From Joy to Fear in Scientific Titles: Automated Emotion Recognition and the Citation Payoff

Nicolas Gerardy[1], Nicolas Kervyn[1], and Vincenzo Verardi[1]

[1]UCLouvain - LouRIM. Place des Doyens 1/L2.01.01. B-1348 Louvain-la-Neuve.

This is a post-peer-review, pre-copyedit version of an article published in Scientometrics. The final authenticated version is available online at: https://doi.org/10.1007/s11192-026-05720-z

Abstract

Citation-based indicators are widely used to assess scholarly impact, yet they may be linked to factors unrelated to scientific contribution. This study investigates whether emotions in article titles might constitute an attentional bias that could be associated with citation outcomes. Focusing on marketing research, we analyze 18,272 article titles indexed in Scopus, of which 15,836 are included in the final analysis after data cleaning. We assess their emotional content with ChatGPT by aggregating the outcomes of multiple runs rather than depending on a single output, and we organize the resulting classifications according to Plutchik's psychoevolutionary model of emotions. Using large language models helps us capture subtle differences in emotional intensity and avoid the limits of traditional lexicon-based methods. We estimate Pseudo Poisson Maximum Likelihood models with journal and publication-year fixed effects to assess associations between emotional signals and citation counts. The results show that certain primary emotions, such as joy, as well as specific emotional combinations, are systematically associated with higher citation performance, while others, like despair, correlate with lower visibility. These results indicate that the emotional complexity of scientific titles is associated with citation patterns, pointing to an affective factor that has received little attention in how scholarly visibility and impact are constructed.

Keywords: Citations, Marketing, Emotion analysis, Article titles, LLM, Scopus dataset

JEL: M31 (Marketing), C55 (Large Data Sets: Modeling and Analysis)

1 Introduction

Titles play a central role in shaping the visibility of scientific work (Russell-Bennett and Baron, 2016). In marketing, emotional signals in product names or advertisements can attract attention, spark curiosity, and influence choice (Daume and Hüttl-Maack, 2020; Vrtana and Krizanova, 2023). In a similar way, emotional elements in academic titles may be associated with how readers engage with scholarly work. As the first and often only element encountered in digital search environments, a title functions as filter: it can invite attention or deflect it.

Since citations serve as indicators of recognition, diffusion, and intellectual acceptance (Pieters and Baumgartner, 2002), any systematic influence of emotion on title appeal raises the possibility that scholarly impact may reflect not only by scientific merit but also by emotional attributes in article titles.

While previous scientometric research has examined structural or stylistic features of titles (Nair and Gibbert, 2016), far less attention has been given to the role of emotional content as a potential driver of citation behavior. Yet psychological research demonstrates that emotions capture attention, modulate cognitive processing, and guide behavioral responses, often automatically and outside conscious awareness (Ellsworth, 1991; Lazarus, 1991; Scherer, 1984). If emotional cues in titles function as subtle attentional biases, they may be associated with the pathways through which scientific work gains visibility and, thereby, be correlated with citation counts independently of scientific contributions. Understanding this mechanism is essential for interpreting citation-based indicators and for assessing the fairness and neutrality of impact metrics.

Advances in deep learning have provided new tools that can detect and measure emotions in large amounts of text. Traditional approaches to emotion analysis are mainly based on lexicons, hand-crafted rules, or classifiers that struggle to capture the complexity of emotional nuance, particularly in short texts such as article titles. Large language models (LLMs), including transformer-based architectures such as GPT, overcome these limitations by leveraging large-scale pre-training to interpret context, detect subtle affective indicators, and model long-range dependencies with high accuracy. These models have demonstrated strong performance in identifying emotional tone even in highly condensed linguistic environments, making them well-suited for scientometric applications requiring precise extraction.

To conceptualize emotions in a systematic manner, this study relies on Plutchik's psychoevolutionary model of emotion. This model identifies eight primary emotions organized in opposing pairs, each varying in intensity and capable of combining into more complex affective states.

This study brings these dimensions together by employing ChatGPT as an emotion classifier based on Plutchik's psychoevolutionary model and applying it to a large corpus of marketing article titles. We aim to assess whether emotional intensities, individually or in combination, are associated with citation counts once journal and year-specific heterogeneity is accounted for. By doing so, we investigate whether emotion constitutes an attentional bias capable of shaping scholarly recognition. Beyond offering new empirical evidence, this work contributes methodologically to the emerging intersection of scientometrics and LLM-based text analytics. It shows how language models can be put into practice to assess affective dimensions on a large scale and, importantly, how these emotional indicators may be linked to the formation of scientific visibility.

We restricted the dataset to articles tagged with "marketing" according to Scopus to maintain a coherent and comparable sample in which titles are deliberately crafted to attract attention. In marketing research, authors are especially aware of how wording can be linked to interest and engagement, so emotional indicators in titles are more likely to be intentional rather than accidental. This makes it a suitable context for examining how emotional content in titles relates to citation rates while also reducing noise that would arise from mixing very different disciplines with different title conventions.

In highlighting the role of emotional markers in title-driven attention dynamics, our study raises broader implications for scientometrics. If emotional content in titles is correlated with citation count, then citation metrics may partially reflect affective leverage rather than purely scientific contributions. Recognizing this possibility is essential for interpreting bibliometric indicators and for promoting greater transparency in how scholarly impact is constructed in an increasingly digital and attention-driven scientific ecosystem.

Throughout the paper, our estimates are interpreted strictly as conditional associations within the 2001–2015 marketing corpus. Any phrasing that may seem to suggest a causal effect should therefore be understood as referring only to an observed association, not to causal evidence.

2 Theoretical Background & Literature Review

2.1 Appraisal Theory

In appraisal theory, emotions are understood as responses to events that disrupt an individual's environment and bear relevance to their personal concerns (Frijda, 1986). Rather than constituting static internal states, emotions are dynamic processes that arise from the way individuals interpret and evaluate such events (Frijda, 2007; Moors et al., 2013). These appraisal processes engage both nervous and somatic systems, underscoring the embodied and physiological nature of emotional episodes (Scherer, 2005). Because of their ephemeral nature, emotions differ from more enduring affective phenomena. For example, Scherer (2005) classifies love as an attitude and depression as a mood, since both last longer and therefore do not qualify as emotions in a strict sense.

As appraisals depend on individual concerns and interpretations, the same event may elicit different emotions across individuals. In contrast, when evaluations of an event are similar, they tend to evoke similar emotional responses (Moors et al., 2013). Through this mechanism, appraisal processes translate environmental stimuli into emotions that guide subsequent action and behavior (Clore and Ortony, 2000; Moors et al., 2013; Roseman and Smith, 2001). Importantly, identical emotions may give rise to different behavioral responses depending on situational interpretation. For example, fear may initially trigger avoidance, yet upon reevaluation of the situation, a more confrontational or assertive response may emerge as a means of regulating that emotion (Ellsworth and Scherer, 2003; Scherer, 1984).

According to Scherer's Component Process Model (Scherer, 2009), emotions arise through a sequential appraisal process. First, there is novelty detection, meaning something changes in a person's environment. If this change is relevant to their concerns, they evaluate whether it is pleasant or unpleasant (intrinsic pleasantness). Then, they assess the cause of the event and whether they can control or cope with it. This appraisal sequence is ultimately associated to the emergence of a specific emotion.

2.2 Relation Between Emotions and Attention

From the perspective of selective attention theory, emotions tend to bias attention. Researchers have shown that people direct their focus to a biased competition mechanism (Desimone et al., 1995): stimuli that attract attention are in competition with one another. Since emotional stimuli are naturally more prominent, they tend to dominate and capture attention more easily (Yiend, 2010).

Studies show that individuals are more efficient at identifying targets when they carry emotional signals, particularly in tasks involving threat detection or visual inspection. When facing emotional stimuli, brain emotion-related regions, such as the amygdala and associated cortical areas, are automatically activated (Carretié, 2014). This suggests that emotional content is processed spontaneously, even when it is irrelevant to the task (Frischen et al., 2008; Vuilleumier et al., 2001).

Along the same lines, using words that carry a negative emotional connotation, as opposed to neutral ones, in an advertisement, enhances the attention individuals pay to the advertisement, as evidenced by higher fixation rates on a webpage (Ferreira et al., 2011). In addition, fear is strongly linked to attentional processes because of its evolutionary relevance for survival-related behaviour (Armony and Ledoux, 1999).

Emotional effects extend to textual information: reading is a complex emotion-inducing process, and emotional words possess privileged access to attentional and cognitive resources compared to their neutral counterparts (Hsu et al., 2015; Kissler et al., 2007).

In scholarly contexts, titles are typically among the first indicators used to judge relevance (Chamorro-Padial and Rodríguez-Sánchez, 2023; Jamali and Nikzad, 2011). For this reason, the emotional properties of titles may be correlated with the attention they receive and could be associated with resulting outcomes such as citation patterns.

2.3 Textual Features of Scientific Articles and Their Correlation with Impact

Several studies investigated the characteristics of effective scientific article titles in various disciplines, including psychology, medicine, and the natural sciences. Research suggests that longer titles are associated with higher citation counts (Habibzadeh and Yadollahie, 2010; Jacques and Sebire, 2010).

The use of split titles, where the first part introduces the main topic followed by additional detail or clarification after a splitting symbol, has also been extensively studied. Some researchers report a positive association between split titles and citation counts (Rostami et al., 2014; Krausman and Cox, 2020), while others have observed a negative one (Jamali and Nikzad, 2011; Paiva et al., 2012).

In the field of management science, Nair and Gibbert (2016) examined the impact of various title components on citation rates. Their analysis considered title length, the use of non-alphanumeric characters (e.g., colons, question marks), and other linguistic features. The findings revealed a negative correlation between the number of non-alphanumeric characters in a title and citation rates, while title length and linguistic elements, such as the use of metaphors, exhibited minimal

impact. Other research revealed that elements of articles, such as the use of attention grabbers, do not significantly correlate with citation rates. However, other aspects, such as the quality of the overall article, the domain of research, the visibility of the authors, and the number of references, are positively correlated with citation count (Peters and van Raan, 1994; Stremersch et al., 2007; Van Dalen and Henkens, 2001; Webster et al., 2009)

Others studied the effect of pleasantness or humor in article titles. The results show that the presence of these emotions is not directly correlated with the citation rate (Sagi and Yechiam, 2008). These studies subjectively assessed pleasantness through individual evaluations using Likert scales. This self-reported measure could be criticized (Ciuk et al., 2015; Karpen, 2018), especially when evaluating emotions that can be considered unconscious reactions through appraisal theory.

Beyond these early attempts to capture affective content through subjective ratings, a broader body of work has examined how linguistic positivity in academic writing relates to citation impact. Yuan and Yao (2022) showed, based on a large-scale analysis of articles published in Science over 25 years, that academic writing has become more positive over time. This positivity marker was studied across eight disciplines, confirming that the trend toward positive language is pervasive and that it is associated with higher citation counts (Liu and Zhu, 2023). Edlinger et al. (2023) reached similar conclusions by analysing over 2.3 million abstracts, finding that the use of positively valenced scientific jargon has grown steadily. Direct evidence provided that positive words serve as a predictor of citation counts across journals of varying quality, although this citation advantage has been declining over time (Wu et al., 2024). Related studies further showed that positive language in titles and abstracts acts as a credible signal of research quality and impacts (Wu et al., 2025; Fronzetti Colladon et al., 2020). In a closely related contribution, Ma et al. (2023) applied BERT-based sentiment analysis to 87 years of articles published in the top four marketing journals and found that a more positive writing style mediates the relationship between author gender and citation impact. Similarly, an analysis of over 130,000 abstracts published in Science, Nature, and PNAS shows that promotional language predicts more citations, more article views, and higher alt-metric scores (Stavrova et al., 2025).

According to Cui et al. (2024), while not directly related to citation count, clicks are another important measure of success. The study finds that the use of strong emotional expressions in YouTube video thumbnails, whether positive or negative (through facial expressions, for instance), can significantly boost viewership. In addition, certain title features, such as the inclusion of capital letters or punctuation marks, are shown to enhance click rates.

This literature suggests that the affective properties of academic texts, whether measured through title, structure, or lexical positivity, are associated with how scholarly work is received. However, existing studies have largely relied on unidimensional measures of sentiment, capturing whether language is positive or negative rather than distinguishing between specific emotional states. Our study contributes to this literature by applying Plutchik's eight-dimensional model of

emotions to article titles, allowing us to identify which particular emotions, and which combinations therefore, are associated with citation outcomes.

2.4 Plutchik's Psychoevolutionary Theory

As the appraisal theory, Plutchik's psychoevolutionary theory explains that an emotion is not just a feeling but a hypothesized construct inferred from several indicators, including what people say, how they act, and how others react (Plutchik, 1980b). This represents a complex, adaptive response to important life challenges. It posits eight basic emotion dimensions (joy, trust, fear, surprise, sadness, disgust, anger, and anticipation) organized into four pairs of opposites (joy–sadness, trust–disgust, fear–anger, anticipation–surprise) that vary in intensity and can combine to form more complex affective states (Plutchik, 1980b).

Plutchik's flower models emotions like colors on a wheel, illustrating their relationships and intensity (see Fig. 1). Plutchik's model arranges the eight primary emotions at the center of the wheel, each of them expressed with weaker and stronger intensity levels (Semeraro et al., 2021). Acceptance, for instance, is presented as the milder form of the trust dimension, while admiration is its more intense variant (Plutchik, 2001)[1]. The complete intensity gradients are shown in Table 1.

