

Rachid Azrout,
Faculty of Social and Behavioural Sciences
Course: Content Analysis



UNIVERSITEIT VAN AMSTERDAM

Laughing in the Digital Age: A Content Analysis of Humor Types in Comments on Reddit

Andjela Mestrovic

12334707

12334707@uva.nl

Inga Vondenhof

12052825

12052825@uva.nl

Yuhui Chen

14172372

14172372@uva.nl

Helge Moes

11348801

11348801@uva.nl

Date: 24.03.2023

Word count: 4596

Table of Content

Introduction	3
Methodology	6
Sample	6
Reliability Test	7
Statistical Tests	8
Analyses and Results	9
Descriptives	9
Testing Hypotheses 1a and 1b	10
Testing Hypothesis 2	11
Exploratory Analysis	11
Conclusion	15
Discussion	15
References	18
Appendices	21
Appendix A - Python script for random post-URL generation on subreddits.	21
Appendix B - Stopwords for the word clouds	22
Appendix C - Additional Data Visualization	22

Introduction

Humor is a vital aspect of our everyday lives that makes us enjoy situations, reduces negative emotions and builds social connections both online and offline (Martin & Ford, 2018). One of the most popular user-generated social media platforms where humor is shared is Reddit. Reddit allows users to add their own posts in the form of a link, photo, video or their own text and to vote on or respond to posts (Anderson, 2015). In doing so, it is possible to remain completely anonymous, as users interact with a self-chosen or randomly generated account name. Upvotes and downvotes allow users to judge posts and responses to increase or decrease their visibility and ordering.

The platform also allows users to subscribe to individual communities, so-called “subreddits”, which are created around specific interests and topics (Anderson, 2015). Subreddits vary in structure since they can differ in the degree of moderation, rules on what to post and comment on, and audiences (Horne, Adalı, & Sikdar, 2017). Such factors affect the amount of humor employed by users in posts and comments. On social media, humor is often used in an organic way with the intention to bring laughter to others. Previous studies about other user-generated platforms (Facebook) showed that users employ humor to build social connections and to cope with personal stressful situations they are encountering (Pennington & Hall, 2014).

In the case of Reddit, these intentions vary. The conversations that take place on Reddit may not mimic conversations that occur in other media, seen in their different structures and affordances (Proferes et al., 2021). Furthermore, other research emphasizes the collection of ‘posts’ from Reddit, yet their figures show both original posts and comment threads rather than solely the post (Proferes et al., 2021). The present research solely focuses on humor types in comments. It is crucial to understand what type of comment attracts people’s attention online and drives discussions. Older research has shown that humorous contents are able to grab more attention (Schmidt, 1994), increase the likeability of the source (Sternthall & Craig, 1973), and promote a pleasant mood for people to be more agreeable (Sinclair et al., 1994).

Currently, over 2 million subreddits exist (Lin, 2022). However, not every subreddit has the same amount of interactions. In this study, Reddit comments in three subreddits (r/funny, r/gaming and r/AskReddit) were analyzed, since they are the biggest active subreddits on Reddit 2023. Moreover, it was specified what *types of humor* a funny comment might display, based on the paper of van der Wal et al. (2020). Ten humor types were used:

aggressive, self-defeating, sexual, irreverent, coping, parody, wordplay, incongruity, hyperbole and absurdity. Overall, the aim was to investigate what role humor plays in the comments from different communities and see how humor differs in comments that get the greatest positive attention in comparison to the comments that are marked as controversial. Based on this, we aim to answer the following research question:

RQ: What is the difference in expressing humor between ‘Top’ and ‘Controversial’ comments in three most popular different subreddits?

One way in which this research contributes to the field is by giving insights into how humor shapes attitudes expressed within specific subreddits. Also, it can foster the understanding of moderators and community managers on how humor is expressed within their comment sections. This should overall contribute to how they can create a positive and engaging atmosphere within their subreddits. Lastly, we aim to contribute to the development of coding typologies of humor within texts on social media websites.

Theoretical Background & Hypotheses

Humor is a concept that has not been elaborately researched within media entertainment, due to the challenge of demarcating a specific characteristic of humor. Van der Wal et al. (2020) attempted to characterize these humor types which they base on three theories: relief, superiority, and incongruity theory.

Relief theory revolves around humor that releases (psychological) tension or arousal (van der Wal et al., 2020). According to the authors, three types of humor fall into this category: sexual, irreverent and coping. Superiority theory (disparagement humor) describes humor where aggression is implemented as a component of humor (Zillmann, 1983). It describes a type of enjoyment that emerges from the sense of superiority one experiences when comparing themselves to others’ misfortunes. According to van der Wal et al.’s typology (2020), two types of humor fall into this category: aggressive and self-defeating. Finally, the incongruity theory explains that the theme of the humor type is not as important here as in the previous two theories. Instead, it is based on the cognitive processes that make people perceive something as funny (Morreall, 1989). Humans find concepts funny if the differences between how we perceive and understand the world and how it is are incongruent (McGhee, 1979). Van der Wal et al. (2020) name five types of humor that fall into this category: parody, wordplay, incongruity and hyperbole and absurdity.

