
	'improvements	'improvements	'improvements	1.3
			'improving	1.8
	'increase	'increase	'increase	1.3
		'increased	'increased	1.1
		'interest	'interest	2.0
		'novel	'novel	1.3
			'optimal	1.5
	'opportunity	'opportunity	'opportunity	1.8
	'promise	'promise	'promise	1.3
			'progress	1.8
			'significance	1.1
	'significantly	'significantly	'significantly	0.8
			'strong	2.3
	'strongly	'strongly	'strongly	1.1
	'successfully	'successfully	'successfully	2.2
			'supported	1.3
		'supporting	'supporting	1.9
	'supports	'supports	'supports	1.5
	'useful	'useful	'useful	1.9
	'valuable	'valuable	'valuable	2.1
			'well	1.1



Fig. 3 The most common bi-grams using the *long* dictionary. Only bi-grams with n > 200 are shown. The thickness of the line reflects the strength of the association. The thicker the linking line, the more frequently these words occur in association. Words marked in bold are words of the *long* dictionary. The analysis was done on 10% of randomly selected abstracts



Years (1980 - 2021)

Fig. 4 The number of abstracts containing the corresponding word of the 35 words in the *long* dictionary at least once, related to the number of abstracts for that year

were calculated individually for each third of each abstract. This analysis was conducted three times, once with each of the dictionaries (short, medium, long).

# Results

This study implemented context-checked dictionaries modified for the context of science. Figure 4 depicts the development of the proportion of abstracts containing the corresponding word of the modified *long* dictionary at least once, relative to the number of papers in that respective year. As the *long* dictionary also contains the 25 words of the medium and the 15 words of the *short* dictionary, the values for those dictionaries can also be read out of Fig. 4. In all cases except for the word "greater" did the frequency of abstracts containing that term at least once, relative to the number of papers published in that respective year, increase over the investigated 40-year time period. Moreover, by investigating the range of absolute frequencies, Fig. 4 shows that certain words appear more frequently than others. For instance the absolute frequency of the word "important" is ranging between 7 and 16%, whereas it is ranging from 0.75 to 1.5% for the word "progress".

An initial finding of this study is that not only the number of abstracts per year increased from approximately 4000 in 1980 to almost 70,000 in 2021 but also that the mean length of the abstracts has increased from approximately 900 characters in 1980 to about 1500 characters in 2021. Figure 5 depicts several aspects of the data set using the *long* dictionary. Figure 5 reveals a substantial decrease in the proportion of neutral abstracts, in all fields. The rubric "neutral abstracts" includes abstracts with a sentiment score of 0, in addition to a small number of abstracts with a negative sentiment score. Since only positive expressions were retained in the modified dictionaries, the negativity in some abstracts came about through negation detection, as VADER scores negations of positive expressions (e.g. "not optimal") as slightly negative. This is an advantage of this approach because negations of positive terms cannot be regarded as + jargon. Furthermore Fig. 5, shows the development of the sentiment value using the positivity score provided by VADER. A substantial increase of the sentiment value can be observed in all scientific areas. Finally, Fig. 5 shows an increase in the use of positive sentiment words towards the end of abstracts compared to the middle and the beginning parts of the abstract text. These apparent differences will be tested statistically to determine whether these findings are reducible to an increase in length of the abstracts or whether there is an independent effect of time. Furthermore, eventual differences between the fields will also be tested statistically. Due to similarities between the results of the three differently sized and modified dictionaries and for the sake of simplicity the following section depicts the results when using the *long* dictionary. Besides, the long dictionary contains the most words and may be most sensitive in detecting positivity in the abstracts. Results for the *short* and *medium* dictionaries can be found in the Supplementary Materials (Tables 5-8).

To examine the change in the ratio of neutral vs. positive abstracts over time and to gauge differences between the scientific fields studied and different parts of abstract texts, a logistic binary regression was calculated. Positive abstracts were defined as abstracts that at least contained one of the + jargon expressions defined in the modified dictionaries. As predictors, in addition to the *Year*, the *Scientific Field* and the different *Parts* of the abstract, the *Abstract Length* was also included. This step is crucial since abstracts have become longer in the last decades and therefore have a higher probability of implementing positive vocabulary. The variables of *Abstract Length* (M = 1047.93, SD = 387.90) and



#### Thirds of the abstracts

**Fig. 5** Several findings using the *long* dictionary are shown: the development of the proportion of neutral abstracts (defined as abstracts that have no positive sentiment value), the development of the positivity value over time and the mean positivity value of the different parts of the abstract. All results are depicted separately for each of the three scientific disciples investigated in this study

