

Humour appreciation in the subreddit r/wheredidthesodago: testing three hypotheses

Célia Schneebeli

Université Bourgogne Europe, France
celia.schneebeli@ube.fr

Abstract

In this study, I investigated users' appreciation of humour in the 361 posts published in 2018 on the subreddit r/wheredidthesodago, which is devoted to posting GIFs extracted from infomercials with funny captions. Specifically, I examined both formal and content-related features that might influence humour appreciation. The first feature I analysed was purely formal. I tested the correlation between post format—caption length, GIF duration and number of frames per second—and post score. The second feature I examined was post structure. Drawing upon the broader incongruity theory of humour, I tested the relationship between the success of a post and the presence (or absence) of an incongruous association between the GIF and the caption. The third feature I investigated was content-based, drawing on the Benign Violation Theory of McGraw and Warren (2010, 2014). I tested whether posts addressing taboo topics achieved higher scores than those that did not. While I found no correlation between the format of posts and their success, I found that posts featuring wholly incongruous GIF/caption pairings and posts broaching taboo topics tended to achieve higher scores. However, further analysis of semi-identical posts revealed that these tendencies did not always hold true. These findings suggest that incongruity and benign violation are important components of humour appreciation in the posts I analysed, but they are neither necessary nor sufficient for humour to succeed.

Keywords: humour appreciation, incongruity, benign violation theory, Reddit, online humour.

1. Introduction

Elliot T. Panek describes Reddit as a “vast ecosystem of online communities hosting the largest conversations in the history of humanity” (Panek, 2022, p. 9). Its defining characteristic, which sets it apart from other popular social media platforms, is its modularity. As Panek explains, from its inception, Reddit has functioned as a content sorter, where users collectively sort posts through upvotes (which push content higher) and downvotes (which lower its visibility). But over time, it has evolved into a large network of online communities, divided into topic-based forums called “subreddits”. On these subreddits, all of which have their own rules and

functionalities set by their moderators, registered users post, vote, and discuss content related to certain specific topics.

r/wheredidthesodago is one such subreddit, which is dedicated to posts that repurpose infomercials in a humorous way. It was created in October 2012, after a contributor posted an infomercial with a funny caption on the subreddit r/funny. Another contributor commented that they would subscribe to a subreddit devoted to taking infomercials out of context. A redditor subsequently obliged and created the subreddit r/wheredidthesodago. They named it after the subreddit's first post, which featured a clip from an infomercial where a man's soda bottle mysteriously disappeared between two shots, accompanied by the caption "Where did the soda go?":

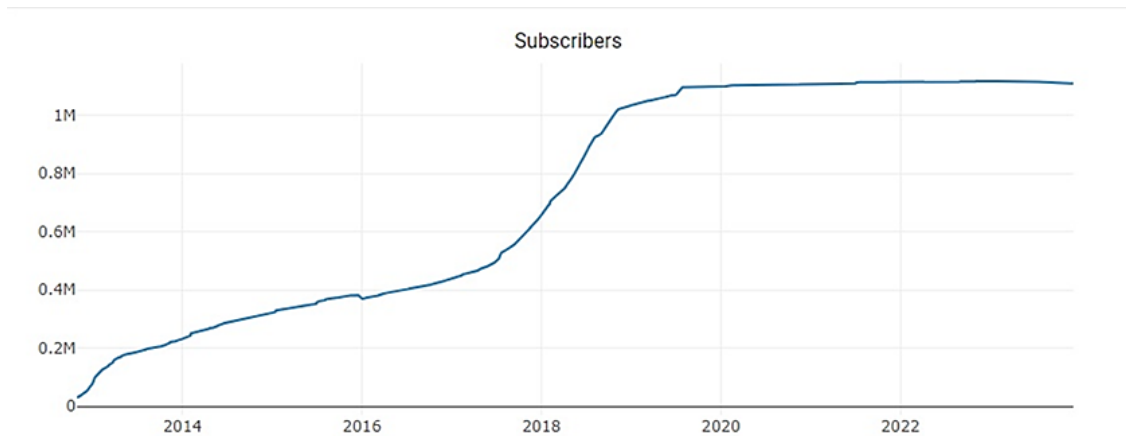


Figure 1. First post on r/wheredidthesodago

The subreddit's description is the following: "Infomercials out of context. A subreddit for infomercials taken OUT OF CONTEXT in animated images. This subreddit is for ads that show real aspects of life but in a completely unrealistic way." However, while posts always have to stick to showing extracts from infomercials, the reinterpretation is not always unrealistic, and captions really range from witty comments on what the GIF shows to complete re-contextualization of its content. The main objective of r/wheredidthesodago, as rule number 7 of the subreddit explicitly claims, is simply "to have fun".

The success of r/wheredidthesodago has been fairly rapid. It had already reached the 200.000 subscriber milestone in August 2013, nine months after its creation. In its first year of existence, media as diverse as Vice, ABC News or Yahoo had already published articles about it. According to subreddistats.com, a website dedicated to collecting subreddit statistics, r/wheredidthesodago's popularity particularly soared between the end of 2015 and the end of 2018, a period during which the number of subscribers almost tripled:

Table 1. Evolution of number of subscribers (source: subredditstats.com)



From 2019 onwards, the number of subscribers began to stagnate, remaining around 1.1 million ever since. Only twenty posts have been published in the last two years (2023 and 2024) and the subreddit is now almost inactive. However, r/wheredidthesodago is still one of the most popular subreddits. At the end of 2024, it was ranked #686 in Reddit’s largest communities ranking, placing it in the top 1%.

