Journal of Geek Studies

Vol. 11(1): 2024.

ISSN 2359-3024

Journal of Geek Studies

The Journal of Geek Studies is a non-peer-reviewed, open-access, non-profit, online biannual publication devoted to the popularization of science.

Journal of Geek Studies
http://jgeekstudies.wordpress.com/
http://jgeekstudies.org /

ISSN: 2359-3024 (online).
Vol. 1 (2014) – present.
São Paulo, SP, Brazil.

1. Science; 2. Technology; 3. Geek Culture.

Each author is responsible for the scientific basis of their articles. The authors' views do not necessarily represent those of the editors. Authors retain the copyright for their articles. Information for authors can be found at http://jgeekstudies.org/guidelines/

Cover art: Final panel of comic strip 'It's a bit of a mess', by Safely Endangered and Rewriting Earth. Credits: David Schneider, Amber Weedon, Paul Goodenough, and Safely Endangered (Chris McCoy). Used with permission.

Journal of Geek Studies

jgeekstudies.org

The Ruins of Tolkien: Archaeology in the world of *The Lord of the Rings*

Leon Corneille-Cowell

Department of Archaeology, University of York, York, UK.
Email: leoncowell96@gmail.com

Tolkien scholarship is defined as the study of J.R.R. Tolkien's personal life, legendarium, or academic work. This is a multidisciplinary area of study encompassing subjects such as literature, linguistics, and gender studies. There are several peer-reviewed journals for Tolkien studies including *Mallorn* by the Tolkien Society, *Tolkien Studies*, and *Journal of Tolkien Research* (Brennan Croft, 2016).

Tolkien himself was an academic, working at the University of Leeds and Pembroke College, Oxford. His research specialised in Anglo-Saxon Literature and Philology. Furthermore, one could argue that fantasy literature and archaeology have an intrinsic link, ruins of lost civilisations and magic items – in fact, the author of the *Malazan Books of the Fallen*, Steven Erikson, was an archaeologist before beginning his literary career (Fantasy Book Review, 2023).

Despite the seemingly obvious connection to the ancient world, very little has been published in the way of archaeological research or interpretations of the world of Middle-earth. Some may argue this to be impossible; however, Tolkien wrote a magical world and mythos which was full of ref-

Figure 1. Weathertop from the game *Lord of the Rings Online* (Daybreak Game Company, 2007). Source: LOTR Fandom 2000 (screen capture from the game).

Cite this article as: Corneille-Cowell, L. (2024) The Ruins of Tolkien: Archaeology in the world of The Lord of the Rings. Journal of Geek Studies 11(1): 1-6. https://doi.org/10.5281/zenodo.10644072

erences from human history and prehistory.

This article aims to examine occurrences of archaeology and archaeological techniques in the world of Middle-earth and demonstrate how archaeological methodology can be used in fantasy literature.

THE RUINS

When one thinks of Archaeology as a discipline people often imagine ruins and standing stones; this idea harkens back to the grand tour of the post-medieval period (Ceserani et al., 2017). In the Narrative, we encounter two main ruins, those of Weathertop and Osgiliath (Tolkien, 2004: 163, 664). Sabo (2007) notes ruins as part of Tolkien's rich cultural landscape and places of memory.

In the case of Weathertop, it is a memory of a fallen kingdom that only a small few (the Dúnedain) can remember. At one point Aragorn (a bearer of cultural burden) educates the hobbits on this fact. Though the idea of the Lost Kingdom sounds fantastical during the tumultuous early medieval kingdom, it was a reality (Williams, 2023). Lost Kingdoms are a reality in Tolkien's legendarium with the city of Gondolin bearing resemblance to Babylon and the tower of Babel (Garth, 2020: 150).

On the other hand, the ruins of Osgiliath are a city of the extant kingdom of Gondor. Though still associated with what is lost, they are much more focused on a kingdom in decline, like the Roman or Byzantine empires (Norwich, 1996). Furthermore, the statues of Gondor's founding father that the party encounters harken back to the statue of the classical and Egyptian civilisations such as those described in *Ozymandias* by P.B. Shelley (Garth, 2020: 143).

OBJECT BIOGRAPHIES

Object biographies are the study of an item's 'life' going from its 'birth' or creation to its 'death' or destruction and in some cases it's 'rebirth' or remaking. The idea dates to Kopytoff (1986) and is a way of examining relationships between people and materials as well as spatial relationships with the landscape (Gosden & Marshall, 1999). In Middle-earth there are several object biographies that can be explored. Middle-earth is based on a fantastical prehistory of Earth (Tolkien et al., 2006); thus, all objects would be considered artefacts by modern archaeologists and scholars – for the purpose of this paper we have selected only the objects with significance to the plot.

The Ring

The most obvious biography is an item central to the plot, namely the One Ring (also called the Ruling Ring and Isildur's Bane). The plot of *The Lord of the Rings* follows a journey, not just of the characters but also of the Ring itself. We are given the Ring's biography with its creation in Mount Doom, its acquisition, and then subsequent loss by Isildur, and then its discovery and loss by Smeagol (Gollum) in chapter three Of *Shadows of the Past* (Tolkien, 2004 [original publication from 1954–1955]: 53–56). The biography of the ring is followed in detail during Tolkien's first publication, *The Hobbit* [1937]. The One Ring is unusual as an object as Tolkien describes it as having malice and even leaving Gollum (Tolkien, 2004: 53; Day, 2020), suggesting a level of personhood in the object, which is unusual for an object biography.

Anduril

Anduril, the sword of Aragorn, is introduced in the chapter *The Council of Rivendell* and is made up of the reforged fragments of Narsil. The sword was forged by a Dwarven smith before being taken by the Númenóreans, when it fell into the possession of Elendil. The sword was then shattered during the battle with Sauron, but the shards were used to kill Sauron by Isildur.

In the Narrative, the sword plays a

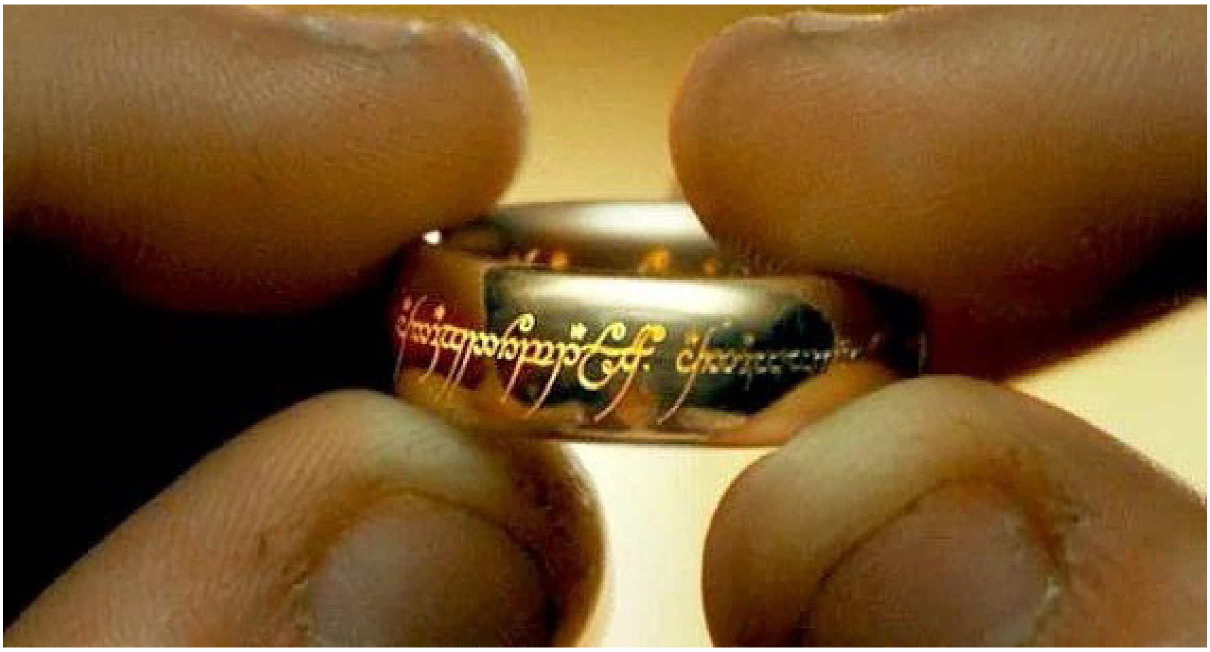


Figure 2. The One Ring from the Peter Jackson Trilogy. Source: LOTR Fandom 2000 (screen capture from the movie *The Fellowship of the Ring*, 2001).

huge role in Aragorn's identity as King of Gondor; he uses it to identify himself and his heritage to Éomer and the other riders of Rohan. Swords are shown as identifiers of kingship and warriorhood and heritage in early medieval literature. Examples include Excalibur in the tales of Arthur and the sword of Wiglaf in Beowulf as opposed to the 'Foreign' weapons such as the spear Gáe Bulga of Cúchulainn (Sayer et al., 2019: 544–545). Sayer et al. (2019) attribute a person-like quality to swords, which in burials are not worn like knives (an indication of utilitarian purpose) but indicates them to be a symbol of masculinity and power.

FUNERARY ARCHAEOLOGY

Funerary Archaeology or Archaeothanatology is the study of mortuary rites, burial practice, and death in an archaeological context (Cooney, 2023). This can relate to anything from grave goods and the landscape of burial to osteological studies of the dead (Stutz & Tarlow, 2013). In *The Lord of the Rings*, there are two instances where funerary archaeological comparisons can be made: the appearance of the Barrow Wight in *Fog in the Barrow-Downs* and the dead men of Dunharrow in *The Passing of the Grey Company* (Tolkien, 2004: 135, 773).

The Barrow Wight and Barrows

Fog in the Barrow-Downs describes an encounter Frodo, Samwise, Pippin and Merry have with an undead creature residing in a Barrow Grave (Tolkien, 2004: 135–148). Here we see Tolkien's early medieval literary origins with the creature resembling that of a draugr from Scandinavian folklore (Chadwick, 1946). Archaeological evidence for the undead in the real world often comprises measures to stop the undead; potential examples of this include cut and burn marks on an assemblage of skeletons found at Wharram Percy, Yorkshire (Mays et al., 2017). Similarly, there are examples of 'Vampire' burials in Poland, though in Tolkien's legendarium a vampire is a distinctly separate creature from a wight (Betsinger & Scott, 2014). The restless dead are also found in the 'Dead Marshes' though this bears more resemblance to no man's land in the Somme, an unlikely type of funerary archaeology (Garth, 2020: 165).

Figure 3. The Dead Men of Dunharrow as they appeared in the Peter Jackson Trilogy. Source: Tolkien Gateway (screen capture from the movie *The Return of the King*, 2003).

Barrows are a common burial practice which dates independently from the Bronze Age, Iron Age, and Early Medieval period (Grinsell ,1984; Noort, 1993; Garrow et al., 2014; Cooney, 2023; Stephens, 2023: 19–20). Barrows are shown again outside the city of Edoras in Rohan; it is a commonly held consensus that the Rohirrim are modelled on the Anglo-Saxons (Bates, 2002; Sabo, 2007: 102). So, it is likely that the Barrows were built in the manner of those in early Medieval England (Noort, 1993). However, the context of burial outside of a settlement resembles that of the Romans (Zachos & Karampa, 2015). In the text it says: "*At the foot of the hill the way ran under the shadow of many mounds, high and green*", "*Seven mounds upon the left, and nine upon the right said Aragorn 'Many long lives of men it is since the golden hall was built*" (Tolkien, 2004: 507).

The Anglo-Saxon comparison is further emphasised by the burial of Théoden's horse, Snowmane, after his demise alongside his master during the Battle of the Pelennor Fields (Sabo, 2007: 104). Horse burials have been found alongside numerous Anglo-Saxon "warrior" graves (Cross, 2011). Horse burials have been found in other sites such as the Chariot Burials of Iron Age East Yorkshire (Stephens 2023).

This contrasts to the description of the Barrow-Downs, which are described differently and in a much more sinister light: "*he turned his glance eastwards and saw that on that side the hills were higher and looked down upon them; and all those hills were crowned with green mounds, and on some were standing stones, pointing like jagged teeth out of green gums*" (Tolkien, 2004: 137). The difference in description may be due to the Rohirrim mounds being that of honoured dead who represent the culture's proud warrior heritage and identity. Whereas the inclusion of standing stones in the Barrow-Downs bears more resemblance to the prehistoric passage tombs from the Neolithic and Bronze Age (Grinsell, 1984: 17; Cooney, 2023: 200). Furthermore, the hobbits fear the barrows, they are 'other' and outside of the hobbit's cultural memory, like the attitude of later cultures to the Neolithic stones at Avebury (Gillings & Pollard, 1999: 186).

The hobbits also awaken covered in jewels and swords, like those found at Sutton Hoo – excavated around a similar time to the writing of *The Lord of the Rings*; connections to Beowulf have also been made, which was a work Tolkien famously translated and published (Lindqvist, 1948). The hobbits are gifted swords and find out the barrows belonged to the Men of Westernesse – who were beaten by the kingdom of Angmar. These men were descended from the Númenóreans, a thalassocracy bearing similarity to the Bronze Age Minoans (Williams, 2020). Here we also see the object biographies of these objects, which are brought back from "death" by the characters using them; one of those swords goes on to help kill the Witch-king of Angmar, fulfilling its purpose symbolically.

The Dead Men of Dunharrow

The Dead Men of Dunharrow (also known simply as Oathbreakers) first appear in the chapter *The Passing of the Grey Company* (Tolkien, 2004: 773). In Appendix F, Tolkien notes them as being kin to the wild men of Drúadan Forest who moved to the mountains and became the men of the mountains (Tolkien, 2004: 1025). These men are described as Pre-Númenóreans, or the indigenous men of Middle-earth, and included the Hill-men and Dunlendings. Comparisons can be found between these groups and Iron Age tribes who lived in Britain when the Romans first arrived; albeit in a rather dated and Roman depiction of them as "savage".

The Dead Men are an unusual depiction of funerary archaeology as no bodies are ever shown. Instead, they appear as wraiths, the unrestful dead. This is due to their breaking of an oath and past worship of Sauron. Breaking of Oaths is a common trope in ancient literature and appears many times in Tolkien's narrative, such as Saruman betraying Gandalf and breaking the law of hospitality, a tradition dating back to Homer (Sinex, 2003; Pitt-Rivers, 2012).

CONCLUSION

In conclusion, we have given a brief outline of archaeology in the worlds and imagination of Tolkien. Toilkien's engagement with the past and classical texts, as well as with archaeological discoveries from his time (Sutton Hoo), is self-evident from his writing and imagery. However, this article aims to highlight how more recent and specialised archaeological areas of study can be applied to Tolkien. Namely, this was done with object biographies and funerary archaeology. We found that Tolkien's rich descriptive prose allows for interpretation of the physical, allowing for easy cultural comparisons to be found.

REFERENCES

Bates, B. (2002) The Real Middle Earth. Sidgwick & Jackson, Basingstoke & Oxford.

Betsinger, T.K. & Scott, A.B. (2014) Governing from the grave: vampire burials and social order in post-medieval Poland. Cambridge Archaeological Journal 24(3): 467–476.

Brennan Croft, J. (2016) Bibliographic resources for literature searches on J.R.R. Tolkien. Journal of Tolkien Research 3(1): a2.

Ceserani, G.; Caviglia, G.; Coleman, N.; De Armond, T.; et al. (2017) British travelers in eighteenth-century Italy: the grand tour and the profession of architecture. The American Historical Review 122(2): 425–450.

Chadwick, N.K. (1946) Norse Ghosts (a study in the Draugr and the Haugbúi). Folklore 57(2): 50–65.

Cooney, G. (2023) Death in Irish Prehistory. Royal Irish Academy, Dublin.

Cross, P.J. (2011) Horse burial in first millennium AD Britain: issues of interpretation. European Journal of Archaeology 14(1-2): 190–209.

Day, D. (2020). Ring Legends of Tolkien. London: Simon and Schuster.

Fantasy Book Review. (2023) Steven Erikson Biography. Available from: https://www.fantasybookreview.co.uk/Steven-Erikson/biography.html (Date of access: 22/Nov/2023).

Garrow, D.; Meadows, J.; Evans, C.; Tabor, J. (2014) Dating the dead: a high-resolution radiocarbon chronology of burial within an early Bronze Age barrow cemetery at Over, Cambridgeshire. Proceedings of the Prehistoric Society 80: 207–236.

Gillings, M. & Pollard, J. (1999) Non-portable stone artefacts and contexts of meaning: the tale of Grey Wether. World Archaeology 31(2): 179–193.

Gosden, C. & Marshall, Y. (1999) The cultural biography of objects. World Archaeology: 31(2): 169–178.

Grinsell, L.V. (1984) Barrows in England and Wales. Bloomsbury Shire Publications, London.

Lindqvist, S. (1948) Sutton Hoo and Beowulf. Antiquity 22(87): 131–140.

LOTR Fandom. (2000) The One Wiki to Rule Them All. Available from: https://lotr.fandom.com/wiki/Main_Page (Date of access: 28/Jan/2024).

Mays, S.; Fryer, R.; Pike, A.W.G.; Cooper, M.J.; Marshall, P. (2017) A multidisciplinary study of a burnt and mutilated assemblage of human remains from a deserted mediaeval village in England. Journal of Archaeological Science: Reports 16: 441–455.

Noort, R.V.D. (1993) The context of early medieval barrows in Western Europe. Antiquity 67(254): 66–73.

Norwich, J.J. (1996) Byzantium: The Decline and Fall. Knopf Publishing Group, New York.

Pitt-Rivers, J. (2012) The Law of Hospitality. Journal of Ethnographic Theory 2(11): 501–517.

Sabo, D.L. (2007) Archaeology and the sense of history in J.R.R. Tolkien's Middle-earth. Mythlore 26(1): a7.

Sayer, D.; Sebo, E.; Hughes, K. (2019). A double-edged sword: swords, bodies, and personhood in early medieval archaeology and literature. European Journal of Archaeology 22(4): 542–566.

Sinex, M.A. (2003). "Oathbreakers, Why Have Ye come?": Tolkien's 'Passing of the Grey Company' and the twelfth century Exercitus Mortuorum. In: Chance, J. (Ed.) Tolkien the Medievalist. Routledge, London. Pp.155–168.

Stephens, M. (2023) Chariots, Swords and Spears. Oxbow Books, Oxford.

Stutz, L.N. & Tarlow, S. (2013) Beautiful things and bones of desire: emerging issues in the archaeology of death and burial. In: Stulz, L.N. & Tarlow, S. (Eds.) The Oxford Handbook of Death and Burial. Oxford University Press, Oxford.

Tolkien Gateway. (2023) Oathbreakers. Available from: https://tolkiengateway.net/wiki/Oathbreakers (Date of access: 28/Jan/2024).

Tolkien, J.R.R. (2004) [1954–1955] The Lord of the Rings. 2nd Ed. Harper Collins, London.

Tolkien, J.R.R.; Carpenter, H.; Tolkien, C. (2006) The letters of J.R.R. Tolkien. Harper Collins, London.

Williams, H. (2020) Tolkien's thalassocracy and Ancient Greek seafaring people. Tolkien Studies 17: 137–162.

Williams, T. (2023) Lost Realms: Histories of Britain from the Romans to the Vikings. William Collins, Glasgow.

Zachos, K.L. & Karampa, M. (2015) The Cemeteries of Nicopolis. The Hellenic Ministry of Culture, Athens.

About the author

Leon Corneille-Cowell is a Bsc student in archaeology at the University of York as well as an avid Tolkien fan who reads the Trilogy once per year. He is starting a MSc in Funerary Archaeology in September 2024. It is his personal ambition to be a Tolkien scholar and contribute archaeologically to the field of Tolkien studies.

Journal of Geek Studies

jgeekstudies.org

The Angry Video Game Model: exploring neural network architectures to predict videogame review ratings

Henrique Magalhães Soares

Independent researcher. São Paulo, SP, Brazil.
Email: hemagso@gmail.com

Videogame reviews are an essential part of the videogame industry today. From reviews by specialized gaming outlets such as IGN or Kotaku to individual reviews on platforms that sell the game themselves such as Steam, reviews are an important part of the gaming ecosystem. They allow players to identify which games are worth investing their hard-earned cash and their scarce leisure time.

Reviews have also been used as a way for customers to draw attention to an issue with the game. This practice, known as review bombing, has increased in recent years. One of the earlier high-profile examples is the review bombing of *Mass Effect 3* in 2012 due to its controversial ending, which led developer Bioware to later release an extended cut as a response (Gelbart, 2019). Other examples include the review bombing of *The Elder Scrolls V: Skyrim* following the announcement in 2015 of paid mods for the game, which ultimately led the game developer Bethesda to postpone this feature until 2017 (Gelbart, 2019), and the review bombing of *Pokémon Sword/Shield* due to the games not including every Pokémon from previous generations, among other issues (Kim & Liao, 2019).

THE ANGRY VIDEO GAME NERD

Reviews can also be the source of a lot of fun. Numerous YouTube channels specialized themselves in making game reviews in an entertaining way, providing (sometimes) insightful commentary on the quality of games.

One of these shows is *The Angry Video Game Nerd*, a YouTube review comedy web series created by James Rolfe. To be clear, despite having increased in popularity on YouTube, the show predates it by more than a year. It was first released on Cinemassacre website on May 25, 2004 (Cinemassacre Productions LLC, n.d.). Its YouTube debut only happened on December 15, 2005 (Wikipedia, 2021).

In the show James Rolfe plays the persona of "The Nerd", an angry, short tempered and foul-mouthed character who reviews bad games (usually from the 16 or 32-bit console era) in an attempt to warn the viewer not to play them. The show is considered one of the pioneers of review videos on the internet, inspiring, for better or worse, many other content creators.

It is very common for reviews to include some kind of numerical or ordinal scale that allows customers to compare reviews or to aggregate reviews from different users into a single score. However, despite being reviews, Angry Video Game Nerd (henceforth AVGN) videos do not feature any numerical ratings that allow us to do that. In this article I aim to investigate whether ma-

Cite this article as: Soares, H.M. (2024) The Angry Videogame Model: exploring neural network architectures to predict videogame review ratings. Journal of Geek Studies 11(1): 7-34. https://doi.org/10.5281/zenodo.10900204

chine learning and sentiment classification techniques can help us overcome this limitation of the show's format.

Note on profanity: Being inspired by Rolfe's show to write this article, I will use many of his quotes to illustrate the techniques being applied here. While I used the actual verbatim text when running the models and predictions (as the removal of an expletive may change the meaning intended for the sentence), the quotes in print will not feature any of the Nerd's trademark profanity. That was done to keep this text accessible to all viewers, so I will instead use less offensive substitutes or *** in place of the expletives. These editorial changes will be marked by a gray highlight over the word.

OBJECTIVES

This paper aims to establish a performance baseline for machine learning methods in predicting video game review scores, offering an introduction to the topic for those new to the field. While the methods described here are definitely outdated compared to recent advances brought by Large Language Models such as GPT-4 or Google Gemini, they still can be useful – both as simple and low-cost solutions for less complex language tasks and as educational tools. I believe the approach used in this article strikes a good balance between achieving a good performance and being easy to understand and follow along – this is a science communication journal, after all. (This still requires some basic understanding of linear algebra to follow along.) To facilitate this goal, all code used in this paper has been made available on Github (https://github.com/hemagso/avgm).

Overview of this article

This article is structured in the following manner:

- First, I will present some common challenges and pitfalls encountered when dealing with natural language.
- Next, I will describe the data used in this article and detail all pre-processing done to it.
- I then discuss model architectures used in this article and present their performance and shortcomings.
- Finally, I present my conclusions on which model architecture works best for this dataset and suggest new lines of research based on this work.

CHALLENGES FOR NLP

Before describing the work done here, I want to briefly discuss some of the challenges of working with Natural Languages from a machine learning perspective. This list is by no means exhaustive, but it should be enough to situate the reader on the challenges that must be overcome to obtain good performance for our machine learning models.

Contextual words

Most languages have words whose meaning changes depending on where they are in the text and what other words accompany them. For example, consider the following sentence:

> *"I **ran** to the game store because we **ran** out of bad games to play."*

The highlighted word, "ran", has two different meanings in this sentence. On the first occurrence it means to quickly move yourself to another location, while in the second occurrence it is part of the "ran out" expression, which indicates that our supply of something was exhausted (in this case, bad games). Another example, more relatable for our sentiment analysis application, can be seen on the following sentence:

> *"You know what? This game is not bad."*

In this example the only way to accurately assess the sentiment of the reviewers towards the game is to take "not bad" as a unit of meaning (probably indicating that the game is mediocre at best).

Machine learning models that analyze the text word by word would be unable to understand the intended meaning in these two examples. Methods that can account for the position of words in a sentence and their relation to each other are necessary when dealing with this kind of data.

World knowledge dependence

"The trophy doesn't fit in the brown suitcase because it's too big."

What is too big? The trophy or the briefcase? This might be a simple question for a human but consider how much knowledge not expressed in the sentence or on the meaning of the words themselves a person needs to answer this question. First, you need to know that briefcases can contain other things, while trophies cannot. Secondly, you need to know that an object can only be contained by a larger object.

This sentence is an example of a Winograd schema, an alternative to the Turing test proposed by Hector J. Levesque as a means to test for machine intelligence (Levesque et al., 2012). The sentences in a Winograd schema are obvious for a human reader, but exceedingly difficult to machines due to the large amount of world knowledge or indirect reasoning necessary to solve their ambiguity. (Humans usually will not even notice that there is any ambiguity at all!)

Although some recent methods have achieved accuracy rates of over 90% by exploiting extremely complicated deep neural networks and pre-trained transformer models (Kocijan et al., 2020), the Winograd schema stands as a good example of the subtle challenges in natural language processing.

