

Research Note

On the use of scale distortion for visual humour: a preliminary analysis

Clara Swaboda

University of Vienna, Austria
swaboda@logological.org

Tristan Miller

Austrian Research Institute for Artificial Intelligence, Austria
University of Manitoba, Canada
Tristan.Miller@umanitoba.ca

Abstract

In contrast to verbal humour, visual humour remains a relatively underdeveloped area of research. In this exploratory study, we investigate whether scale incongruity – i.e., discrepancy between the expected and actual experience of the size of an object – can serve as a source of humour in the visual modality. We adapt a pre-existing visual data set of mundane scenes by altering the size of an individual object in each scene and collecting humorousness ratings from human annotators on the original and scale-distorted versions. Our analysis of these annotations reveals that scenes with distorted objects are perceived to be significantly funnier than the original images.

Keywords: visual humour, humour perception, art, scale distortion

1. Introduction

Humour plays an essential role in everyday communication. Studying humour leads to a better understanding of human communication and culture, which in turn may have important implications for human–machine interaction. With the growing ubiquity of artificial intelligence, particularly in the generation and processing of artistic media, there is a growing need for systems capable of grasping the nuances of humour. Engineering such systems will

require a more fine-grained understanding of potential sources of humour across multiple modalities, and of the contexts that enable these humour sources.

Although verbal humour has long been studied in linguistics and natural language processing (Attardo, 2017; Attardo, 2020), other modalities of humour, including the visual, have attracted considerably less attention and remain a great challenge, particularly in computational applications (Cowie, 2023; Hempelmann & Samson, 2008). This is unfortunate, since visual humour is prominent in contemporary culture, such as Internet memes (Chandrasekaran et al., 2016; Wang & Wen, 2015). Visual humour is moreover different from verbal humour in structure and likely also in the way it is cognitively processed (Hempelmann & Samson, 2008), pointing to the need for specialised theories and methodologies. To date, much visual humour research has concerned itself with semantics, leaving finer formal aspects un- or under-explored. This is surprising, as visual aspects such as form, colour, line, and scale are known to be sources of humour, and moreover set the visual medium apart from the textual realm (Gérin, 2013, p. 168). In this work, we investigate whether scale incongruity – i.e., discrepancy between the expectation and actual experience of the size of an object – is a source of visual humour. To this end, we customise a pre-existing data set of scenes and collect humorousness ratings by human annotators. Our results suggest that scale incongruity can indeed be a source of visual humour.

2. Background

Most visual humour research has focused on higher-level semantics in visual scenes as a whole (Chandrasekaran et al., 2016; Gérin, 2013; Hempelmann & Samson, 2008). Hempelmann & Samson (2008, p. 626) observe that these sorts of studies have an underlying assumption that visual humour can be transferred into verbal humour in the sense of “reading” images. However, they argue that the cognitive processes underlying the experiences of verbal and visual humour are probably different: verbal humour is presented in a linear way, even though the processing is not necessarily linear, whereas visual humour imposes no firm structure of processing on the viewer, even though the composition of an image guides the viewer’s attention. They also argue that visual humour can bear multiple incongruities that may or may not be fully recognised by the viewer. As Gérin (2013, p. 168) notes, while visual humour can convey humorous aspects based on semantics, it has a much broader toolkit that encompasses visual aspects such as “medium, form, color, line, composition, scale or technique”, and sometimes *what* is represented is less significant than *how* it is represented.

We hypothesise that scale incongruity can serve as a source of visual humour. Even though incongruity is thought to be a necessary condition of humour, it is not a sufficient one, because the mere experience of violated expectations could alternatively lead to puzzlement or even aversion (Ruch, 2008, p. 25). Whether or not incongruity leads to a mirthful experience depends on its (partial) resolution (Attardo, 2020, Ch. 4). Since incongruity is an inherently semantic concept, and since semantics goes beyond linguistics, we assume that visual aspects can be a source of visual humour only if they are not merely decorative but alter the semantics of a scene. An example of scale incongruity as a source of visual humour could be a depiction of a child that is about to eat an enormous pile of food. Humour is evoked because the viewer knows that this feat is impossible; the same scene with a modestly sized portion of food would not be considered funny. Distorting the scale of an object to create a humorous effect has a long history in visual art, having been observed, for example, in the wall paintings in Pompeii (Klein, 2014). Although scale distortion seems to be a basic humour technique, it has, to the best of our knowledge, never before been investigated empirically.