2. Data collection

Because of its popularity, its explicitly humorous intent, and the fact that it gives direct access to users’ appreciation of posts through its vote system, r/wheredidthesodago appears to be a good choice for examining humour appreciation, which is what the study presented in this paper was aimed at. The study more precisely explores users’ appreciation of all posts published on the subreddit in 2018 in order to test three hypotheses likely to explain what makes the humour of a post successful. I chose to sample data from 2018 because, according to subredditstats.com, it is one of the subreddit’s peak years of activity and popularity. It is also the year during which the one million subscriber milestone was reached and the year in which four of the five most liked posts in the subreddit’s history were posted. As a result, the posts published in 2018 are numerous enough and have been seen, voted, and commented on by enough users for the study to be relevant.

I used ScreenToGif to save the 361 posts published during that year. ScreenToGif is a screen recorder that allows users to record a selected area of their screen and save it as an animated GIF. It allowed me to save each post in its environment with the complete animated GIF and its score and number of comments. The score is what I used to measure humour appreciation, defined as the “subjective, psychological response characterized by amusement, the tendency to laugh, and the perception that something is funny” (Warren et al., 2021, p. 42). On Reddit, the score of a post is determined by the number of upvotes and downvotes it receives. Users upvote posts they find worth seeing or entertaining and downvote those they consider off-topic or not interesting. The score of a post is therefore simply the number of upvotes minus the number of downvotes, though it can never go below zero. As a result, a higher score indicates a more successful post. In the context of a conversation, upvotes and downvotes also allow users to evaluate other users’ contribution to the topic (a downvote meaning the contribution is judged to be off-topic or not constructive) or whether they agree or not with other users’ contributions. In the context of r/wheredidthesodago, an upvote arguably means the post has achieved the goal of the page, i.e., to take an infomercial out of context and turn it into something funny. This is why I chose to use scores as a measure of humour appreciation.

As Reilly (2021, p. 15) notes, “a literal interpretation of [the] Terms of Service would suggest that users should have no expectation of privacy when posting content” on public social media. This applies to r/wheredidthesodago, a public subreddit visible to everyone. However, as Reilly further emphasizes, even when data is publicly available, it is necessary for researchers to systematically engage in ethical reflection before publishing it. Following the ethical guidelines of the Association of Internet Researchers concerning online data sampling (Franzke et al., 2020), I assessed whether publishing posts for illustrative purposes could potentially harm their authors, and whether verbatim publication was necessary. Regarding the latter, paraphrasing posts would be impractical, as the posts contain images and because the study relies on both their content and form. Publishing them verbatim does mean that posters’ identities could be traced using a search engine, even if anonymized. However, because all posters use anonymous usernames, real-life identification remains difficult. More importantly, the collected posts are neither controversial nor sensitive. While they occasionally touch on taboo subjects, they do so within the strict boundaries of the community’s rules. They contain no problematic statements (such as incitement to violence, hate speech, or ad hominem attacks) that might expose posters to potential harm. As a result, I have decided to publish the posts verbatim, without anonymizing them.

Once the corpus was built, I set out to determine what features to test in order to evaluate potential factors influencing humour appreciation. In line with Weber et al. (2014) and Heintz (2020), who underline the importance of separating structure and content when investigating humour appreciation, I decided to examine both formal features and content features.

The first feature I chose to look at is the format of the post, more precisely the length of the caption, the GIF’s duration and its number of frames per second. The second feature I decided to examine is the structure of the post, focusing on the presence (or absence) of an incongruous association between the GIF and the caption. The third and last feature I chose to investigate is purely content-related, focusing on whether the post broaches a taboo subject or not. For each feature, I formulated a hypothesis and developed a precise methodology to determine whether the data supported or refuted the hypothesis.

In the next section, I outline the three hypotheses I tested, along with the theoretical framework and prior research that informed them. Section 4 details the methodology I used to test the hypotheses. Section 5 finally presents the results of the tests, which are then discussed in section 6.

3. Presenting the three hypotheses

As Heintz et al. (2023) point out, “[a] large host of theories and models [...] have been proposed to explain [...] the processes underlying humour appreciation” (p. 245) and a wide range of stimuli has been explored. Building on this foundation, I selected three potential “driving force[s] in the creation of humor” (Veale, 2004, p. 420) to test, each relating to a different aspect: form, structure, and content. They were also selected because they lend themselves to empirical examination through observable variables. All three hypotheses indeed imply a possible correlation between certain variables and humour appreciation. Even though correlation does not imply causation, the results were expected to provide meaningful insights into possible relationships between the variables under study and the success or failure of the humour of a post.