Ambiguity

Consider the following sentence:

"I went into the forest, where I found a bat."

What did I find in the forest? Was it a small flying mammal, or a long piece of wood? Both answers are possible for this sentence, and without further context no correct answer can be given. Differently from Winograd schemas there is no prior world knowledge that can 100% disambiguate the meaning of this sentence. Humans might disambiguate it based on their prior beliefs on how likely each encounter is,[1] and machines can take a similar approach.

Language detection

Humans around the world speak lots of different languages. It is hard to pin down an exact number, as languages are constantly evolving and the distinction between a language and a regional dialect can be hard to define. As of the writing, there are 7,139 languages recognized by the Ethnologue, a reference publication on the topic (Eberhard et al., 2021).

Although NLP models can be trained on datasets consisting of multiple different languages (and some applications such as Machine Translation in fact require such datasets), it is often useful to split your problem into individual languages and then train specialist models for each. This way, your model does not need to learn how to deal with things like false friends – words that are written or sound similar in two different languages but mean completely different things. For example, "parente" in Portuguese is a false friend for "parent" in English: the former refers to any person belonging to your family, while the latter is more specific, referring only to your mother and father.

This, however, introduces another problem: how can we automatically detect the language of a text? I will discuss this problem in more detail below, but for now, let us

[1] Actually, no. They would probably ask me "What do you mean? Like, the animal?" instead of assuming something and risk making a fool out of themselves.

explore a few more challenges for NLP models.

Spelling errors

People make mistakes, and written language is not an exception. While humans are rather good at correcting these mistakes (by noticing typos, erroneous pronunciation or context), machines are terrible at that. One of the first steps in almost all NLP models is tokenization, in which the text is split into small pieces that are mapped to a predefined set of tokens, and a misspelled word would not map to any of these, causing what is usually called an "Out-of-vocabulary" token – a token that the model never met before. The effect of these can be severe on predictions. Consider the following example:

"This game is awful."

This review might be tokenized into the following tokens, all of which are known by the model, having been assigned a sentiment score during training.

Token	This	game	is	Awful
Score	+0	+5	+0	-70

In this hypothetical example, the presence of the word "awful" allows our simple "Bag-of-words" model to estimate a negative sentiment for this review. But consider what would happen in the next misspelled example:

"This game is aful."

Token	This	game	Is	aful
Score	+0	+5	+0	?

The model has never met the word "aful", so it does not know what to make of it. It might assign it an average score, wrongly classifying the sentence as a neutral sentiment.

This problem can be partly alleviated by increasing the size of your token set to accommodate common misspellings of words. However, it is impossible to account for all occurrences, and there is a trade-off between model training time and performance and vocabulary size. A lot of research has gone into developing tokenization methods that can deal with this kind of issue, and some of them will be explored further below.

Domain-specific vocabulary

Domain specific vocabulary happens when a specific word has a different meaning within a specific domain when compared to its everyday usage. These are quite common in science, requiring a reader to consider the topic of a text when deciding on the meaning of a word. For example, when dealing with set theory an "element" refers to an individual member of a set. If on the other hand we are discussing chemistry, an "element" refers to a chemical element, a substance consisting of atoms that have the same number of protons in their nucleus. When talking about a game, an "element" might refer to elemental spells, such as "Fire", "Air", or "Lightning", a feature common to many magic systems in games.

This problem is more severe when you are trying to build a generalist language model, as the model will need large volumes of data to learn how to differentiate between meanings. Since we are dealing with a narrow application in this article (sentiment classification for game reviews), we will not discuss this problem in depth.

High dimensionality and sparsity

Finally, we discuss a problem that is not exclusive to NLP, but rather it is something that needs to be considered for most machine learning problems. The curse of dimensionality refers to phenomena that arise from dealing with high-dimensional spaces. In machine learning it is usually related to the fact that with an increase in dimensionality of your dataset there is an exponential increase in the amount of data required to cover all the space.

Let us work with a hypothetical example

to illustrate this problem. Imagine a classification model that takes as input 3 nominal variables, each with 10 categories. In this example, we would need at least 1,000 different training examples to cover the whole space. Now consider what would happen if instead we had 10 nominal variables with 10 categories each.

This curse of dimensionality affects text classification in a very particular way. One approach when modelling text data is to create dummy variables for each possible word:

$$D_w = \begin{cases} 1 & \text{If word } w \text{ is present} \\ 0 & \text{Otherwise} \end{cases}$$

The English language, however, has a lot of words.[2] If we consider a small vocabulary set of just 1,000 words, the number of combinations needed to cover this space is greater than 10^{301}.

Language space is also sparse, having most of the dummies described above assigned a value of zero. This happens because most texts use only a small subset of the vocabulary of the English language. Even *The Lord of the Rings*, a masterpiece of 481,103 words famous for its elaborate descriptions and flowery language uses only 15,493 distinct words (LotrProject, n.d.).

This high dimensionality and sparsity provide several challenges for training machine learning models. We will discuss "Word Embeddings", a common method for dealing with this problem further in this article.

DATA & PREPROCESSING

Data description

For this study I collected data from the review aggregator website Metacritic (www.metacritic.com). The data are comprised of 644,268 user reviews for 15,931 different videogames.[3] Each review is also associated with an ID that uniquely identify the user who made the review, as well as the publication date of that review on the website. Table 1 describes the fields available on the dataset.

Table 1. Dataset description.

Name	Type	Description
game_url	URL	URL identifying the game
user	String	Username of the user who wrote the review
date	Date	Date the review was published on Metacritic
userscore	Integer	0 to 10 integer score containing the score assigned by the user to the game being reviewed
review	String	String containing the unformatted review text

Unfortunately, I cannot make the dataset available for further research as the contents of the review themselves are copyrighted by CBS Interactive, who owns Metacritic, as stated in their Terms of Use (https://cbsinteractive.com/legal/cbsi/terms-of-use/).

Data preprocessing pipeline

Language detection

Metacritic is a website with global presence and users from a multitude of nationalities can post reviews there. Although most reviews are written in English, there are other languages represented on the dataset, and the website provides no structured data on what that language is. I utilized the *langdetect* python package (Nakatani, 2010) to identify the languages of the reviews, yielding a total of 46 different languages. The distribution of reviews among the top-10 most common languages on the dataset is shown in Table 2.

As expected, most of reviews are written in English. As I do not have enough data to train our model in multiple languages, I opted to work only with English reviews from this point forward.

[2] There are currently over 550 thousand entries on Wiktionary for English (Wiktionary, 2020), and native speakers usually have a vocabulary of around 10,000 words.

[3] This count considers games available in different platforms as entirely different games. For example, *Skyrim* on PC, PS4, Xbox, and Amazon Alexa counts as four different games.

Table 2. Language distribution of reviews.

Language	#	Percent
English	596,477	92.58%
Spanish	16,25	2.52%
Russian	12,043	1.87%
Portuguese	7,32	1.14%
French	2,202	0.34%
German	1,743	0.27%
Italian	1,234	0.19%
Somali	695	0.11%
Polish	584	0.09%
Turkish	577	0.09%
other	4530	0.70%
unknown	613	0.10%

Train / validation / test split

When training machine learning models, it is important to have a method to estimate your model performance in unseen data, that is, data that was not used to train the model. This avoids overfitting, a common problem when training large models in which the model ends up "learning the training data by heart" and performing poorly in unseen data.

Various methods exist to estimate the performance of the model on unseen data, but the simpler method is to simply holdout a fraction of your data, not using it to train your model parameters. This method, aptly called the holdout method, has the downside of reducing the amount of data available for the model to learn from. However, since I have enough data for my purpose, I decided to use it. I split the data into three different sets:

- Train set: data used to train the model weights through back-propagation;
- Validation set: data used to choose which model architecture is the best for this problem and to calibrate models hyperparameters;
- Test set: data used solely to estimate the performance of the final model on unseen data.

I reserved 10% of the dataset as the Validation set and 10% more as the Test set, leaving 80% of the data for model training. Table 3 shows the number of records in each set.

Table 3. Train / Validation / Test split.

Dataset	Percentage	Samples
Train	80%	417,884
Validation	10%	89,299
Test	10%	89,13

Tokenization

The last step on our pre-processing pipeline is tokenization. As mentioned above, tokenization is the process through which we segment a text into a sequence of meaningful tokens. These tokens (after being converted into numerical id's that can be manipulated with math) are then fed into our machine learning models for training and predictions.

The choice of tokenization method can have a huge effect on how easy a model is to train, as it effectively sets the minimum level of detail from which a model can derive meaning. Hence, it is usually a trade-off between having a token that is large enough to convey sufficient information by itself and the overall number of distinct tokens in your dataset (also known as your vocabulary size). To illustrate this trade-off, consider the following choices for tokenization:

- Letter tokenization: each different letter and digit is a separate token;
- Word tokenization: each group of characters separated by whitespace or punctuation marks are a separate token;
- Sentence tokenization: each different sentence is a different token.

Let us use a quote from AVGN episode "Hong Kong 97" to illustrate the differences between these three methods: "*I've been called upon to take care of business once again. Apparently, there is a game worse than* Big Rigs. *WORSE than* Dr. Jekyll and Mr. Hyde. *WORSE than* CrazyBus *or* Desert Bus. *It is known as* Hong Kong 97*, and I've been getting requests for it up the butt.*"

Table 4 shows the length of the sequences produced by each tokenization method, and the number of distinct tokens produced. Notice how the number of tokens needed to represent the text decreases with token complexity. This means that each token conveys more information.

Table 4. Statistics for different tokenization methods.

Method	Length	Distinct Tokens
Letter	257	32
Word	50	43
Sentence	5	5

On the other hand, the uniqueness of each token also increases with token complexity. Letters will be repeated quite often, as well as most words.[4] On the other hand, it is rare for full sentences to be repeated (and those which are probably phatic constructions or other types of uninformative sentences). This is a problem for us since we need many examples of a token to allow our model to learn how it should deal with it.

A lot of different methods of tokenization have been tested for natural language processing, and today most models use a sub-word unit approach. That method is somewhere between our letter tokenization and word tokenization. Sub-word methods have several useful properties that help us to better deal with misspellings and rare words, as we will see on the next section.

Sub-word units

Let us explain why we would want to use tokens smaller than a word with an example. Consider the following excerpt from AVGN episode "Plumbers don't wear ties": "*Oh, so is he a **plumber**? Well, the game's called **Plumbers** Don't Wear Ties, so I guess it makes sense: he's a **plumber**, and I don't see him wearing a tie…* [Images of John wearing a tie] *...WHAT THE HECK?! You can't even trust the darn title!*"

Take note of the two highlighted words, "plumber" and "plumbers". One tokenization option if to consider both as separate tokens. However, the model would see them as completely unrelated, and would need a lot of data to learn the relationship between them.

Another option is to create two separate tokens: "plumber" and "#s". The first token is just the word plumber by itself, and the second token is just the letter s (The # symbol indicates that this token is appended to another token to form a word). Table 5 compares the tokenization of these words.

Table 5. Word vs. sub-word tokenization examples.

Word	Word units	Sub-word units
Plumber	["Plumber"]	["Plumber"]
Plumbers	["Plumbers"]	["Plumber", "#s"]

In the sub-word representation both words share a token. In this way, the model does not need to learn that both words are related. It only needs to learn that the token "#s" usually means that the previous token is plural. And there are much more examples of plurals for the model to learn this than examples of the words "Plumber" and "Plumbers".[5] This is even more useful for rarer words. Consider the word "supernaturally", for example. There are many more examples of the word "supernatural" than "supernaturally", as can be seen in Figure 1, extracted from Google N-Gram viewer. It is easier for the model to learn the meaning of "supernatural" and then learn the meaning of "#ly" as the adverbial form from all other adverbs on the dataset than trying to learn the meaning of "supernaturally" by itself.

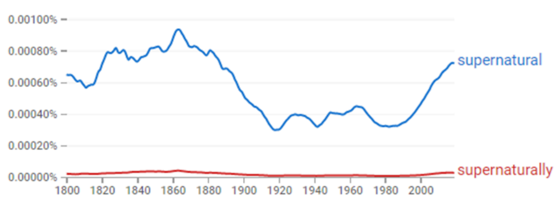


Figure 1. Occurrence over time for "supernatural" and "supernaturally".

[4] Unless we are dealing with rare words such as "gobbledygook" or "winklepicker". Yes, those are real words.

[5] Despite the existence of a very prolific game series with a plumber character.

For this work I opted to use the Word Piece tokenizer model. First proposed by (Wu et al., 2016) to address the problem of segmenting Korean and Japanese text,[6] this method was then adopted to automatically segment text into sub-word units. Its main advantage is being unsupervised, allowing us to learn the best token representation directly from the corpus and without the use of any annotated data. I trained the tokenizer on our train corpus to produce a vocabulary of 30,000 tokens, using the implementation available in the *tokenizers* python library (huggingface, 2021) with the following parameters:

- Normalization: I used the same normalizer as the one used by the BERT language model (Devlin et al., 2018). This normalizer replaces all types of whitespace characters by the common whitespace, replaces accented characters by their unaccented version, and applies lowercasing to all characters;
- Pre-tokenizer: I used the same pre-tokenizer as BERT, splitting on whitespace characters and punctuations to produce the first tokenization.

The details on the tokenization process are beyond the scope of this text. If you are interested in learning more, please check the accompanying *jupyter* notebooks available on Github (https://github.com/hemagso/avgm) where I go in more details about the process. To test the tokenization, let us check how it tokenized the following phrase:

> *"Feast your eyes on this accursed nonsense."*

> ['feast', 'your', 'eyes', 'on', 'this', 'accur', '#sed', 'nonsense', '.']

Everything seems to be working fine. Most common words consist of a single token, but the word "accursed" was split in sub-word units. With this step out of the way we can now proceed to discuss my modelling methodology.

The final step is converting the sequence of tokens into a sequence of numerical IDs, as models need things to be converted into numbers for them to be able to operate on them. For tokenizers, each unique word in the vocabulary is assigned during training a unique ID. In the case of the example above, it is:

> [16746, 1456, 4308, 1360, 1358, 4305, 5417, 6438, 15]

METHODOLOGY

Approaches to text classification

How can computers understand human languages? After all, computers are engineered to deal with numbers, and their language if one of numbers and symbols, rigid. Can computers understand the nuances of human language, with all its intricacies and beauty? Can a machine write a poem? Maybe.[7] But first we will need to help it turn language into math. In this section I will discuss different approaches that can be used to train text classifiers from labelled tokenized text.

Bag-of-words models

Let us go back to the sentence we tokenized above:

> [16746, 1456, 4308, 1360, 1358, 4305, 5417, 6438, 15]

The actual IDs here have absolutely no meaning and are completely arbitrary. The first step when modelling is deriving a useful representation of our data. One of the simplest options is calculating the count on

[6] Tokenization is a challenge in these languages because, in contrast to most languages based on the Latin script, Korean and Japanese words are not whitespace-separated. For example, can you spot the boundaries between words in the following text? 日本語を勉強しましたが本当に大変でした

[7] Surprisingly, yes (Lau et al., 2020), although others will have to judge its quality.

the text for a specific word and using this count as features for a classifier. This approach is called bag-of-words, and it is surprisingly effective for some application. Bag-of-word Naïve Bayes classifiers were one of the first effective spam filtering applications (Delany et al., 2013). This type of classifier works on the assumption that the mere presence of a word is informative about the dependent variable. In our review prediction problem, for instance, we could select words that are known to have a negative or positive sentiment to be part of our bag-of-words and use this to predict the sentiment for reviews:

- Positive words: good, great, excellent, masterpiece, incredible, marvelous;
- Negative words: bad, awful, terrible, trash, stupid.

This approach, however, has some flaws. It cannot take the context of the words into account, as all information about where the word is in the sentence is lost. For example, if we use the bag-of-words listed above we would not be able to properly classify the phrase "*This game is not bad.*" This is particularly important in cases where the word might not be informative by itself but is a powerful predictor when in context. In the sentences "*This game is **very** bad*" or "*This game is **slightly** bad*" the highlighted words are only informative in the presence of the word "bad". This weakness can be mitigated by building not a bag-of-words but a bag-of-n-grams. For example, we could calculate the counts of the 2-grams ("very", "bad") and ("slightly", "bad") and use those counts as features for our model. However, this starts to introduce a whole bunch of new challenges. How do I select the words in my bag of words? How can I find n-grams that are informative and should be included? The bag-of-words is a simple and surprisingly effective approach and you should definitely start with it before trying more complicated approaches – an advice that I will completely ignore in this article as I go forward to talk about Sequence Models.

Sequence models

So, how can I make use of the information provided by the order of the tokens in our sentence? Well, a good place to start is by not throwing it away at all. Sequence models consume the raw sequence of tokens as its input, allowing us to build architectures that can make use of the order information on the sentence.

However, this introduces a new problem. Models need not only be finite, but also of a fixed size. As we need to train the parameters in advance, the number of parameters and how they are related to each other need to be determined ahead of time. Text, however, can be of arbitrary length, and our model need to be able to deal with reviews such as "*It is good*" as well as "*This is an amazing piece of gaming history. The developers were probably inspired by God's angels when they were writing each single line of code of this masterpiece.*"

In this article we make use of recurrent architectures to solve this problem. Recurrent neural networks work by having an internal state of fixed size. An also fixed function is used to update this hidden state based on the current element of the input and the previous value. After applying this function on all elements of the sequence we are left with a fixed size state vector that can then be fed into a classifier to produce predictions. To illustrate how this type of model can work let us use a very simple mock example with a recurrent model composed by the following components:

- A 2-dimensional hidden state

  $$h_t = (h_t^0, h_t^1)^T, h_0 = (0,0)^T$$

- An input stream where:

  $$x_t = \begin{cases} 1 & \text{if } w_t \text{ in } (Good, Amazing, Incredible) \\ 2 & \text{if } w_t \text{ in } (Bad, Awful, Trash) \\ 3 & \text{if } w_t = Not \end{cases} \text{, or } 0 \text{ otherwise}$$

- An update function

  $$h_t = \left(f_1(x_t, h_{t-1}), f_2(x_t, h_{t-1})\right)^T$$ where

$$f_1(x,h) = \begin{cases} 1 & if\ x = 3 \\ 0 & otherwise \end{cases}$$

$$f_2(x,h) = \begin{cases} h^1 + 1 & if\ x_t = 1\ and\ h^0 = 0 \\ h^1 - 1 & if\ x_t = 1\ and\ h^0 = 1 \\ h^1 - 1 & if\ x_t = 2\ and\ h^0 = 0 \\ h^1 + 1 & if\ x_t = 2\ and\ h^0 = 1 \\ h^1 & otherwise \end{cases}$$

Before we go through an example, try to figure out what this recurrent rule does. What does h_0 represents? How about h_1? Let us run through some examples.

Example 1: "*This game is good*"

t	0	1	2	3	4
w_t		This	game	is	good
x_t		0	0	0	1
h_t^0	0	0	0	0	0
h_t^1	0	0	0	0	1

Example 2: "*This game is bad*"

t	0	1	2	3	4
w_t		This	game	is	bad
x_t		0	0	0	2
h_t^0	0	0	0	0	0
h_t^1	0	0	0	0	-1

Example 3: "*This game is not good*"

t	0	1	2	3	4	5
w_t		This	game	is	not	good
x_t		0	0	0	3	1
h_t^0	0	0	0	0	1	0
h_t^1	0	0	0	0	0	-1

Example 4: "*This game is not bad*"

t	0	1	2	3	4	5
w_t		This	game	is	not	bad
x_t		0	0	0	3	2
h_t^0	0	0	0	0	1	0
h_t^1	0	0	0	0	0	1

Notice that both negative sentiment examples got a negative h^1 by the end, and both positive sentiment examples got a positive one, despite examples 3 and 4 expressing those sentiments using a negation clause. The model was able to do that because it used h^0 as a memory of whether or not the previous word in the sequence was a negation word, allowing it to properly assess the sentiment of the words good or bad in context.

Of course, this is a toy example that only serves to illustrate the mechanism through which recurrent models can understand context. In practice, it is almost impossible to interpret the update function and the meaning of each element of the state vector in the way we did here. However, we can learn this function and representations from the data! This is the basic principle behind Recurrent Neural Networks, the method of choice for this article.

Word embeddings

Until now we have been using the index for the word as a categorical feature for our model, representing them by their indices. In practice, categorical features are usually represented by an encoding scheme known as one-hot encoding, where an indicator variable indicates the presence of a category:

$$x_{OH} = (I_1 \quad I_2 \quad \dots \quad I_N)^T$$

$$I_k = \begin{cases} 0 & If\ x \neq k \\ 1 & If\ x = k \end{cases}$$

This can work fine for small vocabulary of tokens, but as the vocabulary increases, we quickly start facing the problems of high-dimensionality and sparsity mentioned above. For our 30,000 words vocabulary, each element of our sequence (that is, every sub-word unit for our samples) would need to be represented by a 30,000-long vector of a single "one" and 29,999 "zeros".

One way of dealing with this problem is by using word embeddings. Word embeddings reduce the size of the representation of each word by replacing the long and sparse Boolean (only zeros and ones) vectors by smaller and dense (containing any real number). The nice thing about embeddings is that not only they can be trained from your data, but they can also be learned from unlabeled data. Below we illustrate an example for an embedding of size 4:

$$w_t = your$$
$$x_t = 1456$$
$$x_{OH} = (0,0,0,\dots,0,1,0,\dots,0)^T$$
$$x_E = (0.24, 3.71, -0.87, -1.33)^T$$

Embedding has many interesting properties, and there is a lot of research on methods to build embeddings. For a detailed explanation of embeddings, I recommend the work of Alammar (2019).

I will start with a simple sequence model. This will establish a baseline performance level for this task and justify some design choices that we will make going forward. It is also good practice when dealing with a new application: we start with the simplest model and build it up to address weak points identified along the way.

Model design

Our first model will be a Vanilla Recurrent Neural Network. The model has the architecture shown in Figure 2. Do not worry right now about what exactly a Recurrent Layer is; we will get into more detail about it later on.

Network description	Output	#Weights
Embedding Layer Vocabulary Size = 30,000 | Embedding Size = 512	**512**	**15,360,000**
Recurrent Layer Hidden Size = 512 | Bidirectional = Yes	**1024**	**1,050,624**
Fully Connected Layer Output Size = 11	**11**	**11,275**
Activation Function Type: Log Softmax	**11**	**0**
		16,421,899

Figure 2. Simple model architecture.

Figure 2 describes the parameters of each layer, such as Embeddings Sizes and Hidden Sizes. I also noted the output tensor size (which is useful to wrap your head around on how each layer transforms its input) and the number of weights on each layer (which will be particularly useful when we are comparing different model architectures).

Model training

I trained this model on the train dataset described above. The training ran for 20 epochs, and at the end of each epoch performance metrics were collected both for the training set and the validation set. The model was trained using Negative Log Likelihood Loss, with the Adam optimizer (Kingma & Ba, 2015) with default parameters (Learning rate = 0.001; β_1=0.9; β_2=0.999) used for gradient descent.

Model evaluation

To evaluate the model, I used the following model level metrics:

- Loss: the Negative Log Likelihood value;
- Exact Accuracy (ACC): the percentage of ratings that were perfectly predicted by the model;
- Accuracy ± 1 (ACC1): the percentage of ratings that were wrong by at most 1 rating;
- Mean Absolute Error (MAE): the average distance between the true rating and the predicted rating.

I also evaluated the following class level metrics to assess the quality of the prediction:

- Recall: the percentage of records with a certain rating that were predicted with said rating;
- Precision: the percentage of records predicted with a certain rating that were indeed of that rating.

All metrics were calculated for both train and validation datasets.

Model level metrics

Let us start by looking at the model level metrics, and how they varied during training (Fig. 3). We can see that on average all metrics improved with training (Fig. 3), although there is a lot of variation on both the training and validation set. This could be an indication that our model is having trouble learning and that we might need a larger or more sophisticated model.

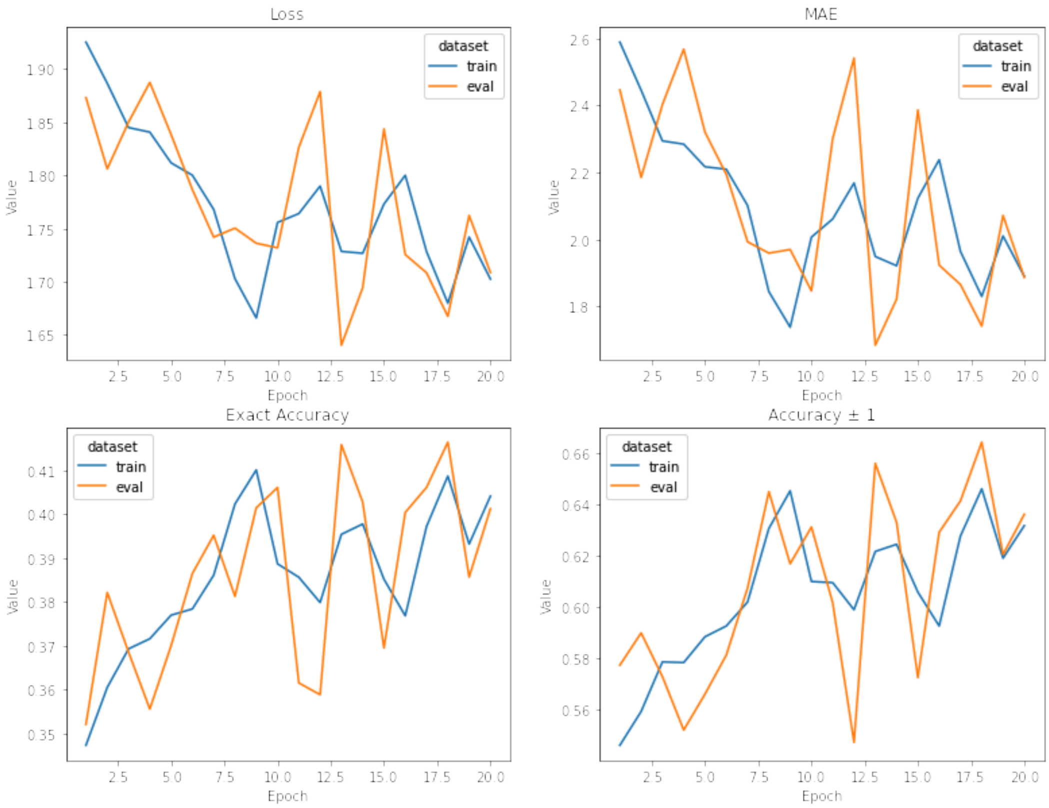


Figure 3. Model level metrics for a simple model over training epochs.