	В		Ζ		<i>p</i> -value		Odds ratio	
Intercept	87	77	- 34.07	- 30.46	<.001	<.001	.42	.46
Abstract Length <sup>a</sup>	.40	.45	30.35	34.11	<.001	<.001	1.49	1.58
Year <sup>a</sup>	.05	.11	3.92	8.24	<.001	<.001	1.05	1.12
Scientific Field								
Psy vs. Bio	13	01	-4.61	-0.46	<.001	.56	.88	.99
Psy vs. Phy	23	88	-6.69	-2.60	<.001	.01	.80	.92
Parts								
1st vs. 2nd Part	.15	.27	4.85	8.59	<.001	<.001	1.16	1.30
1st vs. 3rd Part	.87	.98	28.58	32.15	<.001	<.001	2.37	2.66
Pseudo $R^2 = .097113 (95\% \text{ CI})$								

 Table 2 Results of binary logistic regression using the long dictionary

Bootstrapping results are presented at the 95% confidence interval

<sup>a</sup>The values of this variable have been z-standardized for better interpretability

*Year* (M=2008.96, SD=9.34) were *z*-standardized in favor of a better interpretability of the results. To replace the theoretical sample distribution, with an empirically determined sample distribution, bootstrapping at 1000 replicates and a size of 10,000 data points was selected. The resulting sample distribution was then truncated at both ends at 2.5% of the values, leaving a 95% confidence interval of the empirically determined values. Table 2 illustrates the results obtained from binary logistic regression using the *long*. The results for the *short* and *medium* dictionaries can be found in the Supplementary Materials (Tables 5, 6).

The Abstract Length was the strongest predictor of classifying abstracts as positive and exhibited the highest odds ratio values across all dictionaries. Furthermore, the variable *Year* exhibited a weaker, but still significant influence on the classification of abstracts as neutral or positive in two out of three dictionaries, approaching statistical significance in the third one, the *short* dictionary (Table 6). No consistent picture emerged for the *Scientific Field* because statistically significant differences were only shown in the long dictionary, with the discipline of psychology exhibiting a higher likelihood that their abstracts are being classified as positive, compared to biology and physics. Besides, with the passing of time the likelihood that a randomly selected abstract will be classified as positive increases. Likewise did the *Third Part* of the abstracts exhibit significantly higher rates of being classified as positive when compared to the *First Part* of the abstracts. The same is true for the *Second Part* compared to the *First Part*. The differences concerning the parts of the abstracts were consistent across the three modified dictionaries.

A mixed-effects model was employed to examine the effects of Year, Scientific field and part on the positivity of abstracts. Given the very large number of abstracts, we opted to follow a bootstrapping procedure for the evaluation of the significance of the effects. We generated 1000 samples of abstracts containing each 10,000 different abstracts and estimated the 95% confidence interval for the sum-of-squares, mean-square error, F, and p-values obtained by applying the Satterthwaite's method for computing the denominator degrees of freedom and F-statistics. Table 3 illustrates the results obtained from the subsample mixed effects model using the *long* dictionary. The results for the *short* and *medium* dictionaries can be found in the Supplementary Materials (Tables 7, 8) The effect of time

	NumDf	DenDf	Demue		DC IIIDINI		-		Ь	
Year*	1	29,990	0	.07	0	.07	1.55	25.71	<.001	.214
Scientific field*	2	29,990	.07	.21	.04	.10	14.66	40.80	<.001	<.001
Part*	7	29,990	4.29	5.23	2.14	2.62	847.06	1036.90	<.001	<.001
Part × Scientific field	4	29,990	.18	.40	.05	.10	17.80	39.69	<.001	<.001

 Table 3
 Results of mixed-effects model on the positive abstract subsample using the *long* dictionary

Scientometrics

measured in *Years* on the sentiment score was not significant in the *long* dictionary but significant in the *medium* and *short* dictionary. Regarding the *Scientific Field* a significant difference between the various disciplines was found within all dictionaries. The field of Psychology displayed the highest level of positivity (90% CI of Mean = .0438-.0452), followed by Physics (90% CI of Mean = .0393-.0414), Biology displayed the lowest level of positivity 90% CI of Mean = .0394-.0409). Furthermore, a significant difference in the respective *Parts* of the abstracts was observed, across all dictionaries. The interaction between Part and Field of Science was significant as well across all dictionaries.