3.1. Hypothesis 1: correlation between the format of the post and humour appreciation

My first hypothesis was that there is a correlation between the format of a post, specifically the length of the caption, the GIF’s duration and its number of frames per second, on the one hand,

and the post's success, on the other. As Attardo (2023, p. 28) notes, "the pace of humor has changed" and "Internet humor is faster" and "tends to favor shorter forms of expression". In the online culture of TLDR (Too Long Didn't Read), which favours short, rapidly skimmed through content, and in an era where human attention span is supposed to have been decreasing, particularly online, a logical postulate would be that shorter captions or GIF durations are likely to enhance a post's success. In this case, posts with concise captions or brief GIFs should achieve higher scores on average.

However, in other contexts, scholars have found diverging results when exploring the correlation between the length of a joke and its success. For instance, Dunbar et al. (2016), in their study of 101 jokes labelled the "funniest of all time", observed that "humor rating increases exponentially with word length" (p. 136). They concluded that "the word count of a joke" may be "a significant predictor of its quality" (p. 139). Conversely, Shahaf et al. (2015), in a study of humorous cartoon captions, found that "funnier captions actually use fewer joke words" (p. 1071). One possible explanation for these conflicting findings lies in the difference in format. In Shahaf et al.'s study, humour is conveyed through both the caption and the image, whereas purely verbal jokes, such as those examined by Dunbar et al., rely solely on words, which might necessitate a higher word count. Similarly, Hossain et al. (2019), in their study of funnily edited headlines on Reddit—which also rely solely on words—found that "short headlines did not lend themselves to high humor scores, while longer headlines generally had more potential for humor" (p. 137). They attributed this to the additional context needed to make the humour effective. To make it short, prior research has identified correlations going in both directions (shorter being better, longer being better, within the limits of human cognition), depending on the type and form of humour (verbal, visual, verbo-visual). Given that our dataset is verbo-visual, similar to Shahaf et al.'s, and that humour likewise emerges from both the caption and the GIF, we might expect posts with concise captions to be more successful overall.

As far as GIFs are concerned, research is scarcer than on verbal jokes and humorous captions, particularly regarding the correlation between their duration and success, as well as their frame rate and success. Bakshi et al. (2015) examined factors contributing to GIF engagement in a corpus of Tumblr posts and found that "every added second of longer playback decreases chances of reblogs by 16% and likes by 20%, suggesting that shorter content is more engaging on average" (p. 581). In other words, in their study, posts featuring shorter GIFs tended to be more successful. Additionally, they observed that GIFs with a higher frame rate, which means GIFs with a greater number of images per second, received more likes compared to those with a lower frame rate. Based on these findings, we can expect posts featuring shorter GIFs and GIFs with higher frame rates to achieve greater success overall.

3.2. Hypothesis 2: correlation between the incongruity of the post and humour appreciation

My second hypothesis concerned the structure of the posts, structure being understood as "a cognitive stimulus property" that "refers to how the joke [in this case the humorous post] 'works' (i.e., the mechanisms underlying the punchline)" (Heintz, 2020, p.37). This hypothesis postulated that posts featuring an incongruous association between the GIF and the caption tend to achieve greater success. It draws upon the broader incongruity theory of humour, succinctly described by Morreall (2024) as follows: "the cause of laughter [...] is the perception of something incongruous—something that violates our mental patterns and expectations"

Incongruity lies at the core of many theories of humour, from Beattie's introduction of the concept in 1779 to more modern versions of the incongruity theory focusing on the resolution of incongruity (Shultz 1976; Suls 1972, 1983). According to Warren et al. (2020), it is central to one of the three most important groups of humour theories. It is even sometimes presented as

the most consensual of humour theories. For example, Richie (1999) observes that “[i]f there is one generalisation that can be extracted from the literature about humour, it is that humour involves incongruity” (p.78). Similarly, Veale (2004) writes that “[o]f the few sweeping generalizations one can make about humor that are neither controversial or trivially false, one is surely that humor is a phenomenon that relies on incongruity” (p. 419). Raskin’s script-based semantic theory of humour (SSTH, Raskin, 1985) which introduces the neighbouring notion of script opposition, and its later revision by Raskin and Attardo, the general theory of verbal humour (GTVH, Attardo & Raskin, 1991), are also both based around the notions of incongruity or opposition as Veale (2004, p. 419) points out.

I chose incongruity as my second hypothesis not only because of its widespread popularity but also because it can be tested empirically. In fact, incongruity is one of the three basic humour stimuli assessed in the 3WD humour test (Ruch, 1992), one of the most widely used humour appreciation tests. The concept of incongruity is also particularly relevant when examining posts published on a page which is explicitly aimed at taking infomercials out of context, i.e., giving them an unexpected twist. More broadly, scholars working on humour online (Dynel, 2016; Chovanec & Tsakona, 2018; Vásquez, 2019; Yus, 2021, 2023) have often noted how central to digital humour “the notions of incongruity and script opposition” (Vásquez, 2019, p. 29) are. We might therefore expect incongruity to play a key role in users’ appreciation of the posts published on r/wheredidthesodago.

3.3. Hypothesis 3: correlation between the presence of a benign violation and humour appreciation

My third and last hypothesis was based on the content of the posts. It postulated that posts which broach a taboo topic tend to achieve higher scores than posts which do not. It is based on the Benign Violation Theory (abbreviated below as BVT) of McGraw and Warren (2010, 2014).