Weaker	**Primary**	**Stronger**
Serenity	Joy	Ecstasy
Acceptance	Trust	Admiration
Apprehension	Fear	Terror
Distraction	Surprise	Amazement
Pensiveness	Sadness	Grief
Boredom	Disgust	Loathing
Annoyance	Anger	Rage
Interest	Anticipation	Vigilance

Table 1: Primary emotions and their intensity variations in Plutchik's model Plutchik (2001)

In Plutchik's framework, the eight primary emotions can mix with one another to form more complex emotional states. He described these mixtures as dyads, arranged in successive tiers radiating outward from the wheel (Plutchik, 1980a). Primary dyads arise from the mixing of adjacent emotions, creating states such as joy with trust, which forms love, or fear with surprise, which produces awe. Secondary dyads combine emotions that are separated by one step, leading to more complex feelings: joy together with fear gives guilt, while anger mixed with joy results in

[1]In Plutchik (1980a) acceptance was considered as a primary emotion

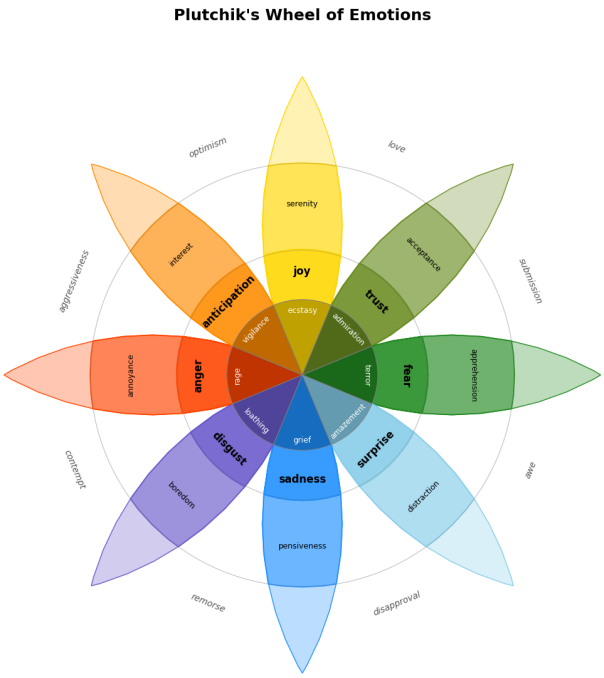

Figure 1: Plutchik's wheel of emotions

pride. Tertiary dyads involve emotions that are three steps apart, such as trust with sadness, which produces sentimentality, or anticipation with fear, which generates anxiety. Opposite emotions, however, neutralize one another rather than forming coherent states, which is why no additional level beyond the tertiary dyads is recognized in Plutchik's model. Following Plutchik's structural assumption that emotional similarity decreases with angular distance on the wheel, we expect dyads formed by more distant emotion pairs to be less conceptually coherent and therefore less well defined, with opposing pairs being the limiting case in which antagonistic components counteract rather than combine into an interpretable state. Table 2 illustrates every dyads[2].

[2]Table 2 is based on Plutchik's original classification of dyads, but the wording of some emotion labels has been slightly adjusted to align with Plutchik (2001).

Table 2: Plutchik's Dyads of Emotions (Primary, Secondary, and Tertiary)

Dyad type	**Emotions combined**	**Name of dyad**
Primary (adjacent)	Joy + Trust	Love
	Trust + Fear	Submission
	Fear + Surprise	Awe
	Surprise + Sadness	Disapproval
	Sadness + Disgust	Remorse
	Disgust + Anger	Contempt
	Anger + Anticipation	Aggressiveness
	Anticipation + Joy	Optimism
Secondary (two apart)	Joy + Fear	Guilt
	Trust + Surprise	Curiosity
	Fear + Sadness	Despair
	Surprise + Disgust	Unbelief
	Sadness + Anger	Envy
	Disgust + Anticipation	Cynicism
	Anger + Joy	Pride
	Anticipation + Trust	Hope
Tertiary (three apart)	Joy + Surprise	Delight
	Trust + Sadness	Sentimentality
	Fear + Disgust	Shame
	Surprise + Anger	Outrage
	Sadness + Anticipation	Pessimism
	Disgust + Joy	Morbidity
	Anger + Trust	Dominance
	Anticipation + Fear	Anxiety

This structure makes the model appealing for linguistics and text-based emotion mining. In particular, the widely used NRC Word–Emotion Association Lexicon encodes word–emotion associations specifically for Plutchik's eight categories (Mohammad and Turney, 2013). The Plutchik model has been revealed to perform well when analyzing online text, reviews, or even posts on social media (Li et al., 2023; Mohsin and Beltiukov, 2019; Motger et al., 2025).

2.5 ChatGPT as a Tool for Emotion Recognition in Text

Recent research provides compelling evidence that ChatGPT demonstrates strong competence in identifying and interpreting emotions in textual material, including condensed forms such as titles, headlines or tweets (Belal et al., 2023; Kang and Choi, 2025). Studies have shown that large language models can capture subtle affective signals embedded in linguistic patterns, tone, and lexical choice with a level of accuracy that approaches human performance in emotional awareness evaluations (Elyoseph et al., 2023).

Wake et al. (2023) observed that ChatGPT 3.5-turbo performs reliably on various emotion recognition benchmarks, though its precision may depend on dataset composition and emotional category balance. In a complementary study, scholars showed that ChatGPT 3.5 can infer fine-grained emotional dimensions such as valence, arousal, and dominance without additional training, revealing a nuanced understanding of affective meaning (Broekens et al., 2023). Moreover, Schaaff et al. (2023) reported that ChatGPT 3.5 identifies emotional states in more than ninety percent of evaluated cases.

Further evidence comes from Yongsatianchot et al. (2023), who applied appraisal theory to assess the perception of emotional context in ChatGPT 3.5 and 4. Their results show that their attributions align closely with established psychological models of emotion generation, and found that LLMs follow the trend of human annotations on emotional valence evaluation. In addition, Fatahi et al. (2024) highlighted ChatGPT 3.5 is biased toward more positive emotions compared to human classification.

Beyond general emotion recognition, recent work has confirmed the model's robustness in more specific domains (Lecourt et al., 2025). Lecourt et al. (2025) conducted large-scale comparative experiments showing that ChatGPT performs competitively on multiple emotion detection datasets, especially when applying prompt engineering techniques. They found that ChatGPT, based on GPT-3.5 and GPT-4o, achieved the highest accuracy among general-purpose models. GPT-4 reached a macro F1 score (i.e., the harmonic mean of precision measured as the proportion of correctly identified emotions among all identified ones and recall measured as the proportion of correctly identified emotions among all actual ones) of about 30.95 %, surpassing competitors such as LLaMA, Gemini, and Mistral.

Butler et al. (2025) applied Plutchik's Wheel of Emotions and their dyadic combinations to compare how different large language models recognize and express emotions in textual contexts. In this comparative analysis, ChatGPT-4 showed a more balanced and contextually coherent mapping across Plutchik's categories than competing models such as Claude 2, Gemini 1.5, and LLaMA 3. Its emotion classification remained stable across both neutral and emotionally charged prompts, and it showed particular strength in detecting mixed affective states, for instance recognizing combinations such as "joy–anticipation" or "sadness–fear." The study concludes that ChatGPT's internal

linguistic representations appear well-suited to implementing Plutchik's multidimensional model, supporting the idea that recent large language models are capable of processing emotion in a way that mirrors classical psychological taxonomies more closely than previous generation systems.

Finally, empirical studies indicate that ChatGPT continues to improve as successive versions are released. Evaluations show that newer models outperform alternative emotion-detection tools such as RoBERTa, and that each generation of the pre-trained transformer exceeds the performance of its predecessors, with GPT-4 consistently outperforming GPT-3 and GPT-3.5 (Amin et al., 2024; Krugmann and Hartmann, 2024).

Taken together, these findings suggest that ChatGPT can successfully interpret emotional undertones not only in extended discourse but also in shorter and more compact linguistic forms. ChatGPT's deep semantic representation allows it to detect such subtle emotional signals with considerable precision. This capacity positions it as a valuable instrument for automated affective analysis across diverse textual genres.

3 Data

We focused on articles classified as marketing by Scopus. We considered the period ranging from January 1, 2001, to December 31, 2015. This is consistent with Webster et al. (2009) as they lagged their analysis in time due to the high variation in citation count in the early stage of the paper's existence. This window is deliberately chosen to ensure that citation counts have reached a stable and comparable level across all observations. The primary objective of this study is to identify structural associations between emotional signals in titles and long-run citation outcomes, rather than to describe current publishing practices. What matters is that the citation accumulation process is sufficiently complete for all articles in the sample. As shown in Appendix 1, citation counts for marketing articles continue to accumulate well beyond five years after publication, and papers published after 2015 would therefore exhibit artificially low citation counts at the time of data collection, introducing a systematic downward bias. Conversely, articles published before 2001 have largely exited the active citation cycle and show signs of saturation. The 2001–2015 window thus represents the period over which citation counts are both sufficiently accumulated and still actively growing, providing the most reliable basis for estimating the structural relationship of interest. The specific Scopus search query was:

```
SUBJAREA(BUSI) AND KEY("Marketing") AND PUBYEAR > 2000 AND PUBYEAR < 2016
AND DOCTYPE(ar)
```

Marketing was selected as the focal discipline for this study because, as an academic discipline, it

is centered on attention, capture, and persuasion competencies that align closely with our research questions about emotional content in titles. The purpose of the academic discipline of marketing is "to develop theoretically informed but practical advice for organizations to help them better navigate constantly evolving markets, with a particular emphasis on understanding customers and transforming this information into action" (Gustafsson and Ghanbarpour, 2022, p. 184). This disciplinary orientation toward audience engagement suggests that marketing scholars may be particularly sensitive to the strategic use of emotional appeals in their own academic communication. This focus on marketing research defined a coherent sample while preserving comparability, which would be more difficult to justify if multiple disciplines were mixed.

The final regression sample is constructed through a sequential cleaning procedure applied to the raw Scopus export. Table 3 summarises each step and the number of observations excluded at each stage. Starting from the 18,272 articles returned by the Scopus query, three observations are removed because the GPT response does not begin with the expected emotion label, indicating a malformed output, and a further 52 are dropped because the response string exceeds 100 characters, suggesting that the model returned an explanation rather than a structured score. Articles with a missing or non-positive page count are then excluded as they likely correspond to coding errors in the Scopus metadata, reducing the sample to 16,865 observations. Twenty additional observations are dropped due to a missing citation count. The resulting eligible sample of 16,845 articles is then passed to the PPML estimator with journal and publication-year fixed effects. As is standard with high-dimensional fixed effects estimators, `ppmlhdfe` automatically drops singleton observations - that is, articles belonging to a journal represented by only a single observation in the dataset - as these cannot contribute to the identification of the fixed effects. This procedure excludes a further 2,358 observations, yielding a final regression sample of 15,836 articles.

Table 3: Sample Construction and Data Cleaning

Step	Observations	Excluded
Total Scopus query	18,272	
Malformed GPT responses	18,269	−3
Responses too long (>100 characters)	18,217	−52
Missing or non-positive page count	16,865	−1,352
Missing citation count	16,845	−20
Singletons excluded by `ppmlhdfe`	15,836	−2,358
Final regression sample	**15,836**	

Notes: Singletons are observations belonging to a journal represented by a single article in the dataset. They are automatically dropped by the `ppmlhdfe` estimator as they cannot contribute to the identification of the fixed effects.

4 Methodology

To analyze the emotions conveyed in the titles of scientific articles in marketing, we run a Python-based code within the Orange Data Mining environment (Demšar et al., 2013), in which queries are submitted to ChatGPT through an API. This API expects input as a list of messages, each defined by a `role` and corresponding `content`. The **system message** specifies the model's expert role, whereas the **user message** provides the actual task instructions and input text. The model then returns a concise, single-line output containing the intensity values of each emotion, expressed as integers ranging from 1 to 10. The resulting emotion scores are then stored and subsequently processed for statistical analysis.

The choice to rely on a large language model for emotion scoring, rather than on human annotation or lexicon-based methods, is grounded in empirical evidence. Large language models have been shown to match or outperform human annotators on text classification and emotion labelling tasks, while offering greater scalability and consistency (Gilardi et al., 2023; Törnberg, 2023). Lexicon-based methods assign emotional scores to individual words independently of their context. While this approach can be effective for longer texts with explicit emotional vocabulary, it is poorly suited to short academic titles.

4.1 System and User Roles in the ChatGPT API

The ChatGPT API operates with structured messages, each defined by a `role` and corresponding `content`. In our configuration, two roles were specified: the `system` role and the `user` role.

The `"system"` message establishes the model's identity and expertise, guiding how it should interpret and respond to the input. In our case, the model was instructed to act as *"an expert at analyzing emotions in the titles of scientific marketing articles."*

The `"user"` message provides the concrete task and the input data. It requests the model to assign an intensity value (from 1 to 10) to each of Plutchik's eight primary emotions (Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, and Trust) based on the emotional tone conveyed in each title.

When producing an answer, ChatGPT considers both messages jointly: the `system` message shapes the analytical framework and role of the model, while the `user` message supplies the specific content and task to be completed. This dual-message structure ensures that outputs are both context-aware and task-relevant.