Within the study, the three subreddits chosen revolve around different topics. The content of r/AskReddit tends to be more objective and takes the form of a formal discussion (Caplan & Purser, 2017), while the style of r/funny is more subjective and casual. The r/gaming can be seen as a mix of the two, including memes but also serious questions (Hernandez, 2018). Based on this, all three subreddits were expected to have different amounts of humor and humor styles within the comments. Since funny content is the core idea for r/funny, we expect this to be the subreddit to have the most humoristic content. The first hypothesis is as follows:

H1a: There is more humor present in the comments of subreddit r/funny than in subreddits r/AskReddit and r/gaming.

Reddit has a couple of ways in which the comments under a post can be sorted. Two types of sortings in this research are focussed on: top and controversial. ‘Top’ comments are the comments shown in the comment section when sorting them by *top*. These are the comments that got the most upvotes from other users. ‘Controversial’ comments are the comments that are shown in the comment section when sorting them by *controversial*. They usually have approximately an equal amount of upvotes/downvotes, which makes them discussion points. The topics of these discussions can vary but opinions are usually split and cause arguments. Based on the nature of the two types of sortings, our second hypothesis is the following:

H1b: There is more humor present in top comments of Reddit posts than in controversial comments across all subreddits.

Moving on, controversial comments tend to have more aggressive language, be insensitive towards (a group of) people and can contain an opinion that is not shared by the majority (Jhaver et al., 2019). This makes that controversial comments can be viewed as more negative and rude and include more socially sensitive aspects, but especially include more “dark humor” than top comments. Here, dark humor is defined by the Merriam-Webster dictionary (2023) as: “humor that is marked by the use of jokes, irony, and sarcasm that are intended to make light of subject matter that is generally considered taboo, particularly subjects that are normally considered serious or painful to discuss.” Based on this description, we appointed aggressive, irreverent and coping humor types for examining dark humor in the comments on Reddit. Based on this, we expect the same when looking at humor types:

H2: Controversial comments contain more dark humor (aggressive, irreverent and coping humor types) compared to top comments.

Methodology

In order to answer the hypotheses, a quantitative content analysis of the humor types within the comments section of three different subreddits was conducted. This section consists of the sample, procedure, a description of the variables and the reliability test.

Sample

To collect as much data as possible in 20 hours, a Python script (Appendix A) was implemented to randomly sample 1004 URLs from each subreddit (r/funny, r/Gaming, r/AskReddit), choosing out of a selection pool of max. 100.000 posts from the category “Top”, sorted by “last year” (February 2022 - February 2023). These URLs are printed out in a .txt file. These were put in an excel sheet, which made it possible for the coder to go through the links and tag them when they were coded. Every researcher got 80 randomized links to post for each subreddit, containing as many links as they need. In total, 1004 comments on whether humor and humor types were present in “top” and “controversial” filtered subreddits were coded. After removing duplicates in the final dataset, 960 comments were used to examine the data.

Procedure and Variables

For every post, the registration unit was the content of the first user-written comment under the sorting categories: “Top” and “Controversial.” If a moderator or the author themselves deleted the comment, the next comment was coded. The context unit was the subreddit (r/Askreddit, r/gaming/ and r/funny) and the post the comments referred to. To make the decision to code humor as present or not, humor was defined based on the dictionary explanation on Merriam-Webster (2023): “a. funny or amusing quality, b. the ability to be funny, c. a sudden, unpredictable, or unreasoning inclination”. The replies to the comment were not taken into account, since these could lead to complications in the coding process. Moreover, the context and tone of the comment were taken into consideration, when deciding whether it should be coded as humor. Exact instructions can be found in the codebook (file provided in the attachment). Additionally, comments may contain cynicism, sarcasm and irony or may include elements of the definitions mentioned in the humor types, but are not intended to be humorous. As stated by van der Wal et al. (2020), the coder may

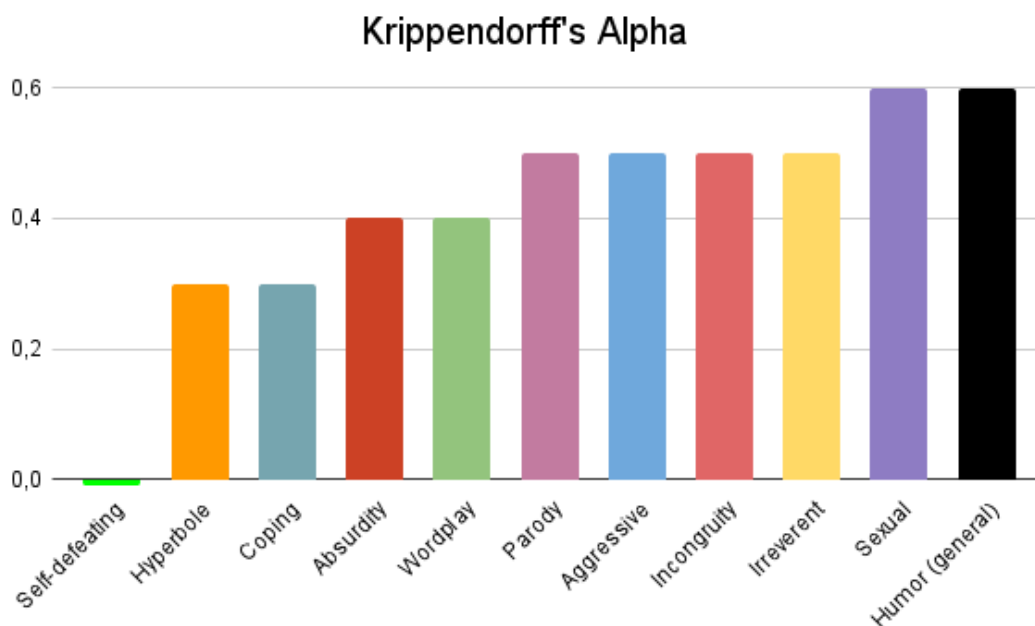
not perceive content as humorous, yet they should still be able to recognize that the intention is meant to be humorous by the person in question or the creator of the comment. Therefore, it was expected to be coded as humorous content. The coder also had to consider the context of the post when making the evaluation of the comment, as it may be a humorous response directly referring to the original post, which can also contain humor.