#### Discussion

In this study, the usage of + jargon to describe one's own research in scientific abstracts was examined in the disciplines of psychology, biology, and physics. In line with previous studies, an increase in positivity was shown over the last four decades, as abstracts became more emotionally positively charged and the rate of exclusively neutrally formulated abstracts steadily decreased. These results hold also after removing the effect of the length of the abstracts and the scientific discipline, suggesting that this may be a cross-disciplinary, systemic process. Interestingly, this trend was especially pronounced in the final part of the abstract texts, so that a strong emotional context for the interpretation of the final part of the abstracts have not only become longer descriptions of research, but have also increasingly used positive language in describing their study results, thereby adopting a more promotional function. Since the concentration of + jargon is increasing particularly in the part of the abstract text, a temporal coincidence between the positive emotional context and the decision making in information seeking behavior emerges with several possible consequences that will be discussed below.

The binary logistic regression showed that the longer the abstract is, the more likely it is to be classified as a positive abstract. This is comprehensible, as abstracts that are composed of more words are subjected to an increased likelihood that + jargon is being used in them. Beyond the effect of abstract length, an effect of time was also evident, i.e. as time progressed, the likelihood of an abstract being judged as positive increases. From 1980 to 2021, every 9.34 years, the odds of a randomly selected abstract from MEDLINE being classified as having positively laden language increased by 8 to 16%, regardless of the length of the abstract and the scientific field (Table 2). The results for shorter dictionaries can be seen in the Supplementary Materials. Differences in the ratio of neutral abstracts between the disciplines were only identified using the long dictionary. Because the long dictionary is the most extensive one of the three dictionaries it may be most sensitive in detecting positivity in the abstracts and detect differences that cannot be observed using smaller sets of 15 and 25 words. Thence, Hypotheses 1 relating to a decreasing ratio of neutral abstracts over time was confirmed in large parts. Hypotheses 2 relating to a lower ratio of neutral abstracts in psychology than in the other disciplines was only partially confirmed and more research is needed in order to make substantiated assertions. The analysis of positivity scores again indicated an effect of abstract length, however without an additional independent effect of time, suggesting that over time there are fewer neutral abstracts, as suggested by the logistic regression, but not that within the group of positive abstracts the sentiment score increases with time (Table 3). This can be explained by the way the positivity score is computed by VADER, namely, by considering the total length of the text. This helps explain the apparent discrepancy to other studies investigating the emotional valence of MEDLINE abstracts, which employed different dictionaries and procedures (e.g. Liu & Zhu, 2023). Hypothesis 4 relating to the differences between the disciplines of *psychology*, *biology* and *physics*, was confirmed in large measure, as *psychology* showed higher positivity scores than the other two disciplines. In view of the literature suggesting that psychology is more exposed to those forces that create hype in science than "harder" sciences (Fanelli, 2009, 2010, 2012; Laws, 2013; Madden et al., 1995) the present results suggest that + jargon may represent a problem for the quality of scientific work. As presented above, + jargon induces a feeling of similarity and familiarity (Alves et al., 2017; Garcia et al., 2012; Unkelbach et al., 2008; Verde et al., 2010) as well as of semantic coherence and conceptual appropriateness which may not be justified (Topolinski & Strack, 2009a). This may worsen the quality of published and cited materials in this discipline plagued by problems with reproducibility (Open Science Collaboration, 2015). Besides, it was shown that the respective parts of the abstracts (first, second or third) differ with regard to their sentiment, with the third and the second part exhibiting significantly higher sentiment values than the first part and the sentiment value gradually increasing over all parts (Fig. 5). Therefore, Hypothesis 5, which stated that the distinct parts of the abstracts, will differ regarding their sentiment was also mostly confirmed. The utilization of + jargon was concentrated at the end of the abstract text, in good temporal coincidence with the decision making regarding further steps of information seeking. As depicted in Fig. 1, a positive emotional context impacts decision making as a cue indicating similarity and familiarity (Alves et al., 2017; Garcia et al., 2012; Unkelbach et al., 2008; Verde et al., 2010) as well as semantic coherence and conceptual appropriateness (Topolinski & Strack, 2009a). The concentration of + jargon at this palace in the abstract may contribute to increase the feeling of appropriateness of the abstract in the reader and strengthen the motivation to skip reading the whole manuscript thereby opening a door for citation- (Simkin & Roychowdhury, 2003) and interpretation errors so frequent in the literature (Pavlovic et al., 2021; Sauder et al., 2022; Teixeira et al., 2013; Horbach & Aagaard, 2021).