In the Python implementation, messages are defined within a list passed to the ChatGPT API:

```
messages = [
    {
        "role": "system",
        "content": "You're an expert at analyzing emotions in the titles "
                   "of scientific marketing articles."
    },
    {
        "role": "user",
        "content": (
            "For each title, provide the intensity scale for each of "
            "Plutchik's eight primary emotions "
            "(Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust), "
            "exactly as Plutchik defined them in his psychoevolutionary theory of emotion. "
            "Assign a number from 1 to 10 quantifying each emotion's intensity. "
            "Return all emotions and their intensities in a single line, "
            "separated by commas:\n\n{text}"
        )
    }
]

response = client.chat.completions.create(
    model="CHOSEN_MODEL_HERE",
    messages=messages,
    seed=SET_SEED_HERE
)
```

However, a practical concern with any LLM-based measurement is that outputs are stochastic: querying the same model on the same title, or using a different models does not guarantee identical scores. To address this, we label the full corpus five times using a deliberately varied set of configurations. Three runs use GPT-5-mini with distinct random seeds (seed =42, seed =1, and seed =2) (OpenAI, 2026b). However, even with a fixed seed, perfect reproducibility remains difficult: the OpenAI API aims to sample deterministically, but determinism is not guaranteed and may vary with backend changes (tracked via `system_fingerprint`) (OpenAI, 2026a). Two additional runs use different model versions: one with ChatGPT-5.4 and one with GPT-5.4-mini, each with its default configuration (OpenAI, 2026c). This design captures two distinct sources of scoring variability: within-model stochasticity, reflected by the three GPT-5-mini runs which differ only in their random seed, and between-model variation, reflected by the two additional model versions. The five resulting annotated datasets share the same papers and differ only in how the LLM scored each paper on the eight Plutchik dimensions. They are, therefore, five noisy readings of the same corpus, not five independent samples.

4.2 Rubin's rules in the context of this paper

The paper treats ChatGPT's eight Plutchik scores as explanatory variables in a PPML citation model. These scores are not directly observed properties of the individual papers, but annotations produced by querying the LLM with each paper's title and interpreting the resulting response through Plutchik's psychoevolutionary model of emotions. A single LLM reading can therefore be seen as one noisy realization of an underlying annotation process. This motivates a *Rubin-style* pooling approach across the five independent readings.

Rubin's framework was originally developed for survey non-response, where multiple imputations represent alternative realizations of an unobserved answer. By analogy, our repeated LLM readings represent alternative annotations of each article's latent emotional profile. We do not, however, treat these readings as formal Bayesian imputations drawn from a common posterior predictive distribution. Rather, we use the Rubin-style decomposition of uncertainty into within-reading and between-reading components as a transparent device for combining estimates and for recognizing that the estimated associations may vary across LLM realizations.

Formally, we combine the variance components as

$$V = W + \left(1 + \frac{1}{m}\right) B,$$

where W denotes the average within-reading variance, B the between-reading variance, and m the number of LLM readings. We treat the readings as noisy measurements of a latent emotional score rather than as Bayesian imputations of missing data, so the $(1+1/m)$ factor is retained heuristically in the spirit of Rubin (1987) rather than derived from a posterior-predictive argument.

A measurement-error treatment views the five LLM passes as noisy replicate scorings of the same paper sample. Averaging the five coefficients reduces the labelling component of the variance but not the sampling component, since every reading is computed on the same papers; the variance of the pooled estimate $\bar{\beta}$ is therefore $W + B/m$. The variance of any single reading $\hat{\beta}_k$, taken on its own as an estimator of β, retains both components in full and equals $W + B$. Our combination sits at or above both, so the reported intervals are conservative relative to either benchmark. For the same reason we do not apply the Barnard and Rubin (1999) small-m degrees-of-freedom correction: that correction is derived under an imputation framework whose assumptions our setting does not satisfy, and importing it would compound small-sample penalties on top of an already conservative variance.

More precisely, for any quantity, the model produces a tier coefficient, a predicted citation count at a grid cell, and we have five annotation-specific estimates $\hat{\beta}_1, \ldots, \hat{\beta}_5$ with cluster-robust PPML

standard errors $\mathrm{SE}_1, \ldots, \mathrm{SE}_5$. The pooled estimate is the arithmetic mean

$$\bar{\beta} \;=\; \frac{1}{m}\sum_{k=1}^{m} \hat{\beta}_k, \tag{1}$$

and its total variance decomposes into a within-annotation and a between-annotation component,

$$W \;=\; \frac{1}{m}\sum_{k=1}^{m} \mathrm{SE}_k^2, \tag{2}$$

$$B \;=\; \frac{1}{m-1}\sum_{k=1}^{m} \big(\hat{\beta}_k - \bar{\beta}\big)^2, \tag{3}$$

$$V \;=\; W \;+\; \Big(1 + \tfrac{1}{m}\Big)\, B, \tag{4}$$

with $m = 5$. Each term has a transparent reading in this setting. W is the sampling uncertainty, how much the estimate would move if the set of papers changed, i.e. the average of the five within-reading PPML sampling variances. B is the annotation uncertainty (how much the estimate itself moves when ChatGPT re-reads the same titles), i.e. a direct measurement of the LLM's scoring reproducibility for that particular quantity. V is their total; confidence intervals are built from $\bar{\beta} \pm z_{1-\alpha/2}\sqrt{V}$.

Two limiting behaviors make clear why this pooling rule is well-suited to this paper. When ChatGPT labels the corpus reproducibly so that $B \approx 0$, we have $V \approx W$, and the pooled interval coincides with a single-annotation interval: re-running the LLM on the same title adds no information, and the pooled inference recognizes this by refusing to advertise extra precision. When ChatGPT disagrees with itself so that B is large, V widens appropriately, and a coefficient that was significant in any one annotation but not reproducible across re-readings correctly loses significance in the pool.

Two alternatives might look tempting but are inappropriate here. Pointwise-averaging the five emotion scores into one regressor and running a single PPML shrinks the noise silently: the standard-error machinery sees a clean regressor and delivers a tighter interval than the data justify, because it does not know the averaged regressor was itself estimated from five noisy readings. Inverse-variance meta-analysis of the five fitted models treats the five datasets as independent studies; the pooled variance is then approximately W/m (distinct from the measurement-error benchmark $W + B/m$ discussed above, which correctly retains the between-reading component) so the confidence interval tightens by a factor of roughly $\sqrt{m}$. That would be correct if the five datasets were five independent samples of papers, but they are one sample of papers re-labelled five times: meta-analysis would manufacture a precision claim on the strength of $(m-1)N$ observations that do not exist. Our decomposition avoids both failures by keeping the two sources of uncertainty explicit.

4.3 Article-Level Controls

The regression model includes three categories of control variables alongside the emotion scores: document-level characteristics, title features, and author-team quality proxies. Appendix 2 provides a synoptic overview of all controls included in the model.

For the document-level controls, page count serves as a proxy for article length and is included as a continuous variable. Open-access status is captured through a categorical indicator derived from the Scopus metadata and reflects the degree to which an article is freely accessible.

Title features are constructed directly from the article title string. Word count captures title length. A binary split-title indicator flags articles whose title contains a colon. The count of non-alphanumeric characters captures punctuation richness.

Author-team quality proxies are constructed from the co-authorship network induced by the corpus itself. Each paper defines a fully connected subgraph among its authors; across all 15,836 papers, this yields an undirected weighted graph in which the edge between two authors carries a weight equal to the number of papers they co-authored in the sample. For each author, weighted degree centrality is computed as the sum of incident edge weights (Freeman, 1978). The variable `author_max_degree` takes the maximum of this measure across the paper's co-author team, capturing whether at least one team member occupies a well-connected position within the marketing literature. The variable `author_max_prior` counts, for each author, the number of their papers in the corpus published strictly before the focal year, and retains the team maximum. This provides a within-corpus measure of the most experienced co-author's track record, while avoiding the mechanical endogeneity that would arise from using career-wide citation-based metrics such as the h-index as regressors in a citation model. The number of authors on the paper (`n_authors`) is included as an additional team-size control.

A natural concern is that our author-team controls are coarser proxies for author quality than the h-index or career citation totals. We prefer the coarser measures for two reasons. First, any author-level bibliometric score that aggregates citations is itself a function of the outcome we are modeling, which introduces a mechanical dependence between the regressors and the dependent variable and would bias the emotion coefficients in a direction we cannot sign a priori. Second, our controls are constructed from the marketing corpus itself, that is, from the set of papers identified by the "marketing" keyword used to define our sample. Weighted co-author degree centrality counts ties within this corpus, and cumulative prior publications count papers placed in it strictly before the focal year. The resulting measures capture each author's standing and connectedness in the marketing literature specifically, which is arguably more relevant to the mechanism under study than a career-wide index built from heterogeneous outlets across many fields.

We count only papers published up to the focal year, rather than the author's entire publication

record. If we used all papers, later publications would influence the controls for earlier papers, which is logically inconsistent and would indirectly reintroduce dependence on the outcome, because authors whose early work is highly cited generally go on to publish more. Restricting the count to publications strictly before year t keeps the control predetermined with respect to the focal paper's citations. Centrality, by contrast, is computed on the full corpus. Rebuilding the co-authorship graph year by year would be statistically unstable in the early sample, where few edges exist and most authors appear as isolates. Centrality is a structural property reflecting long-run collaborative position rather than year by year flow. Using the full graph gives a more stable estimate at the cost of a small look-ahead, which we regard as acceptable because centrality is not built from citations and therefore does not reintroduce a dependence on the outcome.

5 Example of Emotion Recognition by ChatGPT

We select four articles from our database to illustrate the emotional makeup of titles, using the Python visualization tool from Semeraro (2022).

These four Plutchik emotion wheels illustrate the emotional profiles extracted from sample academic paper titles by chatGPT-5 mini (seed = 42), with scores ranging from 1 (minimal) to 10 (maximal intensity) (see Fig. 2).

The title "We love the Gong: A marketing perspective" (Kerr et al., 2012) stands out with the highest joy score (9.00) and strong trust (8.00), creating a distinctly positive emotional profile. The explicit use of "love" directly contributes to elevated positive sentiment, demonstrating how affective language choices in titles can signal the paper's domain and tone.

The three remaining titles show no joy at all (1.00), reflecting their emphasis on negative themes. "Boycott and racism" (Abdul Latif and Abdul Talib, 2015) evoke strong anger (9.00) and disgust (8.00), as terms like "racism" and "boycott" have strongly negative connotations that the sentiment model captures well.

"Casting a nuclear shadow" (Sandoval, 2003) evokes fear (9.00) as its dominant emotion. The metaphorical imagery of "shadow" combined with "nuclear" creates a threat-laden atmosphere.

Finally, "Bad smells from Köln" (Buetow, 2002) is assigned a high disgust intensity score (9.00) and the phrase "bad smells" aligns with an isolated maximum for disgust in the resulting profile.

These results highlight how even short academic titles encode rich emotional information that sentiment analysis can uncover, revealing authors' framing choices and subject matter connotations.

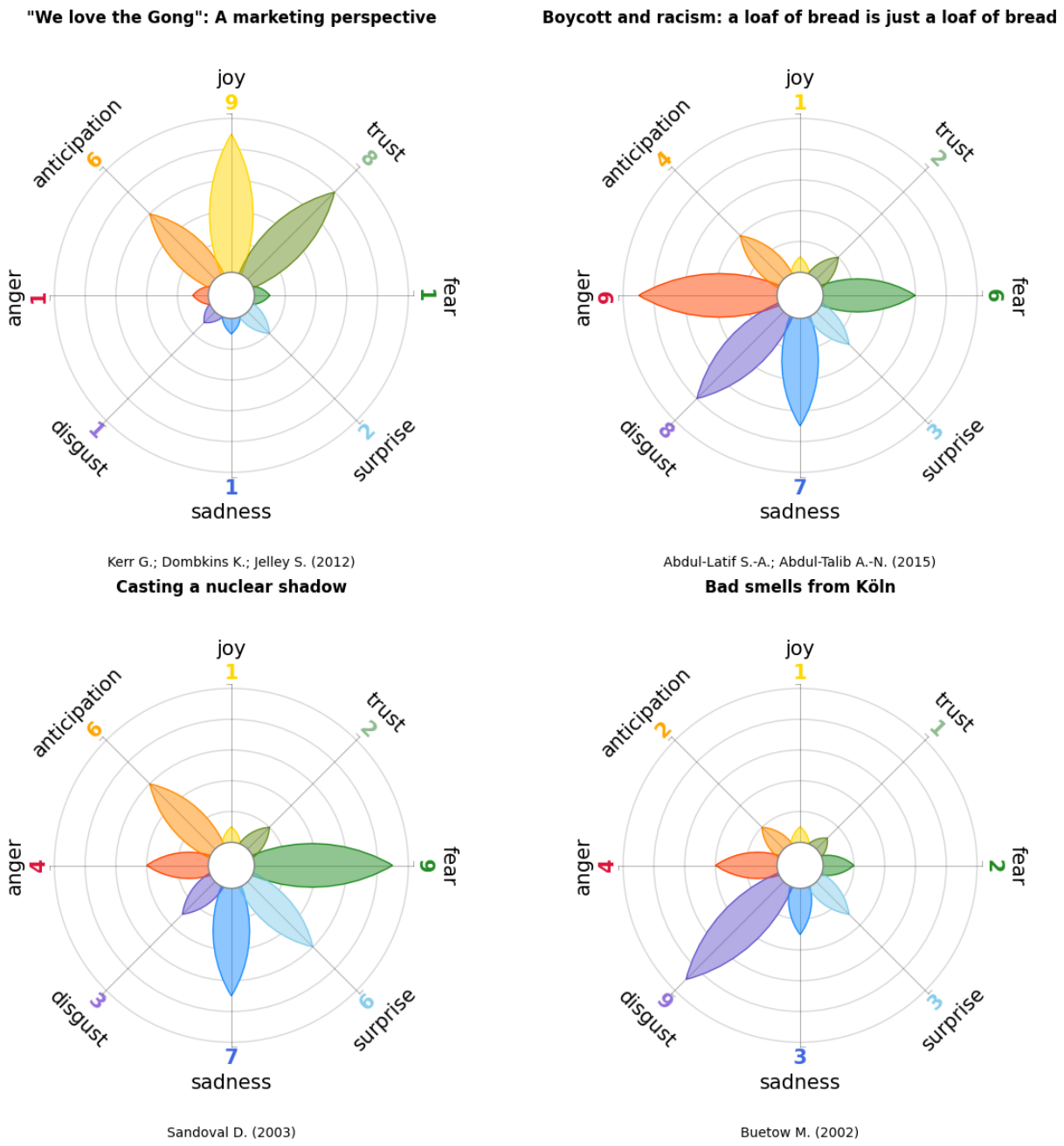

Figure 2: Illustration of emotion recognition by ChatGPT on Plutchik's wheel of emotions

5.1 Distribution of emotions

To visualize how emotions are distributed across the different LLM runs, we present the Density Plot shown below (see Fig. 3). Each ridge represents a Gaussian kernel density (with a bandwidth of 0.5) computed on the subset of the combined dataset corresponding to a single reading.