The BVT is a comprehensive theory of humour that claims that “laughter and amusement result from violations that are simultaneously seen as benign” (2010, p. 1141). It aims at reconciling what is presented as “three conditions that facilitate humor” (2010, p. 1142): a perceived violation, a “safe, playful, nonserious” context, and “that two contradictory ideas about the same situation be held simultaneously [...]” (2010, p. 1142). To achieve this, the BVT begins with the premise that humour requires some form of violation (linguistic, social, moral or otherwise) to succeed. However, while such violations might initially elicit negative emotions, they can also evoke humour if simultaneously appraised as benign, non-threatening, or acceptable. The essence of the BVT therefore lies in reconciling the negative valence of the violation required for humour to arise with the simultaneous amusement elicited by this violation.

Jokes violating taboos are good examples of humorous violations. Broaching a taboo subject inherently involves a form of transgression. Indeed, the term “taboo” refers to “the interdiction of the use or practice of anything, especially an expression or topic, considered offensive and therefore avoided or prohibited by social custom” because it is “perceived as in some way harmful to an individual or to his/her community” (Allan & Burridge, 2006, p. 11). By definition, taboo topics are therefore not meant to be addressed. However, when framed within the context of a page whose primary aim is entertainment, broaching a taboo becomes more benign and acceptable. This is especially true for a page that is moderated and where all posts are reviewed before being published, as is the case with r/wheredidthesodago. This makes posts that deal with taboo subjects an ideal instance to test a potential correlation between the presence of a benign violation and humour appreciation.

In the same way as the presence of an incongruous association, I chose the presence of a taboo subject as the object of my third hypothesis because it can be tested empirically too. In

fact, “sexual humor”, which relies on taboo content, is also one of the three basic humour stimuli assessed in the 3WD humour test.

Of course, this third hypothesis echoes the preceding one since the BVT originates, in part, from the incongruity theory. Broaching a taboo is never expected and does violate our mental patterns and expectations. However, it operates not on a cognitive level but on an emotional one.

4. Testing the three hypotheses: methodology

As I have previously mentioned, each hypothesis was tested by verifying the existence of a correlation between selected variables and the score of the posts. While correlation alone does not prove the existence of a causal relationship, it can suggest that the causal relationship proposed by the theoretical framework and previous research informing each hypothesis does indeed exist. Such a method also has the advantage of being grounded in empirical observation rather than relying solely on interpretation.

4.1. Methodology for hypothesis 1 (format of the post)

Verifying hypothesis 1 involves investigating potential correlations between the length of the caption, the GIF’s duration and / or its frame rate on the one hand, and the post’s success on the other hand. To achieve that, I collected the following data for each post: score, number of words in the caption, GIF duration and GIF frame rate. I then applied Pearson’s correlation coefficient formula to measure the relationship between two pairs of variables in each instance.

Pearson’s correlation coefficient is a widely used statistical method for assessing the strength of the correlation between paired continuous variables that are not related. In this case, I used it to evaluate the following variable pairs: the number of words in the caption and the post’s score, the duration of the GIF and the post’s score, and the frame rate of the GIF and the post’s score. The closer the result is to 1, the stronger the correlation between the two variables.

4.2. Methodology for hypothesis 2 (incongruous association)

Some scholars have already attempted to test script opposition and incongruity in joke or meme corpora. For example, Tanaka et al. (2022) have worked on devising different computational processes to evaluate humour in memes, which involves having computers detect incongruity between an image and a caption in a meme and evaluate humour. In their study of funnily edited headlines on Reddit, Hossain et al. (2019) also tested the relationship between incongruity and humour appreciation. They quantified incongruity using GloVe, a model for distributed word representation, which they used to examine the relationship between the replacement words in the humorous headline and the words in the original headline. They operated under the assumption that greater semantic distance between the original and replacement words would indicate greater incongruity. They then tested whether greater semantic distances were associated with higher humour ratings. Their findings revealed “a relatively high correlation between humor scores and the cosine distance between the added word [...] and the replaced word [...]”, which “suggests that incongruity leads to humor” (p. 138). Unfortunately, neither experiment provides a model that could be directly replicated for the dataset in this study. Therefore, I devised an alternative empirical method, focusing on the relationship between the lexical content of the caption and the visual content of the GIF and how well they align. To carry this out, I manually reviewed all posts and categorized them into the following three groups:

-“0” is the category for posts where the caption aligns with the GIF. In this first category, all the content words in the caption correspond to elements featured in the GIF. This means that the caption adds little to no information or interpretation beyond what is already present in the original infomercial. The following post is an example of a post belonging to the “0” category:



Figure 2. Category “0” post

In this post, the caption describes exactly what the GIF shows: kids getting overly excited about a heated ice-cream scoop.

-“1” is the category for posts where the caption partially aligns with the GIF. In this second category, only some of the content words in the caption correspond to elements present in the GIF. As a result, the caption and the GIF appear to be only partially related, as is the case in this post:



Figure 3. Category “1” post

The girls in the GIF are indeed playing with a broom, but the GIF does not depict a tough round of quidditch (a fictional sport in the Harry Potter universe that involves magical balls and flying broomsticks). As a result, some words in the caption correspond to what the GIF portrays but others do not correspond to anything shown in the GIF, although they do elaborate on what it shows (quidditch is indeed played with flying broomsticks).