The study of scientific texts is a fairly new application area for sentiment analysis, as these tools are mostly applied in the context of social media or in the analysis of customer reviews (Mäntylä et al., 2018). Hence, this study also faced the problem that an analysis with the standard VADER lexicon, a gold-standard tool for sentiment analysis, led to biased results due to differences of everyday language and scientific language. Therefore, the VADER lexicon was adjusted to the context of scientific language usage so that primarily content-related terms (e.g. "cancer", "energy", etc.) were no longer used to determine the sentiment of the abstract texts. As expected due to an increasing publication bias (Fanelli, 2012), an increasing evaluation of scientists' work by citation and publication counts (Meho, 2007), and an increasing competition for funding (Caulfield, 2018), not only the rate of abstracts that did not make use of positive vocabulary in recent decades decreased, but also the positivity in the subsample of positive abstracts increased. This research builds on and extends the research of Vinkers et al. (2015), as the length of abstracts was taken into account and other context-checked dictionaries were used. Their finding that the frequency of + jargon increased sharply is mirrored by the lower rate of neutral abstracts and the increased sentiment score in abstracts and thus strengthens their original result. As this study is partly inspired by Vinkers et al. (2015), we also validated our results using Vinkers' dictionary. The absolute frequencies of their proposed, predefined words were examined in this study, while additionally running a sentiment analysis using their dictionary. Using their terms also leads to an increase in sentiment (Table 9; Figs. 7, 8, Appendix). The results thus suggest that abstracts were short, unemotional accounts of research just a few decades ago, whereas over time they have not only become more detailed, but also increasingly populated with vocabulary designed to lower the critical threshold required to establish semantic coherence, and generate an artificial feeling of familiarity with novel contents, and to lower the proactive cognitive control required to focus on the information seeking behavior (Fröber & Dreisbach, 2014; Topolinski & Strack, 2009a, 2009b; Verde et al., 2010). Thus, it seems that abstracts have increasingly taken on a promotional function and that in order to stand out from the mass of publications, researchers are consulting a certain, more conspicuous vocabulary. This finding is in line with the research of Chiu et al. (2017), Yavchitz et al. (2012) and Boutron et al. (2010) who identified a certain exaggeration in the abstracts of scientific papers, although in the present case not an open exaggeration but a much more subtle emotional tuning of the reader seems to take place. Even though the discipline of psychology seems to be most affected by this trend, showing the lowest rates of neutral abstracts and the highest sentiment scores in abstracts that were not neutral, all three investigated were impacted by this development. This result seems to be in line with the observation, that while psychology as a discipline seems to have more problems with replicability and publication bias than biology and physics, all scientific areas are affected by the same interdisciplinary developments that affect hype in science and the scientific use of language (Fanelli, 2012; Statzner & Resh, 2010).

To be more precise, the increase in sentiment scores in each of the three scientific disciplines observed and the shrinking proportion of neutral abstracts probably represents a competition for the attention of the readership, which emerges as more and more authors compete for limited funding, jobs and publications (Brischoux & Angelier, 2015; Caulfield, 2018; Plume & van Weijen, 2014). Much of the most influential scientific work accumulates in a small set of journals with a high impact factor, which can give a career advantage to scientists who publish in them (Ioannidis, 2006; Tahamtan et al., 2016). In order to keep up with other scientists or to gain an advantage over them, there could be a steady increase in the use of + jargon which may lead to an upward spiral of the sentiment, just to keep up with the current baseline of the sentiment in scientific abstracts. Concomitantly, the use of -jargon does not seem to change in the last decades. Their frequency is also much lower than that of + jargon (Fig. 9, Supplementary Materials). Considering that negative emotions are powerful inducers of critical thinking and scrutiny (Unkelbach et al., 2021), their low and stable numbers in contrast to the much higher and steadily increasing numbers of positive words also reinforce the impact of + jargon on scientific information seeking behavior. From the viewpoint of an individual researcher, the utilization of terms that give the impression of the novelty and importance of the research and may lead the reader to a more benevolent interpretation of the results is desirable as it may substantially impact the success of the article. As described in the introduction, positive terms can subliminally influence intuitive judgements about the familiarity, likeability and truth of a stimulus (Reber & Schwarz, 1999; Topolinski & Strack, 2009a; Verde et al., 2010). Subtle, short positive states of affect, which can be induced by contact with positive objects or affective cue stimuli (e.g. certain positive words) (Verde et al., 2010), can serve as an internal signal for evaluating a range of situations and provide the basis for decision-making (Clore et al., 2013). In this manner, abstracts may get evaluated systemically in a distorted way, based on the vocabulary they implement. In general, emotions influence information processing as contents eliciting positive emotions are processed faster and not as elaborate as negative information (Alves et al., 2017; Garcia et al., 2012; Taylor, 1991; Unkelbach et al., 2008). Abstracts play a special role in this regard, as they act as a figurehead for a paper and often have to convince the reader of the usefulness and the value of the work within a few characters, which is why exaggeratedly positive language is often used in abstracts (Boutron et al., 2010; Chiu et al., 2017). Under certain circumstances, the success of a paper, measured by the publication and the subsequent number of citations, stands or falls with an abstract that can attract the attention of readers, because the abstract is often the first point of contact for assessing a scientific article (2009 et al., 2009). Abstracts that stand out in the editorial process or in the journals may therefore have a competitive advantage over abstracts that do not overly rely on the use of + jargon. In this manner, the goal of science, which consists in an approximation to the truth may be put in the rear while the focus shifts to the pursuit of ever more striking and novel results, and marketing them as such even though it may not be justified in the individual case. However, the rising sentiment within abstracts and the decrease in the proportion of abstracts that do not rely on + jargon could also be a reflection of science getting better and better, with results becoming more certain and well-founded, thus justifying more positive language. Several factors contradict this theory. Firstly, a certain degree of overinterpretation, exaggeration and misrepresentation of scientific findings is well documented in the literature (Chiu et al., 2017; Ochodo et al., 2013; Yavchitz et al., 2012). Secondly, the phenomenon of publication bias, which means that papers that were able to confirm their hypotheses in a statistically significant way are published more often and that researchers subsequently go to great lengths to reach that significance threshold or at least to create the appearance of it through the targeted use of certain terms, speaks against this assumption (Otte et al., 2022; Vinckers et al. 2021). Thirdly, the findings of Fanelli (2009) and John et al. (2012) speak against the image of an increasingly rigorous science, as considerable proportions of researchers admit to having already resorted to questionable research practices at least once.