Anger, Disgust, Fear, Sadness form one natural cluster. All four have a sharply spiked density at score 1 and a long thin tail that doesn't really rise off the floor until the very far right. This is what we would expect of emotions that academic paper titles almost never invoke; most papers get labelled 1 or 2 on these by every run. The five ridges in each of these four panels sit tightly on top of each other: the three GPT-5, ChatGPT-5.4, and GPT-5.4-mini all agree closely both on the mode's location and on the shape of the tail.

Surprise behaves similarly, with a mode at 1 and a thin tail, but the five ridges are visibly less tightly stacked than in the Anger/Disgust/Fear/Sadness cluster; the peak height varies a little more across readings. Probably a slightly noisier emotion to label, though still in the same "rare" regime.

Joy departs from this pattern: the mode sits at 2–3 rather than at 1, the distribution is considerably less peaked, and there's meaningful mass all the way out to 10. Papers do get labeled as expressing joy at a moderate rate, so the distribution has a proper shape rather than being a boundary-spike. The five ridges track each other reasonably well.

Trust and Anticipation are the only two emotions in the group that are truly distinct. For both, the bulk of responses lie in the middle of the 1–10 scale: Trust peaks around 5–7, while Anticipation is more spread out and roughly level from about 3–8. Anticipation, in particular, exhibits a strikingly varied pattern across the five readings, with the ridges less sharply and consistently aligned than for the other emotions.

To summarize, five of the eight emotions (the four "negative-valence" ones plus Surprise) are rare and uniformly labeled across LLM configurations; one emotion (Joy) is moderately common and stable; and two emotions (Trust, Anticipation) are common and show the most annotation-level disagreement across the five LLM readings.

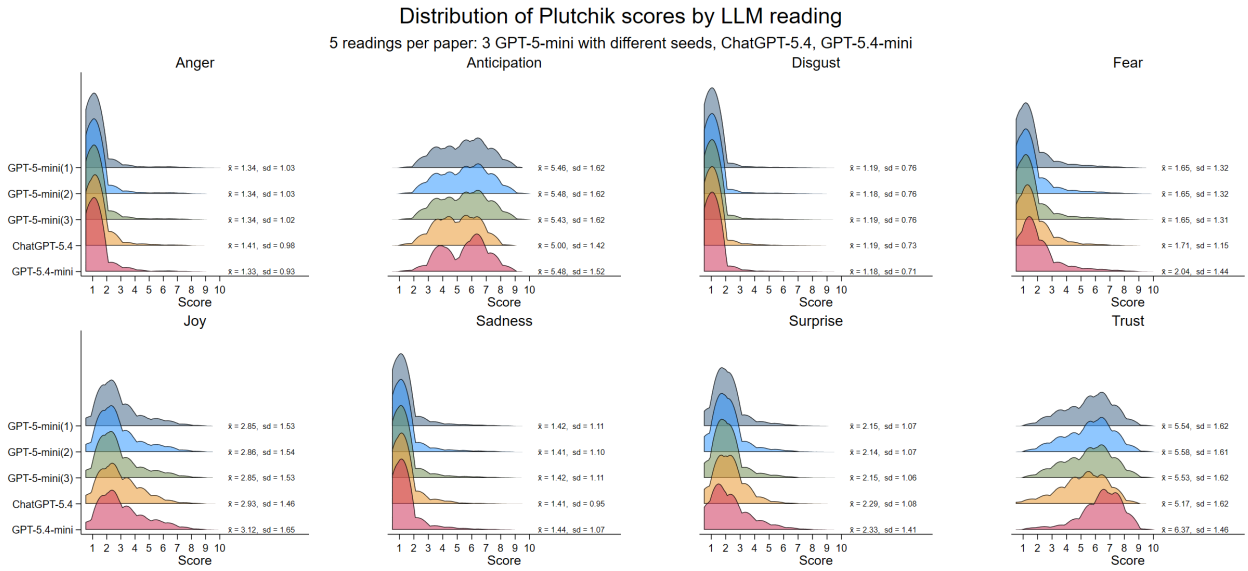

Figure 3: Emotion Density Plot

6 Results

6.1 Regression Model with Primary Emotions and Mean Effect for Their Combinations

We first present a restricted specification in which higher-order emotions enter as simple averages of their constituent primaries. This serves as a benchmark: it is the natural extension of standard sentiment regressions to Plutchik's tiered taxonomy, and its limitations (discussed at the end of this subsection) motivate the interaction model developed below. We model the conditional mean of article citation counts as a function of the different level covariates, journal-specific unobserved heterogeneity, and publication-year effects. Let i index articles, with each article published in journal $j(i)$ and in year $t(i)$. Denote the observed citation count by Y_i. Let $X_i \in \mathbb{R}^K$ denote the vector of emotion intensities and let $W_i \in \mathbb{R}^Q$ collect the remaining controls (excluding fixed effects).

The conditional mean is specified as

$$\mathbb{E}\big[Y_i \mid X_i, W_i, \alpha_{j(i)}, \gamma_{t(i)}\big] = \exp\Big(X_i^\top \beta_X + W_i^\top \beta_W + \alpha_{j(i)} + \gamma_{t(i)}\Big),$$

where $\alpha_{j(i)}$ and $\gamma_{t(i)}$ denote journal and publication-year fixed effects, respectively.

Estimation is carried out with the Poisson pseudo-maximum likelihood estimator with high-dimensional fixed effects (PPMLHDFE), which is consistent under general forms of heteroskedasticity and naturally accommodates zero outcomes.

By absorbing $\alpha_{j(i)}$, the model accounts for persistent heterogeneity across journals, such as differences in reputation, audience size, or editorial policy, while $\gamma_{t(i)}$ controls for time-specific shocks or changes in publication and citation practices. This approach ensures that the coefficients β are identified from within-journal and within-year variation and are not confounded by unobserved factors common to a journal or a particular year. Zero-citation observations are retained, as they contain important information for consistent estimation under the Poisson pseudo–likelihood framework.

We estimate the model using the Stata command `ppmlhdfe` (Correia et al., 2020), which relies on `reghdfe` (Correia, 2017) and `ftools` (Correia, 2016) for the absorption of high-dimensional fixed effects. Standard errors are clustered at the journal level to allow for within-journal dependence[3]

[3]Replication code (Stata) is available at: `https://osf.io/3xazq/overview?view_only=b994f589c2e44ac38d74602d5a545fb4`

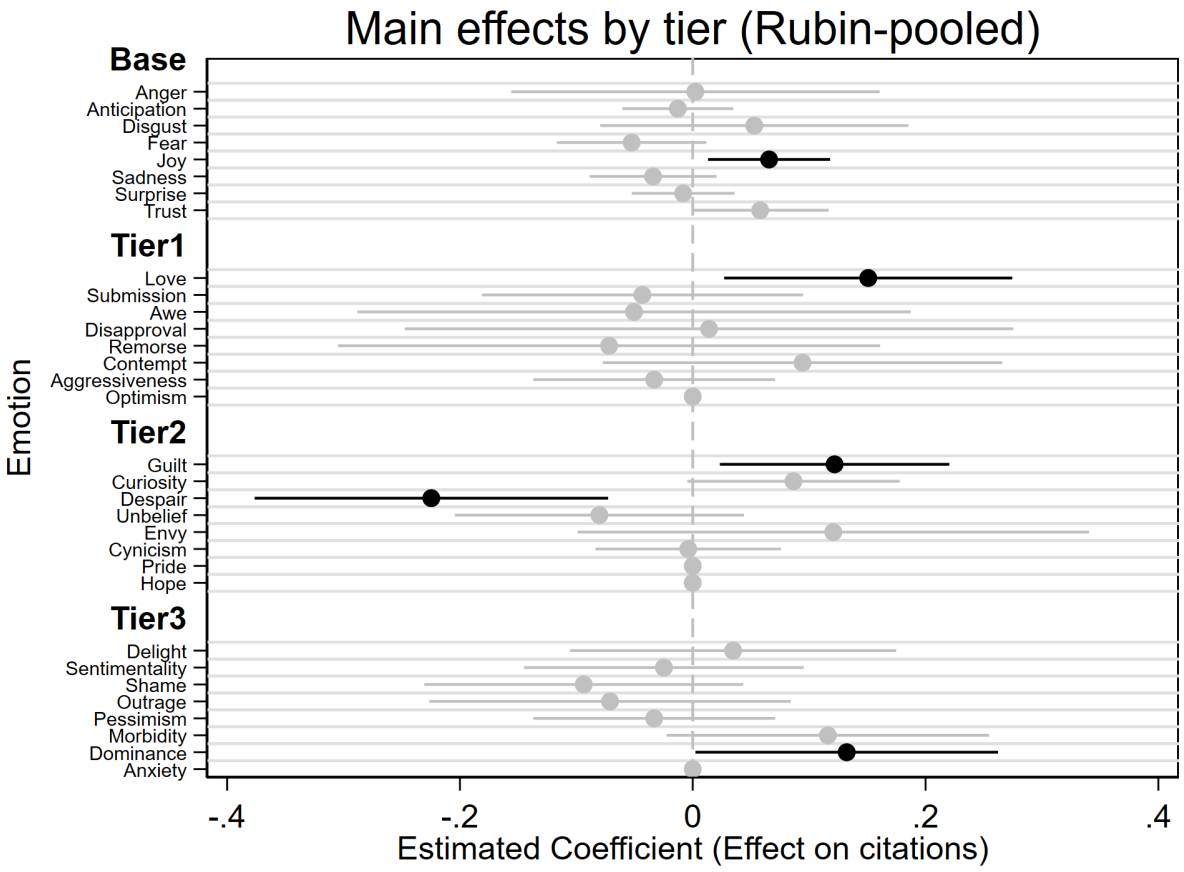

Figure 4: Coefplot illustrating significance and effect of emotions on citation count

Figure 4 presents the estimated coefficients with 95% confidence intervals from the high-dimensional fixed effects Poisson regressions, produced with the `coefplot` package (Jann, 2014). The dashed vertical line at zero indicates the absence of any effect, so only coefficients whose intervals do not overlap zero are statistically significant (Appendix 3 reports the estimated coefficients).

Base illustrates the effects of primary emotions, whereas Tiers 1,2, and 3 reflect first-, second-, and third-order interactions among these primary emotions, as defined in Table 2. These interactions are computed as the average of the two underlying primary emotion intensities. In the case of perfect collinearity, one independent variable is dropped.

In basic emotions, only the emotion *Joy* displays significant associations with citation outcomes.

A one unit increase in the intensity of *Joy* is associated, on average, to an increase of about 6.6% in citation counts.

In Tier 1, only *Love* exhibits a positive and significant coefficient, indicating that a greater intensity of Love is associated with a 15.1% increase in citation counts. This finding is consistent with Plutchik's framework, in which Love emerges as a composite emotion resulting from the combination of Joy and Trust.

Tier 2 yields several notable estimates. *Despair* exhibits a sizeable and statistically significant negative association, corresponding to a decrease of about 22.5% in citation counts. *Guilt* is positively and significantly associated with citation outcomes, with an increase of about 12.0%.

Finally, in Tier 3, only *Dominance* displays significant effects. It is linked to an approximate 14.1% increase in citations. The remaining tertiary dyads do not display interpretable associations with citation performance.

However, we believe that creating Tier 1 or higher-order emotions beforehand and directly inserting them as explanatory variables in a regression framework is a restrictive approach. Plutchik's model is built on the principle that emotions are composed of primary dimensions whose intensity may vary. For instance, Love is a blend of Joy and Trust, but, even considering that a composite emotion is the average of the basic emotions, a score of 5 for Love could represent Joy = 5 and Trust = 5, or alternatively, no Joy and full Trust (or the opposite). Such aggregation masks the underlying variation and distorts the theoretical foundation. Second, the Tier 1, 2, and 3 regressions omit the primary variables themselves. Since the dyads are deterministic linear functions of these primaries, the dyad coefficients absorb variation that the primaries' own partial effects could just as well account for. As a result, a dyad formed from two primaries whose partial associations with citations offset one another will look non-influential, even though each primary is substantively important on its own. To address these issues, we estimate a general specification with interactions between emotions.