To code humor types, a tenth type was added, namely hyperbole. This was done to take into account possible irony and exaggerated statements (going beyond literal meaning) that count as humorous. Moreover, it replaced an omitted humor type 'slapstick', which is a physical type of humor, and is usually presented in visuals. Van der Wal et al. (2020) investigated humor in television shows where this humor type was more relevant but was challenging to retrieve from text in our case. Moreover, the definition of each humor type and humor examples were adapted to fit the characteristics of a textual comment on Reddit. The categories were not meant to be mutually exclusive, which means that a comment may contain more than one humor type. This was also coded accordingly.

Reliability Test

To test the reliability of the variables of the humor types, all coders coded a selection of 80 comments per Subreddit. In contrast to van der Wal et al. (2020), the reliability test in this study showed questionable outcomes (Figure 1). Overall, the humor types scored around 0.5 Kappa which is insufficient. This threatened the reliability of the study and the actual coding. While most of the findings do hint at low reliability, self-defeating gave a negative outcome (-0.02), meaning it was worse than random. After investigation, it was due to the fact that the coders did not find a lot of instances of this type of humor within the reliability sample. Nevertheless, the codebook had to be adjusted. This was done in terms of definitions, coding procedures and coding instructions. The definitions and examples that were initially designed were revised and extended, using sub-questions that define the type of humor more precisely instead of simply indicating whether the type of humor is present. Moreover, all codes of comments where there was disagreement were discussed in coder training sessions to make sure that each coder had the same understanding before data collection.

Figure 1. Reliability scores for humor types (Kalpha).



Statistical Tests

The independent variables in this study were type of comment (top/controversial) and type of Subreddit (r/funny, r/gaming, r/AskReddit). These were used to draw comparisons of different comments, and communities on subreddits. In order to analyze Hypotheses 1a and 1b, a binary logistic regression model was employed to examine the presence of humor (dependent variable) in different subreddits and types of comments. This was done since all variables measured were dichotomous (not continuous), and to control for both independent variables at the same time to get a more nuanced result.

Hypothesis 2 was analyzed with a Chi-square, checking whether comments sorted by “controversial” include more aggressive, irreverent and coping humor types (‘dark humor’) than comments sorted by “top”. These three humor types were chosen to find out why comments under the category may be classified as controversial, and whether they express more controversially deemed types of humor (Hietalahti, 2016). Thus, the dependent variable to test hypothesis 2 was the presence of dark humor, while the predictor was the type of comment. Lastly, exploratory analyses were performed to see if there are differences in the presence of different types of humor between the three subreddits and types of comments.

Analyses and Results

This section comprises several elements: descriptives, the testing of the hypotheses and an exploratory analysis. These findings are supported by tables and figures based on the methods implemented and the results of the statistical tests.

Descriptives

First of all, the raw data was structured, organized and cleaned in order to analyze the information to examine the hypotheses. This consisted of removing duplicates (coding the same comment twice) and invalid or incomplete codes (missing data). In total, 44 comments were excluded based on these criteria. The reason behind this large number is explained by the fact that Subreddit posts are ordered. Therefore, the script randomly picked up the same posts for different coders or the same post with a different URL in two instances.

After data cleaning, computation and recoding of variables to prepare for analyses, descriptive statistics were used to gain an overview of the data. Out of the total 960 comments, 313 included humor. Within these comments, 866 instances of humor types were measured, as they were able to co-occur. In Table 1, the prevalence of humor types within humorous-coded comments is shown in percentages by subreddit and type of comment. Among all types of humor, incongruity humor is found the most in the “top” comments, and in all of the subreddits (Appendix C). This finding was to no surprise, as these types of humor had the broadest definitions (van der Wal et al., 2020) and therefore often co-occurred with other humor types. In total, only 49 comments had no co-occurring humor types, meaning that they were coded with one unique humor type.

Irreverent humor appears the most in the “controversial” comments and they also contain two times the aggressive humor when compared to “top” comments. Lastly, the largest proportion of self-defeating, sexual, irreverent and coping humor was found in r/AskReddit.

Table 1. Prevalence of humor types by subreddits and type of comment.