-“2” is the category for posts where caption does not align with the GIF. In this last category, none of the words in the caption correspond to any element featured in the GIF. The following post serves as a clear example of this category:



Figure 3. Category “2” post

In this post, the GIF does not depict anything mentioned in the caption, be it public urination or the Stay-Puft Marshmallow Man from *Ghostbusters*. Instead, the caption offers a complete reinterpretation and recontextualization of the GIF, which shows foam cleaner being sprayed and rinsed onto various surfaces.

It is important to note that some posts have been classified in the “2” category even when the caption involves the same agents or participants as those featured in the GIF, as long as the action described in the caption has no explicit connection to what the GIF depicts. For example, the quidditch post mentioned earlier falls into category “1”, rather than “2”, because while the girls are not playing quidditch, playing quidditch does require a broom. On the contrary, in the following post, classified as “2”, the caption and GIF feature the same agents, but the action described in the caption is entirely unrelated to what the GIF actually shows:



Figure 4. Category “2” post

The GIF does depict smiling children, but it does not show them being slowly sacrificed and consumed by their alien hosts. This description bears no connection to what the children are really doing in the infomercial, which is reading or playing in a Snuggie tail blanket. Additionally, there are no aliens present in the GIF. Therefore, this post aligns more closely with category “2” than with category “1”. In fact, most GIFs feature humans—women, men and children—who are generally referred to as such in the captions (or not mentioned at all). Recontextualising infomercials typically relies on reinterpreting the context, actions, states, or sequence of actions depicted in the GIF than rather than the individuals or animals featured.

I interpreted these three categories as representing three levels of incongruity. Posts in the “0” category, where caption and GIF align, were assumed to exhibit the lowest level of incongruity. In contrast, posts in the “1” category, where caption and GIF only partially align, were considered moderately incongruous. Finally, posts in the “2” category, where caption and GIF do not align, were assumed to display the highest level of incongruity. Using this framework, I calculated the mean score for each category to determine whether posts that rely on more unexpected associations between the caption and the GIF tend to receive higher ratings

4.3. Methodology for hypothesis 3 (benign violation)

To test the third and final hypothesis, which postulated that posts addressing taboo topics tend to receive higher scores than those that do not, I manually reviewed the 361 posts again, classifying them based on whether they broach a taboo subject or not. I classified as broaching a taboo all posts addressing the following topics, based on Allan & Burridge (2006):

- sex and all forms of sexual and erotic practices
- death, including murder and suicide
- disability and illness, physical or mental
- bodily functions, body effluvia, and excretion

- discriminatory behaviours in general (sexism, racism, homophobia, fat-shaming, age-shaming)
- addiction (smoking, drugs, alcohol)
- violence and aggression (sexual harassment, mugging, child and animal abuse or neglect)
- religion

I also looked for posts whose caption included dysphemistic language, which includes “curses, name-calling, and any sort of derogatory comment directed towards others in order to insult or to wound them” (Allan & Burrige, 2006, p.31). Using such language indeed amounts to violating a linguistic taboo.

To determine whether taboo posts obtained higher scores on average, I then calculated the mean score for posts broaching a taboo, posts using dysphemistic language, and posts doing neither. Before presenting the results, two clarifications are necessary.

First, I chose to consider all posts tackling taboo subjects collectively, rather than examining each type of taboo individually, as some subjects were represented by too few posts to yield statistically significant findings. For example, there are only two posts joking about religion, and five posts addressing disease or disability. The most frequently broached taboo subject is sex, appearing in 43 posts and accounting for more than a third of the posts broaching a taboo subject, followed by addictions (19 posts) and bodily functions (16 posts). However, I treated posts containing dysphemistic language separately, despite their limited number, because in this case the taboo lies in the language itself rather than the subject matter.

Secondly, since the GIFs in the dataset are all extracted from perfectly innocuous and bland infomercials, they never show anything particularly taboo, offensive, or shocking. However, the interpretation derived in conjunction with the caption, or the caption alone, sometimes relates to a taboo topic, as illustrated in the following post:



Figure 5. Post broaching a taboo topic

While the infomercial simply shows a woman applying a charcoal mask on her face, the caption reframes it as her engaging in blackface in the privacy of her home, a practice that is widely

considered taboo due to its offensive and racist implications. This is why the central stimulus of this third hypothesis, as previously mentioned, also relates to script opposition or incongruity.

5. Results

I will now examine the results for each hypothesis, before discussing them in the next section.