6.2 Regression Model with Primary Emotions and Interaction Terms for Their Combinations

As explained by Butler et al. (2025) secondary dyads evaluation by large language models tend to favor one primary emotion over the other. We therefore assume that secondary and tertiary dyads are not merely the result of an average of their underlying emotions, but rather reflect specific combinations of emotional intensities. To allow for non-additive combinations of emotions, we define a restricted set of emotion dyads

$$\mathcal{D} \subset \{(k, \ell) : 1 \leq k < \ell \leq K\},$$

for example motivated by Plutchik's taxonomy (adjacent, secondary, and tertiary pairs), while excluding opposite emotions that do not correspond to meaningful mixes in that framework. Let $m_{\mathcal{D}}(X_i) \in \mathbb{R}^{|\mathcal{D}|}$ denote the vector stacking the selected interaction products,

$$m_{\mathcal{D}}(X_i) = \big(X_{ik}X_{i\ell}\big)_{(k,\ell)\in\mathcal{D}}.$$

The conditional mean becomes

$$\mathbb{E}\big[Y_i \mid X_i, W_i, \alpha_{j(i)}, \gamma_{t(i)}\big] = \exp\Big(X_i^\top \beta_X + W_i^\top \beta_W + m_{\mathcal{D}}(X_i)^\top \theta + \alpha_{j(i)} + \gamma_{t(i)}\Big),$$

where θ collects the interaction effects for the dyads in $\mathcal{D}$. Estimation again proceeds via PPML-HDFE with journal and publication-year fixed effects absorbed.

To examine the joint effect of emotional intensities on citation outcomes, we generated model-based predictions for each relevant pair of emotions by evaluating the fitted model over a two-dimensional intensity grid. Concretely, both emotion dimensions were varied from 1 to 10, yielding predictions across the full 10×10 combination space for every emotion pair. For each dyad, both emotions were allowed to vary from 1 to 10, while all other covariates were held constant at their sample means. Predicted citation counts are obtained from the estimated Poisson model and interpolated on a finer grid with the `bipolate` command (Canner, 2014) to construct smooth response surfaces. Journal and publication-year fixed effects are absorbed during estimation rather than estimated explicitly, and so are not reflected in the predicted counts shown in the surfaces below (Appendix 4 reports the estimated coefficients).

In the count model, the object of interest is the fitted conditional mean (expected citations) as a function of the two emotions,

$$\mu(x_k, x_\ell; z) = \mathbb{E}[Y \mid X_k = x_k,\ X_\ell = x_\ell,\ Z = z], \qquad \widehat{\mu}(x_k, x_\ell; z) = \widehat{\mathbb{E}}[Y \mid X_k = x_k,\ X_\ell = x_\ell,\ Z = z],$$

where Z denotes the remaining covariates (and any fixed effects) held fixed at a chosen reference value $z = z_0$. For notational simplicity, write $\widehat{\mu}(x_k, x_\ell) = \widehat{\mu}(x_k, x_\ell; z_0)$.

For each point on the fitted surface, local changes with respect to both emotions were approximated via numerical partial derivatives. Using forward differences,

$$\frac{\partial \widehat{\mu}}{\partial x_k}(x_k, x_\ell) \approx \frac{\widehat{\mu}(x_k + \Delta_k, x_\ell) - \widehat{\mu}(x_k, x_\ell)}{\Delta_k}, \qquad \frac{\partial \widehat{\mu}}{\partial x_\ell}(x_k, x_\ell) \approx \frac{\widehat{\mu}(x_k, x_\ell + \Delta_\ell) - \widehat{\mu}(x_k, x_\ell)}{\Delta_\ell},$$

where Δ_k and Δ_ℓ are small increments on the respective emotion scales.

The resulting gradient vectors,

$$\nabla \widehat{\mu}(x_k, x_\ell) = \left(\frac{\partial \widehat{\mu}}{\partial x_k}(x_k, x_\ell), \frac{\partial \widehat{\mu}}{\partial x_\ell}(x_k, x_\ell)\right),$$

summarize the local direction and magnitude of change in the fitted conditional mean of citations within the (x_k, x_ℓ) plane, holding $Z = z_0$ fixed.

6.3 Interaction Between Primary Emotions

The following figures of combined basic emotions represent the effect of a composite emotional dyad on predicted citation counts. The luminance gradient reflects predicted citation levels, while the arrows indicate both the direction and steepness of change in citations when the emotional intensities vary. The x- and y-axes in each subplot show the intensity of the two component emotions, both scaled from 0 to 10. The shade contours and numbers within each graph represent the predicted citation count associated with the particular combination of those two emotional scores (independently from the journal and year effect). Darker gray correspond to higher predicted citation counts, while lighter gray indicate lower citation counts. Thus, by following the gradient luminance, one can deduce where combinations of emotions tend to be linked with more or fewer citations. This visualization makes it possible to interpret the estimated interaction terms $\theta_{k\ell}$ in dynamic form, showing how the co-variation of two emotions modifies the predicted scholarly impact.

The arrows represent the direction and strength of the gradient of the predicted citation count across the emotion space. The black polygonal line encloses the convex hull of the observed data points, computed with the `delaunay` command (Naqvi, 2022). This boundary outlines the region of the emotional space that is supported by empirical data. Observations outside the convex hull are only extrapolations on values that are not observed, meaning they are predicted by the model but not directly supported by observed data. Consequently, interpretations beyond this border should be viewed with caution.

To clarify the interpretation, let us take the example of Love, which results from the combination of the emotions Joy (on the vertical axis) and Trust (on the horizontal axis) (see Fig. 5).

Each point inside the graph corresponds to a possible combination of Joy and Trust levels, both measured from 0 to 10. The color background represents the predicted citation count, which is the variable of interest. The numbers shown over the bands report the model's predicted citation counts for each region of the emotional space, with journal and year fixed effects absorbed and all other covariates held at their mean values. In this case, the color scale moves from light (lower predicted citations) to dark (higher predicted citations).

The arrows show the gradient of change in the predicted citation count; they point toward the direction where citations are expected to increase most strongly. Their orientation, therefore, tells us which way in the emotional space is associated to higher predicted citations, while their length indicates the magnitude of this effect. When the arrows are long, small changes in Joy or Trust are strongly associated with the predicted citation count; shorter arrows signal weaker variation. For Love, the gradient arrows mostly point upward, indicating that the fitted conditional mean of citations increases primarily with Joy, holding Trust fixed. The largest fitted values (about 132) occur in a region where Joy is high and Trust is comparatively low. This pattern should

be interpreted cautiously because the region lies outside the convex hull of observed Joy-Trust combinations, so the model is extrapolating beyond the support of the data. Prediction precision is summarized by the relative half-width of the confidence interval $[L_{ij}, U_{ij}]$ around the fitted mean $\widehat{\mu}_{ij}$,

$$r^*_{ij} = \frac{(U_{ij} - L_{ij})/2}{|\widehat{\mu}_{ij}|}.$$

A natural criterion for flagging a cell as precisely estimated is $r^*_{ij} \leq 0.2$, equivalent to requiring the point estimate to be at least five half-widths away from zero. Under a single-LLM analysis the half-width in the numerator is a pure sampling quantity and this criterion is unambiguous. Under the Rubin pool, by contrast, the pooled standard error carries both sampling and annotation uncertainty, $\mathrm{SE}_{\mathrm{pool}} = \mathrm{SE}_{\mathrm{within}}/\sqrt{1-\bar{\gamma}}$, where $\bar{\gamma} = (1+1/m)\, B/V \in [0,1)$ is the average across grid cells of the Rubin fraction of missing information, i.e. the share of pooled variance attributable to between-reading disagreement across the $m = 5$ LLM annotations. The raw r^*_{ij} computed on the pooled interval therefore overstates imprecision by the same factor relative to the single-LLM benchmark. We therefore apply the adjustment to the threshold rather than to the measurement: in the figures, arrows are shown in a lighter shade when

$$r^*_{ij} \;>\; \frac{0.2}{\sqrt{1-\bar{\gamma}}}.$$

Equivalently, defining the signal-to-noise ratio

$$\mathrm{SNR}_{ij} = \frac{|\widehat{\mu}_{ij}|}{(U_{ij} - L_{ij})/2} = \frac{1}{r^*_{ij}},$$

the shading rule triggers when $\mathrm{SNR}_{ij} < 5\sqrt{1-\bar{\gamma}}$.

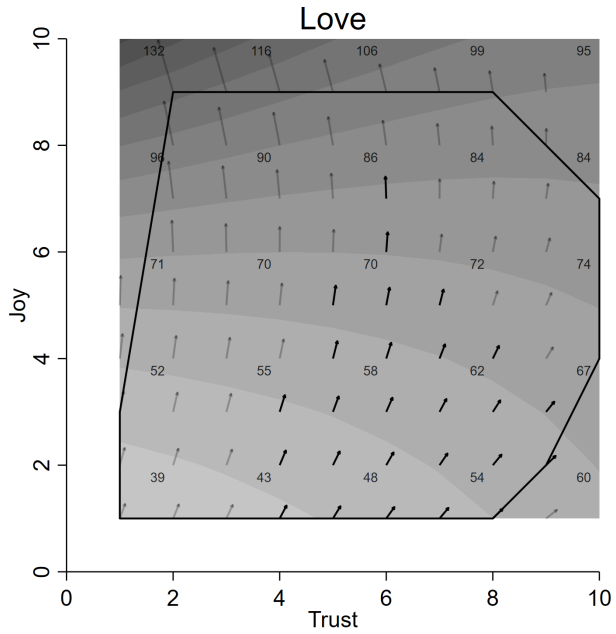

Figure 5: Visualization of the effect of love made by the interaction between joy and trust on citation count

6.4 Primary Dyads

Figure 6 illustrates the predicted effects of primary emotional dyads on citation counts, derived from the full interactions PPML model. The following explanations are summarized in Table 4 but we will provide a qualitative reading of these graphs.

Two dyads - *Love* and *Optimism*, - exhibit moderate certainty, indicating locally reliable predictions in regions close to their respective optima. For *Love*, higher predicted citation levels are associated with configurations combining high *Joy* and low *Trust*. *Optimism* is characterized by low *Anticipation* and high *Joy*, suggesting that positive affect dominates the association with citation outcomes.

Three dyads - *Submission*, *Awe*, and *Disapproval* - display high certainty, with bright arrows concentrated in their optimal region, indicating more stable and reliable associations between emotional configuration and predicted citation outcomes. *Submission* reaches its optimum with high level of *Trust* and low level of *Fear*. Maximum output of *Awe* is associated with low level of *Fear* and high or low level of *Surprise*. Finally, *Disapproval* must be combined via a low level of *Surprise* and *Sadness*.

By contrast, *Remorse* (*Disgust* × *Sadness*), *Contempt* (*Disgust* × *Anger*), and *Aggressiveness* (*Anger* × *Anticipation*) are characterized by low certainty, with faint arrows throughout the emotional space. For these dyads, prediction quality is insufficient to draw robust conclusions regarding their relationship with citation performance.

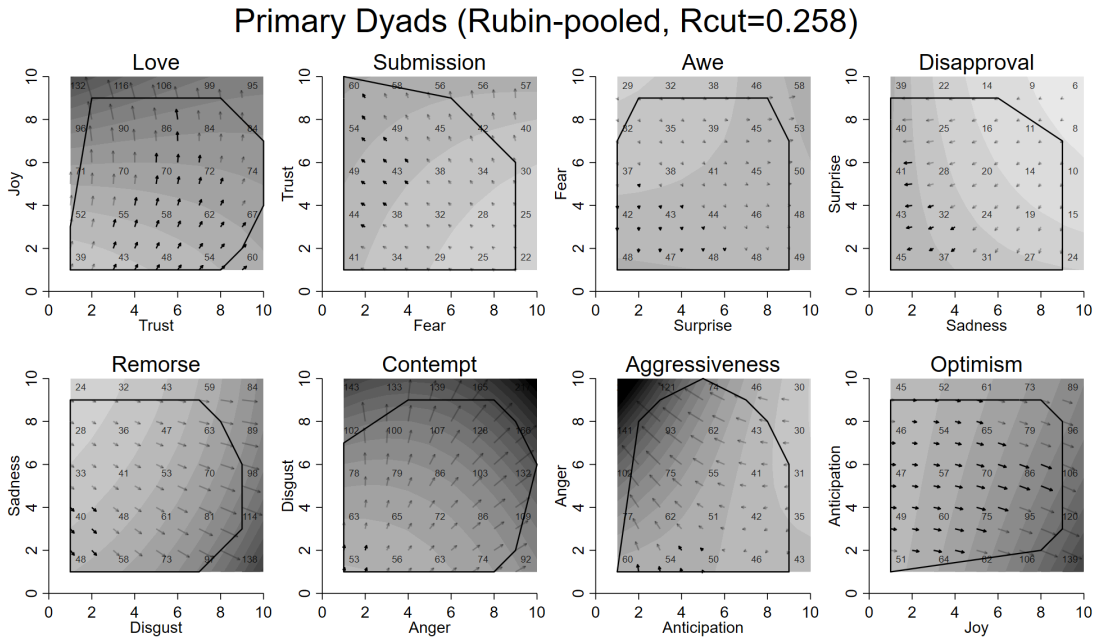

Figure 6: Primary Dyads

6.5 Secondary Dyads

Figure 7 illustrates the predicted effects of secondary emotional dyads on citation counts, derived from the full interactions PPML model. The following explanations are summarized in Table 4 but we will provide a qualitative reading of these graphs.