Types of humor	Subreddits			Type of comment	
	r/funny	r/Askreddit	r/gaming	Top	Controversial
Aggressive	9.40%	8.50%	5.90%	5.50%	10.30%
Self-defeating	6.00%	7.90%	4.60%	9.30%	3.10%
Sexual	6.30%	8.80%	3.70%	6.10%	6.40%
Irreverent	9.70%	10.40%	5.60%	3.60%	13.40%
Coping	5.00%	6.00%	4.00%	7.40%	2.70%
Parody	11.30%	8.50%	9.60%	12.80%	6.80%
Wordplay	9.70%	7.20%	6.20%	11.20%	4.30%
Incongruity	24.80%	16.40%	16.00%	29.30%	9.10%
Absurdity	10.40%	4.40%	5.60%	8.60%	4.90%
Hyperbole	18.60%	10.70%	9.90%	17.90%	8.20%

Testing Hypotheses 1a and 1b

To test whether the presence of humor would significantly differ between the subreddits r/funny, r/AskReddit, r/gaming and between “top” and “controversial” comments, a logistic regression model was run. For the independent variable type of subreddit, dummy variables were created with the reference category of r/funny. The model with presence of humor as the dependent variable and type of subreddit and type of comment as independent variables in a sample of 960 comments had a good fit, $\chi^2(960, 3) = 63.77, p < .001$.

Holding all other variables constant, r/funny had the highest odds of humor being present, as the chance of humor being present in the other two subreddits decreased significantly, with r/AskReddit, $B = -0.70, SE = 0.17, p < 0.00$, and r/gaming, $B = -0.81, SE = 0.17, p < .001$. As shown in Table 2, the odds of humor being present in comments from the subreddit "funny" were found to increase by 44.62% (95% CI [.32, .63]) for r/AskReddit, and for r/gaming by 49.7% (95% CI [.36, .70]). As such, *H1a* was supported.

Furthermore, the model showed that when controlling for the type of subreddit, there was a significant increase in the odds of humor being present in top comments as compared to controversial comments, $B = .88, SE = 0.44, p < .001$. Notably, the expected log odds of

humor being present increased by 2.42 when type of comment increased by one unit (1 = top, 0 = controversial) while holding other variables constant. Therefore, *H1b* was also supported.

Table 2. Logistic regression to test *H1a* & *H1b*.

	<i>B</i>	<i>SE</i>	<i>Exp (B)</i>	95% CI for <i>Exp(B)</i>	
				Lower	Upper
Constant	-.71***	.14	.49		
r/AskReddit	-.70***	.17	.50	.36	.70
r/gaming	-.81***	.17	.47	.32	.63
Top comment	.88 ***	.14	2.42	1.82	3.21

N = 960.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Note. Reference categories are r/funny and controversial comments.

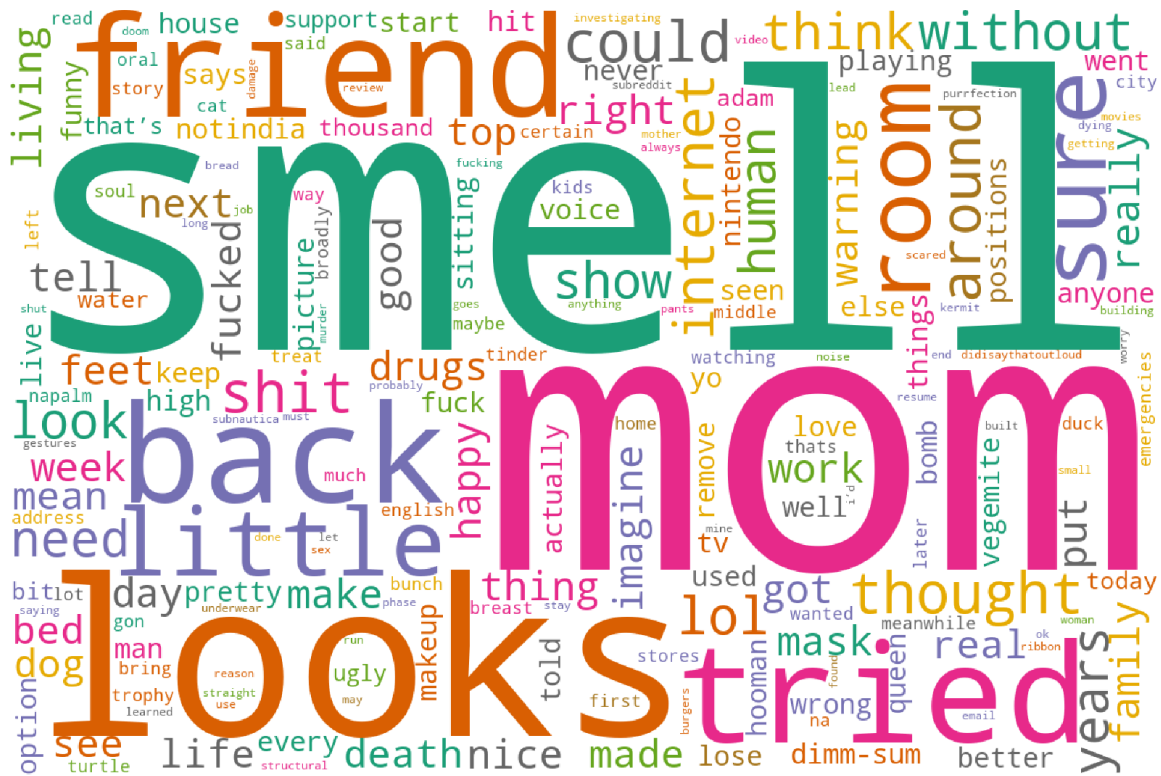
Testing Hypothesis 2

In order to evaluate the second hypothesis, a chi-square test of independence was conducted to examine the potential differences in the presence of dark humor between top and controversial comments. The chi-square test indicated that there was no significant association between the presence of dark humor and the type of comment, $\chi^2(1, N = 960) = 0.89, p = .35$. Therefore, *H2* was rejected. Controversial comments do not significantly inhibit more dark humor than top comments.