#### Study limitations

One important limitation of the present study is the exclusive reliance on the MEDLINE as a source of abstracts in the fields of psychology, biology and physics. As can be appreciated in the Supplementary Materials, the Abstracts included in the present study do not cover evenly all subdisciplines of psychology, biology or physics. Therefore, some bias in the estimation of field specific jargon cannot be excluded.

# **Conclusion and prospect**

In this study, the scientific communication and language use of scientists was examined in different scientific areas, revealing an increasing usage of words with positive connotations to describe their own research in abstracts. It was shown that as time progressed over the last four decades, abstracts became more emotionally positively charged and that the rate of neutrally formulated abstracts steadily decreased, even after controlling for the length of the abstracts. This trend was especially pronounced in the discipline of psychology. All disciplines however were affected by this trend, suggesting that this may be a cross-disciplinary, systemic process. By triggering a low-threshold, positive affect through an excessive use of terms with positive connotations this trend can systematically impact the judgments of the respective abstracts in a way that is counterproductive to rigorous scientific practices. Future research in this area could turn to the construction and validation of science-specific lexicons so that sentiment analyses can be conducted across sciences and linguistic spinning can be detected in abstracts extensively without much effort. Likewise, in addition to the abstract, the rest of the article, or at least the introduction and the discussion section,

could be analyzed as well. Furthermore, an attempt to replicate the results on a database other than MEDLINE would be useful, as MEDLINE is characterized by a focus on biomedical literature and may not accurately represent the full range of other disciplines.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s11192-023-04864-6.

Funding Open access funding provided by University of Graz.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

#### References

- Abola, M. V., & Prasad, V. (2016). The use of superlatives in cancer research. JAMA Oncology, 2(1), 139– 141. https://doi.org/10.1001/jamaoncol.2015.3931
- Aldhous, P. (2011). Journal rejects studies contradicting precognition. New Scientist. Retrieved May 11, 2023, from https://www.newscientist.com/article/dn20447-journal-rejects-studies-contradicting-precognition
- Alves, H., Koch, A., & Unkelbach, C. (2017). Why good is more alike than bad: Processing implications. Trends in Cognitive Sciences, 21(2), 69–79. https://doi.org/10.1016/j.tics.2016.12.006
- Atkin, P. A. (2002). A paradigm shift in the medical literature. BMJ, 325(7378), 1450–1451. https://doi.org/ 10.1136/bmj.325.7378.1450
- Baes, N., Speagle, H., & Haslam, N. (2022). Has psychology become more positive? Trends in language use in article abstracts. *Frontiers in Psychology*, 13, 870549. https://doi.org/10.3389/fpsyg.2022.870549
- Boutron, I., Dutton, S., Ravaud, P., & Altman, D. G. (2010). Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA, 303(20), 2058– 2064. https://doi.org/10.1001/jama.2010.651
- Brischoux, F., & Angelier, F. (2015). Academia's never-ending selection for productivity. Scientometrics, 103(1), 333–336. https://doi.org/10.1007/s11192-015-1534-5
- Bubela, T. M., & Caulfield, T. (2010). Role and reality: Technology transfer at Canadian universities. *Trends in Biotechnology*, 28(9), 447–451. https://doi.org/10.1016/j.tibtech.2010.06.002
- Caulfield, T. (2018). Spinning the genome: Why science hype matters. Perspectives in Biology and Medicine, 61(4), 560–571. https://doi.org/10.1353/pbm.2018.0065
- Caulfield, T., & Condit, C. (2012). Science and the sources of hype. Public Health Genomics, 15(3–4), 209–217. https://doi.org/10.1159/000336533
- Chen, T. J. (2023). ChatGPT and other artificial intelligence applications speed up scientific writing. Journal of Chinese Medical Association, 86(4), 351–353. https://doi.org/10.1097/jcma.000000000000000
- Chiu, K., Grundy, Q., & Bero, L. (2017). 'Spin' in published biomedical literature: A methodological systematic review. PLoS Biology, 15(9), e2002173. https://doi.org/10.1371/journal.pbio.2002173
- Citron, F. M. (2012). Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language*, 122(3), 211–226. https://doi. org/10.1016/j.bandl.2011.12.007
- Clore, G. L., Wyer, R. S., Dienes, B., Gasper, K., Gohm, C., & Isbell, L. (2013). Affective feelings as feedback: Some cognitive consequences. In *Theories of mood and cognition* (pp. 27–62). Psychology Press. Retrieved May 11, 2023, from https://www.researchgate.net/profile/JrWyer/publication/22905 9906\_Affective\_Feelings\_as\_Feedback\_Some\_Cognitive\_Consequences/links/574f88ac08ae188 0a82289de/Affective-Feelings-as-Feedback-Some-Cognitive-Consequences.pdf