Using loss as our selection criteria, we see that the model achieved the best generalization (performance on unseen data) on Epoch 13:

Loss	MAE	Exact Accuracy	Accuracy ± 1
1.640418	1.683351	41.6%	65.6%

So, how good is this model? Since this is the first application on this dataset, we do not have any established benchmarks. In this case, it is useful to look at the performances achieved by other models on similar datasets.

State-of-the-art (SOTA) performance on the IMDB dataset (a dataset with movie reviews and associated sentiment) showed 96.21% accuracy (NLP-progress, 2021). So, our model is awful, right? Wow, not so fast! The IMDB dataset collapses the rating measurement scale, classifying all ratings 6 and below as negative, and all ratings 7 and above as positive. That reduces the task to a binary classification! So, our accuracy metrics are not comparable with the IMDB dataset.

The most comparable benchmark I could find was the Yelp dataset (a dataset with reviews extracted from www.yelp.com), which has a 5-level measurement scale for ratings. SOTA for this application achieves 72.8% accuracy. This indicates that, yes, my model is probably bad and that we should probably use a better architecture. (This was already indicated by the volatile loss training curve, but it is always nice to have further evidence.) Before trying to build a new model, however, let us explore a bit more this first attempt – we might learn some other useful things to incorporate into new attempts.

Class level metrics

Let us now look on class level metrics. These metrics will allow us to know if our model has a good performance predicting all ratings, or if for some reason it predicts some ratings better than others.

Figure 4 shows both Recall and Precision metrics for each of our classes, and we can immediately notice something funky is going on. The model seems to perform way better when predicting ratings 0 or 10. Recall for both these ratings is high, indicating that we correctly "retrieve" 80% of these ratings. However, Precision is way lower, indicating that this outstanding recall may in fact be caused by the model favoring these two ratings instead of the other 9 possibilities. Why would that be?

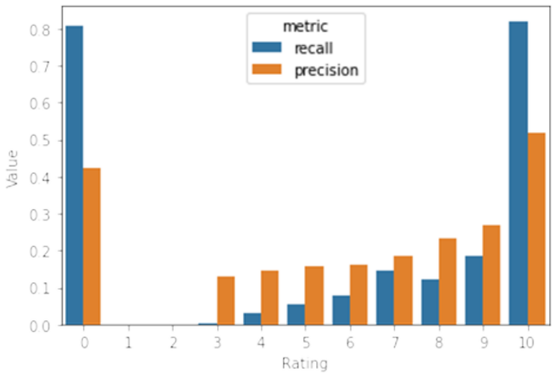


Figure 4. Precision and recall metrics for the simple model.

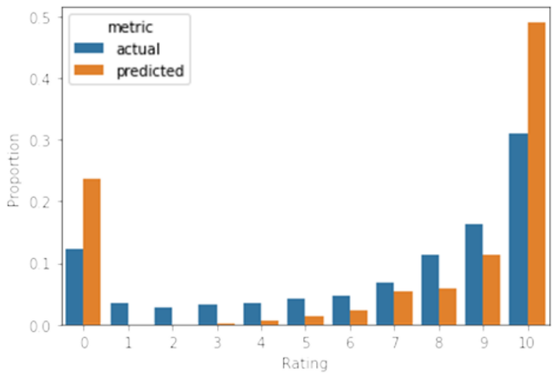


Figure 5. Actual and predicted class distribution for the simple model.

A quick look on the distribution of our data reveals the issue. Figure 5 shows the distribution of ratings in our dataset along the distribution of our model's predictions. The thing is: review ratings usually have an unbalanced distribution. In our case, over half the reviews is either 10 ("*This is the best game ever.*") or 0 ("*I hate this game with the power of a thousand suns.*"). Consequently, our optimization process ends up prioritizing getting those two ratings in detriment of the others, polarizing our reviews even more.

Situations like this are an example on why we should not only analyze model level metrics but also use class level metrics in the analysis. I will discuss which design decisions we can make to avoid this issue in later sections.

Individual predictions

It is also always useful to look at individual predictions made by our model. You might get qualitative insights that you would not notice from the aggregated data. To that end, I picked three sentences from three different AVGN episodes.

Dr. Jekyll and Mr. Hyde: "*You'd think I'm jokin', like I'm trying to be funny or somethin'. But, no, the fact that that game exists is a horrible abomination of mankind. That game is so freaking horrible, and I am not kidding*" (AVGN, 2010).

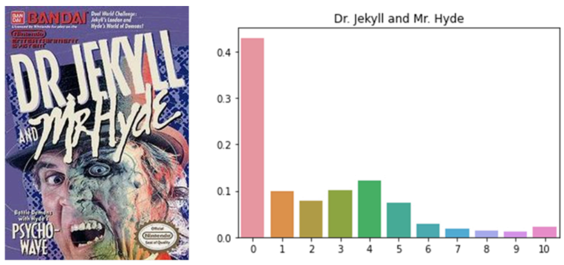


Figure 6. Cover art (source: Wikipedia) and predicted scores for the game *Dr. Jekyll and Mr. Hyde*.

Here we can see that the model captured the overall sentiment of the sentence, with the largest probability being assigned to rating 0, with a longer tail towards intermediate ratings.

Earthbound: "*I am blown away. That was one of the craziest games I've ever played. Sure, it has flaws but I think it does belong on the list of mandatory Super Nintendo games*" (AVGN, 2018).

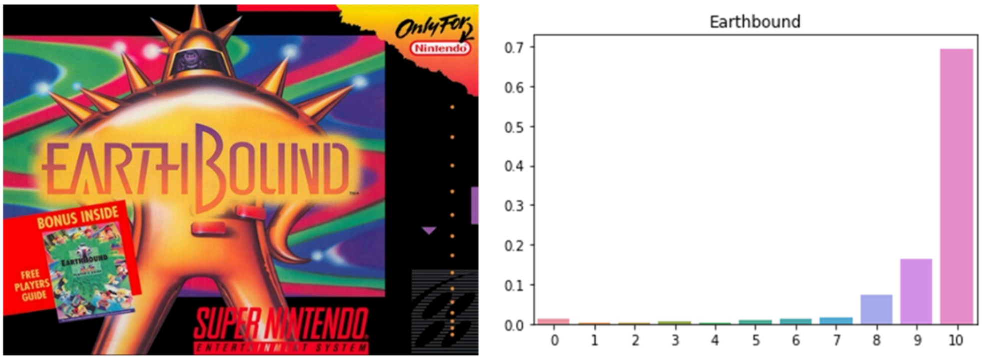


Figure 7. Cover art (source: Wikipedia) and predicted game scores for the game *Earthbound*.

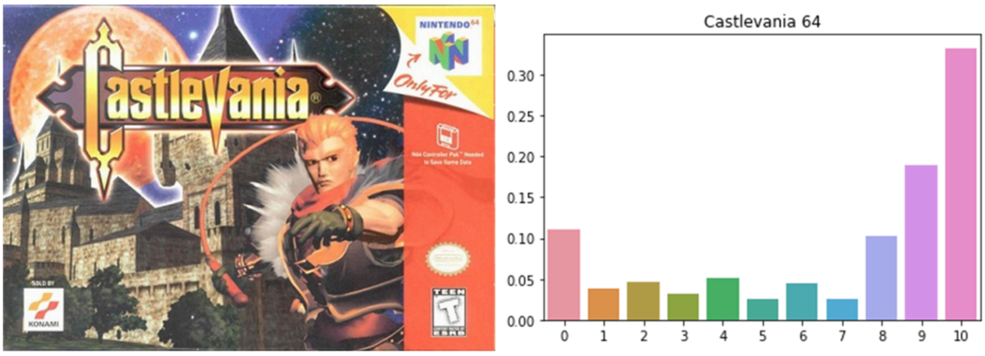


Figure 8. Cover art (source: Wikipedia) and predicted game scores for the game *Castlevania 64*.

Here we can see that the model is very sure of the positive sentiment of the review, assigning most of the probability to a 10 rating.

Castlevania 64: "*The graphics are good, for Nintendo 64 standards, but I find them unappealing, because it's the beginning of the 3D age, and they haven't perfected it yet. It's that awkward period between the old and the new*" (AVGN, 2009).

This is a mixed review, and it shows one of the flaws that the model has right now. Note how there are peaks on both rating 10 and rating 0, with a valley in between. Isn't this weird? How can the model assign a high probability for both 10 and 0, and not to anything in between?

This happens because the model has no idea that there is an order associated to the ratings. It has no idea that if the probability for a rating 10 is high, the probability for a rating 0 should be low. It treats the ratings as an unordered categorical scale, also known as a nominal scale. Below, I discuss how we can make the model aware of the order of the scale, and what trade-offs that entails.

Model design choices

As we saw in the previous section there are several issues with the current model: poor performance overall vs. SOTA benchmarks; poor Recall and Precision on intermediate ratings; unawareness of the ordinal nature of the ratings.

All these issues stem from design decisions we made when building our model. In this section, I will present which decisions those are and discuss options to improve the model.

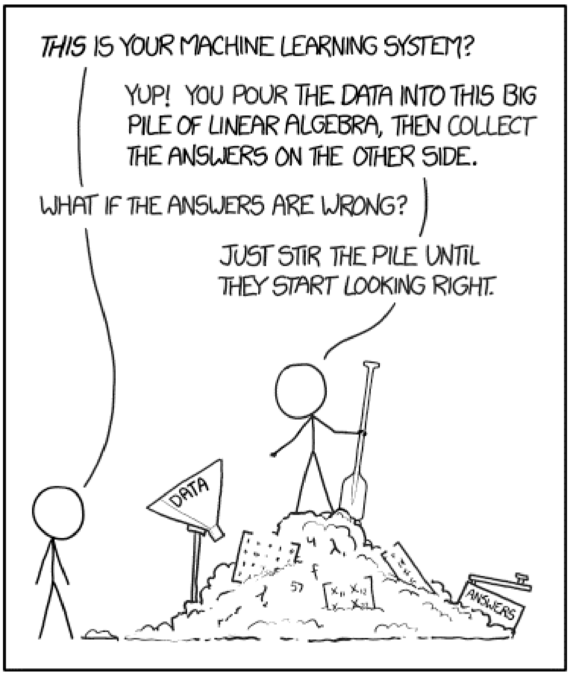


Figure 9. The XKCD model design approach (source: https://xkcd.com/1838/).

Target variable measurement scale

Not all measurements are created equal. Consider the following measurements associated to myself:

- my country of residence is Brazil;
- I am the eldest son in my family;
- I live near latitude -23.6 and longitude -46.7.
- as of the writing of this article, I am 32 years old.

There are different things that I can do with each of those measurements. I can compare my country of residence to another person, but I cannot calculate what "twice my country of residence" would be. You can know that my age is greater than my brothers', but without any extra information you cannot know by how much. These are examples that show that there are different types of measurements, and it is useful to be aware of that when building machine learning models.

There has been some work in statistics and measurement theory to create definitions for the different types of measurement. For instance, Stevens (1946) proposed a four-level measurement scale (Nominal, Ordinal, Interval, and Ratio). Other researchers, such as Mosteller & Tukey (1977) and Chrisman (1998), proposed more sophisticated classifications, with 7 and 10 different levels, respectively. I found in practice however that Stevens' taxonomy works well to discuss machine learning. But what exactly is each kind of measurement?

- **Nominal:** nominal measurement scales differentiate between values based on their identity. Other than that, no other comparisons can be done on measurement scales. For example, you cannot rank order them or calculate the difference between them. In the examples above "Country of residence" is a nominal scale variable. You can say that Brazil is different from the United States of America, but you cannot rank order them[8] or calculate the difference between Brazil and USA.[9] Other examples of nominal scales are Gender, Language, and Favorite Book.

- **Ordinal:** Ordinal measurement scales are like nominal scales, but they have an intrinsic order associated to them. You still cannot calculate the difference between them, but you can determine if one is greater than another, allowing one to rank order them. In the examples above "eldest" is an ordinal measurement scale. You know that by being the eldest my age is greater than my middle and youngest brothers', but you cannot know by how much. Other examples of ordinal scales are Likert scales that are commonly used in surveys to measure agreement level, and star ratings on Amazon.com reviews.

- **Interval:** Interval measurements are something that most of us might call a numerical measurement. We can not only compare and rank them, but also calculate the difference between two

[8] You can rank order them on other associated measurements, such as GDP, population, or HDI, but in those cases the measurements being ranked are those indices, not the countries themselves.

[9] Although you could argue that this difference is at least a couple of caipirinhas.

values. However, interval scales have an arbitrary zero value and, as such, their ratios are not meaningful. My location in latitude and longitude is an example of an interval scale. You can say that the difference between my latitude and someone located in Cambridge, MA, is 66 degrees. But it makes no sense to say that the ratio between my latitude and the latitude of someone in Cambridge, MA, is -0.56. Other examples of interval scales include temperatures on both Fahrenheit and Celsius scales.[10]

- **Ratio:** Ratio measurements are like interval measurements, but their scale has a well-defined and usually non-arbitrary zero scale so that calculating ratios make sense. In the examples above, my age is a ratio scale. It makes sense to say that I am twice as old as my brother. Other examples of ratio measurements include income and temperatures measured on the Kelvin scale.

Now that we know the four types of measurement scales, which one do you think best applies to videogame ratings in our dataset? We can say that one rating is greater than another, so nominal is out of the picture. But is the difference between two ratings meaningful? And more than that, is it consistent across the scale? Consider the following two completely unrelated and hypothetical cases:

- after much deliberation, you decided to increase your rating for this article from 2/10 to 3/10;
- after much deliberation, you decided to increase your rating for this article from 9/10 to 10/10.

Do you feel that the increase in rating in both cases is the same? Most people would argue no. The first increase changed the article from a bad article to a slightly "less bad" article. The second case, on the other hand, elevated it from a very good article to perfection! (Thank you, by the way.) However, people calculate metrics such as average scores all the time and, strictly speaking, you should never do that to ordinal scales! What gives?

The fact of the matter is that this is a controversial topic (see Knapp, 1990) into which we will not delve further. In this article we will compare the choice between modelling ratings as a nominal variable (in which the model is unaware of order) and as an ordinal variable (in which order is considered). It is possible to also model this target variable as an interval scale, although we need to take some extra care to avoid out of domain problems for our predictions (for instance, our model assigning a rating of 13 or -3 for a game).

Class weights

Review ratings (outside specialized media) have an exceedingly unbalanced distribution, as it is quite common for people to give a game a 10 if they liked it, or a 0 if they disliked it. This makes our model care more about getting 0's and 10's right than getting other ratings right, as we saw in the precision and recall metrics for our simple model (Fig. 5). Although this is the approach that maximizes overall accuracy, this can sometimes lead to useless models (see Box 1 for an example).

One way to correct for this is to assign weights for each observation, increasing the importance of the ones that are less frequent. This way, even though there are less reviews with a rating of 7, getting one of them wrong will "hurt" more (from a loss function standpoint) than getting a score of 10 wrong.

Take note that this will decrease our model's accuracy on the unbalanced dataset, but it will probably yield a more useful model in the end.

[10] And the Rankine, Rømer, Delisle, Réaumur or any other weird temperature scale that you might be using and that is not Kelvin.

Box 1. When can maximizing accuracy be bad for your model?
We discussed above how maximizing accuracy might not be in our best interests and might even produce useless models. But how come is that? Isn't higher accuracy always better? Not in the case of highly unbalanced datasets. Imagine you are building a machine learning classification model to detect a rare pulmonary disease based on thorax X-ray images. Less than 0.1% of people examined have the disease, which is a highly unbalanced dataset, but you decide to ignore my warning and just feed the raw data to the model. You get a model with 99.9% accuracy, which seems amazing! Until you realize it is just predicting that no one has the disease. This model might have amazing accuracy, but it is completely useless for diagnostics.

Model architecture

Another choice when designing our models is the architecture that will be used for our neural network. The architecture describes how each individual neural connects to one another. Good architecture allows us to reduce the model's size by exploiting some feature of the problem being addressed.

In previous sections, we hand-waved our simple model architecture, just saying it was a Recurrent Neural Network (RNN). In this section we explain this architecture in more detail and also introduce two other architectures: the Long Short-Term Memory (LSTM) and the Gated Recurrent Unit (GRU).

- There will be a little bit of math on this section, so we better get our notations straight:
- upper case letters represent 2-D tensors (also known as matrices);
- lower case letters represent 1-D tensors (also known as vectors);
- the operator represents the Hadamard product (also known as element-wise multiplication);
- the symbol represents the hyperbolic tangent function;
- the symbol represents the sigmoid function.

With that out of the way let us describe our model architectures.

Recurrent Neural Networks (RNN)

The Vanilla RNN is one of the simplest examples of a sequence model there is. The model has a hidden state h_t which is updated at each time step of the sequence, based on the input value at that time (x_t: for language models, this is usually some form of embedding representation of the input tokens) and on the value of the hidden state from the previous step (h_{t-1}). This update is done by the following expression:

$$h_t = \phi_h(W_{ih} \cdot x_t + b_{ih} + W_{hh} \cdot h_{t-1} + b_{hh})$$

where: W_{ih} and b_{ih} are tensors that describe how the input updates the hidden state; W_{hh} and b_{hh} are tensors that describe how the previous hidden state updates the current one.

These tensors are shared among all time steps in the sequence. This update rule can be represented by the following computation graph (Fig. 10; the bias terms b_{ih} and b_{hh} were omitted for brevity).

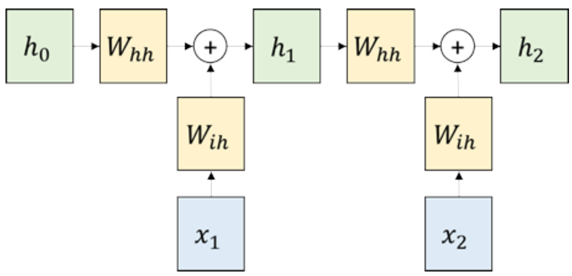


Figure 10. RNN computation graph.

Note that at each time step, the new state is calculated as a combination of the current input and the previous state. This way, the model has an internal memory that allows it to remember elements seen in the past. However, Vanilla RNNs have poor performance on long sequences due to its inability to "judge" if an input needs to be remembered or not.[11] For a Vanilla RNN both the word "and" (a common uninformative stop word) and the word "awful" (a highly informative word for game reviews) are the same in terms of whether they should be remembered by the internal hidden state.

Long Short-Term Memory (LSTM)

LSTM neural networks are a recurrent architecture proposed by Hochreiter & Schmidhuber (1997) to improve on the long-term dependencies problem seen in vanilla RNNs. This is done by introducing an internal memory cell c_t and some update gates:

- the input gate i_t produces a scalar between 0 and 1 that judges how much influence the input should have on the internal memory cell. One can interpret this value as "what percentage of the input should I keep on the internal cell state?"

- the forget gate f_t produces a scalar between 0 and 1 that judges how much influence should the previous cell memory state have on the new internal memory cell state. One can interpret this value as "what percentage of the previous cell state should I keep?"

- the output gate o_t produces a scalar between 0 and 1 and judges how much influence should the internal memory cell have on the output value (the hidden state h_t). One can interpret this value as "what percentage of the cell state should I output?"

Box 2. Sigmoids and hyperbolic tangents? What the heck are those?

If you crossed paths with neural networks before you might have noticed that hyperbolic tangents and sigmoids show up a lot. Ever wondered why that is?

Hyperbolic tangents

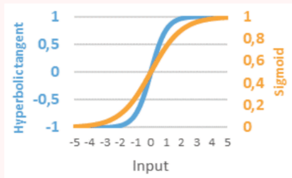
$$\phi_h(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Neural networks work due to the presence of non-linearities between neurons. One common non-linearity (Inspired by the way neurons on the brain work) is an activation non-linearity – That is, the neuron does not respond to a stimulus until it reaches a certain threshold. The hyperbolic tangent function has this behavior and is differentiable and because of that it is widely used in neural networks.

Sigmoids

$$\sigma(x) = \frac{e^x}{1 + e^x}$$

Although sigmoid have a similar shape to hyperbolic tangents their use in neural networks is due to their [0, 1] domain, being the activation function of choice when dealing with probabilities.



These gates and their dynamic can be represented by the following expressions. Note how the use of the sigmoid function guarantees the scalar [0, 1] domain on the output of each gate.

$$i_t = \sigma(W_{ii} \cdot x_t + b_{ii} + W_{hi} \cdot h_{t-1} + b_{hi})$$
$$f_t = \sigma(W_{if} \cdot x_t + b_{if} + W_{hf} \cdot h_{t-1} + b_{hf})$$
$$o_t = \sigma(W_{io} \cdot x_t + b_{io} + W_{ho} \cdot h_{t-1} + b_{ho})$$
$$g_t = \phi_h(W_{ig} \cdot x_t + b_{ig} + W_{hg} \cdot h_{t-1} + b_{hg})$$
$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$
$$h_t = o_t \odot \phi_h(c_t)$$

I will not describe all tensors here, as the notation is analogous to the one used for RNNs. Again, all weight tensors W and b are shared among all time steps. However, as the gates depend on the input and on the hidden state, the LSTM can learn to weight the importance of different inputs and

[11] This is an oversimplified analogy, but I find it helpful to understand how clever design of networks and exploiting characteristics from your application can facilitate training. In theory, Vanilla RNNs can model arbitrarily long-term dependences on the input sequence. However, the finite precision of computers leads to numerical problems when training them via back-propagation, as the error gradients tend to vanish or explode as we move through time.

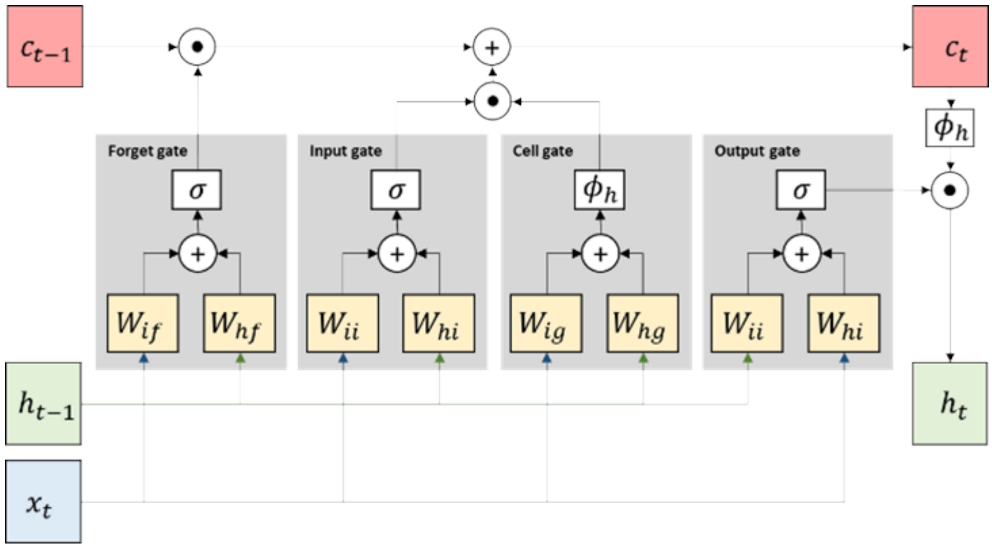


Figure 11. LSTM computation graph. For brevity I represent only a single time step here and, as before, omit the bias tensors.

which inputs are worth remembering. As with RNNs, we can represent the update rule with a slightly more complicated computation graph (Fig. 11). LSTMs have seen a lot of success in a wide range of applications, from speech recognition to beating human players in the popular game Dota 2 (OpenAI, 2019).

Which models are you training, after all?

After all the considerations made in this section, I am finally ready to present which models I ran for this article. I decided to work with the following options for the design decisions we just discussed:

- **Target measurement scale:** Nominal and Ordinal;
- **Class weights:** Unbalanced and Balanced;
- **Model architecture:** Vanilla RNN (hidden size = 256) and LSTM (hidden size = 128).

I will try all combinations between these design decisions, yielding a total of 8 different models, training every combination for 20 epochs. Note that I am using different hidden sizes between the RNN and LSTM. I did this to keep model capacity constant between architectures so that we can attribute any improvement to the change in architecture itself. If we do not do this, we would not be able to distinguish between an improvement due to the architecture and an improvement due to the increase on the number of weights of the model.

RESULTS

After training all 8 models I can pick the best one before scoring the Nerd's reviews. To that end, I will inspect the model level metrics and class level metrics we discussed above.