5.1. Results for hypothesis 1 (format of the post)

The correlation coefficients for the three pairs of variables I tested are as follows:

Table 2. Pearson's coefficient for each tested pair of variables

Pair of variables	Pearson's correlation coefficient
Number of words in caption / score of post	0,062729999
GIF duration / score of post	0,245217709
GIF frame rate / score of post	0,230213255

To interpret the results, I used the following scale for Pearson's correlation coefficient ("r") provided by Whitford (2005, p. 525):

- little to any correlation ($0.00 \leq |r| \leq 0.30$)
- low correlation ($0.30 \leq |r| \leq 0.49$)
- moderate correlation ($0.50 \leq |r| \leq 0.69$)
- high correlation ($0.70 \leq |r| \leq 0.89$)
- very high correlation ($0.90 \leq |r| \leq 1.00$)

Based on this ordinal scale for the strength of correlation, the three coefficients obtained indicate no meaningful relationship between the different pairs of variables. In the case of the correlation between the number of words in the caption and the score of the post, the coefficient is very close to zero, suggesting that the relationship between the two variables is virtually non-existent. The coefficient is slightly higher for the correlation between the duration of the GIF and the post score, but it still falls short of even constituting a low correlation. Similarly, the coefficient for the correlation between the frame rate of the GIF and post score is very close to the previous one and also insufficient to meet the threshold for a low correlation on Whitford's scale.

It can therefore be concluded that the study revealed no correlation between the number of words in the caption of a post, the duration of the GIF or its frame rate, and the success of the post. This indicates that, contrary to my expectations and the findings of the studies referenced earlier, there is no evidence to suggest that the success of a post's humour depends on its format, nor that its format has any influence on its score.

5.2. Results for hypothesis 2 (incongruous association)

Testing the second hypothesis involved examining the correlation between the incongruity of the caption / GIF pairing and the success of the post. Below are the mean scores for each category of posts, from least to most incongruous:

Table 3. Mean score for each level of incongruity

Category of post	Mean score
All posts	9044,487535
0 (caption aligning with GIF)	7437,478261
1 (caption partially aligning with GIF)	8932,704819
2 (caption not aligning with GIF)	11522,86076

The mean score for posts in the “0” category is 7437. Compared to the mean score for the entire dataset (all categories combined), posts where the caption adds little to no elements to the original infomercial score, on average, about 18% lower.

The mean score for posts in the “1” category is 8933. It is higher than the mean score for the “0” category and very close to the overall dataset mean score—just slightly over 1% lower.

The mean score for posts in the “2” category is 11523. This is nearly 55% higher than the score for the “0” category”, 26% higher than the score for the “1” category, and a little over 27% higher than the mean score for all posts in the dataset.

These results lead to the following conclusions: first, posts that feature an incongruous association between the caption and the GIF tend to achieve higher scores than other posts. Second, posts featuring a mildly incongruous pairing tend to perform better than those with a fully congruent pairing, the latter scoring significantly below average. This suggests that the greater the incongruity in a post, the more successful it is likely to be.

However, these results need to be put into perspective in the light of a phenomenon that emerges on closer examination of the posts. Indeed, 17 GIFs are used in more than one post within the dataset, as exemplified by this pair of posts:



Figure 6. Semi-identical posts

Posts which feature the same GIF form sets of two to four semi-identical posts, characterized by identical GIF but different captions. These sets of posts allowed me to compare different caption / GIF associations and assess whether, as anticipated, the posts with the most unexpected pairings were systematically the most successful. The results are as follows:

Table 4. Post score / incongruity in semi-identical posts

Type of set (score / incongruity)	Number of sets	Proportion (% of semi-identical sets)
All posts in the same category (non-significant)	5	29,4%
Most incongruous post = most successful	6	35,3%
Least incongruous post = most successful	6	35,3%

The post with the most incongruous caption / GIF pairing is the most successful within its set in only 35% of cases, the same proportion as cases in which the post with the least incongruous pairing is the most successful. This clearly indicates that the most successful post among semi-identical posts is not systematically the one where the association between the GIF and the caption is the least expected.

Combining the results of both tests suggests that, while there is a general tendency toward incongruous posts to achieve higher scores, this trend does not always hold and cannot be considered the sole factor contributing to the success of a post.

5.3. Results for hypothesis 3 (benign violation)

The final stimulus tested was the presence of a benign violation, specifically in the form of a taboo subject being broached. The test aimed to determine whether posts displaying such violations achieved higher scores than those that did not. The results are as follows:

Table 5. Mean score for taboo and non-taboo posts

Category of post	Mean score
Posts broaching a taboo subject	11408,95763
Posts not broaching a taboo subject	7831,84898
Posts with dysphemistic language	20940

The data reveals a clear tendency for posts addressing taboo topics to be more appreciated than those that do not. What is more, when considering the average score for all posts in the corpus (9044), posts not broaching a taboo subject score about 13% lower on average, while those broaching a taboo topic score approximately 26% higher than the mean. Posts containing dysphemistic language, though few (likely due to moderation), are particularly successful: the

mean score of the 10 posts featuring dysphemistic language is over 131% higher than the mean score for the entire dataset.

However, an analysis of semi-identical posts, posts with identical GIFs but different captions, once again demonstrates that this is merely a tendency, not a strict rule:

Table 6. Post score / presence of taboo in semi-identical posts

Types of set (score / presence of taboo)	Number of sets	Proportion (% of semi-identical sets)
All posts in the same category (non-significant)	8	47%
Taboo post = most successful	4	23,6%
Non-taboo post = most successful	5	29,4%

The post that broaches a taboo subject is the most successful within its set in only 23,6% of cases, a proportion that is actually lower than the 29,4% of sets where the post not addressing a taboo subject is the most successful. Thus, the most successful post among semi-identical posts is far from consistently being the one that addresses a taboo subject.