3. Experiment

Our basic approach to investigating the contribution of scale distortion to humour involves altering the size of a single object in a scene while keeping all other elements constant. Some previous studies have used a similar approach, but have concentrated on other formal elements such as the complexity of composition (Huber & Leder, 1997), the effect of action lines on memory and understanding (Brooks, 1977), and the amount and shape of hair (Karabas, 1990). In our study, we took mundane scenes as stimuli that, as is often the case in visual humour, are not funny in and of themselves except through the introduction of incongruities perceived as “non-serious and not consequential” (Ruch, 2008, p. 20). In our case, we introduced a distortion of scale of an object in the scene with the expectation that it would significantly change its perceived relation to the other objects. This could generate an incongruity and in turn lead to a humorous experience.

3.1. Data collection

Our study uses visual scenes from the Abstract Visual Humor (AVH) data set (Chandrasekaran et al., 2016), itself incorporating part of the Visual Question Answering data set by Antol et al. (2015). AVH, created for one of the few previous studies on visual computational humour, consists of 6400 scenes that crowdsourced workers composed by selecting and arranging objects from a fixed catalogue of clip art images. For half of the scenes, the crowd workers were instructed to produce something humorous. Each scene was then presented to a separate set of crowd workers who applied humorousness ratings using a five-point Likert scale; scenes were then assigned a label of either “funny” or “not funny” using a threshold applied to the average humorousness rating.

We randomly selected 200 of the AVH scenes labelled as “not funny”. We made two copies of each scene, one in which a single object was distorted via up-scaling, and another in which the same object was distorted via down-scaling. (See Fig. 1) Our raw data set therefore consists of 600 scenes: 200 original, and 200 each with up-scaling or down-scaling. Similar to Chandrasekaran et al. (2016), we sourced humorousness ratings for the scenes by asking five human annotators to rate them on a five-point Likert scale. Unlike Chandrasekaran et al. (2016), however, we used the same five annotators to rate all 600 scenes (in random order).

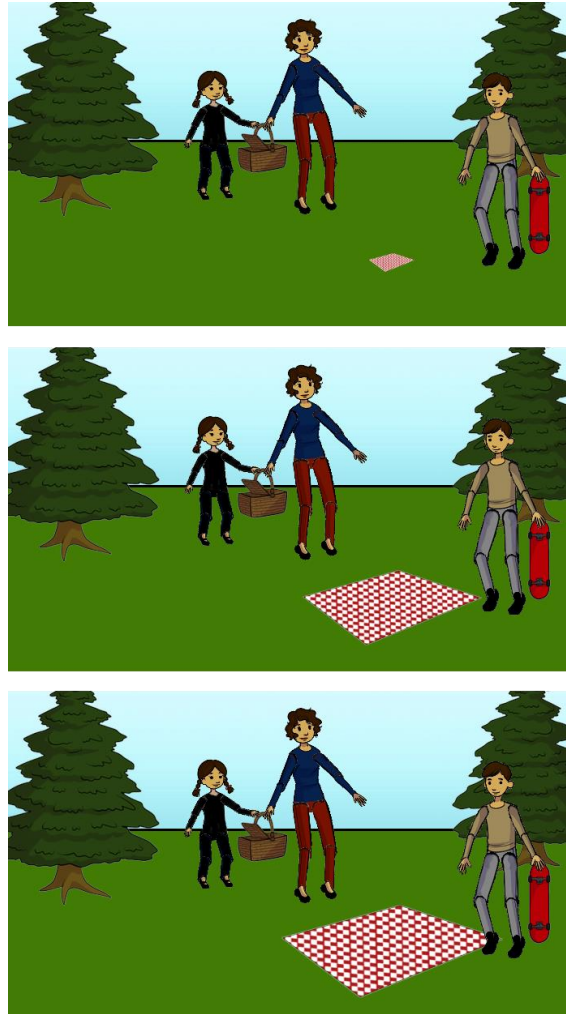


Figure 1. An original scene (middle) and its down-scaled (top) and up-scaled (bottom) variants.

3.2. Results

The distribution of average humorousness ratings for each scene across the five annotators was right-skewed with most images rated between 1 (“not funny”; 1945 scenes) and 2 (“a little funny”; 622 scenes). The overall distribution (see Fig. 2) shows that distorted scenes tend to have higher funniness ratings than the original ones.

To assess whether the differences in humorousness ratings between original and distorted images are significant, we performed a Mann–Whitney U test, a non-parametric test for differences between two groups. We find a significant difference ($p < .001$) between the group means of distorted ($\bar{x} = 1.59, \sigma = 0.84$) and original ($\bar{x} = 1.38, \sigma = 0.70$) scenes. This finding supports our hypothesis that scale distortion can indeed serve as a source of visual humour.