Model performance

Selecting the best epoch

For each model, I needed to select the best epoch before comparing their performance. This happens because although the performance on the training set will usually get better and better as you train your model, the same cannot be said about the performance on the validation set, which is the one that matters. After a while, performance on the validation set can start to de-

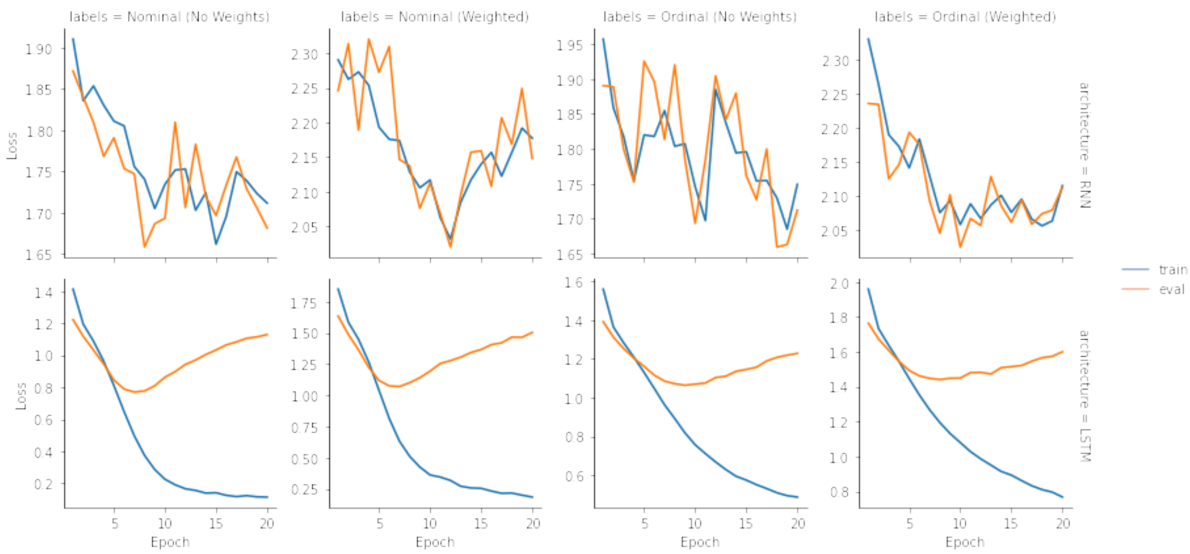


Figure 12. Loss over training time for RNN and LSTM.

grade, which is an indicative that you might be overfitting your training data that and getting worse at generalizing to unseen data.

With that in mind, I used the Negative Likelihood loss value for that epoch as my selection criteria for the model. Figure 12 shows how this metric behaved over training time.

As we've seen with the simple model before, both training and validation Loss are highly volatile for models using the Vanilla RNN architecture (Fig. 12). This is a strong indicative that this model is too simple for our problem, and that it might take too much training time and data for it to achieve a good performance. The LSTM architecture, on the other hand, fared way better, displaying a trend that is common for machine learning models: a constant decline on the training Loss as the model gets better and better at predicting the data it already saw and a V-shaped behavior for the validation loss as after a point the model starts to overfit the data (Fig. 12).

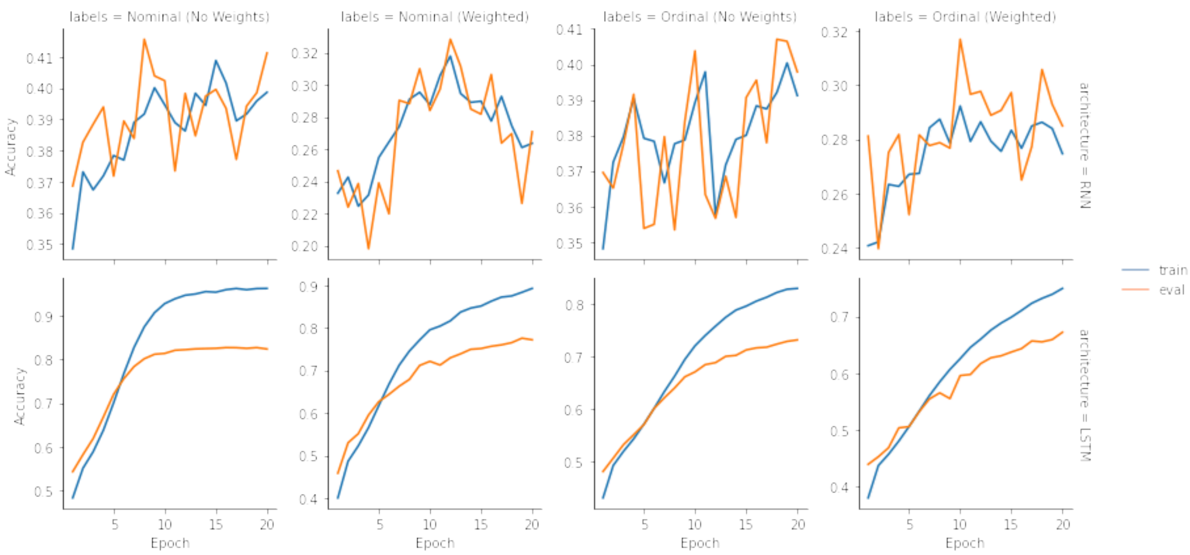


Figure 13. Accuracy over training time for RNN and LSTM.

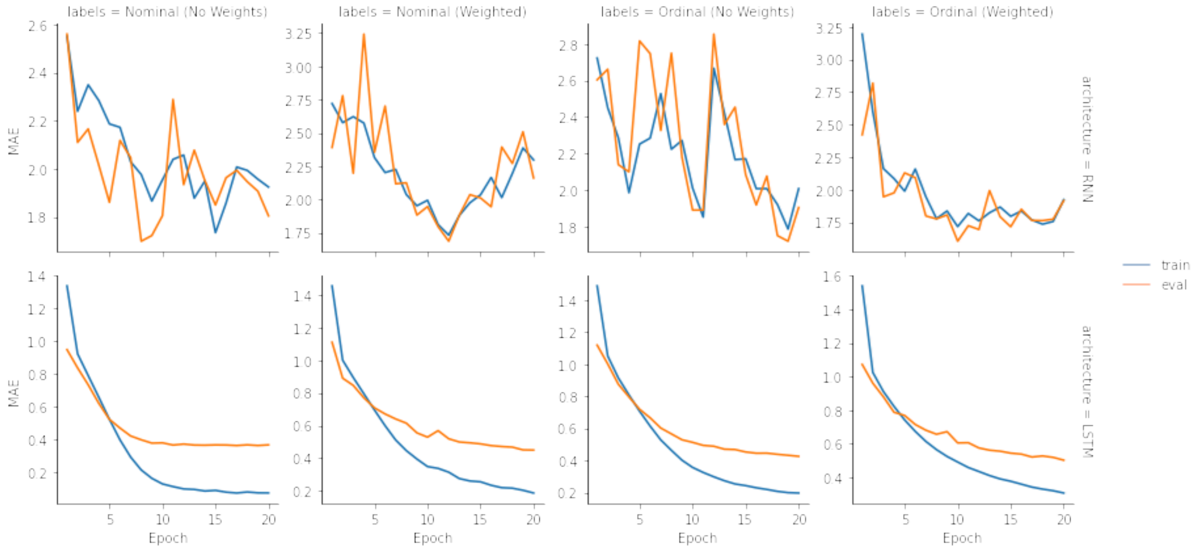


Figure 14. Mean Absolute Error (MAE) over training time for RNN and LSTM.

The best model will be the one with the lowest validation loss. So, let us also take a look on the Accuracy and Mean Absolute Error (MAE) metrics before continuing (Figs. 13 and 14, respectively).

Notice that, unlike what was seen for loss both accuracy and MAE are stable or even improve for the validation set beyond the point in which the model started overfitting. This happens because Accuracy and MAE take into account only the final predicted value, while the loss also considers how confident the model is on the prediction. Take the following examples:

- true review rating is 7, and the model predicted 9 with 55% accuracy;
- true review rating is 7, and the model predicted 9 with 98% accuracy.

Accuracy and MAE would be equally impacted by these examples, while the Loss would be much more affected by the second example then by the first. As the model overfits the training data it tends to get ever more confident in its predictions, which in turn makes its wrong predictions "hurt" more and more. For this article I opted to pick the epoch with the best loss for each model. The results are summarized in Table 6.

Selecting the best model

Now that we have the best epoch for each model, we can proceed to select the best overall model. Since we want it to have good performance predicting all classes, I will use the precision and recall metrics,

Table 6. Best epoch for each model.

Architecture	Scale	Balanced?	Epoch	Loss	Acc.	Acc+1	MAE
RNN	Nominal	No	8	1.66	41.6%	65.0%	1.70
RNN	Nominal	Yes	12	2.02	32.8%	61.0%	1.69
RNN	Ordinal	No	18	1.66	40.7%	63.1%	1.75
RNN	Ordinal	Yes	10	2.05	27.9%	54.0%	1.78
LSTM	Nominal	No	7	0.77	78.4%	91.3%	0.42
LSTM	Nominal	Yes	7	1.07	66.3%	87.2%	0.64
LSTM	Ordinal	No	10	1.07	67.2%	90.5%	0.52
LSTM	Ordinal	Yes	8	1.44	56.5%	88.2%	0.66

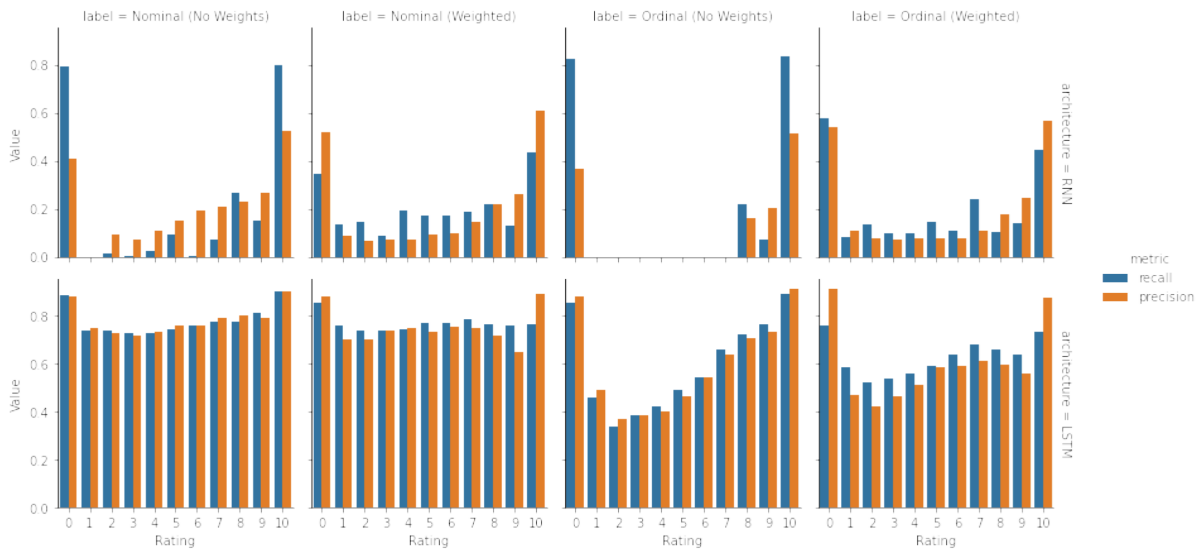


Figure 15. Precision and recall for each model.

combined into a single metric using a Macro-averaged F1-score described by the formula below:

$$F_1 = \frac{2}{N} \cdot \sum_{i=1}^{N} \frac{P_i \cdot R_i}{P_i + R_i}$$

where *Pi* and *Ri* are the precision and recall for score class *i*, and *N* is the overall number of classes. (The macro-averaged F1-score is essentially the simple average between the individual F1- scores for each class, which in turn is the harmonical mean between precision and recall.) To calculate that let us check what are the values for precision and recall for all the models (Fig. 15).

We can immediately see – even before calculating any metrics – that the LSTM Nominal models perform better than all other variations (Fig. 15). Interestingly this architecture seems to have been less affected by the unbalanced dataset than the others, though it is hard to tell by inspecting the chart whether the balanced or unbalanced model performs better. Table 7 shows the F1-score for each model.

The unbalanced LSTM Nominal had the best performance based on the F1 criteria (Table 7). But the difference between it and the ordinal model was not that big; and remember, our motivation for testing an ordinal model was to make our predictions reflect better the ordinal nature of reviews. Let us revisit those three examples from above before making any decisions.

Table 7. F1-score by model. (For some models, recall or precision were both 0, which makes the F-1 undefined. In this case I considered the F-1 score to be zero for that class).

Model architecture	Scale	Balanced?	F-Score
RNN	Nominal	No	0.19
RNN	Nominal	Yes	0.19
RNN	Ordinal	No	0.13
RNN	Ordinal	Yes	0.18
LSTM	Nominal	No	0.78
LSTM	Nominal	Yes	0.76
LSTM	Ordinal	No	0.59
LSTM	Ordinal	Yes	0.61

Dr. Jekyll and Mr. Hyde (Fig. 16): "*You'd think I'm jokin', like I'm trying to be funny or somethin'. But, no, the fact that that game exists is a horrible abomination of mankind. That game is so freaking horrible, and I am not kidding*" (AVGN, 2010).

Earthbound (Fig. 17): "*I am blown away. That was one of the craziest games I've ever played. Sure it has flaws but I think it does be-*

long on the list of mandatory Super Nintendo games" (AVGN, 2018).

Castlevania 64 (Fig. 18): "*The graphics are good, for Nintendo 64 standards, but I find them unappealing, because it's the beginning of the 3D age, and they haven't perfected it yet. It's that awkward period between the old and the new*" (AVGN, 2009).

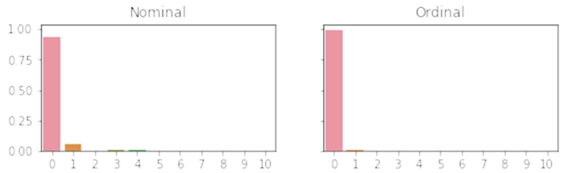


Figure 16. Predicted scores for *Dr. Jekyll and Mr. Hyde* on both the nominal and ordinal model.

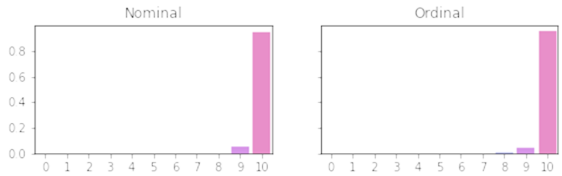


Figure 17. Predicted scores for *Earthbound* on both the nominal and ordinal model.

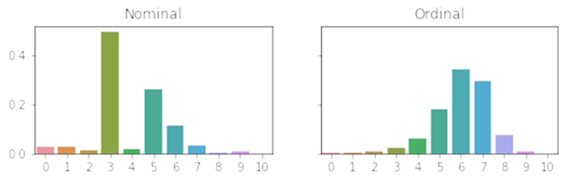


Figure 18. Predicted scores for *Castlevania 64* on both the nominal and ordinal model.

For both *Dr. Jekyll and Mr. Hyde* (Advance Communication Co., 1988) and *Earthbound* (Ape / HAL Laboratory, 1994) we see remarkably similar output from both models. That is to be expected; both are very polarized reviews. However, the output for *Castlevania 64* (Konami Computer Entertainment Kobe, 1999) is different, and we see from this example that the model is better able to take into account the ordinal nature of the data. Although this did not improve the accuracy of the model (in fact, it decreased exact accuracy by almost 10 percentage points), I decided that it was a price worth paying, especially when considering that our off-by-one metric is very similar on both models.

Predicting ratings for *Angry Video Game Nerd* reviews

Well, it has come to this. After all this work we will finally be able to assign proper review scores to the reviews made by the *Angry Video Game Nerd*. For this section, I used the episode transcripts available on the AVGN Wiki (https://avgn.fandom.com/). As for which episodes to test, I opted to score the Top 10 AVGN episodes, as selected by the Nerd himself (Table 8). I also added some of my favorite episodes that were not on the top 10 (Table 9).

So, how did it go? Figure 19 shows the predicted probability for each review, as well as the expected value calculated by averaging the classes.[12]

Table 8. Top 10 AVGN episodes.

#	Episode name	URL
1	R.O.B.	https://www.youtube.com/watch?v=vYm_UaYVSzc&t=0s
2	Mario 3	https://www.youtube.com/watch?v=Y5wK5Z23hoo&t=0s
3	Mega Man	https://www.youtube.com/watch?v=Q3iEn5rzMnw&t=0s
4	Jekyll and Hyde Re-Revisited	https://www.youtube.com/watch?v=EjXn5qiM8Zw&t=0s
5	Crazy Castle	https://www.youtube.com/watch?v=wai9CnvatBo&t=0s
6	Ninja Gaiden	https://www.youtube.com/watch?v=6t2YvyLqw3c&t=0s
7	Seaman	https://www.youtube.com/watch?v=-IV8hCvsXy0&t=0s
8	How the Nerd Stole Christmas	https://www.youtube.com/watch?v=iMINBv_Dqvs&t=0s
9	Berestein Bears	https://www.youtube.com/watch?v=LB3CybXl8rs&t=0s
10	Dick Tracy	https://www.youtube.com/watch?v=t9nxIUhZCCw&t=0s

[12] Yes, exactly what I said you cannot do with an ordinal variable. But I did it anyway. It is a useful metric! Now you see why this is such a hotly debated topic.

Table 9. Some of my favorite AVGN episodes, listed in no particular order.

#	Episode name	URL
11	Hong Kong 97	https://www.youtube.com/watch?v=M_aXcH1zDEE
12	Plumbers Don't Wear Ties	https://www.youtube.com/watch?v=DyaF_gCKWsI
13	Castlevania (Part 1)	https://www.youtube.com/watch?v=Hfo6hoN0PUw
14	Battletoads	https://www.youtube.com/watch?v=UD7k4mTThLY
15	Ghost N' Goblins	https://www.youtube.com/watch?v=94Y6y1MOoEo
16	Earthbound	https://www.youtube.com/watch?v=LZ5nX0FTH6Q
17	Silver Surfer	https://www.youtube.com/watch?v=gvnRBywkUZ0
18	Ikari Warriors	https://www.youtube.com/watch?v=bByE7n2AJj4
19	Big Rigs: Over the Road Racing	https://www.youtube.com/watch?v=h6DtVHqyYts
20	Tiger Electronic Games	https://www.youtube.com/watch?v=_u5dtBtG9yU

Figure 19. Angry Video Game Model predictions for *Angry Video Game Nerd* reviews.

Well, most things seem to make sense. Games such as *Mario 3* (Nintendo R&D4, 1988), *Mega Man* (Capcom, 1987) and *Earthbound* have almost perfect scores, which make sense given the praise the Nerd gives them in the reviews. On the other hand, games such *Hong Kong 97* (HappySoft, 1995) and *Big Rigs: Over the Road Racing*

(Stellar Stone, 2003) have terrible scores. But wait… what is this? *Dr. Jekyll and Mr. Hyde* has a score of 5.4? This cannot be right. And what about *Castlevania* (Konami, 1986) getting a score of 1.2? The Nerd had mostly praise for this game. What is happening here? WHAT IS THE MODEL THINKING?

Figure 20. What were they thinking? (source: AVGN).

To answer this question, I had to dig deeper. Something in those reviews is throwing off the model.

What is the model thinking?

Machine learning models' predictions are extremely complicated to explain and this one is no different. The model is essentially a black box, and a lot of effort has been put into it to understand why it made a prediction. Machine learning is being used to allocate investments, predict credit risk, score test results, and even to drive cars. In all these applications, being able to explain a model decision is highly desirable, be it from a legal standpoint (to explain to a customer why they were denied credit) or from a safety standpoint (to understand why an autonomous vehicle thought it could drive through a barn[13]).

There is a lot of research being done in model explainability. Some works, such as the LIME method proposed by (Ribeiro et al., 2016), work by deriving proxy models; that is, simpler, linear and locally bound approximations of the full model that can be easily explained. Another approach is to use salience mapping, in which parts of the input are occluded from the model, and the variations in the predicted output can give us some insights on what is influencing the prediction. The latter approach works well for our problem, as we can feed parts of a longer review to the model to get a localized sentiment for a sentence.

To investigate the predictions of our model, I ran them again multiple times. Each time, a small sliding window of 5 lines was cut out from the review, and the expected score was calculated by the model using only the text in the sliding window ("Only") and using everything in the review but the text on the sliding window ("Except"). Figure 21 shows the results of this analysis for *Dr. Jekyll and Mr. Hyde* and *Castlevania* (part 1).

We can see that the predicted sentiment can vary a lot throughout the review, as

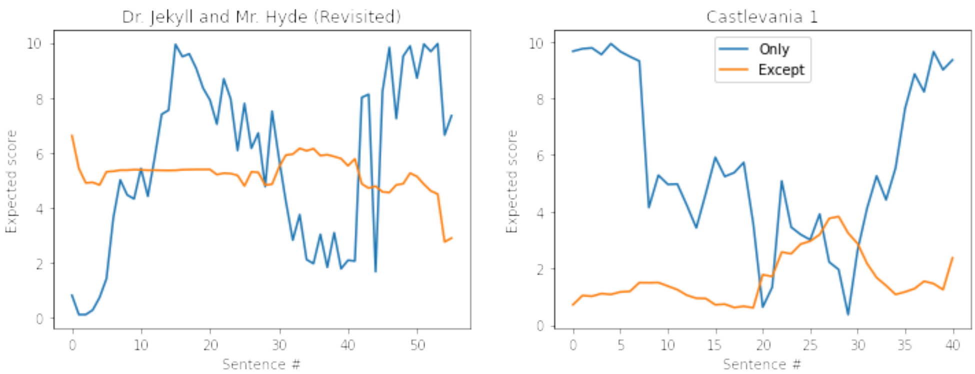


Figure 21. Salience mapping for two AVGN episodes.

[13] Maybe it was trained using *Big Rigs: Over the Road Racing*?

shown by the wild variations observed in the blue line. However, the removal of a couple of five lines usually has little effect on the review, with a few exceptions. In *Dr. Jekyll and Mr. Hyde*, removing the beginning of the review significantly increased the score, while removing the ending significantly decreased it. Let us take a close look on those sentences: "[The episode begins with a black-and-white clip; the first few seconds from the original Dr. Jekyll and Mr. Hyde review from 2004. The Nerd's voice can be heard over this.] *In May of 2004, I gave a warning about a game called Dr. Jekyll and Mr. Hyde. I made it perfectly clear: DO NOT PLAY THIS GAME. But from what I understand... people have played it! They didn't listen. But it wasn't their fault... I only showed about one minute of footage from the game, and even though I talked about it at great length, it didn't do any good...* [The Nerd drinks some Rolling Rock.] *I called it a piece of ****. I called it an awful pile of steaming goat ****. But that was honoring it. I could've said anything, it wouldn't have mattered. I could've taken a **** on it, but my own **** would have been offended to lay on this loathsome piece of FILTH! Just the thought of covering this thing in doo-doo is like encasing it in gold! I curse the day I ever laid eyes on it. I curse the plastic that holds this abomination. My words are insufficient in describing the total insult to humanity that this "game" has provided! Everything that I've ever said and anything that anybody else has ever said is NOT enough! It MUST be shown.* [He drinks more Rolling Rock.]" (AVGN, 2010).

Well, okay. I see why this might have such a large effect on the expected score. No surprises here. Now let us look at the ending sentences: "*No...!* [The scene fades to black and fades back in a blur. The Jekyll-to-Hyde transformation music from the game plays as the Nerd wakes up, back in his room, where he first transformed. He resumes playing the game, and has an epiphany.] *The Nerd: I think I get it. Why, it's the best game ever made. It's more than a game... it exposes the dual nature of the human spirit. The only way to win the game is to be Jekyll, but you wanna be Hyde so you can shoot ****. You see, it's a constant battle between good and evil, and Jekyll must stay farther along his path than Hyde. If Hyde gains the lead, then evil will triumph over good, and that's the true conflict to the human soul. And to deny the evil completely, would only force it into the subconscious mind, like a city broken into different social classes. People don't wanna step outside their own boundaries, like Jekyll wandering into the wrong section of town. He's unwelcome. Nevertheless, he must abide by his own good nature. No wonder the cane doesn't work. The game does not reward you for acting upon your malevolent intentions. It's a proposed guideline for a set of morality rules to be programmed in real life! It uses the Victorian era as a fundamental depiction of outward respectability and inward lust. It's a metaphor for social and geographical fragmentation. It eludes the Freud theory of repression, in which unacceptable desires or impulses are excluded from the conscious mind, and left to operate on their own... in the unconscious.*" (AVGN, 2010).

Here we can see something interesting. Taken at face value, the Nerd's words seem to sing high praise for the game. He makes it sounds almost like a transcendent experience. But it is all good old sarcasm. Apparently, the model did not learn enough to be able to identify the true meaning of the Nerd's words.

The *Castlevania* example is less interesting. The Nerd starts talking about how good it was when it was launched, and the impact it had in his life. But after that he proceeds to talk about all the frustrating parts of the game, and the model seem to have found that part more relevant. The bump in the orange line correspond to comments by the Nerd on puns made on the game credits sequence. Apparently, the model really hates puns and the reviews are being pun-inshed for it. "*Hmm...Trans Fisher? It reminds me of Terence Fisher, the director of many of the* Hammer Horror *films. That's a funny coincidence. Oh wait... Vran Stoker? Like Bram Stoker, the author of* Dracula*? Wha-- Christopher Bee?! Is it a joke? I don't get it. Are they saying Christopher Lee is like a bee?* [Bee with a face like Christopher Lee's comes buzzing by] *No, they can't mean that. This is probably just a series of strangely coincidental typos.* [The Nerd notices another name] *Belo Lugosi? Boris Karloffice? They're just ******* around. Love Chaney Jr.? Mix Schrecks? Green Stranger?! Is this supposed to be funny? Like just take a celebrity's name and change it around? That's like if I took the name 'Stephen Spielberg', and called him 'Stephen Jeelberg'. Like, that's not funny, that's kindergarten level! No, kindergarten students don't find that funny! Aliens don't find that funny! Well anyway, that's Castlevania for you. Good game, but holy **** is it hard. Now as promised, we're gonna plow through the rest of 'em, all the old-school Castlevania games. The ones that I grew up with –* [The 'What a horrible night to have a curse' box from *Castlevania II: Simon's Quest* appears in front of the Nerd, interrupting him. The box disappears a few seconds later, and a day-to-night transition in the style of said game is shown. The nighttime music plays and the Nerd's room looks darker than before. The Nerd notices the *Castlevania II: Simon's Quest* cartridge]" (AVGN, 2009).