Among secondary dyads, three combinations - *Guilt*, *Pride*, and *Hope* - exhibit moderate certainty, indicating locally reliable predictions near their optima. *Guilt* shows its optimal configuration at high *Joy* and low *Fear*. *Pride* is composed of a low level of *Anger* and high level of *Joy* to reach its optimum. The best return of *Hope* is characterized by low *Anticipation* and high *Trust*.

Only one dyad - *Curiosity* - displays high certainty, with bright arrows concentrated in its optimal region, defined by high *Trust* and low *Surprise*, indicating a stable and reliable association

between emotional configuration and predicted citation outcomes.

By contrast, *Despair* (*Fear* × *Sadness*), *Unbelief* (*Surprise* × *Disgust*), *Envy* (*Sadness* × *Anger*), and *Cynicism* (*Disgust* × *Anticipation*) are characterized by low certainty, with faint arrows throughout the emotional space. For these dyads, prediction quality is insufficient to draw robust conclusions regarding their relationship with citation performance.

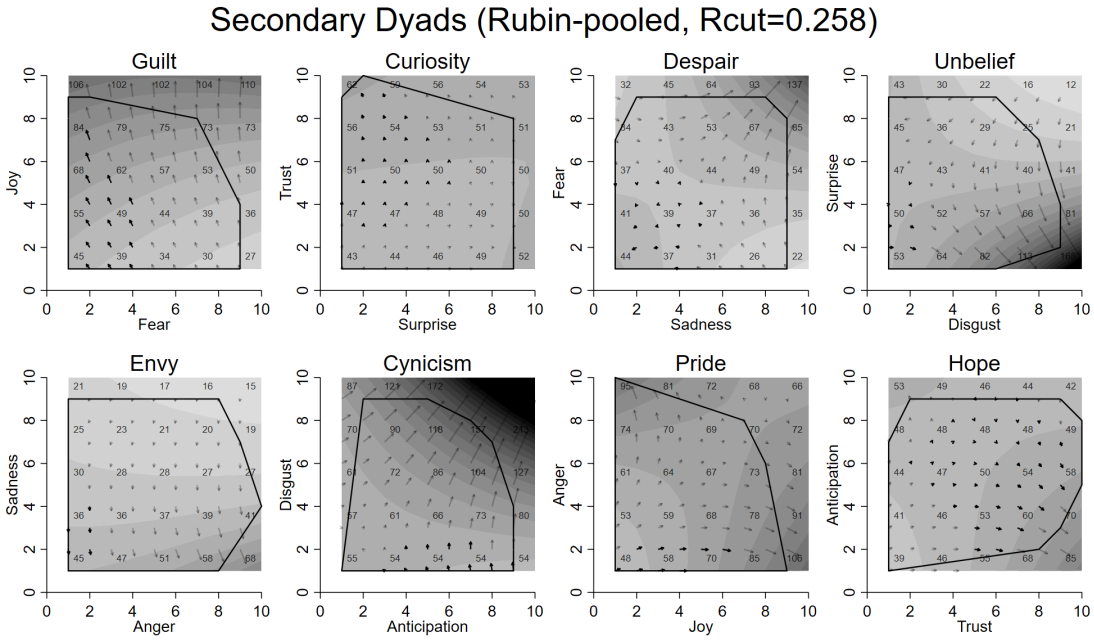

Figure 7: Secondary Dyads

6.6 Tertiary Dyads

Figure 8 illustrates the predicted effects of tertiary emotional dyads on citation counts, derived from the full interactions PPML model. The following explanations are summarized in Table 4 but we will provide a qualitative reading of these graphs.

Among tertiary dyads, one combination - *Delight* - exhibit moderate certainty, indicating locally reliable predictions near its optima with a high level of *Joy* and a low level of *Surprise*.

Three combinations - *Sentimentality*, *Pessimism*, and *Anxiety* - display high certainty, with bright arrows concentrated in their optimal regions, indicating stable and reliable associations be-

tween emotional configuration and predicted citation outcomes. *Sentimentality* reaches its optimum at high *Trust* and low *Sadness*. *Pessimism* is characterized by low *Sadness* combined with either high or low *Anticipation* while *Anxiety* is defined by low *Anticipation* and *Fear*.

By contrast, *Shame* (*Fear* × *Disgust*), *Outrage* (*Surprise* × *Anger*), *Morbidity* (*Disgust* × *Joy*), and *Dominance* (*Anger* × *Trust*) are characterized by low certainty, with faint arrows throughout the emotional space. For these dyads, prediction quality is insufficient to draw robust conclusions regarding their relationship with citation performance.

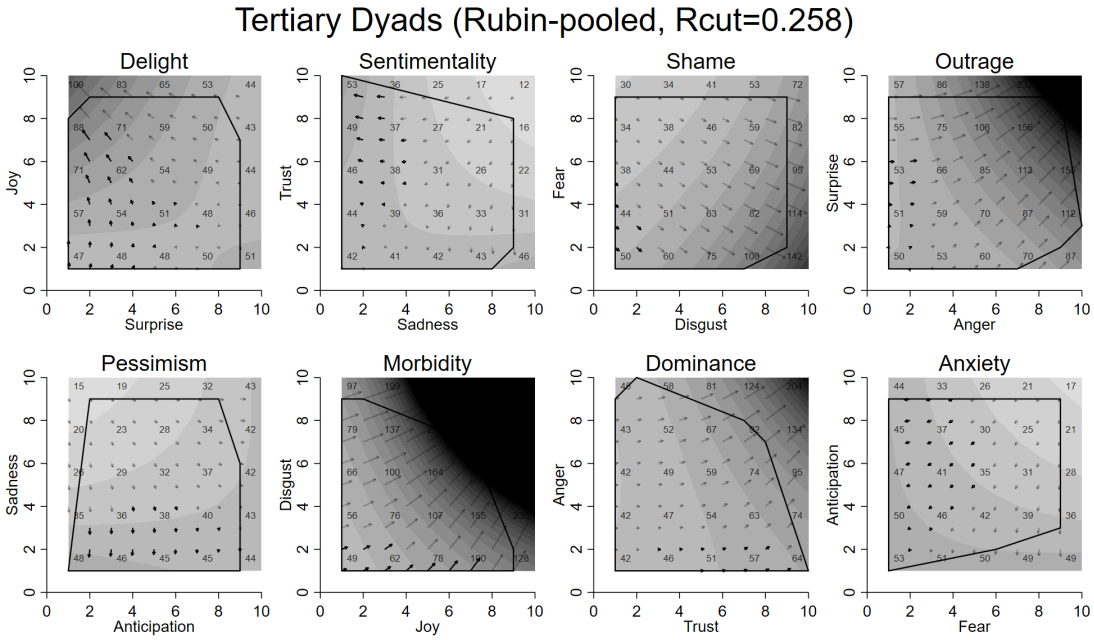

Figure 8: Tertiary Dyads

Table 4: Synoptic overview of emotional dyads and their association with citation impact. Certainty reflects arrow luminescence in high-impact regions.

Dyad	**Component Emotions**	**Optimal Config.**	**Certainty**
Primary Dyads			
Love	Joy × Trust	High Joy, Low Trust	Moderate
Optimism	Anticipation × Joy	Low Anticipation, High Joy	Moderate
Aggressiveness	Anger × Anticipation	High Anger, Low Anticipation	Low
Submission	Trust × Fear	High Trust, Low Fear	High
Awe	Fear × Surprise	Low Fear, High or Low Surprise	High
Disapproval	Surprise × Sadness	Low Surprise, Low Sadness	High
Remorse	Disgust × Sadness	Low Sadness, High Disgust	Low
Contempt	Disgust × Anger	High Disgust, High Anger	Low
Secondary Dyads			
Guilt	Joy × Fear	High Joy, Low Fear	Moderate
Cynicism	Disgust × Anticipation	High Disgust, High Anticipation	Low
Pride	Anger × Joy	Low Anger, High Joy	Moderate
Hope	Anticipation × Trust	Low Anticipation, High Trust	Moderate
Curiosity	Trust × Surprise	High Trust, Low Surprise	High
Despair	Fear × Sadness	High Fear, High Sadness	Low
Unbelief	Surprise × Disgust	Low Surprise, High Disgust	Low
Envy	Sadness × Anger	Low Sadness, High Anger	Low
Tertiary Dyads			
Morbidity	Disgust × Joy	High Disgust, High Joy	Low
Outrage	Surprise × Anger	High Surprise, High Anger	Low
Delight	Joy × Surprise	High Joy, Low Surprise	Moderate
Shame	Fear × Disgust	Low Fear, High Disgust	Low
Sentimentality	Trust × Sadness	High Trust, Low Sadness	High
Pessimism	Sadness × Anticipation	Low Sadness, High or Low Anticipation	High
Dominance	Anger × Trust	High Anger, High Trust	Low
Anxiety	Anticipation × Fear	Low Anticipation, Low Fear	High

Notes: Certainty levels derived from arrow luminescence: High = bright arrows in optimal region; Moderate = bright arrows nearby but fading at optimum; Low = faint arrows throughout.

7 Conclusion

In this study, we quantify the emotional signals embedded in the titles of 15,836 marketing articles. We operationalize Plutchik's psychoevolutionary model of emotions through ChatGPT and show that emotions in scientific titles are linked to citation performance. This relationship remains after controlling for journal fixed effects, publication-year effects, document characteristics, title features, author-team quality, and access conditions. This suggests that emotional tone is associated with scholarly visibility beyond what can be attributed to an article's scientific contribution alone.

Beyond significant main effects for joy, our analysis indicates that emotional dynamics are multilevel. The interaction model shows that certain emotional combinations capture patterns of scholarly attention that would remain hidden in models limited to primary emotion scores. These results suggest that citation counts are associated not only by emotional valence, but also by the structure of emotional composition, intensity, and co-activation. This extends scientometric work on other linguistic determinants of impact (Jamali and Nikzad, 2011; Nair and Gibbert, 2016) by showing that emotional complexity in titles carries information for citation rates.

Methodologically, our work illustrates how large language models can be integrated into scientometric pipelines to extract affective information from compact textual units such as titles. Whereas earlier research on sentiment in science has often relied on lexicon-based approaches or manual coding (Subotic and Mukherjee, 2014), large language model-based emotion scoring offers a way to measure more subtle affective cues that traditional tools do not easily capture.

Our findings also imply that emotions, often seen as cosmetic or stylistic, may be associated with unintended biases in evaluation, attention, and recognition within science. If titles expressing certain emotions are systematically associated with higher citation counts, they may be associated with perceived impact in ways that do not align with research quality. This raises questions about visibility in scholarly communication, including the possibility that emotional design in titles could function as a strategic lever for academic success.

This interpretation does not imply that emotions causally determine citation outcomes, nor that scholars deliberately respond to affective cues when citing. Rather, emotions can be seen as one contextual factor that shapes the probabilistic environment in which citations emerge. In cumulative advantage systems, even modest associations between emotional tone and early visibility may correspond to larger citation differences over time.

Taken together, our results show that emotions are not peripheral to scientific writing but an under-recognized dimension of scholarly communication with measurable links to the diffusion of knowledge. By showing how emotional cues in titles align with citation patterns, this study contributes to emerging efforts to understand the affective infrastructure of science. Recognizing the role of emotion matters not only for interpreting citation-based indicators, but also for developing a

more nuanced and transparent understanding of how scientific visibility and impact are constructed.

7.1 Limitation and Future Research

In this article, we focus solely on citation count as the key indicator of scholarly impact. While widely used, this metric is not without limitations, as citation numbers can be correlated with various strategies (Ale Ebrahim et al., 2013) or by the Matthew effect, where already well-known scientists gain disproportionate recognition for their work (Merton, 1968). Relying exclusively on this measure, therefore, risks providing a partial view of a scholar's true influence. To obtain a more comprehensive and reliable assessment, it might be interesting to use alternative indicators such as in-press citations, article downloads, or even online engagement metrics. These additional measures could offer valuable insights into both the immediate reach and the broader academic importance of scholarly work.

A key limitation is that we restricted our analysis to articles published in the field of marketing. This disciplinary focus may limit the generalizability of our findings. In other fields, such as medicine, engineering, or physics, the role of emotions in shaping citation impact could differ due to variations in disciplinary norms and publishing practices. Future research could therefore extend this work by examining whether the effect of emotional cues in article titles is consistent across disciplines or whether it reflects field-specific dynamics.

A further limitation concerns topic heterogeneity within the marketing field. Certain sub-fields may attract structurally larger citation pools than others, and articles belonging to more active research areas may accumulate citations independently of their emotional framing. Our analysis does not explicitly control for topic-level variation, yet the journal fixed effects absorbed in our model capture a share of this variance.

Future research should investigate these dynamics across disciplines, languages, and publication types, as emotional norms may vary widely between scientific fields. Extending the analysis to abstracts, keywords, and full texts could also illuminate whether emotion-driven attention persists beyond the title level or whether titles serve primarily as initial attentional gateways. Moreover, comparative evaluations of alternative large language models, or other emotion-detection tools, could further strengthen the methodological foundations of emotion detection in the service of scientometrics.

It is important to emphasize that our analysis of emotions in article titles and their association with citation counts represents a snapshot of a specific period. The observed relationships are prevalent between 2001 and 2015, and thus cannot be assumed to remain stable over time. Therefore, while our findings provide valuable evidence of the role of emotions in marketing ti-

tles in influencing academic visibility during the studied timeframe, they should be interpreted as temporally bound and subject to change as emotional expression in titles continues to evolve.