- Deutsch, R., & Strack, F. (2008). Variants of judgment and decision-making: The perspective of the reflective-impulsive model. In H. Plessner, C. Betsch, & T. Betsch (Eds.), *Intuition in judgment and decision making* (pp. 39–53). Erlbaum.
- Dreisbach, G., & Goschke, T. (2004). How positive affect modulates cognitive control: Reduced perseveration at the cost of increased distractibility. *Journal of Experimental Psychology: Learning, Memory,* and Cognition, 30(2), 343. https://doi.org/10.1037/0278-7393.30.2.343
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS ONE, 4(5), e5738. https://doi.org/10.1371/journal.pone.0005738
- Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. PLoS ONE, 5(4), e10068. https://doi.org/10.1371/journal.pone.0010068
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. Scientometrics, 90(3), 891–904. https://doi.org/10.1007/s11192-011-0494-7
- Fraser, A. G., & Dunstan, F. D. (2010). On the impossibility of being expert. BMJ. https://doi.org/10.1136/ bmj.c6815
- Fröber, K., & Dreisbach, G. (2014). The differential influences of positive affect, random reward, and performance-contingent reward on cognitive control. *Cognitive, Affective, and Behavioral Neuroscience,* 14(2), 530–547. https://doi.org/10.3758/s13415-014-0259-x
- Garcia, D., Garas, A., & Schweitzer, F. (2012). Positive words carry less information than negative words. EPJ Data Science, 1(1), 1–12. https://doi.org/10.1140/epjds3
- Horbach, S., Aagaard, K., & Schneider, J. W. (2021). Meta-Research: How problematic citing practices distort science. https://doi.org/10.31222/osf.io/aqyhg
- Hosseini, M., & Horbach, S. P. (2023). Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. *Research Integrity Peer Review*, 8(1), 4. https://doi.org/10.1186/s41073-023-00133-5
- Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8(1), pp. 216–225). Retrieved May 11, 2023, from https://ojs.aaai.org/index.php/ICWSM/article/view/14550
- Ioannidis, J. P. A. (2006). Concentration of the most-cited papers in the scientific literature: Analysis of journal ecosystems. *PLoS ONE*, 1(1), e5. https://doi.org/10.1371/journal.pone.0000005
- Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology*, 19(5), 640–648. Retrieved May 11, 2023, from http://www.jstor.org/stable/25662607
- Islamaj Dogan, R., Murray, G. C., Névéol, A., & Lu, Z. (2009). Understanding PubMed user search behavior through log analysis. *Database*. https://doi.org/10.1093/database/bap018
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524–532. https://doi.org/10.1177/ 0956797611430953
- Laws, K. R. (2013). Negativland—A home for all findings in psychology. BMC Psychology, 1(1), 1–8. https://doi.org/10.1186/2050-7283-1-2
- Li, X. A., Wu, Q. L., Hubbard, K., Hwang, J., & Zhong, L. (2023). Information-seeking behavior for COVID-19 boosters in China: A cross-sectional survey. *Vaccines*, 11(2), 323. https://doi.org/10.3390/ vaccines11020323
- Liu, X., & Zhu, H. (2023). Linguistic positivity in soft and hard disciplines: Temporal dynamics, disciplinary variation, and the relationship with research impact. *Scientometrics*, 128(5), 3107–3127. https:// doi.org/10.1007/s11192-023-04679-5
- Madden, C. S., Easley, R. W., & Dunn, M. G. (1995). How journal editors view replication research. Journal of Advertising, 24(4), 77–87. https://doi.org/10.1080/00913367.1995.10673490
- Mahrukh, R., Shakil, S., & Malik, A. S. (2023). Sentiments analysis of fMRI using automatically generated stimuli labels under naturalistic paradigm. *Scientific Reports*, 13(1), 7267. https://doi.org/10.1038/ s41598-023-33734-7
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, 7(6), 537–542. https://doi.org/10.1177/17456 91612460688
- Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, 27, 16–32. https://doi.org/10. 1016/j.cosrev.2017.10.002
- Meho, L. I. (2007). The rise and rise of citation analysis. *Physics World*, 20(1), 32. https://doi.org/10.1088/ 2058-7058/20/1/33
- Moyle, B. D., Weaver, D. B., & McLennan, C. L. (2022). Citation quality and knowledge creation in tourism. *Tourism Recreation Research*, 47(1), 105–110. https://doi.org/10.1080/02508281.2020.1862980