CONCLUSION

Can a model truly capture the full emotion of an AVGN review? Maybe time will tell. In this article, however, I demonstrated that Recurrent Neural Networks, in particular the LSTM architecture, have a good performance on the videogame review rating prediction task when compared to other sentiment analysis benchmarks. I also demonstrated that the model trained on this dataset can produce coherent ratings[14] for the reviews performed by the Angry Video Game Nerd, barring issues of dealing with sarcasm.

FURTHER WORK

As discussed in the start, this work focused on simpler model architectures to make it more approachable. More advanced architectures such as Transformers might be able to better deal with long term dependences. I also only trained this model on the review dataset data, and I did not make use of transfer learning at all. Pretraining on a larger and more general dataset might improve the quality of the embeddings. Testing contextual embeddings such as BERT (Devlin et al., 2018) might also improve how the model deals with context dependence ambiguity. Finally, training the model on longer reviews might improve its performance on the AVGN dataset, as the length of an AVGN episode is several times longer than the average length of a user review.

REFERENCES

Anonymous. (2021) Sentiment analysis on IMDB. Papers with Code. Available from: https://paperswithcode.com/sota/sentiment-analysis-on-imdb (Date of access: 03/Jan/2022).

Alammar, J. (2019). The Illustrated Word2vec. GitHub. Available from: https://jalammar.github.io/illustrated-word2vec/ (Date of access: 03/Jan/2022).

Chrisman, N.R. (1998) Rethinking levels of measurement for cartography. Cartography and Geographic Information Systems 25: 231–242.

Cinemassacre Productions LLC. (n.d.) Cinemassacre FAQ. Available from: https://cinemassacreold.website-us-east-1.linodeobjects.com/faq/ (Date of access: 06/Apr/2021).

Delany, S.J.; Buckley, M.; Greene, D. (2013) SMS spam filtering: methods and data. Expert Systems with Applications 39: 9899–9908.

Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. (2018) BERT: pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv. Available from: https://arxiv.org/abs/1810.04805 (Date of access: 03/Jan/2022).

Eberhard, D.M.; Simons, G.F.; Fennig, C.D. (2021) Ethnologue: Languages of the World. SIL International, Dallas.

Gelbart, B. (2019) A history of review bombing. Gamerant. Available from: https://gamerant.com/mass-effect-borderlands-3-modern-warfare-review-bombs/ (Date of access: 06/Apr/2021).

Hochreiter, S. & Schmidhuber, J. (1997) Long

[14] This is opinion. Unfortunately, without ground truth ratings provided by the Nerd himself we might never be able to go beyond such qualitative statements.

short-term memory. Neural Computation 9: 1735–1780.

Kim, A. & Liao, S. (2019) Why fans aren't happy with Pokémon Sword and Shield developer Game Freak. CNN. Available from: https://edition.cnn.com/2019/11/16/tech/nintendo-pokemon-sword-shield-trnd/index.html (Date of access: 06/Apr/2021).

Kingma, D.P. & Ba, J. (2015) Adam: a method for stochastic optimization. ICLR, San Diego.

Knapp, T.R. (1990) Treating ordinal scales as interval scales: an attempt to resolve the controversy. Nursing Research 39: 121–123.

Kocijan, V.; Lukasiewicz, T.; Davis, E.; Marcs, G.; Morgenstern, L. (2020) A review of Winograd schema challenge datasets and approaches. arXiv. Available from: https://arxiv.org/abs/2004.13831 (Date of access: 03/Jan/2022).

Lau, J.; Cohn, T.; Baldwin, T.; Hammond, A. (2020) This AI poet mastered rhythm, rhyme, and natural language to write like Shakespeare. IEEE Spectrum. Available from: https://spectrum.ieee.org/artificial-intelligence/machine-learning/this-ai-poet-mastered-rhythm-rhyme-and-natural-language-to-write-like-shakespeare (Date of access: 03/Jan/2022).

Levesque, H.J.; Davis, E.; Morgenstern, L. (2012) The Winograd schema challenge. Thirteenth International Conference on Principles of Knowledge Representation and Reasoning 2012: 552–561.

LotrProject. (n.d.) Word count and density. LotrProject. Available from: http://lotrproject.com/statistics/books/wordscount (Date of access: 06/Jun/2021).

Mosteller, F. & Tukey, J.W. (1977) Data analysis and regression: a second course in statistics. Addison-Wesley, Boston.

Nakatani, S. (2010). Language Detection Library for Java. SlideShare. Available from: https://www.slideshare.net/shuyo/language-detection-library-for-java (Date of access: 03/Jan/2022).

NLP-progress. (2021) Sentiment analysis. NLP-progress. Available from: http://nlpprogress.com/english/sentiment_analysis.html (Date of access: 03/Jan/2022).

OpenAI. (2019) Dota 2 with large scale deep reinforcement learning. arXiv. Available from: https://arxiv.org/abs/1912.06680 (Date of access: 03/Jan/2022).

Ribeiro, M.T.; Singh, S.; Guestrin, C. (2016) "Why Should I Trust You?": explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2016: 1135–1144.

Stevens, S.S. (1946) On the theory of scales of measurement. Science 103: 677–680.

Wikipedia. (2021). YouTube. Wikipedia. Available from: https://en.wikipedia.org/wiki/YouTube (Date of access: 06/Apr/2021).

Wiktionary. (2020) Statistics. Wiktionary. Available from https://en.wiktionary.org/wiki/Special:Statistics (Date of access: 06/Jun/2020).

Wu, Y.; Schuster, M.; Chen, Z.; Le, Q.V.; Norouzi, M.; Macherey, W.; et al. (2016) Google's neural machine translation system: bridging the gap between human and machine translation. arXiv. Available from: https://arxiv.org/abs/1609.08144 (Date of access: 03/Jan/2022).

Acknowledgements

I want to thank James Rolfe for creating the web series that inspired this work, and for providing hours of quality entertainment while I was growing up. I am still subscribed to his channel, and I do not aim to change that any time soon.

About the author

Henrique Soares is an engineer and machine learning enthusiast who, fortunately, grew up playing better games than the ones featured in AVGN. When he is not working on unconventional applications of machine learning, Henrique spends way too much time watching YouTube web series.

Journal of Geek Studies

jgeekstudies.org

Death, burial, and commemoration in *The Elder Scrolls*

Leon Corneille-Cowell

Department of Archaeology, University of York, York, UK.
Email: leoncowell96@gmail.com

Fantasy video games often find the player exploring and looting tombs of long-forgotten cultures and fighting otherworldly guardians. In this respect, they contain elements of archaeology albeit in an unscientific manner. *The Elder Scrolls* franchise is no different. This article will analyse the various burial practices in the *Elder Scrolls* games, finding parallels between several real-life cultures from medieval Europe to Ancient Egypt.

THE ELDER SCROLLS

The *Elder Scrolls* franchise has been running since 1994 and consists of eight games developed and published by Bethesda Softworks (1994–1998) and Bethesda Game Studios (2002–present). It is set in the fictional continent of Tamriel on the world of Nirn and has a rich lore and history behind it, with races and cultures adjacent to real-world cultures – a typical feature in the fantasy genre. The games follow an open-world roleplaying style in which the player character typically faces existential threats in the provinces of Tamriel, usually of divine origin from the lore's pantheon of deities.

The Elder Scrolls are full of historical comparisons: the Nords resemble the Viking age Scandinavia, the Imperials are the Romans, and the Breton are medieval Celtic peoples. The Wood-Elves and High-Elves have been compared to indigenous North

Figure 1. Concept art of a Nordic tomb (Nordic Ruins, by Adamowicz, 2008; Bethesda Softworks). Source: Bethesda Softworks, via The Unofficial Elder Scrolls Pages (https://en.uesp.net/).

Cite this article as: Corneille-Cowell, L. (2024) Death, burial, and commemoration in The Elder Scrolls. Journal of Geek Studies 11(1): 35-43. https://doi.org/10.5281/zenodo.10900278

Americans/ Mongolians and Imperial China respectively, with some Tolkien influence, while the Dark Elves bear a resemblance to the Ancient Near East, and the Red Guards to medieval North Africa. The Argonions resemble Mesoamerican cultures living in jungles and the Khajit are like a variety of cultures from nomadic cultures in Eurasia and settled ones from the Indian subcontinent (Schuhart, 2021).

Archaeology and roleplaying games are inextricably linked – player characters will enter ancient ruins and caves to seek ancient treasures and mysteries. These often include tombs and burial grounds (Fig. 1) full of the undead and draugr. In this respect, funerary archaeology or the burial of the dead is present everywhere in the game (Parker Pearson, 2010). This article will examine the funerary and burial practices in the *Elder Scrolls* games and look at real-world parallels. Archaeology is not new to the franchise, with both the in-canon excavations and texts as well as projects like the Skyrim Archaeological Survey (2020).

MATERIALS AND METHODS

This study was undertaken using the games *The Elder Scrolls: Morrowind* (2002), *Oblivion* (2006) and *Skyrim* (2011) as well as the associated download content (DLC) of each game. Comparison of in-game burial practices was done with real world practices using archaeological and historical literature.

DRAUGR AND DRAGON CULTS

The Draugr (Fig. 2) are undead warriors of the Nord race who most commonly inhabit Skyrim and culturally resemble early medieval Scandinavian cultures. They are followers of the dragon priests or nobility who are sealed inside tombs with them. An in-game book, *Amongst the Draugr* by Bernedette Bantien, which resembles a piece of anthropological/archaeological research on the Draugr observes these beings waking up daily to clean and worship the dragon priest, transferring energy to the priests to maintain their eternal afterlife (Bantien, 2012). This is a form of retainer sacrifice or the act of killing servants and burying them with their lords so they might join them in the afterlife. While human sacrifice was found in Norse burials, it is more common on a smaller scale than observed in Skyrim, and sacrifices were usually young women and often decapitated (Price, 2012: 266). Egyptian burials often contained figurines that functioned as servants for the afterlife, known as *shabti* (Taylor, 2001: 112). Another non-biological example of servants accompanying their overlords is the terracotta warriors of Qin Shi Huang in northern China (Man, 2007; Rawson, 2023: 348).

Figure 2. A Draugr during battle. Source: UESPWiki (screen capture from the game).

In the game, the dragon priests and levelled draugr bosses inhabit the main antechamber of the tombs are the remains of cult leaders or local nobility; Several of these are named and have associated quests. They are often guarding a wall covered in script resembling Norse runes, known as a 'Word Wall' (Williams, 2012).

In the in-game book *Cadaver Preparation Findings* by an Unnamed Necromancer, it is noted that the Draugr are mummified by removing certain organs that "no longer serve any function and only encumber the cadaver" and the skin is removed to prevent rot. Organ or viscera removal is a feature of Ancient Egyptian mummification; burial urns resembling Egyptian canopic jars can be found commonly in the tombs of Skyrim (Taylor, 2001: 54). The method of skin removal resembles no real-world methods of mummification; besides, Skyrim's cold climate would have contributed to the preservation of the skin as it did with the Iron Age Scythian Pazyrk burials, in which the mummies still had their skin and even tattoos (Parker Pearson, 2010: 61; Cunliffe, 2019: 199).

Draugr are typically found laid in grouped chambers horizontally in shelving built into the wall of tombs; the monumental tombs vary in form and size. They carry corroded iron axes and swords, resembling those used by Norse warriors (Pederen, 2012). There are 44 tombs on the map in Skyrim and Solstheim, with forms resembling temples (Bleak Falls Barrow), cities (Saarthal), and barrows (Shroud Heath Barrow) (Skyrim Archaeological Survey, 2020). Norse chambered tombs have been found but many of the more prominent examples are from the Neolithic and Bronze Age (Parker Pearson, 2010; Price, 2012: 263; Cooney, 2023). The complexity of the tombs found in Skyrim is likely to be a decision based on gameplay rather than practicality, but the closest real-world examples are that of the tomb of Qin Shi Huang and those of the Ancient Egyptians (Taylor, 2001: 150; Rawson, 2023: 353).

Red Eagle

Most of the uniquely named draugr or dragon priests (quest bosses) are of Nordic descent. The one exception to this is one named 'Red Eagle', associated with the quest 'The Legend of Red Eagle' (Elder Scrolls Fandom, 2023). Red Eagle is from an ethnic group known as the Reachmen from western Skyrim, a group who has clashed with the empire and the Nords on several occasions and followed their own cultures and practices – namely, the veneration of Hargravens (witch-like creatures) and the ritual raising of Briar hearts, a form of undead. Red Eagle bears similarity to the tale of King Arthur, who fought off invaders and promised to return as well.

Overall, despite the Nord culture's resemblances to early medieval Scandinavian and Germanic cultures. The combination of mummification, large-scale monumental tombs, and master-retainer sacrifice (notably only observed in Early Dynastic tombs), bears more in common with the burial practices of the Ancient Egyptians (Taylor, 2001).

MORROWIND ANCESTRAL TOMBS

Ancestor Tombs are found around Morrowind, usually guarded by Ancestor Ghosts or Undead. The player can access them to find loot. Ancestor worship is noted in most cultures in Tamriel but in Morrowind it is a particular point of contention between the tribunal religion of the settled towns (of whom the gods were once ancestors themselves) and the nomadic Ashlander tribes (UESPWiki, 2024a).

Ancestor worship is something that has been found in various cultures varying chronologically and geographically from the Romans to groups in Africa, the Americas and Asia (Steadman et al., 1996; Janelli & Janelli, 2000).

Falkreath Graveyard

The graveyard at Falkreath (Fig. 3) is an in-game contemporary cemetery that is still in use and maintained by Runil, the local priest of Arkay (God of the Cycle of Birth and Death) and local monk, Kust; both these NPCs reside in the Hall of the Dead (see below). It is one of two graveyards found in game, with the other being Hamvir's Rest. When entering Falkreath for the first time, the player can witness a funeral taking place; the funeral is of a young child and consists of Runil and the two parents. It is difficult to ascertain whether this small funeral is a norm or whether it is due to the young age of the person. Archaeologically, child burials are often seen as "deviant"

Figure 3. The graveyard at Falkreath. Source: Elder Scroll Fandom (screen capture from the game).

and to be avoided, possibly explaining the small funeral (Millett & Gowland, 2015; Evans, 2020: 58).

The graveyard bears a resemblance to medieval European cemeteries and contains many worn graves but includes those of heroes such as Hoag Merkiller thus it is considered an honour to be buried here (Elder Scrolls Fandom, 2024; Evans, 2020). Nord beliefs of the afterlife pertain that warriors travel to Sovengarde after death, an afterlife like that of the Norse Valhalla (Schodt, 2012: 220). The player can enter here in the quests 'The World-Eater's Eyrie' and 'Sovngarde'. When entering, the player can encounter several NPCs who have died in the game as well characters from the game lore and Tsun, Nord god of trials.

Hall of the Dead

Halls of the Dead are buildings encountered in most major settlements in Skyrim. These mausoleums are built for the dead to be laid to rest and be prepared to undergo funerary rites; they are usually run by a priest of Arkay. The Hall of the Dead in Whiterun and Solitude has access to the catacombs (UESPWiki, 2022). When an NPC who is associated with the settlement dies, the player can locate their body in this area afterwards.

The burial rites and treatment of the (in-game) contemporary Nordic dead vary across Skyrim from shared tombs such as in Whiterun/Solitude to inhumations such as those in Falkreath Cemetery. The reason for this difference is unclear and never explained in the game but we can hypothesise it being due to reasons such as lack of space or unsuitable geography for burials in these locations. However, Whiterun is located on vast grasslands, so this reason is improbable. It may be a cultural reason with some parts of Skyrim having more Imperial influence than others.

Tombs of Emperors and Holy Knights

In *The Elder Scrolls IV: Oblivion*, the player undertakes a quest 'Blood of the Divines', to find the sacred armour of the Tiber Septim in Sancre Tor (Golden Hill) (UESPWiki, 2024c). The site has a long relationship with the imperial civilisation, dating back to the founding of the first empire by St Alessia. The site was eventually abandoned by Tiber Septim, founder of the third empire and contains the tombs of Reman (second Empire) emperors (in-game book *The Legendary Sancre Tor*). The tombs, of which there are five, belong to Reman I, Reman II and Reman III, plus two unmarked tombs (Fig. 4). The tombs are inscribed as such (UESPWiki, 2021):

> Reman — "*Here lies Reman of Cyrodiil. He defeated the Akaviri Horde and brought peace to Tamriel. 2762.*"
>
> Reman II — "*Here lies Reman II of Cyrodiil, crowned Emperor of Tamriel in the year 2812. He fell in battle against the Dark Elves, in the fifty-seventh year of his age, after a reign of thirty-nine years and eight months wanting a day.*"
>
> Reman III — "*Here lies Reman III, last Emperor of the Cyrodiils, the scourge of the Dark Elves, who was cruelly slain by treachery, in the year 2920. He reigned forty-three years.*"

Sarcophagi like this appeared in the early Christian Byzantine transition and were typically reserved for the imperial family or martyrs (Chalkia, 2015: 57).

Figure 4. The tombs of the Reman emperors in Sancre Tor. Source: UESPWiki (screen capture from the game).

This matches with the next set of Sarcophagi found from the *Knights of the Nine* DLC (2006). The knights they contain are part of the eponymous Knights of the Nine, an order dedicated to gathering the relics of the crusader, a set of armour worn by Pelinal Whitestrake, a legendary warrior with parallels to Achilles. The knights themselves bear a resemblance to the real-life holy orders the Templars combined with elements of the Arthurian knights of the Round Table (Jones, 2018). Templar burials have been found in England in 2021 (Holder, 2021). The tombs here are effigy tombs, a common occurrence during the high/later medieval period (1000–1500 CE). Effigy tombs are monuments to the elite dead, depicting the deceased in a supine position holding an object or artefact; in this case a sword (Orme, 2021: 348; Williams, 2014). It is unclear who made the tombs, but it was likely they were commissioned by Sir Amiel Lannus, the last surviving member of the order.

NECROMANCY

It is impossible to discuss death in a fantasy setting without mentioning necromancy. Necromancy is defined as the practice of communicating with the dead to foresee the future or, more generally, black magic (Cambridge Dictionary, 2024). This practice dates to antiquity and has been found in Ancient Egypt, Babylonia, Greece, Rome, and China. In Ancient Greece, people would travel to the Neckromanteion, where they took part in complex rituals involving psychological and physical preparation including a specialised diet (broad beans, milk, oysters, honey, pork, barely bread), as well as chanting and rituals; then the people entering a soundless room where a figurine was manipulated with pulleys and levers to speak to them (Aggaili, 2015: 31). There are also accounts of it dating from the medieval and renaissance eras and relying on plants from the nightshade family to create hallucinogenic affects. In the contemporary world, the use of Ouija boards and séances continues the practice.

In modern fantasy, necromancy concerns the raising of dead thralls using black magic and achieving immortality through lichdom. Immortality has famously been sought by philosophers and emperors such as Qin Shi Huang (Man, 2007). In *The Elder*

Scrolls, in-game attitudes to necromancy vary from province to province, with some areas being accommodating while others treat it as a taboo. *The Elder Scrolls IV: Oblivion* depicts necromancy as taboo and involves a quest whereupon the player faces the king of the necromancers Mannimarco during the Mages Guild questline (UESPWiki, 2023a). The player is unable to use necromancy without obtaining a special item, the Staff of Worms. In *Skyrim*, necromancy can be used by the player as part of the conjuration magic school, though there are several quests and encounters in which necromancers are villains (UESPWiki, 2023b). In the real world, there is archaeological evidence of preventing the dead from resurrecting; for instance, the practice of mutilating corpses to stop them from rising as revenants has been noted in a possible case study in the deserted medieval village of Wharram Percy in England (Mays et al., 2017).

Figure 5. Cadlew Chapel, a chapel that has been desecrated by necromancers for rituals. Source: UESPWiki (screen capture from the game).

City of the Dead

During the *Shivering Isles* DLC, the player can encounter the ruined city of Vitharn. The ruin is inhabited by the ghosts of its past inhabitants being attacked by ghosts of a tribe of Fanatics (in-game book *Fall of Vitharn* by an Unknown Author). These are cursed by the island's mad god Daedra. The main zones of the city are its Keep, Bailey, Mausoleum, Sump and Reservoir (UESPWiki, 2024b). In the centre of the Mausoleum sits a brazier (marked by an 'E' in Fig. 6). In the Mausoleum's western wing, there are the coffins for Countess Sheen-In-Glade and Count Csaran Vitharn, separated by a glass display with jewellery (Fig. 6: C). The eastern wing contains coffins for Countess Jideen and Count Cesrien Vitharn, with a bit more jewellery (Fig. 6: D).

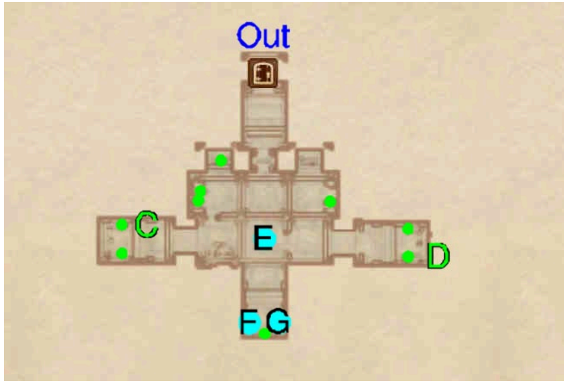


Figure 6. A map and guide to Vitharns mausoleum. Source: UESPWiki (2024c).

The richness of the tombs (Fig. 7) and the jewellery found is indicative of the noble birth of the deceased within. The style of the coffins bears similarity to the imperial tombs of the Reman emperors mentioned before. The two possible explanations of this may be due to meta reasons related to game design or it could be postulated that the rulers of Vitharn saw themselves as equal to or higher than the god ruler of the Shivering Isles, Sheogorath. In world history, many emperors were thought to be divine, and this may have explained this similarity with the count's attempt at usurpation and hubris (Lieven, 2022).

Mammoth Graveyards

In *Skyrim*, players can encounter roving bands of giants herding mammoths. Giants are tall humanoids who stand between 11 and 12 feet tall. They are not a playable race but are encountered by the player in the plains of White Run (UESPWiki, 2024d). The player can encounter the Camp of Secunda's Kiss, located southwest of Whiterun in which several Mammoth skulls can be found with carvings on them, and it is mentioned in-game as being a place where rituals take place. It can be seen in random encounters that giants will mourn the death of mammoths, which is indicative of complex cosmological beliefs held by giants concerning their mammoths (Varnson, 2021). This suggests that giants saw mammoths as more than just livestock but held them in high regard. Archaeological studies of animal burials have found them to be part of mourning as well as ritualistic purposes (Morris, 2016).

Figure 7. The mausoleum of Vitharn. Source: UESPWiki (screen capture from the game).

CONCLUSION

This article has covered the wide range of burial practices shown in the games *Morrowind*, *Oblivion* and *Skyrim*. We have analysed them anthropologically and archaeologically and found various real-world comparisons with surprising results, such as the dragon cult burials resembling those of the Ancient Egyptians rather than those of the Norse, on which their culture is based. For the most part, the other burial types analysed resemble the real-world cultural parallels they were intended to, albeit with a fantastical twist. The practice of necromancy likewise bears little resemblance to the real-world practice of communicating with the deceased, instead focusing on the raising of undead minions and the search for immortality. This article has not discussed the MMORPG *The Elder Scrolls Online* (2014), however, which would provide a whole new dataset of burial practices. If one were to look at other games too, an entire volume could be written on burial practices in fantasy video games, as the interactive nature of the role-playing genre naturally lends itself to anthropological analysis.

REFERENCES

Aggaili, A. (2015) The archaeological site of Neckromanteion. In: Aggaili, A. (Ed.) The Archaeological Sites of Neckromnateion. Ministry of Culture and Sports, Athens. Pp. 20–32.

Cambridge Dictionary. (2024) Necromancy. Available from: https://dictionary.cambridge.org/us/dictionary/english/necromancy (Date of access: 17/Mar/2024).

Chalkia, E. (2015) Basilica D of Nicopolis. Ministry of Culture and Sports, Athens.

Conte, M. & Kim, J. (2016) An economy of human sacrifice: the practice of sunjang in an ancient state of Korea. Journal of Anthropological Archaeology 44: 14–30.

Cunliffe, B. (2019) Scythians: Nomad Warriors of the Steppe. Oxford University Press, Oxford.

Elder Scrolls Fandom. (2023) Red Eagle. Available from: https://elderscrolls.fandom.com/wiki/Red_Eagle (Date of access: 14/Feb/2024).

Elder Scrolls Fandom. (2024) Falkreath Graveyard. Available from: https://elderscrolls.fandom.com/wiki/Falkreath_Graveyard (Date of access: 17/Feb/2024).

Evans, L. (2020) Burying the Dead: An Archaeological History of Burial Grounds, Graveyards and Cemeteries. Pen & Sword History, Yorkshire.

Holder, B. (2021) Forgotten Graves of the Knights Templar Rediscovered in Village

Churchyard. Stourbridge News. Available from: https://www.stourbridgenews.co.uk/news/19710449.historic-graves-knights-templar-discovered-enville/ (Date of access: 28/Feb/2024).

Janelli, R.L. & Janelli, D.Y. (2000) Ancestor Worship and Korean Society. Stanford University Press, Stanford.

Jones, D. (2018) The Templars: The Rise and Spectacular Fall of God's Holy Warriors. Penguin Books, New York.

Man, J. (2007) The Terracotta Army. Bantam Press, London.