Exploratory Analysis

To further investigate how humor within two types of comments might differ, an automated text visualization approach was chosen to generate word clouds for the two categories. Each word is displayed in a size that is representative of the *term frequency · inverse document frequency* ($tf \cdot idf$), weighing each occurrence by its raw frequency within a comment, corrected for how often it occurs in all comments (Van Atteveldt et al., 2022). Figure 2 displays all comments which were coded as funny in the “top” category (N = 199).

Figure 2. Word cloud for humoristic top comments.

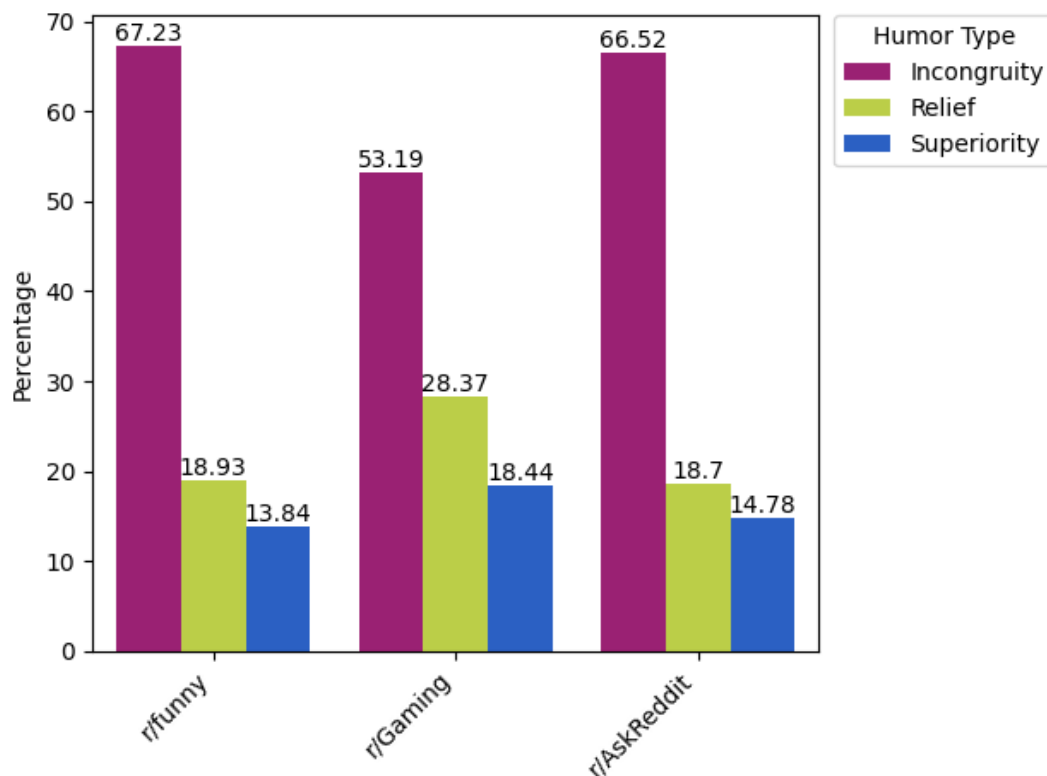


Moreover, Figure 3 displays the $tf \cdot idf$ for all comments in the controversial section coded as funny ($N = 114$). While both clouds display some overlapping terms, such as “mom” or “lol”, it can be seen that in the controversial cloud, there are more words that could be used as swearwords (“fuck”, “fat”, “balls”) and references to gender and sexuality (“dick”, “sex”, “heterosexual”). In Appendix B there is a list of stopwords which were not considered by the pre-defined stopwords list from the English language package of the TreebankTokenizer (van Atteveldt et al., 2022). This list of additional stopwords includes terms that would overlap between these subreddits which were not automatically caught, such as “someone”. These clouds provide a more qualitative picture of the content within the two comment types. Both clouds and Table 1 give reason to assume that the way humor is used in top vs. controversial comments may differ after all.

used the same predictor and control variables, with the dependent variable being incongruity humor, $\chi^2(960, 3) = 76.53, p < .001$.