Similar to the results observed when testing incongruity, combining the findings of both tests leads to the conclusion that, while there is a general tendency for posts addressing taboo subjects to achieve higher scores, this trend does not always hold true. Like incongruous caption / GIF pairings, it cannot be considered the sole factor contributing to a post's success either.

6. Discussion and conclusion

My research question was to explore potential factors influencing humour appreciation in a corpus of digital posts and to test three hypotheses relative to the format, structure, and content of these posts. My first hypothesis, which focused on the format of posts, was proven incorrect. Contrary to my expectations and the findings of other studies, I found no relationship between caption length, GIF duration, or frame rate, and the score of a post. In contrast, my second and third hypotheses, which addressed the structure and content of posts, yielded more conclusive results, although they were not definitive. On average, posts featuring wholly incongruous GIF/caption pairings and broaching taboo topics tended to achieve higher scores. However, an analysis of semi-identical posts in both cases revealed that these tendencies required nuance, as they did not always hold true.

So, while incongruity and benign violation appear to be critical components of humour appreciation in the posts I analysed, they are certainly neither necessary nor sufficient for humour to succeed, despite sometimes being presented as comprehensive theories of humour. As Ruch and Hehl (1998) write, “[t]he process of perceiving, processing, and responding to a humorous stimulus is complex and involves many areas of psychic functioning, such as cognition, emotion, and motivation.” A “more comprehensive account of humor appreciation” has to consider “many factors relating to the humorous message, the sender, the receiver, their relationship, physical and social factors of the situation, etc” (p. 110). The factors tested in this study are message-oriented, focusing solely on how the humorous message, in this case a digital

post, is built, structured, and what it is composed of, rather than on receiver- or context-oriented factors.

One other of the three main groups of humour theories, the superiority theory, could provide an interesting and complementary perspective on humour appreciation in r/wheredidthesodago. The superiority theory, which traces back to Plato, Aristotle, and Hobbes' Leviathan, is succinctly summarized by Morreall (2024) as follows: “[O]ur laughter expresses feelings of superiority over other people or over a former state of ourselves”. In other words, humour is rooted in ridicule, where we see ourselves as superior to the object of amusement. Although the superiority theory has been heavily criticized since the 18th century and cannot be treated as a comprehensive theory of humour, as Lintott (2016) explains, it “can still play an important role in humor theory” (p. 356). This theory is particularly relevant to r/wheredidthesodago, where taking infomercials out of context also often amounts to mocking them. For example, it could help explain why a post such as the heated ice-cream scoop post (see 3.2) is so successful. As we observed, this post features no unexpected caption/GIF pairing and does not address a taboo subject, yet its score (11000) is still significantly higher average—more than 20% above the mean score. Perhaps this success stems from the implication that it is ridiculous to get overly excited over a heated ice cream scoop or from the overexaggerated reactions of the children in the GIF. In other words, users may find the post funny because they feel superior to the individuals shown in the GIF, or those who created the infomercial. There are many other posts in the dataset that similarly mock or ridicule the actions depicted in the infomercial, such as stacking two plates on top of each other in a microwave or being annoyed after accidentally spilling paint in an area that is going to be painted anyway. However, this aspect of humour, rooted in superiority and ridicule, is challenging to annotate and quantify as it depends largely on affective response rather than the measurable elements of content, structure or format. A potential area for future research could involve developing observable variables to make the concepts of superiority and ridicule compatible with empirical examination. This could form the basis for a fourth hypothesis, allowing us to achieve a more comprehensive understanding of humorous stimuli in r/wheredidthesodago and other accounts, groups and sites dedicated to creating humour by recontextualizing content.

References

- Allan, K., & Burrige, K. (2006). *Forbidden words: Taboos and the censoring of language*. Cambridge University Press.
- Attardo, S., & Raskin, V. (1991). Script theory revis(it)ed: Joke similarity and joke representation model. *Humor. International Journal of Humor Research*, 4(3-4), 293–347. <https://doi.org/10.1515/humr.1991.4.3-4.293>
- Attardo, S. (2023). *Humor 2.0: How the internet changed humor*. Anthem Press.
- Bakhshi, S., Shamma, D.A., Kennedy, L., Song, Y., De Juan, P., & Kaye, J. (2016). Fast, cheap, and good: Why animated GIFs engage us. *CHI'16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 575-586. <https://doi.org/10.1145/2858036.285853>
- Beattie, J. (1779). Essay on laughter and ludicrous composition. In *Essays*, 3rd ed., London.
- Chovanec, J., & Tsakona, V. (2018). Investigating the dynamics of humor: Towards a theory of interactional humor. In V. Tsakona & J. Chovanec (Eds.), *The dynamics of interactional humor* (pp. 1-28). John Benjamins. <https://doi.org/10.1075/thr.7.01cho>
- Dunbar, R.I.M., Launay, J., & Curry, O. (2016). The complexity of jokes is limited by cognitive constraints on mentalizing. *Human Nature*, 27(2), 130-40. <https://doi.org/10.1007/s12110-015-9251-6>.