Mays, S.; Fryer, R.; Pike, A.W.G.; Cooper, M.J.; Marshall, P. (2017) A multidisciplinary study of a burnt and mutilated assemblage of human remains from a deserted Mediaeval village in England. Journal of Archaeological Science 16: 441–455.

Millett, M. & Gowland, R. (2015) Infant and child burial rites in Roman Britain: a study from East Yorkshire. Britannia 46: 171–189.

Morris, J. (2016) Mourning the Sacrifice. Behaviour and meaning behind animal burials. In: DeMello, M. (Ed.) Mourning Animals: Rituals and Practices Surrounding Animal Death. The Animal Turn. Michigan State University Press, East Lansing. Pp. 11-20.

Orme, N. (2021) Going to Church in Medieval England. Yale University Press, New Haven.

Parker Pearson, M. (2010) The Archaeology of Death and Burial. The History Press, Stroud.

Pederen, A. (2012) Viking Weaponry. In: Brink, S. & Price, N. (Eds.) The Viking World. Routledge, Oxon. Pp. 204–211.

Price, N. (2012) Dying and the dead: Viking Age mortuary behaviour. In: Brink, S. & Price, N. (Eds.) The Viking World. Routledge, London. Pp. 257–271.

Rawson, J. (2023) Life and Afterlife in Ancient China. University of Washington Press, Seattle.

Schodt, J.P. (2012) The Old Norse Gods. In: Brink, S. & Price, N. (Eds.) The Viking World. Routledge, Oxon. Pp. 219–222.

Schuhart, J. (2021) Which real-world cultures inspired the Elder Scrolls' fictional races. ScreenRant. Available from: https://screenrant.com/elder-scrolls-races-real-world-cultural-inspirations-fictional (Date of access: 17/Mar/2024).

Skyrim Archaeological Survey. (2020) The Skyrim Archaeological Survey. Available from: https://skyrimarchaeologicalsurvey.wordpress.com/ (Date of access: 13/Feb/2024).

Steadman, L.B.; Palmer, C.T.; Tilley, C.F. (1996) The Universality of Ancestor Worship. Ethnology 35(1): 63–76.

Taylor, J.H. (2001) Death and the Afterlife in Ancient Egypt. University of Chicago Press, Chicago.

UESPWiki. (2021) Oblivion: Sancre Tor. The Unofficial Elder Scrolls Pages. Available from: https://en.uesp.net/w/index.php?title=Oblivion:Sancre_Tor&oldid=2438946 (Date of access: 17/Mar/2024).

UESPWiki. (2022) Skyrim: Hall of the Dead (Whiterun). The Unofficial Elder Scrolls Pages. Available from: https://en.uesp.net/w/index.php?title=Skyrim:Hall_of_the_Dead_(Whiterun)&oldid=2695608 (Date of access: 26/Feb/2024).

UESPWiki. (2023a) Lore: Mannimarco. The Unofficial Elder Scrolls Pages. Available from: https://en.uesp.net/wiki/Lore:Mannimarco (Date of access: 16/Feb/2024).

UESPWiki. (2023b) Skyrim: Conjuration. The Unofficial Elder Scrolls Pages. Available from: https://en.uesp.net/wiki/Skyrim:Conjuration#Necromancy (Date of access: 16/Feb/2024).

UESPWiki. (2024a) Lore: Ancestor Worship. The Unofficial Elder Scrolls Pages. Available from: https://en.uesp.net/w/index.php?title=Lore:Ancestor_Worship&oldid=2925916 (Date of access: 16/Feb/2024).

UESPWiki. (2024b) Shivering: Vitharn. The Unofficial Elder Scrolls Pages. Available from: https://en.uesp.net/wiki/Shivering:Vitharn (Date of access: 06/Mar/2024).

UESPWiki. (2024c) Oblivion: Sancre Tor. The Unofficial Elder Scrolls Pages. Available from: https://en.uesp.net/w/index.php?title=Oblivion:Sancre_Tor&oldid=2438946 (Date of access: 28/Feb/2024).

UESPWiki. (2024d) Lore: Giant. The Unofficial Elder Scrolls Pages. Available from: https://en.uesp.net/w/index.php?title=Lore:Giant&oldid=2949081 (Date of access: 17/Mar/2024).

Williams, H. (2012) Runes. In: Brink, S. Price, N. (Eds.) The Viking World. Routledge, Oxon. Pp. 281–291.

Williams, H. (2014) Speaking with Effigy Tombs. ARCHAEOdeath. Available at: https://howardwilliamsblog.wordpress.com/2014/07/30/speaking-with-effigy-tombs/ (Date of access: 17/Mar/2024).

Varnson, F. (2021) Skyrim's mournful giant reveals they aren't just monsters. ScreenRant. Available from: https://screenrant.com/skyrim-giants-smart-race-culture-lore-elder-scrolls/ (Date of access: 17/Mar/2024).

About the author

Leon Corneille-Cowell is finishing his BSc in Archaeology and starting an Msc in Funerary Archaeology in September. His academic interest are in death, burial, osteology, and paleopathology. Since he was young, he loved fantasy and fantasy video games such as *The Elder Scrolls*.

Journal of Geek Studies

jgeekstudies.org

Are Japanese anime robots isometric or allometric?

Sangjin Ryu[1] & Kiyotaka Obunai[2]

[1]Department of Mechanical and Materials Engineering, University of Nebraska-Lincoln, Lincoln, NE, USA.
[2]Department of Mechanical and Systems Engineering, Doshisha University, Kyoto, Japan.
Emails: sryu2@unl.edu; kiobunai@mail.doshisha.ac.jp

Robots have been a very popular theme of Japanese animation, or anime, since the success of *Mazinger Z* in the early 1970's. This genre of anime is called robot anime or mecha anime. Well-known robot anime series include Nagai Go's *Mazinger* series, Tomino Yoshiyuki's *Gundam* series, Kawamori Shoji's *Macross* series, and Anno Hideaki's *Evangelion* series. These anime robots take the general humanoid form, which resembles the human body because they usually have a head with eyes (sometimes even nose and mouth), one torso, two arms, and two legs. Anime robots are categorized into two groups: super robot (Fig. 1A) and real robot (Fig. 1B) (Wikipedia, 2024). In general, super robots are imaged as gigantic superheroes with incredible powers, and real robots are treated as mass-produced weapons governed by seemingly realistic technological limitations.

The two anime robot groups contrast to each other quite distinctively from mechanical engineering perspectives as follows (TV Tropes 2024a, 2024b). The common features of super robots include: (1) super robots are often the creations of mad scientists and engineers who work alone or as a small team, aliens, or an ancient civilization; (2) super robots are usually gigantic and powered by mystical energy sources; (3) super robots are controlled by a relatively simple system such as a handful of buttons and a joystick, motion capture, voice command, or direct link to the mind or brain of the pilot; and (4) the motion of super robots is unrealistically humanlike, such as running, jumping, and kicking (Fig. 1A).

By contrast, realism is emphasized for the real robot group, although they are still fictional. The common features of real robots are: (1) real robots are designed, developed, tested, and mass-produced by governments or large commercial corporations; (2) real robots are often developed through stages of prototype, test-type, and mass-production types; (3) real robots rely on largely conventional, yet futuristic, power sources; (4) real robots require periodic extensive maintenance and the logistics to supply fuel, ammunition, and spare parts; (5) real robots happen to malfunction and break down; (6) real robots use scaled-up and advanced versions of infantry ranged weapons, which require ammunition cartridges; and (7) real robots use an elaborate system of controls for manual operation. As such, real robots are usually smaller and less powerful than super robots.

Japanese anime robots have various sizes, ranging from a few meters to several hundred meters (Fig. 1C). Similarly, their mass varies over a wide range. As researchers and educators in mechanical engineering, we question whether there exists any relationship between the height and mass of anime robots. As such, the main research question of this study is whether anime ro-

Cite this article as: Ryu, S. & Obunai, K. (2024) Are Japanese anime robots isometric or allometric? Journal of Geek Studies 11(1): 45-54. https://doi.org/10.5281/zenodo.10963038

Figure 1. Anime robots. **A.** Examples of super robots. From left to right: Combattler V, Getter Robo, Mazinger Z, Great Mazinger, Daimos, and Voltes V. Background: Daiku-Maryu. Source: TV Tropes (2024b). **B.** Realistic (yet still fictitious) battle scene with Mobile Suit Zaku II as an example of the real robot. Source: TV Tropes (2024a). **C.** Height comparison of various anime robots. Source: Cattic's Workshop (https://blog.naver.com/whitesav/221907083584).

bots show any scaling relationship between their height and mass and whether anime robots are isometric or allometric.

ISOMETRY VS ALLOMETRY

A scaling relationship between size and mass can be exemplified by the following cube example (McMahon & Bonner, 1983; Schmidt-Nielsen, 1984; Vogel, 2013). In this example, we are going to find a relationship between the size and mass of cubes of various sizes. As shown in Fig. 2A, we start with a cube with the height of 1. The volume of this unit cube is 1, and if the density of the cube is assumed to be 1, then the mass of the cube is 1. The next step in this example is to double the height of the cube by stacking up eight unit cubes. Simply, the second cube is two times larger than the unit cube in terms of the size. However, its volume, and thus mass, have increased by a factor of 8 (= 2^3). As the cube size is tripled, the height and mass of the cube becomes 3 and 27, respectively.

Figure 2B visualizes how the mass of the cube (m) increases with the height (h). As h increases from 1 to 10, m increases from 1 to 1,000. Since the data points form a curvy alignment, it is hard to find any clear relationship between m and h. When the same data are plotted in the logarithmic scale as shown in Figure 2C, all the data points are aligned on a single straight line. Because the slope of the line is 3, this line means $\log(m) \sim 3\log(h)$, which is equal to $m \sim h^3$. Thus, the scaling relationship between the mass and height of the cubes is found to be $m \sim h^3$. This scaling relationship is known as the cube law (Froese, 2006).

The found relationship means that as the size of the cube increases, its mass also increases when a constant density is assumed. This finding is quite obvious because the mass of a cube with the height h is proportional to its volume h^3. Thus, it can be easily found that when the cube size increases by a factor of n, its mass increases by a factor of n^3. The found relationship is called geometric isometry, and the cubes are isometric because their shapes are identical with the same proportion regardless of their size.

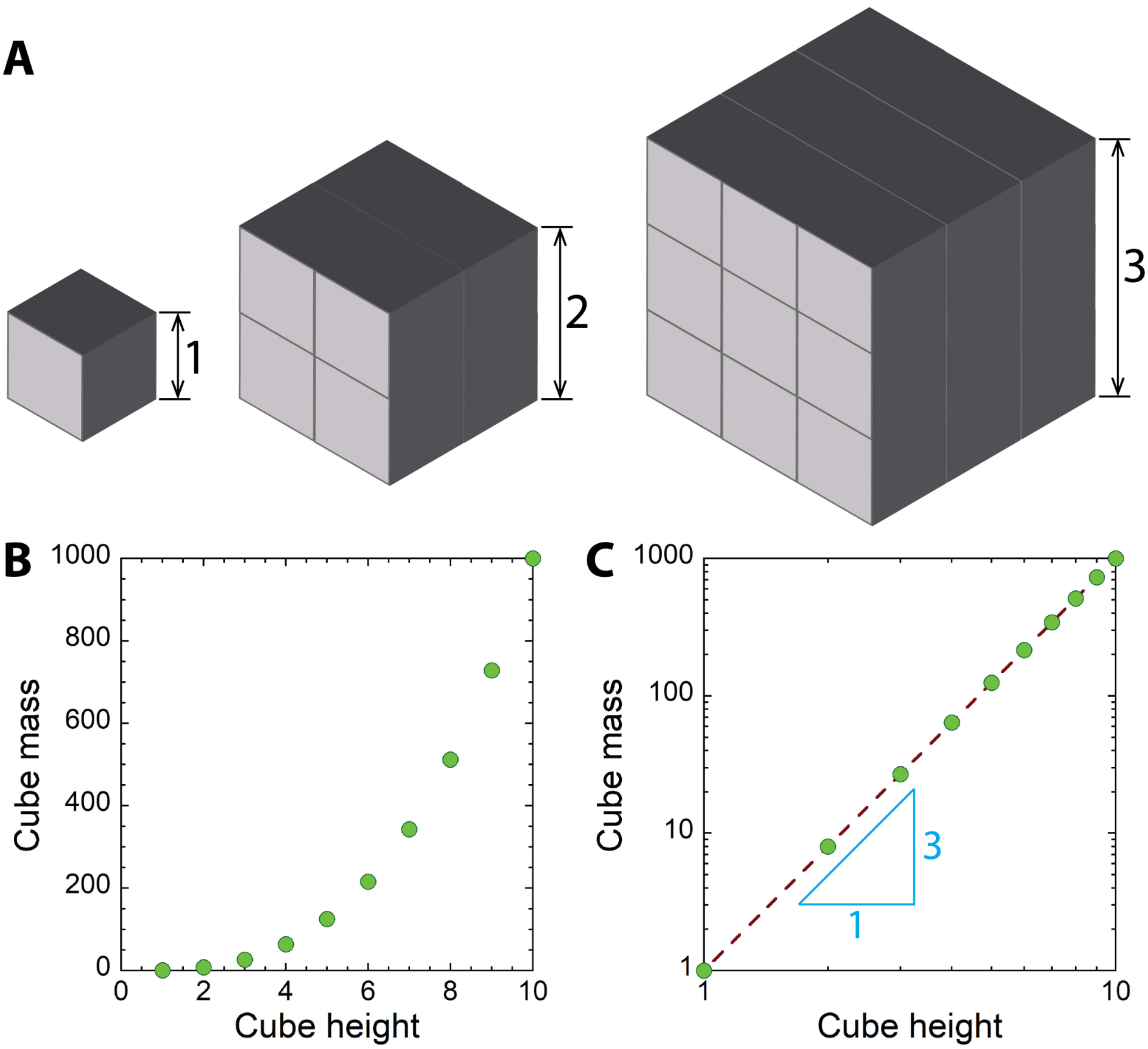


Figure 2. Isometric cubes. **A.** Cube stacks of increasing height with the same proportion. **B.** Linear scale plot of height and mass. **C.** Logarithmic scale plot of height and mass.

Isometry is also found for the human body. According to data collected by various life insurance companies, the following relationship was found between the height and mass of humans: $m \sim h^{2.9}$ (McMahon & Bonner, 1983). It is noticeable that the scaling exponent is very close to that of the cube example. Because our bodies have similar proportions and similar density, human bodies follow the isometric relationship between the height and mass.

Now, we are going to change the way we stack up the unit cubes. As Figure 3A shows, each layer has a different number of the unit cubes, and the number increases from the top to the bottom layer by one, like an arithmetic sequence. The stacks in Figure 3A have the same heights as Figure 2A, but their mass differs because the total number of the unit cubes is different. When the mass of the cube stacks is plotted with their height in the logarithmic scale, all the data points appear to be aligned on a single straight line as shown in Figure 3B. The slope of the data points is found to be 1.75, which shows that the mass of the cube stack increases with their height following $m \sim h^{1.75}$. In this case, the scaling exponent is not 3. Because the proportion of the cube stacks changes as they get higher, as shown in Figure 3A, the found scaling exponent is different from that of the isometric cubes shown in Figure 2. This non-isometric scaling relationship found for the second case is called allometry (Schmidt-Nielsen, 1984).

Most anime robots have similar body structures to the human body, as shown in Figure 1, and thus they are expected to have similar proportions regardless of whether they are super robots or real robots. Then, it can be hypothesized that Japanese anime robots are isometric like humans. This hypothesis can be examined by finding proportions of anime robots, but unfortunately only their height information is available, whereas other dimensions such as leg length and body width are unknown. This is understandable because the height of robots is the physical dimension that viewers can feel on screen. Therefore, this study examined that hypothesis by using currently available height and mass data of anime robots, and by finding scaling exponents between them.

MATERIALS AND METHODS

One can find specification information of anime robots on the internet, but often those values could vary depending on sources. Instead of using values found in online sources, we used published reference books about anime robots (Studio Hard MX, 1997; Media Works, 2003, 2004; Office J.B, 2012a, 2012b; Sunplant, 2013).

From these books, we chose 190 super robots and 210 real robots that appeared in anime from 1963 to 2003 and organized

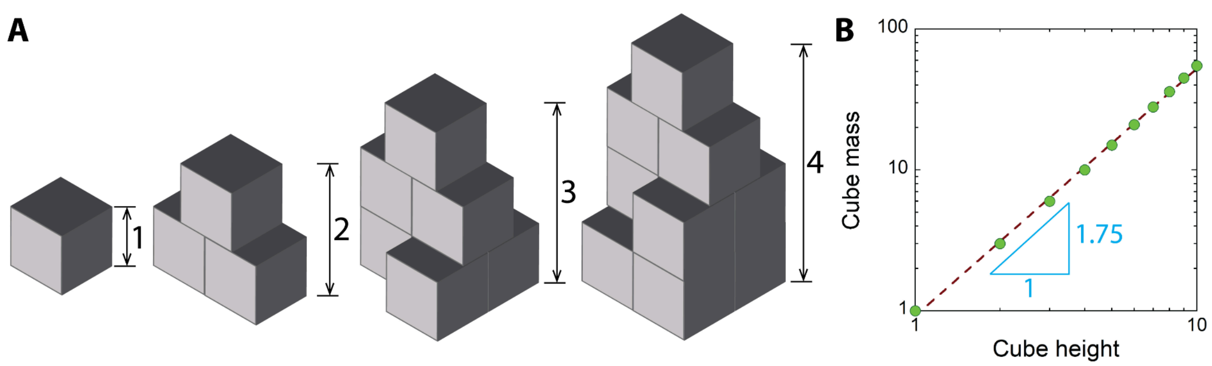


Figure 3. Allometric cubes. **A.** Cube stacks of increasing height with different proportions. **B.** Logarithmic scale plot of height and mass.

their data of height and mass. Some robots had different values between reference books. In such cases, the average value was found and used in this study. The following units were used in this study: meter (m) for height and ton (t) for mass (1 t = 1,000 kg). The specifications of some robots are given in fictional units, and they are included in this study if conversion factors are known. For certain real robots such as Gundam, two different mass values are shown in the reference books: body weight and gross weight. Because gross weight includes the mass of weapons and accessory equipment, only body weight values were considered in this study.

The scaling exponent between height and mass was found by fitting $m = ah^b$ against the data using the method of least squares. Here, b is the scaling exponent of our interest, and if the found value of b is close to 3, anime robots will be found to be isometric.

RESULTS AND DISCUSSION

Figure 4A compares the height and mass distribution between the super robot group and the real robot group. Although the ranges of height and mass differ between the two groups, they agree well with each other for the height range of 5–100 m. Overall, their mass increases with height.

Table 1 compares the statistics of height and mass between the two groups. The super robot group has wider ranges both for height and mass than the real robot group, as shown by the respective maximum and minimum values. This is mainly because of one outlier in the super robot group: Macross from *Super Dimension Fortress Macross*. This humongous robot is actually an alien space battleship that fell to Earth from outer space. When this battleship transforms into the humanoid form, which is known as the Storm Attacker mode, it becomes a robot that is 1,800 m in height and 18 million tons in mass. One may argue that the real robot group also has one outlier, which is Iron-Gear from *Xabungle*. This gi-

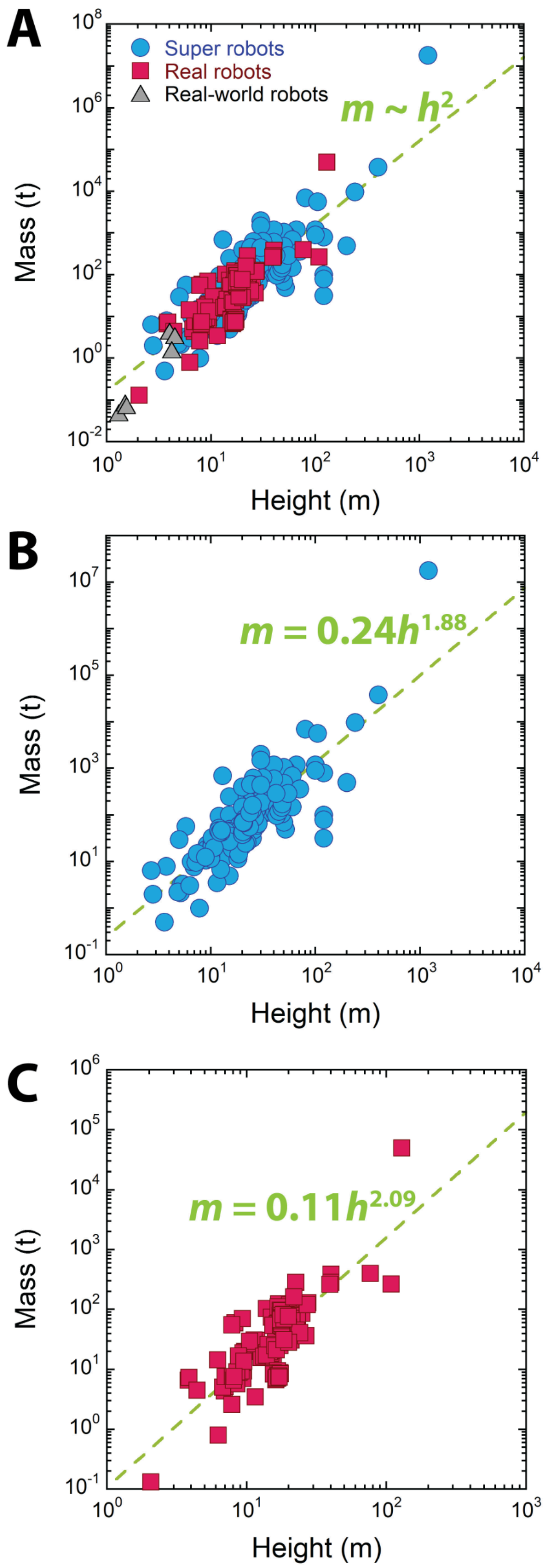


Figure 4. Scaling relationship between height and mass of robots. **A.** Comparison between super robots, real robots, and real-world robots. **B.** Super robots (N = 190). **C.** Real robots (N = 210). The green dashed line is a power function ($m = ah^b$) fitted against the data, and its slope corresponds to the scaling exponent between the height and mass of robots.

gantic hovercraft-type battleship can also transform to a humanoid form, and it is 128.6 m in height and 49,890 t in mass. Although Iron-Gear is gigantic, it is much smaller than Macross since the former is 14 times shorter and 360 times lighter than the latter.

Table 1. Height and mass of anime robots.

	Super robots (*N* =190)		Real robots (*N* =210)	
	Height (m)	Mass (t)	Height (m)	Mass (t)
Max	1200	$1.8x10^{7}$	128.6	49890
Min	2.7	0.5	2.05	0.132
Mean	38.77	$9.52x10^{4}$	17.88	291.76
Standard deviation	93.69	$1.31x10^{6}$	11.95	3439.44
Skewness	10.58	13.78	6.49	14.49
Mode	25	150	18	74.5
Median	23.5	87.5	17.5	55.4

It is harder to grasp a physical sense of mass than height because we can easily sense length dimensions using our vision. The following machines can serve as references. Typical passenger cars weigh 2 t in average, and fully loaded dump trucks mass about 40 t. Boeing 777 series are one of the biggest aircrafts in the world, and their empty operating mass ranges from 135 t to 168 t. Then, what is the heaviest movable machine of human civilization? The answer is aircraft carriers. The U. S. Navy Nimitz class aircraft carrier is one of the largest ships, and each ship is about 325 m long and 99,000 t heavy. Since Iron-Gear is shorter and lighter than Nimitz class aircraft carriers, its size and mass seem somewhat realistic although it is still questionable to fly such a heavy machine in the air. In contrast, Macross is very unrealistic because Macross is equal to a pile of 182 Nimitz class aircraft carriers in terms of mass.

In general, super robots are taller and heavier than real robots in all metrics, which are mean, mode, and median. Because of the high standard deviation and skewness of height and mass, it appears more reasonable to use mode and median rather than mean. Both super and real robot groups have similar mode and median values in each group for height. Super robots are mostly 24–25 m tall while real robots are mostly about 18 m tall. However, the two groups are different in terms of mass because the real robot group has closer mode and median than the super robot group does. It is probably because 69% of the examined real robots are from the *Gundam* series. In this most famous real robot series, many robots, which is called Mobile Suits, are about 18 m in their height.

In Figure 4A, both groups exhibit roughly linear trends between height (h) and mass (m) when plotted in the logarithmic scale. As shown in Figure 4B and C, the obtained values of m and h are similar between the two groups: $m = 0.24h^{1.88}$ for super robots, and $m = 0.11h^{2.09}$ for real robots. This similarity was expected from the good agreement between the two groups shown in Figure 4A. These curve fitting results show that super robots roughly follow $m \sim h^{1.88}$ and real robots $m \sim h^{2.09}$. Since the exponent values are close to 2, the mass of robots in Japanese animation can be thought to increase with their height following $m \sim h^2$ regardless of types. Truly the line of $m \sim h^2$ well represents all the data collected for this study as shown in Figure 4A. Because the found scale exponents are not 3, anime robots exhibit allometric relationships between their height and mass regardless of their type, and the main hypothesis of this study is found to be wrong.

Here, it needs to be reminded that a constant density was assumed for finding the isometry of cubes in the same shape: the cubes were made of the same material and thus they have the same density. For the aforementioned isometry of human height and mass (i.e., $m \sim h^{2.9}$), it is reasonable to assume that the body density of humans is constant because the materials that make up the human body are the same and their compositions are similar across individuals. The same goes for the isometry of various animals. Regarding anime robots, however, their body density appears to vary significantly because those robots are constructed with all different types of substances. The following are some examples: Super-Alloy Z for Mazinger Z, Gundarium for First Gundam, and Ideonite for Ideon. More surprisingly, Aurora Battlers like Dunvine are made up of materials obtained from animals. Since materials for anime robots are all different, it is reasonable to assume that their densities are also different.