The results of the analysis indicate that r/funny ($B = 0.49, SE = 0.22, p = 0.03$) and r/AskReddit ($B = 0.49, SE = 0.22, p = 0.03$) are more likely to include relief humor than r/gaming. Similarly, both r/funny ($B = 0.52, SE = 0.25, p = 0.03$) and r/AskReddit ($B = 0.59, SE = 0.24, p = 0.02$) are more likely to include superiority humor than r/gaming. Additionally, r/funny is more likely to have incongruity humor than other subreddits ($B = 0.79, SE = 0.18, p < 0.001$), while there is no significant difference between r/AskReddit and r/gaming ($B = 0.04, SE = 0.19, p = 0.82$) for this humor type. However, it should be noted that the p -value higher than 0.05 in the relief humor model (Model 1) suggests that the model may not fit the data well. This could be due to the fact that there were relatively few instances of relief humor coded in the dataset, resulting in limited statistical power to detect differences across subreddits as portrayed in Figure 4.

Figure 4. Humor theories in subreddits.



Conclusion

In this research, we tried to examine an answer to the research question concerning the difference in expressing humor between ‘Top’ and ‘Controversial’ comments in three most popular different subreddits. The analyses conducted show that, as we hypothesized, there is more humor present in r/funny in comparison to the other two subreddits and that “top” comments contain more humor than “controversial” comments. For each subreddit, the nuances in humor types are evident as can be seen in three pie charts in Appendix C.

According to Appendix C it was apparent that the most represented humor types were incongruity and hyperbole. This implies that these humor types were employed the most across the three subreddits in ‘Top’ and ‘Controversial’ comments. Consequently, this indicates that if one desires to create a top or controversial comment that grabs the attention of a subreddit, these two humor types seem to have the most beneficial results.

From the exploratory analysis, it can be seen that top and controversial comments differ in their humoristic content. As specified before, controversial comments contain more reference to gender and sexuality and swearwords. In addition, while the humor types of the incongruity theory were the biggest in both types of comments, the top comments contained a significantly higher amount. The content of these humor types is usually paired with other humor types. That is something which future research could elaborate on, as it was not analyzed in this study.

The expectation that “controversial” comments contain more dark humor than “top” comments was not met, indicating that dark humor is also considerably present in ‘top’ comments. Although not all findings seemed to be regarded as successful, it implies intriguing findings on conducting a quantitative content analysis on humor types on online social media platforms.

Moreover, this research was built upon the theory and measurements presented by van der Wal et al. (2020) for assessing humor types on Reddit. We did not attempt to test the particular theory, but to apply it and examine how this played out on text-based social media platforms, such as Reddit. Therefore, we contribute to the conception of humor and to the measurement and typology of van der Wal et al. (2020) of implementing a humor scale.

Discussion

Assessing our hypotheses, H1a and H1b may be regarded as obvious. However, it argues that the statistical test that is run is reliable and that the data portrayed representable information. Moreover, these hypotheses were integral in order to answer the research question and formulate the argumentation of this study. Nevertheless, for further analysis of humor on Reddit, other and more complex hypotheses and subquestions can be formulated, due to the fact that the data is considerably rich and can raise many questions that may be answered.

The controversial comments did not contain more dark humor (aggressive, irreverent and coping humor types) compared to top comments in this research, our analysis shows potential for further research. One may question if the same findings may be applicable to other similar platforms, as the generalizability may vary depending on the specific characteristics of the platform in question. For instance, the ways in which humor is used on Twitter or Facebook may differ from how it is used on Reddit, due to differences in the structure, moderation, posting rules and culture of these platforms.

Although practices in the content analysis were automated, it was integral to check the raw data, since 44 duplicates were removed from the final dataset. This was due to the fact that Subreddit posts are ordered in the span of February 2022 to 2023. Therefore, the script was limited to the rules and acceptance criteria of a post, which was also the case in previous studies (Proferes, et al., 2021). Consequently, the code randomly picked the same posts for multiple coders or the same post with a different URL twice.

With regard to the methodology and the reliability test that was conducted for this study, we would recommend for future research to invest more time and resources towards the reliability test, as this is the most pressing limitation of the study. Due to time constraints and different cultural backgrounds of coders, the reliability test displayed in Figure 1 shows questionable results. In addition, one of the coders had different outcomes than the others on several items. To account for this, we implemented sequential coder training consisting of multiple meetings going through the codebook and discussing alterations and coding behavior. Here, more intensive coder training is advised, especially for analyzing such latent concepts as humor types. A follow-up reliability test was not able to take place due to time and resource constraints.

An important factor in this study was the implementation of the scale of van der Wal et al. (2020). Although these humor types were tested before in a different setting, the

research still experienced a degree of overlap between the humor types. Since humor is subjective, it is challenging to formulate clear differences in the definition of the humor and the experience of it. Therefore, future research could carry out an analysis of the co-occurrences of humor types, since the humor types were not mutually exclusive and comments could contain different types of humor. Consequently, we suggest further research to build on our research and tailor more precise definitions for humor types and the design of scales.

Furthermore, we implemented automated data-collecting methods in order to avoid manual mistakes that emerged during the reliability test. This allowed us to code a post in less time, and for the coder to get more time per post to evaluate the content of the post in relation to the comment. Additionally, we were better equipped to assess nuances such as sarcasm and cynicism, since the coder had more time to re-evaluate whether there was actual humorous intent. When the coder was unsure of humor or the meaning behind the comment, the comment was not considered humorous. This may have influenced the reliability of the study significantly.