- Dynel, M. (2016). "I has seen image macros!": Advice animal memes as visual-verbal jokes. *International Journal of Communication*, 10, 660-668.
- Franzke, A. S., Bechman, A., Zimmer, M., C. Ess, & the Association of Internet Researchers (2020). *Internet Research: Ethical Guidelines 3.0*. <https://aoir.org/reports/ethics3.pdf>
- Heintz, S. (2020). Separating content and structure in humor appreciation. A bimodal structural equation modelling approach. *Journal of Individual Differences*, 41(1), 37-44. <https://doi.org/10.1027/1614-0001/a000301>
- Heintz, S., Forabosco, G., Dionigi, A., & Cioni, F. (2023). Humor comprehension and appreciation: an analysis of Italian jokes. *Humor: International Journal of Humor Research*, 36(2), 245-262. <https://doi.org/10.1515>
- Hossain, N., Krumm, J., & Gamon, M. (2019). President vows to cut taxes hair: Dataset and analysis of creative text editing for humorous headlines. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, ACL, 133-142. <https://doi.org/10.18653/v1/N19-1012>
- Lintott, S. (2016). Superiority in humor theory. *The Journal of Aesthetics and Art Criticism*, 74(4), 347-358. <https://doi.org/10.1111/jaac.12321>
- McGraw, P., & Warren, C. (2010). Benign violations: Making immoral behavior funny. *Psychological Science*, 21(8), 1141-1149. <https://doi.org/10.1177/0956797610376073>
- McGraw, P., & Warren, C. (2014). Benign violation theory. In S. Attardo (Ed.), *Encyclopedia of Humor Studies* (pp. 75-77). Sage Publications.
- Morreall, J. (2024). Philosophy of humor. In E.N. Zalta & U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy*. Retrieved December 5, 2024, from <https://plato.stanford.edu/archives/fall2024/entries/humor>
- Panek, E. (2022). *Understanding Reddit*. Routledge.
- Raskin, V. (1985). *Semantic mechanisms of humor*. Reidel.
- Reilly, P. (2021). *Digital contention in a divided society. Social media, parades and protests in Northern Ireland*. Manchester University Press.
- Ritchie, G. (1999). Developing the incongruity-resolution theory. *Proceedings of the AISB Symposium on Creative Language: Stories and Humour* (pp. 78-85). The Society for AISB.
- Ruch, W. (1992). Assessment of appreciation of humor: Studies with the 3 WD Humor Test. In C. D. Spielberger & J. N. Butcher (Eds.), *Advances in personality assessment*, 9 (pp. 27-75). Lawrence Erlbaum Associates.
- Ruch, W., & Hehl, F.-J. (1998). A two-mode model of humor appreciation: Its relation to aesthetic appreciation and simplicity-complexity of personality. In W. Ruch (Ed.), *The sense of humor: Explorations of a personality characteristic* (pp. 109-142). De Gruyter. <https://doi.org/10.1515/9783110804607.109>
- Shahaf, D., Horvitz, E., & Mankoff, R. (2015). Inside jokes: Identifying humorous cartoon captions. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)*. ACM Digital Library, 1065-1074. <https://doi.org/10.1145/2783258.2783388>
- Shultz, T. R. (1976). A cognitive-developmental analysis of humour. In A. J. Chapman & H. C. Foot (Eds.), *Humor and laughter: Theory, research, and applications* (pp. 11-36). Transaction Publishers.
- Suls, J.M. (1972). A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. In J. Goldstein & P. McGhee (Eds.), *The psychology of humor* (pp. 81-99). Academic Press.
- Suls, J.M. (1983). Cognitive processes in humor appreciation. In P. McGhee & J. Goldstein (Eds.), *Handbook of humor research*, (pp. 39-58). Springer Verlag.

- Tanaka, K., Yamane, H., Mori, Y., Mukuta, Y., & Harada, T. (2022). Learning to evaluate humor in memes based on the incongruity theory. *Proceedings of the Second Workshop on When Creative AI Meets Conversational AI*, (pp. 81–93). AC.
- Vásquez, C. (2019). *Language, creativity and humour online*. Routledge.
- Veale, T. (2004). Incongruity in humor: Root cause or epiphenomenon, *Humor: International Journal of Humor Research*, 17(4), 419-428. <https://doi.org/10.1515/humr.2004.17.4.419>
- Warren, C., Barsky, A., & McGraw, A. P. (2021). What makes things funny? An integrative review of the antecedents of laughter and amusement. *Personality and Social Psychology Review*, 25(1), 41-65. <https://doi.org/10.1177/1088868320961909>
- Weber, M., Ruch, W., Riemann, R., Spinath, F.M., & Angleitner, A. (2014). A twin study on humor appreciation: The importance of separating structure and content. *Journal of Individual Differences*, 35(3), 130-136. <https://doi.org/10.1027/1614-0001/a000136>
- Whitford, A.B. (2005). Interpretation of Pearson's r. In K. Kempf-Leonard (ed.), *Encyclopedia of social measurement* (pp. 523-529). Elsevier.
- Yus, F. (2021). Incongruity-resolution humorous strategies in image macro memes. *Internet Pragmatics*, 4(1), 131-149. <https://doi.org/10.1075/ip.00058.yus>
- Yus, F. (2023). *Pragmatics of internet humour*. Palgrave MacMillan.