Moreover, anime robots usually have cavities or empty spaces in their body to accommodate pilots and to store weapons and fuel, similar to real-world vehicles (of course, no weapons in real-world vehicles). In fact, the relationship between the mass and length of typical passenger vehicles is approximately $m \sim h^{1.85}$, which is very similar to the scaling relationship found for anime robots. Because most cars have a stressed skin structure (i.e., monocoque), their mass is mainly proportional to the body surface area. Let's return to the isometric cube example in Figure 2. If these cubes are assumed to have a shell structure with a constant thickness, their mass increases with the surface area, which results in $m \sim h^2$. The similarity in the scaling relationships between anime robots, passenger vehicles, and empty isometric cubes suggests that the cavities in anime robot's body is responsible for the found allometry of anime robots, despite their similar shapes and proportions, along with variations in the material density.

Our finding suggests how Tomotani can improve his estimation of the mass of anime robots (Tomotani, 2016). Tomotani estimated the mass of AMT-09-ST Scopedog (h = 3.8 m) and RX-78-2 First Gundam (h = 18 m) using the cube law assuming isometry of robots. His estimations quite differed from the official specifications: 3.9 t vs. 6.4 t for Scopedog, and 445 t vs. 60 t (gross weight) for First Gundam. Tomotani attributed the discrepancy to "unknown material" used for robots, as we pointed out. Tomotani may be able to estimate the mass of anime robots and Tengen Toppa Gurren Lagann more reasonably by using the scaling relationship of $m \sim h^2$ found in this study.

ARE ANIME-ROBOT-LIKE ROBOTS POSSIBLE IN THE REAL WORLD?

One may question whether robots like anime robots are possible in the real world. Japan has a few real scale (i.e., 1:1 scale) Gundams at various cities including First Gundam in Yokohama, Unicorn Gundam in Tokyo, and ν Gundam in Fukuoka, as shown in Figure 5. These real scale Gundams can move parts of their body. As shown in Figure 5A, Unicorn Gundam transforms between the unicorn mode and the destroy mode. ν Gundam in Figure 5B can move its head and one arm. However, they are fixed on the ground. In other words, they are more like moving statues

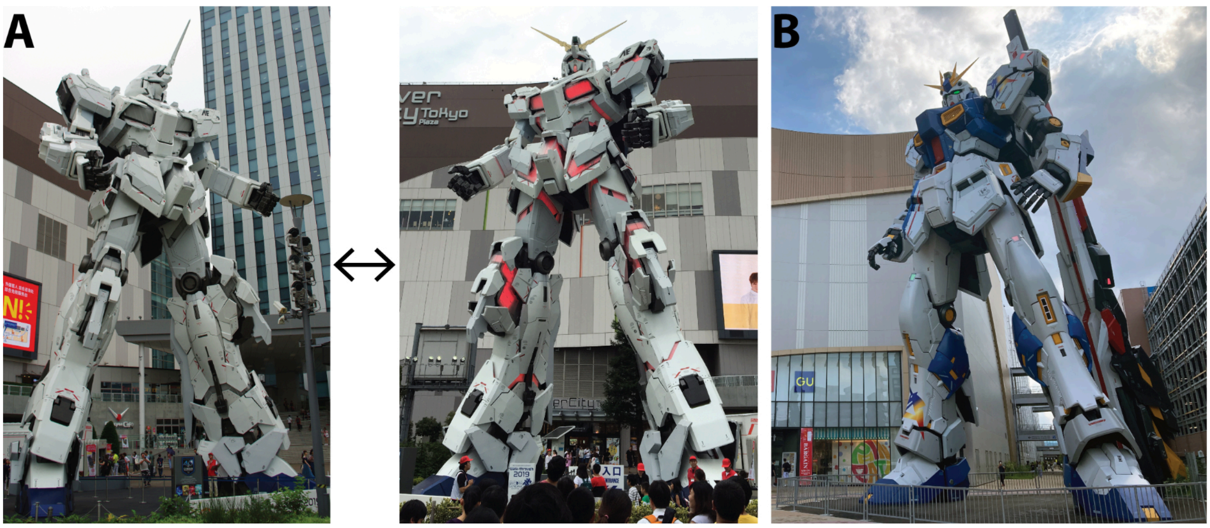


Figure 5. Real-scale moving statues of Gundam in Japan. **A.** Unicorn Gundam in Tokyo, which transforms between the unicorn mode (left) and the destroy mode (right). **B.** ν Gundam in Fukuoka.

rather than robots.

It is too early to be disappointed because several humanoid-type robots have been developed by various industry and academia sectors. Popular two-legged real-world robots include Asimo of Honda (Japan), Atlas of Boston Dynamics (USA), T-HR3 of Toyota (Japan), and Apollo of Apptronik (USA) as shown in Figure 6A–D. Since these robots are 1.3–1.7 m tall and 50–90 kg heavy, they are smaller than average adults but still as heavy as average adults. So, these humanoid robots are like Astro Boy Atom.

Bigger robots do exist in the real world. Suidobashi Heavy Industry (Japan) invented Kuratas (Fig. 6E), which is about 4 m in height and about 4.5 t in mass. Kuratas had a price tag of about USD 1.4 million in 2013. Similar robots were made by MegaBots (USA), and MegaBots robots challenged Kuratas for duels. Recently, Tsubame Industry (Japan) introduced Archax (Fig. 6F). This 3-million-dollar robot is 4.5 m in height

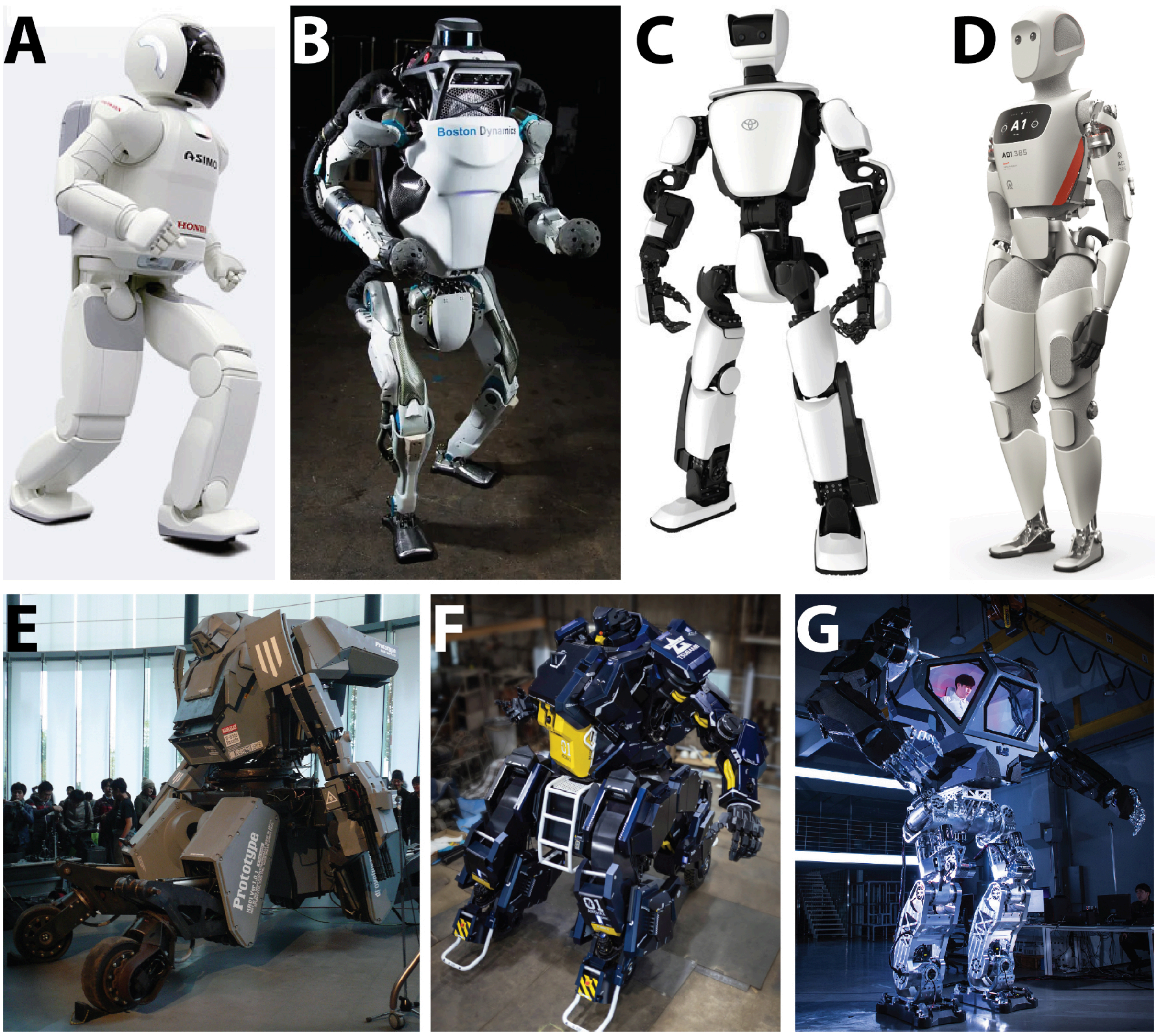


Figure 6. Real-world robots. Images used under Fair Use (limited non-profit and educational use). **A.** Asimo of Honda. Source: Honda (https://asimo.honda.com/asimo-specs). **B.** Atlas of Boston Dynamics. Source: Boston Dynamics (https://bostondynamics.com/resources/). **C.** T-HR3 of Toyota. Source: Toyota (https://global.toyota/en/detail/19666346). **D.** Apollo of Apptronik. Source: Apptronik (https://apptronik.com/apollo). **E.** Kuratas of Suidobashi Heavy Industry. Source: Nandemo-Tsukuruyo (https://nandemotsukuruyo.com/450/). **F.** Archax from Tsubame. Source: Tsubame (https://tsubame-hi.com). G. Method-2 of Korea Future Technology. Source: Korea Future Technology (http://www.method-2.net/).

and 3.5 t in mass. These robots are much heavier than humanoid robots. Considering their heavy mass, it seems unrealistic for them to walk. As such, these robots are wheeled for mobility like Guntank and Getter-3.

Method-2 of Korea Future Technology (South Korea) can walk (Fig. 6G). Being 4.2 m in height, Method-2 is similar to Kuratas and Archax in size. However, Method-2 is much lighter because it weighs about 1.6 t including a human pilot. Although Method-2 is a real-world robot most similar to anime robots so far, it is actually more like Amplified Mobility Platform featured in the movie *Avatar* (20th Century Fox, 2009).

These examples of real-world robots are compared with anime robots in Figure 4A. Similar to anime robots, real-world robots get heavier as they get bigger. It is quite surprising that the real-world robot group agrees relatively well with the anime robot groups. This agreement suggests that for the height range of 1–5 m, the height and mass of anime robots are somewhat realistic.

CONCLUSION

The super and real robots featured in Japanese animation are mostly of humanoid form, mimicking the human body. Their similar proportions suggest that these anime robots are isometric. This hypothesis was examined by finding power-law scaling relationships between the height and mass of anime robots. Such analyses found that the mass of anime robots roughly increases with their height with a power of 2. Because the found scaling exponent significantly differs from 3, which is the expected exponent value for isometry, anime robots are allometric regardless of their type. Considering that isometry between mass and height assumes constant density, the found allometry of the anime robots suggests that their body densities are not constant, which is supported by the variety of materials used for anime robots.

REFERENCES

Froese, R. (2006) Cube law, condition factor and weight–length relationships: history, meta-analysis and recommendations. Journal of Applied Ichthyology 22(4): 241–253.

McMahon, T. & Bonner, J.T. (1983) On Size and Life. W. H. Freeman & Co., New York.

Media Works. (2003) 機動戦士ガンダム MS大全集2003. Mobile Suit Illustrated 2003. Kadokawa Shoten, Tokyo.

Media Works. (2004) スーパーロボット大鑑 Ver.2004. Super Robot Encyclopedia Ver.2004. Kadokawa Shoten, Tokyo.

Office J.B. (2012a) 戦え!全日本アニメロボ大全集 70~80年代篇 [All Japan Anime Robo Complete Works 70-80's Edition]. Futabasha, Tokyo.

Office J.B. (2012b) 戦え!全日本アニメロボ大全集 90年代篇 [All Japan Anime Robo Complete Works 90's Edition]. Futabasha, Tokyo.

Schmidt-Nielsen, K. (1984) Scaling. Why is Animal Size so Important? Cambridge University Press, Cambridge.

Sunplant. (2013) スーパーロボット画報2 — 巨大ロボットアニメ三十五年の歩み. The Super Robots Chronicles Ver.2. The History of Japanese Super Robots Animations, 1998-2013. Takeshobo, Tokyo.

Studio Hard MX. (1997) スーパーロボット画報—巨大ロボットアニメ三十五年の歩み. The Super Robots Chronicles. The History of Japanese Super Robots Animations, 1963-1997. Takeshobo, Tokyo.

Tomotani, J.V. (2016) Is the great attractor a Tengen Toppa Gurren Lagann? Journal of Geek Studies 3(1): 35–44.

TV Tropes. (2024a) Real Robot Genre. Available from https://tvtropes.org/pmwiki/pmwiki.php/Main/RealRobotGenre (Date of access: 17/Feb/2024).

TV Tropes. (2024b) Super Robot Genre. Available from https://tvtropes.org/pmwiki/pmwiki.php/Main/SuperRobotGenre (Date of access: 17/Feb/2024).

Vogel, S. (2013) Comparative Biomechanics. Life's Physical World. Princeton University Press, Princeton.

Wikipedia. (2024) Mecha anime and manga. Available from https://en.wikipedia.org/wiki/Mecha_anime_and_manga (Date of access: 17/Feb/2024).

ACKNOWLEDGEMENTS

SR appreciates Madoka Wayoro and the Kawasaki Reading Room of the University of Nebraska-Lincoln for their support for his Professor Otaku lecture series. We appreciate John J. Socha and Carson Emeigh for proofreading and comments.

ABOUT THE AUTHORS

Dr. Sangjin Ryu is a researcher and educator in mechanical engineering and studies various fluid phenomena at various scales. He is also known as Professor Otaku for his informal lectures on mechanical engineering using examples found in Japanese animation.

Dr. Kiyotaka Obunai is currently a Professor in the Faculty of Science and Engineering at Doshisha University and a Visiting Adjunct Faculty Member in the Department of Mechanical and Materials Engineering at the University of Nebraska-Lincoln. His research interests include materials science and mechanical engineering, and how mechanical engineering is treated in Sci-Fi works, including anime.

Journal of Geek Studies

jgeekstudies.org

Rewriting Earth: environmentally-minded comics to move all audiences

Interview with Paul Goodenough

It is undeniable that we as a species are altering the environment on our planet. Ecosystems are failing and collapsing and species are going extinct under our watch. Little by little, our world is made poorer and more difficult to live in. Solutions to these problems depend on everyone, from policymakers to scientists and to every single one of us.

One particular place where we often stumble is in communicating with the many audiences out there.[1] How to get the message – and the scientific consensus – across and how to turn it into something that people will care about, act upon, and demand measures from their elected officials?[2]

The way the message is built, framed and delivered matters.[3] Enter Rewriting Earth (www.rewriting.earth), a charity dedicated to raise awareness about (and money for) the climate and biodiversity crises.

Rewriting Earth employs powerful storytelling to deliver their message, making extensive use of what's perhaps the best medium to do so: comics. We interviewed Paul Goodenough, one of the founders of Rewriting Earth, to better understand their venture and also to unveil a little bit of all the science, art, and general hard work behind it.

'The Talk', by Nicholas Gurewitch (*The Perry Bible Fellowship*).

[1] The literature on this topic is vast, but see, for instance: Lu, H. (2023) Comparing the effectiveness of different consensus messages in communicating global environmental issues: the role of referent groups, emotions, and message evaluation. Journal of Environmental Psychology 88: 102025.
[2] See, e.g.: Kazdin, A.E. (2009) Psychological science's contributions to a sustainable environment: extending our reach to a grand challenge of society. American Psychologist 64: 339–356.
[3] See, e.g.: Kolandai-Matchett, K. & Armoudian, M. (2020) Message framing strategies for effective marine conservation communication. Aquatic Conservation 30: 2441–2463.

Cite this article as: Journal of Geek Studies. (2024) Rewriting Earth: environmentally-minded comics to move all audiences. Journal of Geek Studies 11(1): 55-60. https://doi.org/10.5281/zenodo.11409445

Rewriting Earth is a massive initiative and fantastic news for nature conservation. We cannot avoid imagining there is a bittersweet origin story for it. So, what sparked the creation of Rewriting Earth?

I'm an animal lover! Animals > people IMO! As a long-time writer, filmmaker and (I kinda hate this term) entrepreneur, I've spent many years ghost writing comedy for comedians, and I thought I could apply that ability for the planet – basically taking science and environmental issues to the mainstream through truly entertaining stories. Basically, making it more fun!!

Why were comics chosen as the main medium to get your message across?

As soon as I saw Jenny Jinya's 'Little Mo' comic,[4] I knew comics offered a completely new way of reaching audiences and making them care about animals. Comics – especially Jenny's – have this incredible ability to turn massive facts and numbers into heartfelt stories that often have me bawling my eyes out.

For example, if I tell you over 1 billion seabirds suffer and die from eating ocean plastics every year,[5,6] that's hard to imagine – hard to 'feel'. But I challenge anyone not to be affected by Jenny's comic, and I've had people from all over the world tell me that it was this comic that changed the way they felt about single use plastics and cigarettes.

Once I saw what Jenny was doing, I decided that turning climate and animal facts into comics was the way to go. So, I conceived the idea of making a comic anthology whereby the world's biggest stars, scientists, activists, NGO's and more could tell their story with the world's best comic creators - and that became our best-selling and Sunday Times Book of the Year, *The Most Important Comic Book on Earth*.[7]

From that, I realised that the format of creating climate stories as comics really worked, but I wanted to reach more people – people who couldn't afford to buy our book. So, we transitioned into creating comics and content for social media – and Rewriting Earth (formerly Rewriting Extinction) was born.

***The Most Important Comic Book on Earth* was organized in main themes such as 'Protect' and 'Inspire'. How did the creation process for the comics worked? Did the creatives have full freedom or were some topics specifically requested? Also, did the creatives had help from scientists?**

All of our content is backed by science and experts in their field. We work with voices from all over the planet, from climate scientists, to UN Ambassadors, to spiritual leaders, to politicians, to those struggling to survive as their climate breaks down… All the stories we make are honest collaborations.

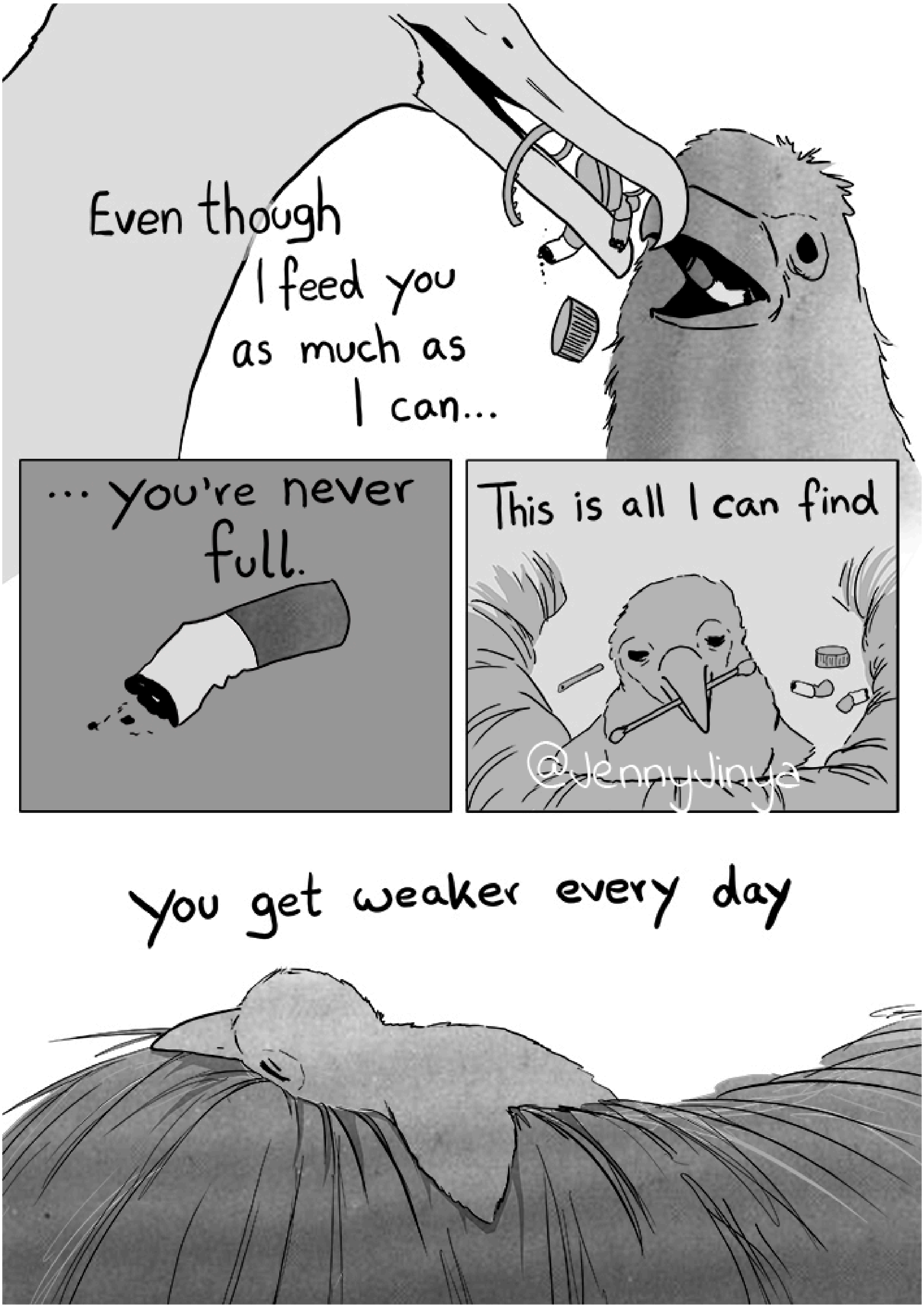
The stars involved were all speaking on subjects they have authority on. Things they care about and want the world to know. We undertook a brainstorm for each comic (over 100 actually) where we'd get on a Zoom meeting with the stars, relevant experts, writers and artists, and between us create the core concept, live. Each comic represents what the collective creators wish the world would know and understand about a particular environmental matter, or

[4] Please read it in full at: https://www.instagram.com/p/CuXRey3Kc9x/

[5] Nine out of ten seabirds have plastics in their digestive system. By 2050, all of them will have plastics. Reference: Wilcox et al. (2015) Threat of plastic pollution to seabirds is global, pervasive, and increasing. Proceedings of the National Academy of Sciences 112: 11899–11904.

[6] For seabirds, there is a 20% chance of lifetime mortality from ingesting a single piece of plastic. Reference: Roman et al. (2019) A quantitative analysis linking seabird mortality and marine debris ingestion. Scientific Reports 9: 3202.

[7] If you're interested, please check https://linktr.ee/mostimportantcomic



'Little Mo' (first page), by Jenny Jinya.

species. Critical to this is that our stories are NOT 'educational' or 'worthy', they're wonderful, stupid, ridiculous, hilarious, terrifying… all the good things a story should be.

How did you manage to bring so many different and influential people together under the Rewriting Earth banner?

Honestly…? I just annoyed them until they gave in. Like most people, I started with the people I knew. Charlie Adlard and Rob Williams deserve a lot of credit, they really came to bat for this. And Rob introduced me to Will Dennis who then opened the project up to the US creators. Managing 300 creators was really tough, but I have to just say, they were amazing. They sacrificed so much, and gave so much of themselves to this. It's truly humbling. I owe them a greater debt than I could ever repay.

Getting 15 charities and 300 contributors to work together was probably the hardest thing I've ever done. It was constant plate-spinning whereby I'd bring on one celeb or one charity, then use that as credibility to talk to the next person, and again and again and again until we had a massive collection of unbelievable names and people involved. To give you some idea, I've racked up over 2,000 hours of video calls, and 22,000 emails to get us to where we are now… I've basically hassled, pressured, begged, pleaded, and blackmailed Rewriting Earth into existence.

Rewriting Earth was a recent re-branding of the original name: Rewriting Extinction. What was the reason for this change? Does that mean a shift in focus or is it a strategic decision for attracting more attention to the project and the cause?

To be honest, it was because people were confusing us with Extinction Rebellion.[8] So, it was a name change only – our mission and purpose remain exactly the same.