Finally, although the exploratory analysis was used for further understanding of the humor present across the subreddits, it also contained limitations, since the results portrayed an insignificant p -value of 0.52 in the relief humor model (Model 1). An initial explanation for this could be that there were relatively few cases of relief humor coded in the dataset. This resulted in limited statistical power to detect differences across subreddits and suggests that the model may not fit the data well and should be further researched.

References

- Anderson, K. E. (2015). Ask me anything: What is Reddit? *Library Hi Tech News*, 32(5), 8–11. <https://doi.org/10.1108/LHTN-03-2015-0018>
- Caplan, M. A., & Purser, G. (2017). Qualitative inquiry using social media: A field-tested example. *Qualitative Social Work*, 18(6). <https://doi.org/10.1177/1473325017725802>
- Dark Humor—Definition and More from the Free Merriam-Webster Dictionary*. (n.d.). Retrieved March 23, 2023, from <https://www.merriam-webster.com/dictionary/dark%20humor>
- Hernandez, P. (2018, April 4). Reddit's r/gaming is a goldmine of humor and insight. *The Verge*. <https://www.theverge.com/2018/4/4/17190944/reddit-gaming-humor-memes-jokes>
- Hietalahti, J. (2016). Humor and disobedience : Understanding controversial humor. *Filosofiska Notiser*, 3(3), 23-44. http://filosofiskanotiser.com/Hietalahti16_3_1.pdf
- Horne, B. D., Adali, S., & Sikdar, S. (2017). Identifying the social signals that drive online discussions: A case study of Reddit communities. *Rensselaer Polytechnic Institute*. <https://doi.org/10.1109/ICCCN.2017.8038388>
- Humor Definition & Meaning—Merriam-Webster*. (n.d.). Retrieved March 23, 2023, from <https://www.merriam-webster.com/dictionary/humor>
- Jhaver, S., Birman, I., Gilbert, E., & Bruckman, A. (2019). Human-machine collaboration for content regulation: The case of Reddit automoderator. *ACM Transactions on Computer-Human Interaction*, 26(5), 1-35. <https://doi.org/10.1145/3338243>
- Lin, Y. (2022, October 3). *10 reddit statistics every marketer should know in 2023*. Oberlo. <https://www.oberlo.com/blog/reddit-statistics#:~:text=There%20are%20currently%20more%20than%202.8%20million%20subreddits,mornings%20on%20weekends%20and%20Mondays>
- Martin, R. A., & Ford, T. (2018). *The Psychology of Humor: An Integrative Approach*. Academic Press.
- Martin, R. A. (2007). *The psychology of humor: An integrative approach* (pp. xviii, 446). Elsevier.

- McGhee, P. E. (1979). *Humor: Its origin and development*. WH Freeman.
- Morreall, J. (1989). Enjoying incongruity. *De Gruyter Mouton*, 2(1), 1-18.
<https://doi.org/10.1515/humr.1989.2.1.1>
- Oppliger, P. A., & Zillmann, D. (1997). Disgust in humor: Its appeal to adolescents. *Humor*, 10(4), 421–437. <https://doi.org/10.1515/humr.1997.10.4.421>
- Park, R. (1977). A study of children's riddles using Piaget-derived definitions. *The Journal of Genetic Psychology*, 130(1), 57–67. <https://doi.org/10.1080/00221325.1977.10533230>
- Pennington, N., & Hall, J. A. (2014). An analysis of humor orientation on Facebook: A lens model approach. *Humor*, 27(1), 1–21. <https://doi.org/10.1515/humor-2013-0053>
- Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M. (2021). Studying Reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media + Society*, 7(2), 20563051211019004. <https://doi.org/10.1177/20563051211019004>
- Schmidt, S. R. (1994). Effects of humor on sentence memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 953–967.
doi:10.1037/0278-7393.20.4.953.
- Sinclair, R. C., Mark, M. M., & Clore, G. L. (1994). Mood-related persuasion depends on (mis)attributions. *Social Cognition*, 12, 309–326.
<https://doi.org/10.1521/soco.1994.12.4.309>
- Sternthal, B., & Craig, C. S. (1973). Humor in advertising. *Journal of Marketing*, 37, 12–18.
<https://doi.org/10.1177/002224297303700403>
- Van Atteveldt, W., Trilling, D., & Calderón, C. A. (2022). Chapter 10: Text as data. In *Computational Analysis of Communication*. Wiley Blackwell.
- van der Wal, A., Piotrowski, J. T., Fikkers, K. M., & Valkenburg, P. M. (2020). More than just a laughing matter: A coding framework of humor in media entertainment for tweens and Teens. *Journal of Broadcasting & Electronic Media*, 64(3), 478–498.
<https://doi.org/10.1080/08838151.2020.1796389>
- Zillmann, D. (1983). Disparagement humor. *Handbook of Humor Research*, 85–107.
https://doi.org/10.1007/978-1-4612-5572-7_5