At the same time, our discussion of reproducibility highlights important caveats: despite fixed seeds, stable prompts, and repeated runs, large language model-generated emotion scores involve uncertainty. This calls for the development of best practices for the responsible use of large language models in quantitative research, a methodological frontier that the scientometric community will increasingly need to address.

References

Abdul Latif, S. A. and Abdul Talib, A. N. (2015). Boycott and racism: A loaf of bread is just a loaf of bread. *Emerald Emerging Markets Case Studies*, 5(6):1–19.

Ale Ebrahim, N., Salehi, H., Embi, M. A., Habibi, F., Gholizadeh, H., Motahar, S. M., and Ordi, A. (2013). Effective strategies for increasing citation frequency. *International Education Studies*, 6(11):93–99.

Amin, M. M., Mao, R., Cambria, E., and Schuller, B. W. (2024). A wide evaluation of chatgpt on affective computing tasks. *IEEE Transactions on Affective Computing*, 15(4):2204–2212.

Armony, J. L. and Ledoux, J. E. (1999). How danger is encoded: towards a systems, cellular, and computational understanding of cognitive-emotional interactions. *MIT Press*.

Barnard, J. and Rubin, D. B. (1999). Small-sample degrees of freedom with multiple imputation. *Biometrika*, 86(4):948–955.

Belal, M., She, J., and Wong, S. (2023). Leveraging chatgpt as text annotation tool for sentiment analysis. *arXiv preprint arXiv:2306.17177*.

Broekens, J., Hilpert, B., Verberne, S., Baraka, K., Gebhard, P., and Plaat, A. (2023). Fine-grained affective processing capabilities emerging from large language models. *arXiv preprint arXiv:2309.01664*.

Buetow, M. (2002). Bad smells from köln. *Printed Circuit Fabrication*, 25(12):29–40. EID: 2-s2.0-0036934965.

Butler, R., SentiMentality, Ward, D., Jenkins, D., Lantrip, A., Armstrong, E., Butler, R., Driza, P., Fields, J., Levario, R., Miller, K., Plessala, B., Sigman, N., Slater, L., Vassallo, E., Vivekanandan, A., and Yildirim, L. (2025). Quantitative analysis of sentiment expression across large language models: A comparative study using plutchik's wheel of emotions parts i & ii.

Canner, J. (2014). Bipolate: Stata module to perform bivariate interpolation and smooth surface fitting for values given at irregularly distributed points. Statistical Software Components, Boston College Department of Economics.

Carretié, L. (2014). Exogenous (automatic) attention to emotional stimuli: a review. *Cognitive, Affective, & Behavioral Neuroscience*, 14(4):1228–1258.

Chamorro-Padial, J. and Rodríguez-Sánchez, R. (2023). The relevance of title, abstract, and keywords for scientific paper quality and potential impact. *Multimedia Tools and Applications*, 82(15):23075–23090.

Ciuk, D., Troy, A., and Jones, M. (2015). Measuring emotion: Self-reports vs. physiological indicators. *Physiological Indicators (April 16, 2015)*.

Clore, G. L. and Ortony, A. (2000). Cognition in emotion: Always, sometimes, or never. *Cognitive neuroscience of emotion*, pages 24–61.

Correia, S. (2016). ftools: Stata module to provide alternatives to common Stata commands optimized for large datasets. Statistical Software Components S458213, Boston College Department of Economics.

Correia, S. (2017). reghdfe: Stata module for linear and instrumental-variable/GMM regression absorbing multiple levels of fixed effects. Statistical Software Components S457874, Boston College Department of Economics.

Correia, S., Guimarães, P., and Zylkin, T. (2020). Fast Poisson estimation with high-dimensional fixed effects. *The Stata Journal*, 20(1):95–115.

Cui, G., Chung, S. Y.-H., Peng, L., and Wang, Q. (2024). Clicks for money: Predicting video views through a sentiment analysis of titles and thumbnails. *Journal of Business Research*, 183:114849.

Daume, J. and Hüttl-Maack, V. (2020). Curiosity-inducing advertising: How positive emotions and expectations drive the effect of curiosity on consumer evaluations of products. *International Journal of Advertising*, 39(2):307–328.

Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., et al. (2013). Orange: data mining toolbox in python. *the Journal of machine Learning research*, 14(1):2349–2353.

Desimone, R., Duncan, J., et al. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, 18(1):193–222.

Edlinger, M., Buchrieser, F., and Wood, G. (2023). Presence and consequences of positive words in scientific abstracts. *Scientometrics*, 128(12):6633–6657.

Ellsworth, P. C. (1991). *Some implications of cognitive appraisal theories of emotion*. Springer.

Ellsworth, P. C. and Scherer, K. R. (2003). *Appraisal processes in emotion*. Oxford University Press.

Elyoseph, Z., Hadar-Shoval, D., Asraf, K., and Lvovsky, M. (2023). Chatgpt outperforms humans in emotional awareness evaluations. *Frontiers in psychology*, 14:1199058.

Fatahi, S., Vassileva, J., and Roy, C. K. (2024). Comparing emotions in chatgpt answers and human answers to the coding questions on stack overflow. *Frontiers in Artificial Intelligence*, 7:1393903.

Ferreira, P., Rita, P., Rosa, P., Oliveira, J., Gamito, P., Santos, N., Soares, F., and Sottomayor, C. (2011). Grabbing attention while reading website pages: the influence of verbal emotional cues in advertising. *Journal of Eye Tracking, Visual Cognition and Emotion*, (1):64–68.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239.

Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

Frijda, N. H. (2007). *The laws of emotion*. The laws of emotion. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US. Pages: xiv, 352.

Frischen, A., Eastwood, J. D., and Smilek, D. (2008). Visual search for faces with emotional expressions. *Psychological bulletin*, 134(5):662.

Fronzetti Colladon, A., D'Angelo, C. A., and Gloor, P. A. (2020). Predicting the future success of scientific publications through social network and semantic analysis. *Scientometrics*, 124(1):357–377.

Gilardi, F., Alizadeh, M., and Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30):e2305016120.

Gustafsson, A. and Ghanbarpour, T. (2022). Challenging the troubled status of the marketing discipline. *AMS Review*, 12(3):184–187.

Habibzadeh, F. and Yadollahie, M. (2010). Are shorter article titles more attractive for citations? crosssectional study of 22 scientific journals. *Croatian medical journal*, 51(2):165–170.

Hsu, C.-T., Jacobs, A. M., Citron, F. M., and Conrad, M. (2015). The emotion potential of words and passages in reading harry potter–an fmri study. *Brain and language*, 142:96–114.

Jacques, T. S. and Sebire, N. J. (2010). The impact of article titles on citation hits: an analysis of general and specialist medical journals. *JRSM short reports*, 1(1):1–5.

Jamali, H. R. and Nikzad, M. (2011). Article title type and its relation with the number of downloads and citations. *Scientometrics*, 88(2):653–661.

Jann, B. (2014). Plotting regression coefficients and other estimates. *The Stata Journal*, 14(4):708–737.

Kang, J.-W. and Choi, S.-Y. (2025). Comparative investigation of gpt and finbert's sentiment analysis performance in news across different sectors. *Electronics*, 14(6):1090.

Karpen, S. C. (2018). The social psychology of biased self-assessment. *American Journal of Pharmaceutical Education*, 82(5):6299.

Kerr, G., Dombkins, K., and Jelley, S. (2012). "We love the Gong": a marketing perspective. *Journal of Place Management and Development*, 5(3):272–279.

Kissler, J., Herbert, C., Peyk, P., and Junghofer, M. (2007). Buzzwords: early cortical responses to emotional words during reading. *Psychological science*, 18(6):475–480.

Krausman, P. R. and Cox, A. S. (2020). Writing an effective title. *The Journal of Wildlife Management*, 84(6):1029–1031.

Krugmann, J. O. and Hartmann, J. (2024). Sentiment analysis in the age of generative ai. *Customer Needs and Solutions*, 11(1):3.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.

Lecourt, F., Croitoru, M., and Todorov, K. (2025). 'only chatgpt gets me': An empirical analysis of gpt versus other large language models for emotion detection in text. In *Companion Proceedings of the ACM on Web Conference 2025*, pages 2603–2611.

Li, Y., Chan, J., Peko, G., and Sundaram, D. (2023). Mixed emotion extraction analysis and visualisation of social media text. *Data & Knowledge Engineering*, 148:102220.

Liu, X. and 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.

Ma, Y., Teng, Y., Deng, Z., Liu, L., and Zhang, Y. (2023). Does writing style affect gender differences in the research performance of articles? An empirical study of BERT-based textual sentiment analysis. *Scientometrics*, 128(4):2105–2143.

Merton, R. K. (1968). The matthew effect in science: The reward and communication systems of science are considered. *Science*, 159(3810):56–63.

Mohammad, S. M. and Turney, P. D. (2013). Nrc emotion lexicon. *National Research Council, Canada*, 2:234.

Mohsin, M. A. and Beltiukov, A. (2019). Summarizing emotions from text using plutchik's wheel of emotions. In *7th scientific conference on information technologies for intelligent decision making support (ITIDS 2019)*, pages 291–294. Atlantis Press.

Moors, A., Ellsworth, P. C., Scherer, K. R., and Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion review*, 5(2):119–124.

Motger, Q., Oriol, M., Tiessler, M., Franch, X., and Marco, J. (2025). What about emotions? guiding fine-grained emotion extraction from mobile app reviews. In *2025 IEEE 33rd International Requirements Engineering Conference (RE)*, pages 6–18. IEEE.

Nair, L. B. and Gibbert, M. (2016). What makes a 'good'title and (how) does it matter for citations? a review and general model of article title attributes in management science. *Scientometrics*, 107:1331–1359.

Naqvi, A. (2022). Delaunay v1.2: A stata package for delaunay triangles, convex hulls, and voronoi tessellations. `https://github.com/asjadnaqvi/stata-delaunay-voronoi`. GitHub repository, version 1.2.

OpenAI (2026a). Chat completions api reference. `https://platform.openai.com/docs/api-reference/chat`. See the `seed` and `system_fingerprint` fields. Accessed 2026-01-02.

OpenAI (2026b). Gpt-5 mini model. `https://platform.openai.com/docs/models/gpt-5-mini`. Accessed 2026-01-02.

OpenAI (2026c). Responses api reference. `https://platform.openai.com/docs/api-reference/responses`. Accessed 2026-01-02.

Paiva, C. E., Lima, J. P. d. S. N., and Paiva, B. S. R. (2012). Articles with short titles describing the results are cited more often. *Clinics*, 67:509–513.

Peters, H. P. and van Raan, A. F. (1994). On determinants of citation scores: A case study in chemical engineering. *Journal of the American Society for Information Science*, 45(1):39–49.

Pieters, R. and Baumgartner, H. (2002). Who talks to whom? intra-and interdisciplinary communication of economics journals. *Journal of Economic Literature*, 40(2):483–509.

Plutchik, R. (1980a). *Emotion: A Psychoevolutionary Synthesis*. Harper & Row, New York.

Plutchik, R. (1980b). A general psychoevolutionary theory of emotion. In *Theories of emotion*, pages 3–33. Elsevier.

Plutchik, R. (2001). The nature of emotions. *American Scientist*, 89(4):344–350.

Roseman, I. J. and Smith, C. A. (2001). Appraisal theory. *Appraisal processes in emotion: Theory, methods, research*, pages 3–19.

Rostami, F., Mohammadpoorasl, A., and Hajizadeh, M. (2014). The effect of characteristics of title on citation rates of articles. *Scientometrics*, 98:2007–2010.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York.

Russell-Bennett, R. and Baron, S. (2016). the importance of the snappy title. *Journal of Services Marketing*, 30(5):477–479.

Sagi, I. and Yechiam, E. (2008). Amusing titles in scientific journals and article citation. *Journal of Information Science*, 34(5):680–687.

Sandoval, D. (2003). Casting a nuclear shadow. *Recycling Today*, 41(12):42–46. Online version published December 15, 2003.

Schaaff, K., Reinig, C., and Schlippe, T. (2023). Exploring chatgpt's empathic abilities. *arXiv preprint arXiv:2308.03527*.

Scherer, K. R. (1984). On the Nature and Function of Emotion: A Component Process Approach. In *Approaches To Emotion*. Psychology Press. Num Pages: 25.

Scherer, K. R. (2005). What are emotions? and how can they be measured? *Social science information*, 44(4):695–729.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and emotion*, 23(7):1307–1351.

Semeraro, A. (2022). pyplutchik (version 0.0.11) [python package]. Python Package Index (PyPI). Released 2022-04-11. Accessed 2026-01-07.

Semeraro, A., Vilella, S., and Ruffo, G. (2021). Pyplutchik: Visualising and comparing emotion-annotated corpora. *Plos one*, 16(9):e0256503.

Stavrova, O., Kleinberg, B., Evans, A. M., and Ivanović, M. (2025). Scientific publications that use promotional language in the abstract receive more citations and public attention. *Communications Psychology*, 3(1):118.

Stremersch, S., Verniers, I., and Verhoef, P. C. (2007). The quest for citations: Drivers of article impact. *Journal of Marketing*, 71(3):171–193.

Subotic, S. and Mukherjee, B. (2014). Short and amusing: The relationship between title characteristics, downloads, and citations in psychology articles. *Journal of information science*, 40(1):115–124.