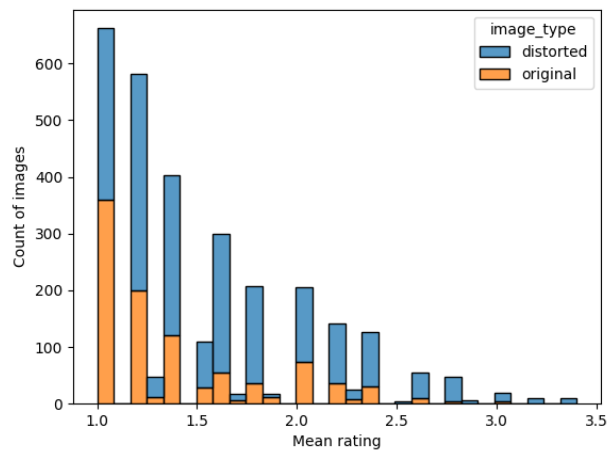


Figure 2. Mean humorousness rating by scene type (distorted vs. original).

4. Conclusion

This research note has presented, to our knowledge, the first empirical investigation of scale incongruity in visual humour. We applied isolated scale distortions to scenes known to lack humour, collected humorousness ratings on the original and distorted scenes, and found a small but significant increase in average humorousness ratings of the distorted scenes. This suggests that scale distortion can play a role in visual humour. Admittedly, our study and analysis was exploratory in nature and thus limited in scope. Further research is needed in order to investigate under which conditions scale incongruity is perceived as humorous, and whether the same effect and conditions hold for other types of visual distortions.

Acknowledgements

This work was supported in part by the Austrian Science Fund (FWF) under project M 2625-N31.

References

- Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Zitnick, C. L., & Parikh, D. (2015). VQA: Visual question answering. *Proceedings of the 2015 IEEE International Conference on Computer Vision*, 2425–2433. <https://doi.org/10.1109/ICCV.2015.279>
- Attardo, S. (Ed.). (2017). *The Routledge handbook of language and humor*. Routledge. <https://doi.org/10.4324/9781315731162>
- Attardo, S. (2020). *The linguistics of humor: An introduction*. Oxford University Press. <https://doi.org/10.1093/oso/9780198791270.001.0001>
- Brooks, P. H. (1977). The role of action lines in children’s memory for pictures. *Journal of Experimental Child Psychology*, 23(1), 98–107. [https://doi.org/10.1016/0022-0965\(77\)90076-5](https://doi.org/10.1016/0022-0965(77)90076-5)
- Chandrasekaran, A., Vijayakumar, A. K., Antol, S., Bansal, M., Batra, D., Zitnick, C. L., & Parikh, D. (2016). We are humor beings: Understanding and predicting visual humor.

- Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 4603–4612. <https://doi.org/10.1109/CVPR.2016.498>
- Cowie, R. (2023). Computational research and the case for taking humor seriously. *Humor: International Journal of Humor Research*, 36(2), 207–223. <https://doi.org/10.1515/humor-2023-0021>
- Gérin, A. (2013). A second look at laughter: Humor in the visual arts. *Humor: International Journal of Humor Research*, 26(1), 155–176. <https://doi.org/10.1515/humor-2013-0009>
- Hempelmann, C. F., & Samson, A. C. (2008). Cartoons: Drawn jokes? In V. Raskin (Ed.), *The primer of humor research* (pp. 609–640). De Gruyter Mouton. <https://doi.org/10.1515/9783110198492.609>
- Huber, O., & Leder, H. (1997). Are more compact cartoons more humorous? *Humor: International Journal of Humor Research*, 10(1), 91–104. <https://doi.org/10.1515/humr.1997.10.1.91>
- Karabas, S. (1990). Hairy Turkish cartoons. *Humor: International Journal of Humor Research*, 3(2), 193–216. <https://doi.org/10.1515/humr.1990.3.2.193>
- Klein, S. R. (2014). Art and visual humor. In S. Attardo (Ed.), *Encyclopedia of humor studies* (pp. 63–68). Sage Publications.
- Ruch, W. (2008). Psychology of humor. In V. Raskin (Ed.), *The primer of humor research* (pp. 17–100). De Gruyter Mouton. <https://doi.org/10.1515/9783110198492.17>
- Wang, W. Y., & Wen, M. (2015). I can has cheezburger? A nonparanormal approach to combining textual and visual information for predicting and generating popular meme descriptions. *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 355–365. <https://doi.org/10.3115/v1/N15-1039>