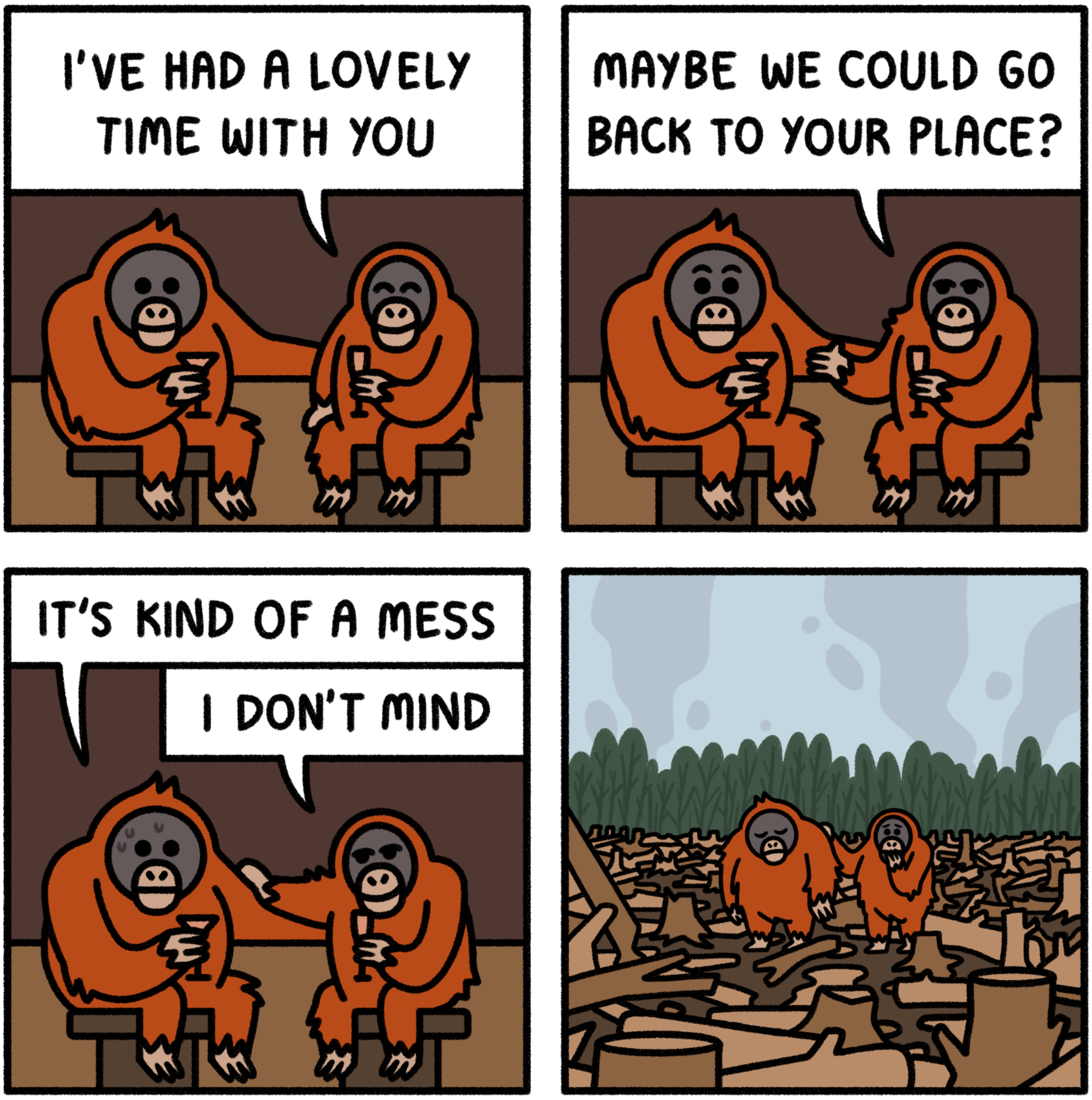
On your website, it is said that Rewriting Earth aims to reach across political and cultural divides. Things have become overly polarized of late and anything that is even slightly nature-friendly is now seen as "leftist". How do you manage to balance your message and reach across audiences? We're asking for a friend, of course.

We have always been a-political and non-judgemental. Being preachy is counter-productive for the stories we are telling. For us it's all about emotion, about moving people to feel something about an environmental issue perhaps they've never heard about before. If we can create mass public engagement and a groundswell of awareness, we could start seeing environmentalism in the same way as the pride movement and BLM, where supermarkets and shops and schools are devoting time and love to it.

The Rewriting Earth campaign is built around a focus on "getting it done" and Laguna Grande Reserve in Guatemala is not only an example of this but also a major success story.[9] Could you tell our readers how that story unfolded and if there are similar projects ongoing or planned?

From the beginning we have enjoyed an incredible partnership with The World Land Trust. Together we managed to raise around $2m and use that to buy Laguna Grande in Guatemala, gifting it back to the families of indigenous peoples who lived there, so it, and the 625 endangered species who live there, will be safeguarded forever. I still get a bit emotional about it.

[8] Here's their website, if you're interested: https://rebellion.global/
[9] You can see the original project at: https://rewriting.earth/projects/world-land-trust/



'It's a bit of a mess', by Safely Endangered and Rewriting Earth.

Still using Laguna Grande as an example, we see that there was a large focus on the protection of species of mammals and birds. We understand that is considered the most surefire way to get people's attention, but the vast majority of biodiversity is made up of other organisms. What kind of strategy do you apply for dealing with the preservation of "uncharismatic" species such as invertebrates and fungi?

If they're important to the planet,[10] they're important to us. It goes back to being non-judgmental and wanting to highlight the best projects out there, resulting in the best solutions for the planet.

So, what's next for Rewriting Earth? How to keep the ball rolling? And how can each of us "get it done" and help our fellow species and our only planet?

[10] Biologist E. O. Wilson used a very apt term to refer to them: "the little things that run the world" (Wilson, 1987: Conservation Biology 1: 344–346).

Follow *@rewritingearth* on Instagram, Facebook and X/Twitter! We'll keep showing content from the best comic creators, comedians, experts, and activists across the world.

A lot of people probably know us for telling a celebrity's story through comics, which was natural for some celebrities and not natural for others. Now we're moving into telling stories through video, having fun finding new ways of getting into the media cycle, like the *Oblivia Coalmine* video for Make My Money Matter.[11]

For me, I think we need to learn from things like the gym industry, because when I was growing up, a gym was where fit people went to get fitter. Over the last 20 years, the gym has become somewhere you go to be healthy, and it doesn't matter if you're going on a treadmill or you're basically bench pressing 300, you're all welcome. Not everyone has to lift the same weight, but everyone does have to be in the gym.

People nowadays tend to be time-poor and would almost rather give us their money than their attention. If we want to reach the biggest possible audience, we have to basically give people as much value for their attention as we physically can, like an exchange at a shop. I work on the smallest exchanges on the spectrum, the 40 seconds or less, the quick, pick-up-read-it-on-the-train-forward-it-in-your-WhatsApp groups. Video, music, computer games, board games, anything that can basically put a welcoming arm around people and make it easy for them to come into environmentalism.

About the interviewee

Paul Goodenough is an award-winning, purpose-driven storyteller and entrepreneur, working in the environmental, charity, peacebuilding and entertainment sectors.

For over two decades, Goodenough has founded and led topflight creative agencies. He launched his first digital agency, Aerian Studios, at the age of 18, enlisting older family and friends to pose as him at meetings, convinced potential clients would be put off by his youth. Aerian Studios went on to become one of the most respected agencies in the UK, creating industry-leading campaigns, websites, products and mobile apps for the BBC, CNN, Warner Bros, Sky, Vodafone, BAFTA, the UK Government and many more.

In 2011, Goodenough founded the production company GBK Hybrid with *Star Wars* producer Gary Kurtz and *Iron Giant* and Disney animator Richard Bazley. Goodenough was nominated for an Emmy in 2015 for his work as a producer on the 2015 multi-award-winning TV Movie *Lost Treasure Hunt*.

In 2019, Goodenough co-founded Rewriting Earth (formerly Rewriting Extinction). Now an independent charity, Rewriting Earth began as a 12-month cross-charity project raising money and awareness for the climate and biodiversity crisis, using the powerful storytelling medium of comics and videos. Goodenough single-handedly forged collaborations between 300+ celebrities, comic artists and experts to create high-impact environmental stories delivering clear information to mainstream audiences. Within 12 months, Rewriting Earth raised £1.54m for the different charities involved, helping to protect 662 species with important milestones, like helping the World Land Trust purchase and protect Guatemala's Laguna Grande Reserve. The project was recognised for reaching a wide, global audience, attaining over 150m engagements with people not traditionally engaged in environmental issues.

The Most Important Comic Book on Earth (Dorling Kindersley) – an anthology of all Rewriting Extinction comics – was published in October 2021, becoming an instant bestseller and a Sunday Times Book of the Year 2021.

Image usage

The comics used in this article were kindly provided by Rewriting Earth. Please do not reproduce them without first checking for permission.

[11] Great video btw: https://makemymoneymatter.co.uk/oblivian/

Journal of Geek Studies

jgeekstudies.org

A queer messiah of the desert: Frank Herbert's *Dune* and the Kwisatz Haderach's challenge to the gender binary

Christine M. E. Hansen

University of Nebraska-Kearney, Kearney, NE, USA.
Email: cmehans2020@gmail.com

Frank Herbert's *Dune* (1965) created a stir and surge of interest with how it pictured a feudal, mystical, and religious future of the universe, and its influences have been incalculable, leaving clear impact on George Lucas' *Star Wars* and beyond. As a result, there is a rather diverse amount of literature on the book and its sequels, with numerous potential reads having been offered on various aspects of it. However, one aspect which has been neglected is queer study of *Dune*. This is, no doubt, in part due to the fact that aspects of the franchise have been rightly noted as homophobic or explicitly anti-queer in nature, and Herbert's own treatment of his son Bruce, a gay activist, have essentially made it clear that any queer sentiments or possible reads of *Dune* were far from the intention of the author.[1] But regardless of his underlying intentions, the Kwisatz Haderach in *Dune* emerges as an entity not bound by the traditional confines of the gender binary of men and women.

The queerness of the Kwisatz Haderach, the messianic and prophesied figure of *Dune*, has not been that fully explored by academics, and this concept is worth considering in more depth, as the Kwisatz Haderach is specifically messianic in function by defying the binary expectations of gender (i.e., by crossing the boundary between the masculine and the feminine). It is via this crossing of normalized boundaries that the Kwisatz Haderach is not only queer, but it is this queerness which gives the Kwisatz Haderach their messianic status. As such, this concept is important to elucidate in depth, as it is only through a queer outlook that the role and importance of the Kwisatz Haderach can be thoroughly understood.

QUEERNESS AND *DUNE*'S BINARY

Queerness is a particularly nebulous term with a rather shrouded history. At one point it was a derogatory slur used against homosexual people, particularly gay men and lesbians. However, in recent years it has become reclaimed by the LGBTQ+ com-

[1] Space does not permit a full discussion of Frank Herbert's homophobic attitudes, but it is pertinent to point to a number of aspects of it, such as the negative stereotype of gay men as pedophilic in the Baron Vladimir Harkonnen, and his comments about homosexuality and the military in *God Emperor of Dune* (Herbert, 2019: 108–109). For his treatment of his son, see Brian Herbert's biography *The Dreamer of Dune* (2004). For an in-depth discussion of this issue, see Kara Kennedy's PhD thesis (Kennedy, 2018: 206–209).

Cite this article as: Hansen, C.M.E. (2024) A queer messiah of the desert: Frank Herbert's Dune and the Kwisatz Haderach's challenge to the gender binary. Journal of Geek Studies 11(1): 61-67. https://doi.org/10.5281/zenodo.12167689

munity but has an array of meanings,[2] which means that it is necessary not only to elucidate what is meant by "queer" in this article, but further how to conceptualize queerness.

For Judith Butler (1988, 2006), gender is not something innate that we are born with. Though there are things which we are born with, we are not born with any innate gender or gender expression, and as a result, gender is seen as performance. As Butler writes, "[…] gender is in no way a stable identity or locus of agency from which various acts proceed; rather, it is an identity tenuously constituted in time — an identity instituted through a stylized repetition of acts" (Butler, 1988: 519). Thus, for gender to be performed is also for it to only have the "appearance of substance" (Butler, 1988: 520), as it is a constructed identity which actor and audience both come to believe as real. With this, the normative configurations of gender in society are not relaying something innate or real and so the opportunity for the subversion of these gender norms is possible and further, even not necessarily subversive acts can then illuminate the illusory nature of gender. I would contend that these subversions of binary gender, the lack of identification with and performance which undermines the binary, is, then, genderqueer. Surya Munro (2019: 126) defines genderqueer as "any type of trans identity that is not always male or female. It is [also] where people feel they are a mixture of male and female" and then non-binary in terms of genderqueer and also extending beyond, to include nonbinary identities that fall entirely outside of the binary and resist description in male or female terms, such as agender people. For this essay, I will specifically make use of Munro's description of genderqueer.

Gender is typically binary as presented in Frank Herbert's *Dune*, and not only binary but it has real and actual qualities to certain extents. Gender is not reified performance, but we are born gendered in the Dune universe. There are qualities of men and of women which are intrinsic and therefore Herbert's world reifies gender (and sex) as having a factual quality to it, conceptualizing gendered differences as biologically founded and deterministic to varying extents,[3] though Herbert conceptualizes some of these as being within our control (Kennedy, 2018: 3–4). Thus, in *God Emperor of Dune* we see how Leto conceptualizes the differences in an all-female and all-male militaristic force, the former being calm, the latter being predatory and self-destructive (and this is linked to homosexuality). Kennedy (2018) argues at length that while there are explicit and at times deterministic differences between genders, there is a degree to which women are given a wider role in how they have some agency (even reproductive autonomy to a certain extent) within the Dune universe. Regardless, however, the series reifies a normative view of gender as binary and with innate qualities, which is also particularly the problem that the Bene Gesserit face, as women are unable to access their male genetic memories (discussed below). This is where the Kwisatz Haderach comes into play, as they are a prophesied being that will be conceptualized as male but be able to bridge the masculine and feminine memories, in short, a unification of masculine and feminine in a single being, transgressing the limitations of gender in normative society. And it is via this transgression that the Kwisatz Haderach is not only exceptional, but messianic (i.e., their messianic qualities are defined by their queerness).

THE KWISATZ HADERACH AND MALENESS

The Kwisatz Haderach conceptualized is first described in Frank Herbert's *Dune* in a

[2] For instance, within LGBTQ+ communities a queer person can be one who defies heteronormative interpretations of sexuality and sexual orientation, or it can relate to one's not fitting into the normative gender-binary of male-female, or any combination thereof. On reclamation of these terms, see Butler (2006: 166–167).
[3] For more on gender essentialism and essentialism in general see Manicom (2010).

Reverend Mother Mohiam holds the Gom Jabbar at Paul Atreides' throat. Screen capture from *Dune: Part One* (2021; directed by Denis Villeneuve).

dialogue between Paul Atreides and the Reverend Mother Gaius Helen Mohiam. There are three aspects in the Reverend Mother's interactions with Paul which demand attention, though we shall handle them slightly out of order. The starting point for us will be the dialogue between the Reverend Mother and Paul:

> *He [Paul] leveled a measuring stare at her, said: "You say maybe I'm the… Kwisatz Haderach. What's that, a human gom jabbar?"*
>
> *"Paul," Jessica said. "You mustn't take that tone with – "*
>
> *"I'll handle this, Jessica," the old woman said. "Now, lad, do you know about the Truthsayer drug?"*
>
> *"You take it to improve your ability to detect falsehood," he said. "My mother's told me."*
>
> *"Have you ever seen truthtrance?"*
>
> *He shook his head. "No."*
>
> *"The drug's dangerous," she said, "but it gives insight. When a Truthsayer's gifted by the drug, she can look many places in her memory – in her body's memory. We look down so many avenues of the past… but only feminine avenues." Her voice took on a note of sadness. "Yet, there's a place where no Truthsayer can see. We are repelled by it, terrorized. It is said a man will come one day and find in the gift of the drug his inward eye. He will look where we cannot – into both feminine and masculine pasts."*
>
> *"Your Kwisatz Haderach?"*
>
> *"Yes, the one who can be many places at once: the Kwisatz Haderach.* […]
>
> – (Herbert, 2005: 13)

This is all extremely important. We have here a figure who is meant to subvert the established feminine and masculine norms and find a way in which they will be unified in a single body. The role of the Kwisatz Haderach person is to perform subversive bodily (and mental) acts, acts which defy gendered expectations (Butler, 2006: 199–200). What is notable is that this is both a masculine/feminine and bodily memory that is being accessed by the Kwisatz Haderach. They are not just in many places but many *gendered* places. Thus, what makes the Kwisatz Haderach special in conceptualization is that they are to be genderqueer in their memory access. They are a body that has access to both realms, normally excluded from those that are locked into the binary; thus, many men failed and died in an attempt to be the Kwisatz Haderach and women are "repulsed" and "terrorized" by the masculine realm (Herbert, 2005: 13).

This is perhaps where there is an antagonism in the text, which allows us to showcase the illusory nature of gender within *Dune*. Women are restricted to their biological sex's memory, and men cannot access the same powers, except for the Kwisatz Haderach. The Kwisatz Haderach, here, is able to access the memories of men *and* women, male and female. Thus, there is a distinction that defies category in Herbert's world. The Kwisatz Haderach accesses the biological arenas of male and female, and further, is trained and taken up as a man within women's spaces, teachings, and ideas. The Bene Gesserit, an all-woman organization, has the goal of bringing in this "male" who has defied and subverted the sexed and gendered configurations of *Dune*'s universe. As the Reverend Mother says, "*Young man, as the Proctor of the Bene Gesserit, I seek the Kwisatz Haderach, the male who truly can become one of us*" (Herbert, 2005: 27). Thus, the "maleness" of the Kwisatz Haderach can be considered queer in more ways than one, as they deconstruct both the binaries of gender and sex in the universe and demonstrate the illusory nature of these categories. The Kwisatz Haderach is a "male," able to transverse the worlds of male and female and become accepted in female spaces as one of them, explicitly, even accessing the arenas locked off in Herbert's gender essentialist world. The Kwisatz Haderach is then the genderqueer messiah of the universe.

What is more, this is also explicitly what makes the Kwisatz Haderach important and gives them their messianic qualities. The Kwisatz Haderach, in being this nonbinary figure, is now able to see all possible futures and gains prescience, which in turn gives them unprecedented power, but this power is contingent on their queerness. For Paul Atreides, accessing the memories and spaces of women, being "one of us" (as the Reverend Mother says), and therefore deconstructing the binary is where his power is found. As one article notes:

> *Paul Atreides is the prophet who will have both a female and male consciousness and bridge time and space. With the rupture of Paul's gender also comes insurmountable knowledge and power; The key lies in the dismantling of the binary, in the merging of the genders.*
>
> — (jalvarez, 2021)

Paul then demonstrates something rather curious, which is that while Herbert's universe is explicitly gender essen-

The Mother Mohiam and Lady Jessica discuss Paul. Screen capture from *Dune: Part One* (2021; directed by Denis Villeneuve).

tialist, it implicitly has now created a vast array of ways in which maleness and femaleness are no longer "factic" sex (to borrow from Butler, 2006: 199), and man and woman are no longer concrete categories, but are instead shown to be constructions themselves, as the Kwisatz Haderach is not limited by the biological or psychological confines previously thought to restrict them.

Particularly curious is that in employing the Kwisatz Haderach's conception as the product of an egregiously long process of eugenics, we essentially find also that sex is no longer an immutable fact in the *Dune* universe but is one subject to both conceptual challenge and biological alteration. As a result, the gender essentialism of the *Dune* universe finds itself collapsing and demonstrated as a cultural construction due to the very existence of its own messiah. To allude to Butler's discussions again, if one sees a Drag Queen/King and then we conceptualize this as "a man dressed as a woman or a woman dressed as a man," we reify the former "man/woman dressed" as the real gender, and the other as unreal via our perceptions (Butler, 2006: xxiii). What this demonstrates, however, is the illusory nature of gender itself, such that drag is not a parody of an "original" gender, but of the fact that there is none (Butler, 2006: 188). The Kwisatz Haderach does similarly. By paradoxically unifying the realms of male and female, man and woman, into one body, they create a situation in which the concept of any original is found illusory, and this even becomes more than just a cultural element as well.

As the Kwisatz Haderach is the product of a breeding program, the universe implicitly contradicts its own essentialism, as we find that sex is not an immutable reality, but is one that can be altered, present differently, and therefore have variant biological and psychological qualities. As such, when Paul takes on the role of the Kwisatz Haderach, he also takes on a genderqueer role that demonstrates that sex and gender are constructs, even within his highly binary universe. Maleness in the *Dune* universe, then, is a mutable, pliable, and constructed concept in the end.

It is by this queerness that Paul is then powerful. Despite being assigned male at birth, Paul takes upon himself the feminine and masculine, and unifies the female and male worlds to create his own queer consciousness and prescience. In a way, then, we can mirror some of his experiences with that of queer and trans individuals. Taking upon this role, he "comes out" as queer (specifically a queer messiah), and this comes with both benefits of him acknowledging his identity, while also making him a target, an outcast, and many wish to exploit him; in fact, the Bene Gesserit see their queer messiah as a tool, wishing to access that male past which has been withheld from them. In this light, we then see the politicalness of Paul's identity, and the ways in which others seek to either destroy or, in a sense, commodify it for their own ends, reducing his humanity, which comes into play earlier, as one of the inciting events for Paul to come into his identity is to "test" whether he is or is not "human" (Herbert, 2005: 8–10), a test where the Reverend Mother inflicts great illusory pain upon his hand which is placed in a box. If he removes his hand from the box, she shall kill him as an "animal." Pain, as a result, becomes an essential part of his queer experience. In many ways, then, we can see how, as an abject identity, Paul is alienated from many of his peers, family, and has many enemies, in no small part because of his queerness as the Kwisatz Haderach.

This queerness should not be seen though as an affirmation of queer identity by Herbert nor as intentional. The world of *Dune* is, as noted before, essentialist in nature, and so the Kwisatz Haderach should arguably be seen as an unintentional aberration. Thus, Herbert still conceptualizes Paul as "male," failing to see the fact that the Kwisatz Haderach actually redefines or, more accurately, defies definition as male or female, as both literally they are the result of a breeding program to create a new gender capable of unifying the male, and female worlds, and further, in terms of gender performance the Kwisatz Haderach performs gendered acts and works in gendered spaces of both, thus, demonstrating the phantasmic qualities of both categories. The Kwisatz Haderach is queer unintentionally on Herbert's part.

gender. The Kwisatz Haderach unintentionally serves to elucidate the illusion that gender is, while existing in a universe that consistently attempts to recapitulate the immutable and almost deterministic qualities of gender, and sex.

CONCLUSIONS

Queerness in the *Dune* universe is by no means something that was intentional on Frank Herbert's part, probably. Herbert's life shows a staunch dislike of queer communities, including the ostracization of his son Bruce. However, unintentional consequences of his conceptualization of the Kwisatz Haderach makes a queer reading of the text not only possible, but even mandatory, as Herbert inadvertently conceptualizes the Kwisatz Haderach, and therefore Paul, as a figure who defies what is otherwise believed to be a factual and deterministic sexual and gendered binary in his world. While sex seems immutable and gender seems stable, the Kwisatz Haderach reveals not only that the biological world is mutable and subject to change, so that any binary conceptualization of sex is invalid, but that the gendered world is also subject to similar changes, in fact, it is revealed to be a performance. Taking on the role of the Kwisatz Haderach is to take on a queer role which defies the binary, and demonstrates that no one is born gendered, but that we come into gender.

This does come with caveats. The queerness of the Kwisatz Haderach can be then seen as affirming of queer identity, but as noted above, the texts of the *Dune* universe are anything but this, particularly in the conceptualization and discussion of non-heterosexual identities. Homosexuality is believed to be self-destructive in *God Emperor of Dune*, while in *Dune* itself, the Baron Vladimir Harkonnen recapitulates harmful stereotypes about gay men. Thus, instead of seeing the Kwisatz Haderach as affirming of queerness, we should see it as an inadvertent and accidental affirmation of the performative and constructed nature of

REFERENCES

Butler, J. (2006) Gender Trouble: Feminism and the Subversion of Identity. Routledge, London.

Butler, J. (1988) Performative acts and gender constitution: an essay in phenomenology and feminist theory. Theatre Journal 40(4): 519–531.

Herbert, B. (2004) The Dreamer of Dune: The Biography of Frank Herbert. TOR Books, New York.

Herbert, F. (2005) Dune. ACE Books, New York. [Supermarkets Hardcover Ed.]

Herbert, F. (2019) God Emperor of Dune. ACE Books, New York. [ACE Premium Ed.]

jalvarez. (2021) The Queerness Behind Dune's Kwisatz Haderach. Science Fiction: Humanity Technology, the Present, the Future. Archived from: https://web.archive.org/web/20240313104713/https://kimon.hosting.nyu.edu/sites/science-fiction/2021/03/02/the-queerness-behind-dunes-kwisatz-haderach/ (Date of access: 14/Apr/2022).

Kennedy, K. (2018) The agency of women in Frank Herbert's Dune. University of Canterbury, Christchurch. [PhD thesis.]

Manicom, D.P. (2010) Gender essentialism: a conceptual and empirical exploration of notions of material essence as a framework for explaining gender difference. University of KwaZulu-Natal, Durban. [PhD thesis.]

Munro, S. (2019) Non-binary and genderqueer: an overview of the field. International Journal of Transgenderism 20(2–3): 126–131.

About the Author

Christine M. E. Hansen is a graduate student in English at the University of Nebraska-Kearney. She is an avid reader and writer of all things fantasy and science fiction. Her favorite book is *The Lord of the Rings*, which she reads ev-

ery year alongside *The Silmarillion*. Her second favorite book is *Dune*. She loves writing papers, and has published on topics ranging from Tolkien's representation of women to New Testament history.

- **Corneille-Cowell, L. ________________ Pp. 1–6.**
  The Ruins of Tolkien: Archaeology in the world of *The Lord of the Rings*
- **Soares, H.M. ________________ Pp. 7–34.**
  The Angry Videogame Model: exploring neural network architectures to predict videogame review ratings
- **Corneille-Cowell, L. ________________ Pp. 35–43.**
  Death, burial, and commemoration in *The Elder Scrolls*
- **Ryu, S. & Obunai, K. ________________ Pp. 45–54.**
  Are Japanese anime robots isometric or allometric?
- **Interview ________________Pp. 55-60.**
  Rewriting Earth: environmentally-minded comics to move all audiences
- **Hansen, C.M.E. ________________Pp. 61-67.**
  A queer messiah of the desert: Frank Herbert's *Dune* and the Kwisatz Haderach's challenge to the gender binary