Törnberg, P. (2023). Chatgpt-4 outperforms experts and crowd workers in annotating political twitter messages with zero-shot learning.

Van Dalen, H. and Henkens, K. (2001). What makes a scientific article influential? the case of demographers. *Scientometrics*, 50(3):455–482.

Vrtana, D. and Krizanova, A. (2023). The power of emotional advertising appeals: Examining their influence on consumer purchasing behavior and brand–customer relationship. *Sustainability*, 15(18):13337.

Vuilleumier, P., Armony, J. L., Driver, J., and Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fmri study. *Neuron*, 30(3):829–841.

Wake, N., Kanehira, A., Sasabuchi, K., Takamatsu, J., and Ikeuchi, K. (2023). Bias in emotion recognition with chatgpt. *arXiv preprint arXiv:2310.11753*.

Webster, G. D., Jonason, P. K., and Schember, T. O. (2009). Hot topics and popular papers in evolutionary psychology: Analyses of title words and citation counts in evolution and human behavior, 1979–2008. *Evolutionary Psychology*, 7(3):147470490900700301.

Wu, D. et al. (2025). Authenticity or self-advocacy? Identifying the credibility of positive words in scientific titles and abstracts. *Scientometrics*.

Wu, D., Wu, H., and Li, J. (2024). Citation advantage of positive words: predictability, temporal evolution, and universality in varied quality journals. *Scientometrics*, 129(7):4275–4293.

Yiend, J. (2010). The effects of emotion on attention: A review of attentional processing of emotional information. *Cognition and emotion*, 24(1):3–47.

Yongsatianchot, N., Ghanad Torshizi, P., and Marsella, S. (2023). Investigating large language models' perception of emotion using appraisal theory. *arXiv preprint arXiv:2310.04450*.

Yuan, Z.-m. and 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.

8 Appendix

8.1 Appendix 1

To estimate the time trend in citations, we fit a Poisson regression of citation counts on a full set of year indicators, controlling for article length (pages) and journal fixed effects, and we plot the resulting year-by-year adjusted predictions with 95% confidence intervals.

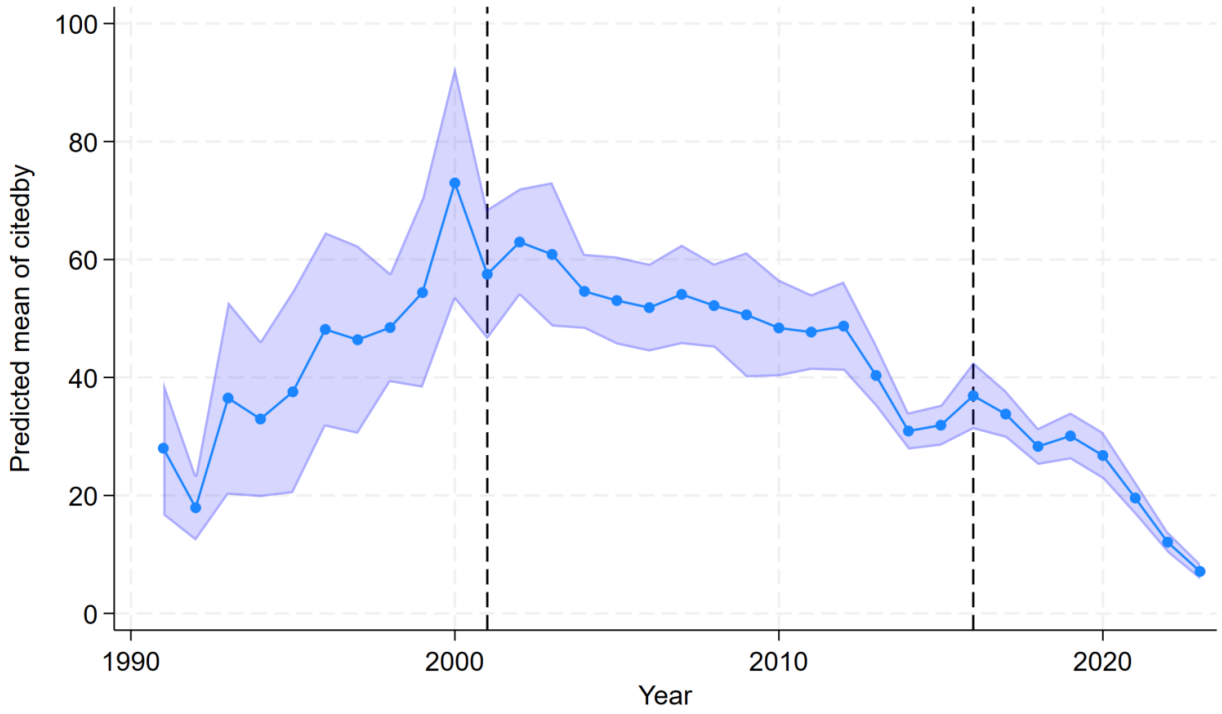

Figure 9: Citation count trend

Why exclude papers published after 2015 (right of the second dashed line): The steep decline in predicted citations after 2015 reflects *incomplete citation accumulation* rather than true differences in paper quality or impact. Recent papers simply haven't had enough time to be discovered, read, and cited. Since citations typically accumulate over approximately 10 years, papers published after 2015 would show artificially low citation counts, biasing the analysis downward.

Why exclude papers published before 2001 (left of the first dashed line): The rising

pattern before 2001, culminating in a peak around 1999-2000, reflects *citation saturation*. Very old papers have essentially "aged out" of the active citation cycle, they've stopped accumulating new citations as the literature has moved on. The high values in the late 1990s likely represent a distorted peak where papers had maximum time to accumulate citations before this saturation effect took hold.

The 2001-2015 window represents a stable measurement period where you observe a relatively consistent, gradual trend. Papers in this range have had sufficient time ($\tilde{1}0+$ years) to accumulate a representative citation count, but are not so old that citation accumulation has ceased. This provides the most reliable and comparable data for assessing the relationship between page count and citations. In any case, years fixed-effects are considered in the analysis.

8.2 Appendix 2

Table 5: Overview of article-level control variables

Variable	Type	Category	Description
Document characteristics			
`Pagecount`	Continuous	Document	Number of pages of the article
`open` (0–7)	Categorical	Document	Open-access status as reported in Scopus metadata
Title features			
`title_words`	Count	Title format	Number of words in the title
`title_split`	Binary	Title format	Equal to 1 if the title contains a colon (split title), 0 otherwise
`title_special`	Count	Title format	Number of non-alphanumeric characters in the title
Author-team quality			
`n_authors`	Count	Author quality	Number of authors listed on the paper
`author_max_degree`	Continuous	Author quality	Maximum weighted co-authorship degree centrality across the team, computed on the full corpus co-authorship graph
`author_max_prior`	Count	Author quality	Maximum cumulative publications in the corpus strictly before the focal year, across the co-author team

8.3 Appendix 3

Table 6: Coefficients of Regression Model with Primary Emotions and Mean Effect for Combinations

	Base	Tier 1	Tier 2	Tier 3
Anger	0.002			
	(0.081)			
Anticipation	-0.013			
	(0.024)			
Disgust	0.053			
	(0.068)			
Fear	-0.052			
	(0.033)			
Joy	0.066**			
	(0.027)			
Sadness	-0.034			
	(0.028)			
Surprise	-0.008			
	(0.022)			
Trust	0.058			
	(0.030)			
Love		0.151**		
		(0.063)		
Submission		-0.043		
		(0.070)		
Awe		-0.050		
		(0.121)		
Disapproval		0.014		
		(0.133)		
Remorse		-0.072		
		(0.119)		
Contempt		0.094		
		(0.088)		

	Base	Tier 1	Tier 2	Tier 3
Aggressiveness		-0.033		
		(0.053)		
Optimism		0.000		
		(.)		
Guilt			0.122**	
			(0.050)	
Curiosity			0.087	
			(0.047)	
Despair			-0.225***	
			(0.077)	
Unbelief			-0.080	
			(0.063)	
Envy			0.121	
			(0.112)	
Cynicism			-0.004	
			(0.041)	
Pride			0.000	
			(.)	
Hope			0.000	
			(.)	
Delight				0.035
				(0.071)
Sentimentality				-0.025
				(0.061)
Shame				-0.094
				(0.070)
Outrage				-0.071
				(0.079)
Pessimism				-0.033
				(0.053)
Morbidity				0.116
				(0.071)

	Base	Tier 1	Tier 2	Tier 3
Dominance				0.132**
				(0.066)
Anxiety				0.000
				(.)
Pagecount	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
1.open	-0.036	-0.034	-0.032	-0.034
	(0.100)	(0.100)	(0.098)	(0.100)
2.open	0.538	0.534	0.506	0.534
	(0.720)	(0.721)	(0.730)	(0.721)
3.open	0.592***	0.600***	0.603***	0.600***
	(0.199)	(0.200)	(0.202)	(0.200)
4.open	1.073***	1.085***	1.087***	1.085***
	(0.226)	(0.226)	(0.228)	(0.226)
5.open	0.261***	0.263***	0.262***	0.263***
	(0.071)	(0.071)	(0.071)	(0.071)
6.open	-0.331	-0.325	-0.334	-0.325
	(0.305)	(0.303)	(0.303)	(0.303)
7.open	0.134	0.139	0.139	0.139
	(0.228)	(0.227)	(0.227)	(0.227)
n_authors	0.024	0.024	0.025	0.024
	(0.019)	(0.019)	(0.019)	(0.019)
title_words	-0.009*	-0.009*	-0.009	-0.009*
	(0.005)	(0.005)	(0.005)	(0.005)
title_split	0.100***	0.101***	0.102***	0.101***
	(0.037)	(0.037)	(0.037)	(0.037)
title_special	0.025**	0.025**	0.023*	0.025**
	(0.012)	(0.012)	(0.012)	(0.012)
author_max_degree	0.010***	0.010***	0.010***	0.010***
	(0.003)	(0.003)	(0.003)	(0.003)
author_max_prior	-0.016*	-0.016*	-0.016*	-0.016*

	Base	Tier 1	Tier 2	Tier 3
	(0.009)	(0.009)	(0.009)	(0.009)
Observations	15836	15836	15836	15836

Notes: All coefficients are Rubin-pooled across five LLM readings (GPT-5-mini seeds 42, 1, 2; ChatGPT-5.4; GPT-5.4-mini). Pooled SE = $\sqrt{W + (1 + 1/m)B}$ where W is within-annotation variance and B between-annotation variance ($m = 5$).
Emotions dropped due to perfect collinearity (Optimism, Pride, Hope, Anxiety) shown as 0.
All models absorb journal and publication-year fixed effects.
Standard errors clustered at journal level in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

8.4 Appendix 4

Table 7: Coefficients of Regression Model with Primary Emotions and Interaction Terms for Their Combinations

	Unified model
Joy	0.228*
	(0.132)
Trust	0.187
	(0.116)
Fear	-0.116
	(0.166)
Surprise	0.110
	(0.143)
Sadness	-0.050
	(0.155)
Disgust	-0.033
	(0.192)
Anger	0.142
	(0.192)
Anticipation	0.106
	(0.148)
Primary dyads (Tier 1)	
Joy × Trust	-0.013
	(0.014)
Trust × Fear	0.008
	(0.024)
Fear × Surprise	0.008
	(0.023)
Surprise × Sadness	-0.026
	(0.023)
Sadness × Disgust	0.007
	(0.024)

	Unified model
Disgust × Anger	0.001
	(0.018)
Anger × Anticipation	-0.031
	(0.025)
Anticipation × Joy	-0.006
	(0.017)
Secondary dyads (Tier 2)	
Joy × Fear	0.005
	(0.023)
Trust × Surprise	-0.007
	(0.021)
Fear × Sadness	0.034**
	(0.013)
Surprise × Disgust	-0.031
	(0.024)
Sadness × Anger	-0.009
	(0.026)
Disgust × Anticipation	0.022
	(0.027)
Anger × Joy	-0.014
	(0.030)
Anticipation × Trust	-0.016
	(0.019)
Tertiary dyads (Tier 3)	
Joy × Surprise	-0.018
	(0.022)
Trust × Sadness	-0.024
	(0.022)
Fear × Disgust	-0.004
	(0.027)
Surprise × Anger	0.021

	Unified model
	(0.031)
Sadness × Anticipation	0.017
	(0.018)
Disgust × Joy	0.016
	(0.043)
Anger × Trust	0.006
	(0.025)
Anticipation × Fear	-0.014
	(0.018)
Pagecount	0.000
	(0.000)
1.open	-0.041
	(0.102)
2.open	0.565
	(0.730)
3.open	0.572***
	(0.200)
4.open	1.063***
	(0.231)
5.open	0.261***
	(0.070)
6.open	-0.385
	(0.307)
7.open	0.120
	(0.229)
n_authors	0.022
	(0.018)
title_words	-0.008
	(0.005)
title_split	0.101***
	(0.036)
title_special	0.025**

	Unified model
	(0.012)
author_max_degree	0.009***
	(0.003)
author_max_prior	-0.015*
	(0.009)
_cons	2.979***
	(0.679)
Observations	15836

Notes: All coefficients are Rubin-pooled across five LLM readings (GPT-5-mini seeds 42, 1, 2; ChatGPT-5.4; GPT-5.4-mini). Pooled SE = $\sqrt{W + (1 + 1/m)B}$ where W is within-annotation variance and B is between-annotation variance ($m = 5$). Interaction terms represent products of standardised emotion scores. Model absorbs journal and publication-year fixed effects. Standard errors clustered at journal level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.