- National Library of Medicine. (n.d.). *MEDLINE: Overview*. Retrieved May 11, 2023, from https://www. nlm.nih.gov/medline/medline\_overview.html
- Neuliep, J. W., & Crandall, R. (1993). Reviewer bias against replication research. *Journal of Social Behavior and Personality*, 8(6), 21. Retrieved May 11, 2023, from https://www.proquest.com/scholarly-journ als/reviewer-bias-against-replication-research/docview/1292304227/se-2
- Niu, X., & Hemminger, B. M. (2012). A study of factors that affect the information-seeking behavior of academic scientists. *Journal of the American Society for Information Science and Technology*, 63(2), 336–353. https://doi.org/10.1002/asi.21669
- Ochodo, E. A., de Haan, M. C., Reitsma, J. B., Hooft, L., Bossuyt, P. M., & Leeflang, M. M. (2013). Overinterpretation and misreporting of diagnostic accuracy studies: Evidence of "spin." *Radiology*, 267(2), 581–588. https://doi.org/10.1148/radiol.12120527
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716. https://doi.org/10.1126/science.aac4716
- Otte, W. M., Vinkers, C. H., Habets, P. C., van Ijzendoorn, D. G., & Tijdink, J. K. (2022). Analysis of 567,758 randomized controlled trials published over 30 years reveals trends in phrases used to discuss results that do not reach statistical significance. *PLoS Biology*, 20(2), e3001562. https://doi.org/ 10.1371/journal.pbio.3001562
- Pavlovic, V., Weissgerber, T., Stanisavljevic, D., Pekmezovic, T., Milicevic, O., Lazovic, J. M., Cirkovic, A., Savic, M., Rajovic, N., Piperac, P., Djuric, N., Madzarevic, P., Dimitrijevic, A., Randjelovic, S., Nestorovic, E., Akinyombo, R., Pavlovic, A., Ghamrawi, R., Garovic, V., & Milic, N. (2021). How accurate are citations of frequently cited papers in biomedical literature? *Clinical Science*, 135(5), 671–681. https://doi.org/10.1042/CS20201573
- Plume, A., & van Weijen, D. (2014). Publish or perish? The rise of the fractional author.... Research Trends, 1(38), 5. Retrieved May 11, 2023, from https://www.researchtrends.com/researchtrends/vol1/iss38/5? utm\_source=www.researchtrends.com%2Fresearchtrends%2Fvol1%2Fiss38%2F5&utm\_medium= PDF&utm\_campaign=PDFCoverPages
- Powell, K. (2015). The future of the postdoc. Nature, 520(7546), 144-148. https://doi.org/10.1038/520144a
- Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. Consciousness and Cognition, 8(3), 338–342. https://doi.org/10.1006/ccog.1999.0386
- Sauder, M., Newsome, K., Zagales, I., Autrey, C., Das, S., Ang, D., & Elkbuli, A. (2022). Evaluation of citation inaccuracies in surgical literature by journal type, study design, and level of evidence: Towards safeguarding the peer-review process. *The American Surgeon*, 88(7), 1590–1600. https://pubmed.ncbi. nlm.nih.gov/35130090/
- Savolainen, R. (2014). Emotions as motivators for information seeking: A conceptual analysis. Library and Information Science Research, 36(1), 59–65. https://doi.org/10.1016/j.lisr.2013.10.004
- Schwieder, D. (2016). Low-effort information searching: The heuristic information-seeking toolkit. Behavioral and Social Sciences Librarian, 35(4), 171–187. https://doi.org/10.1080/01639269.2017.1289019
- Scott, G. G., O'Donnell, P. J., Leuthold, H., & Sereno, S. C. (2009). Early emotion word processing: Evidence from event-related potentials. *Biological Psychology*, 80(1), 95–104. https://doi.org/10.1016/j. biopsycho.2008.03.010
- Simkin, M. V., & Roychowdhury, V. P. (2003). Read before you cite! *Complex Systems*, 14(3), 269–274. https://doi.org/10.25088/ComplexSystems.14.3.269
- Statzner, B., & Resh, V. H. (2010). Negative changes in the scientific publication process in ecology: Potential causes and consequences. *Freshwater Biology*, 55(12), 2639–2653. https://doi.org/10.1111/j.1365-2427.2010.02484.x
- Tahamtan, I., Safipour Afshar, A., & Ahamdzadeh, K. (2016). Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics*, 107, 1195–1225. https://doi.org/10.1007/ s11192-016-1889-2
- Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin*, 110(1), 67. https://doi.org/10.1037/0033-2909.110.1.67
- Teixeira, M. C., Thomaz, S. M., Michelan, T. S., Mormul, R. P., Meurer, T., Fasolli, J. V. B., & Silveira, M. J. (2013). Incorrect citations give unfair credit to review authors in ecology journals. *PLoS ONE*, 8(12), e81871. https://doi.org/10.1371/journal.pone.0081871
- Topolinski, S., & Strack, F. (2009a). The architecture of intuition: Fluency and affect determine intuitive judgments of semantic and visual coherence and judgments of grammaticality in artificial grammar learning. *Journal of Experimental Psychology: General*, 138(1), 39–63. https://doi.org/10.1037/a0014 678
- Topolinski, S., & Strack, F. (2009b). Scanning the "fringe" of consciousness: What is felt and what is not felt in intuitions about semantic coherence. *Consciousness and Cognition*, 18(3), 608–618. https://doi. org/10.1016/j.concog.2008.06.002

- Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: The density hypothesis. *Journal of Personality and Social Psychology*, 95(1), 36. https://doi.org/10.1037/0022-3514.95.1.36
- Unkelbach, C., Koch, A., & Alves, H. (2021). Explaining negativity dominance without processing bias. *Trends in Cognitive Sciences*, 25(6), 429–430. https://doi.org/10.1016/j.tics.2021.04.005
- Verde, M. F., Stone, L. K., Hatch, H. S., & Schnall, S. (2010). Distinguishing between attributional and mnemonic sources of familiarity: The case of positive emotion bias. *Memory and Cognition*, 38(2), 142–153. https://doi.org/10.3758/MC.38.2.142
- Vinkers, C. H., Lamberink, H. J., Tijdink, J. K., Heus, P., Bouter, L., Glasziou, P., Moher, D., Damen, J. A., Hooft, L., & Otte, W. M. (2021). The methodological quality of 176,620 randomized controlled trials published between 1966 and 2018 reveals a positive trend but also an urgent need for improvement. *PLoS Biology*, 19(4), e3001162. https://doi.org/10.1371/journal.pbio.3001162
- Vinkers, C. H., Tijdink, J. K., & Otte, W. M. (2015). Use of positive and negative words in scientific Pub-Med abstracts between 1974 and 2014: Retrospective analysis. *BMJ*. https://doi.org/10.1136/bmj. h6467
- Whittlesea, B. W. (1993). Illusions of familiarity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19(6), 1235–1253. Retrieved May 11, 2023, from https://psycnet.apa.org/
- Winter, D. J. (2017). rentrez: An R package for the NCBI eUtils API (No. e3179v2). PeerJ Preprints. https://doi.org/10.7287/peerj.preprints.3179v2
- Yavchitz, A., Boutron, I., Bafeta, A., Marroun, I., Charles, P., Mantz, J., & Ravaud, P. (2012). Misrepresentation of randomized controlled trials in press releases and news coverage: A cohort study. *PLoS Medicine*, 9(9), e1001308. https://doi.org/10.1371/journal.pmed.1001308
- Young, N. S., Ioannidis, J. P. A., & Al-Ubaydli, O. (2008). Why current publication practices may distort science. *PLoS Medicine*, 5(10), e201. https://doi.org/10.1371/journal.pmed.0050201
- Yuan, Z. M., & Yao, M. (2022). Is academic writing becoming more positive? A large-scale diachronic case study of Science research articles across 25 years. *Scientometrics*, 127(11), 6191–6207. https://doi.org/ 10.1007/s11192-022-04515-2
- Zhang, M., & Jansen, B. J. (2009). Influences of mood on information seeking behavior. In CHI'09 extended abstracts on human factors in computing systems (pp. 3395–3400). https://doi.org/10.1145/1520340. 1520492