Appendices

Appendix A - Python script for random post-URL generation on subreddits.

```
import praw
import random

# Define Reddit instance
reddit = praw.Reddit(
    client_id="please fill in your own credentials",
    client_secret="please fill in your own credentials",
    user_agent="please fill in your own credentials",
)

# Function: Get subreddits
def subreddit_variables(name):
    return reddit.subreddit(f"{name}")

# Function: Get SubReddit post URLs according to the RQ (sorted by top/one year), select a
# poolsize to determine the size of the pool to draw from:
def random_posts_collection(subreddit_list, poolsize):
    posts = {}
    for subreddit in subreddit_list:
        posts[f"{subreddit}"] = []
        for submission in subreddit.top(time_filter = 'year', limit = poolsize):
            posts[f"{subreddit}"].append("https://reddit.com/" + submission.permalink +
            "?sort=top" + "\n" +
            "https://reddit.com/" + submission.permalink +
            "?sort=controversial")
    return posts

# Function: Randomize URL selection front he pool
def select_posts(dictionary, number_of_posts):
    randomized_posts = {}
    random_numbers = []
    for subreddit in dictionary:
        randomized_posts[subreddit] = []
        for number in range(0, number_of_posts):
            random_numbers.append(random.randint(0, len(dictionary[subreddit])-1))
        for index in random_numbers:
            randomized_posts[subreddit].append(dictionary[subreddit][index])
    return randomized_posts

# Function: Export links into a .txt
def create_txt_file(sub_name, text_name, dictionary):
    with open(f"{text_name}, {sub_name}.txt", "w") as file:
        for post in dictionary[sub_name]:
            file.write(f"{post}" + "\n")
```

```
##### RUNNING THE FUNCTIONS:
# Define subreddits
funny = subreddit_variables("funny")
Gaming = subreddit_variables("Gaming")
Ask_Reddit = subreddit_variables("AskReddit")

# Create selection pool from selected Subreddits to draw from (100k)
collection = random_posts_collection([funny, Gaming, Ask_Reddit], 100000)

# Make a selection of 80 for each coder per Subreddit
I_G_selection = select_posts(collection, 80)
H_M_selection = select_posts(collection, 80)
Y_C_selection = select_posts(collection, 80)
A_M_selection = select_posts(collection, 80)
test_selection = select_posts(collection, 80)

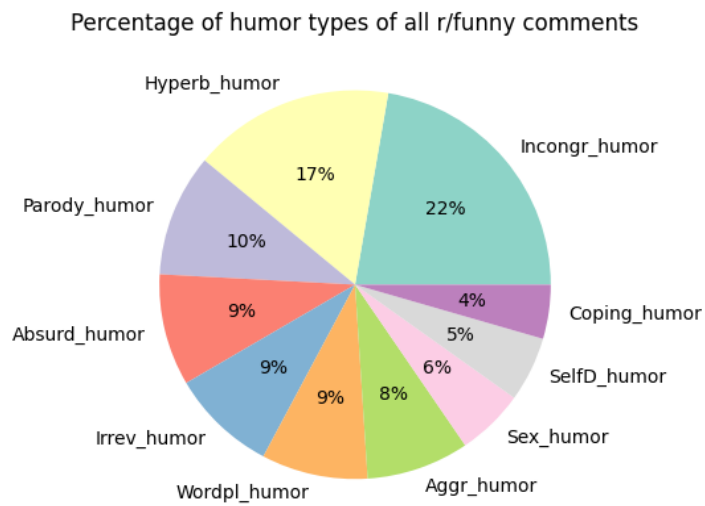
#RUN the txt function to generate a file with links for each subreddit, e.g.:
create_txt_file("AskReddit", "YuhuiAR", Y_C_selection)
```

Appendix B - Stopwords for the word clouds

```
["n't", "pppp", "roo", "oh", "yes", "something", "cid", "b", "someone", "guy", "guys", "f",
"and/or", "s", "m", "n't", "did", "like", "don't", "re", 'don't', "get", "didn't", "'ll", "\ll", "\ll",
"ll", "you'll", "you're", "i'm", "going", "go", "one", "even", "d", "ve", "know", "still", "it's",
"it's", "people", "would", "would ", " would", "people ", " people", "time", "times", "game",
"games"]
```

Appendix C - Additional Data Visualization

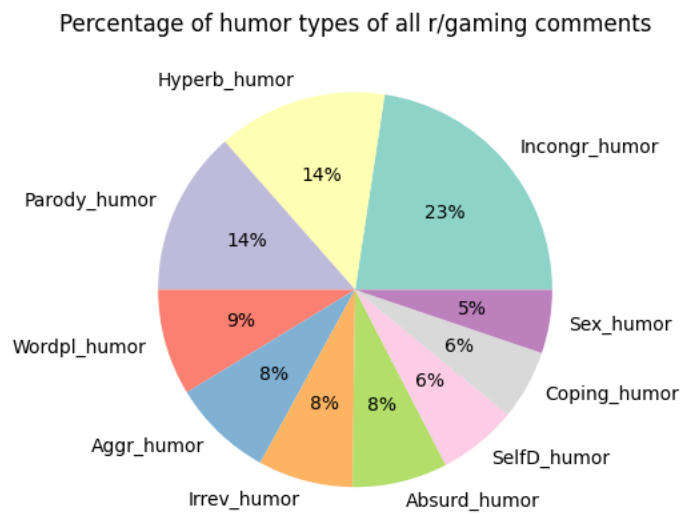
Pie Chart 1. Humor types in r/funny.



Pie Chart 2. Humor types in r/AskReddit.



Pie Chart 3. Humor types in r/gaming.



Bar Chart 1. Humor theories in top and controversial comments.

