HUSO 2026

The Twelfth International Conference on Human and Social Analytics

ISBN: 978-1-68558-353-8

March 8th –12th, 2026

Valencia, Spain

HUSO 2026 Editors

Fernanda Eliott, Grinnell College, USA

HUSO 2026

Forward

The Twelfth International Conference on Human and Social Analytics (HUSO 2026), held between March 8-th, 2026 and March 12-th, 2026 in Valencia, Spain, continued a series of events bridging the concepts and the communities dealing with emotion-driven systems, sentiment analysis, personalized analytics, social human analytics, and social computing.

The recent development of social networks, numerous ad hoc interest-based virtual communities, and citizen-driven institutional initiatives, raise a series of new challenges in considering human behavior, both in personal and collective contexts.

There is a great possibility to capture particular and general public opinions, allowing individual or collective behavioral predictions. This also raises many challenges, on capturing, interpreting, and representing such behavioral aspects. While scientific communities now face new paradigms, such as designing emotion-driven systems, dynamicity of social networks, and integrating personalized data with public knowledge bases, the business world looks for marketing and financial prediction.

We take here the opportunity to warmly thank all the members of the HUSO 2026 technical program committee, as well as all the reviewers. The creation of such a high-quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to HUSO 2026. We truly believe that, thanks to all these efforts, the final conference program consisted of top-quality contributions. We also thank the members of the HUSO 2026 organizing committee for their help in handling the logistics of this event.

We hope that HUSO 2026 was a successful international forum for the exchange of ideas and results between academia and industry for the promotion of progress in the field of human and social analytics.

HUSO 2026 Chairs

HUSO 2026 Steering Committee
Christian Bourret, University of Paris East - Marne la Vallée (UPEM), France
Nitin Agarwal, University of Arkansas at Little Rock, USA
Dennis J. Folds, Lowell Scientific Enterprises (LSE), USA

HUSO 2026 Publicity Chairs
Francisco Javier Díaz Blasco, Universitat Politècnica de València, Spain
Ali Ahmad, Universitat Politècnica de València, Spain
José Miguel Jiménez, Universitat Politècnica de València, Spain
Sandra Viciano Tudela, Universitat Politècnica de València, Spain

HUSO 2026
Committee

HUSO 2026 Steering Committee

Christian Bourret, University of Paris East - Marne la Vallée (UPEM), France
Nitin Agarwal, University of Arkansas at Little Rock, USA
Dennis J. Folds, Lowell Scientific Enterprises (LSE), USA

HUSO 2026 Publicity Chairs

Francisco Javier Díaz Blasco, Universitat Politècnica de València, Spain
Ali Ahmad, Universitat Politècnica de València, Spain
José Miguel Jiménez, Universitat Politècnica de València, Spain
Sandra Viciano Tudela, Universitat Politècnica de València, Spain

HUSO 2026 Technical Program Committee

Nitin Agarwal, University of Arkansas at Little Rock, USA
Harry Agius, Brunel University London, UK
Jessica Barfield, University of Tennessee - Knoxville, USA
Balbir Barn, Middlesex University, London, UK
Christian Bourret, Université Gustave Eiffel, France
Anita Chandra, IIT Patna, India
Dickson Chiu, The University of Hong Kong, Hong Kong
Claudia Da Re, Gustave Eiffel University - Marne-la-Vallée Campus, France
Chen Ding, Ryerson University, Canada
Birgitta Dresp-Langley, Centre National de la Recherche Scientifique, France
Thierry Edoh, RFW-Universität Bonn, Germany
Fernanda M. Eliott, Noyce Science Center - Grinnell College, USA
Diana Florea, Lucian Blaga University of Sibiu, Romania
Silvia Florea, Lucian Blaga University of Sibiu, Romania
Dennis J. Folds, Lowell Scientific Enterprises, USA
Matteo Francia, University of Bologna, Italy
Enrico Gallinucci, University of Bologna, Italy
Luca Giraldi, EMOJ srl, Ancona, Italy
Sergiy Gnatyuk, National Aviation University, Kyiv, Ukraine
Damian Gordon, Technological University Dublin, Ireland
Denis Gracanin, Virginia Tech, USA
Hafida Hammadi, Gustave Eiffel University, France
Tzung-Pei Hong, National University of Kaohsiung, Taiwan
Rodrigo Huerta-Quintanilla, Cinvestav, Mexico
Domenico Talia, Università della Calabria, Rende, Italy
Emilio Insfran, Universitat Politecnica de Valencia, Spain
Yasushi Kambayashi, Sanyo-Onoda City University, Japan
Dimitris K. Kardaras, Athens University of Economics and Business, Greece

Copyright Information

Table of Contents

Analyzing Narrative Evolution about South China Sea Dispute

Nitin Agarwal*†
*COSMOS Research Center, University of Arkansas at Little Rock, Arkansas, USA
†International Computer Science Institute, University of California, Berkeley, California, USA
e-mail: nxagarwal@ualr.edu

***Abstract*—YouTube has become the leading global platform for video sharing, yet most research on YouTube Narratives focuses mainly on metadata rather than video content. To address this gap, our study examines narratives embedded in 9,000 YouTube videos about the South China Sea Dispute using Generative Pre-trained Transformer (GPT-5). We also developed a visualization tool to depict connections between keywords and narratives. Our findings show that attention to the dispute surged in 2016 following China's construction and militarization of artificial islands. In 2022, misinformation-driven anti-Western narratives increased five-fold, portraying the United States as exploiting the dispute. This indicates efforts to manipulate the information environment through narrative amplification. Additionally, the dominant narrative cluster incorporated themes related to the Russia-Ukraine war and the COronaVIrus Disease of 2019 (COVID-19) pandemic.**

Keywords-South China Sea; GPT-5; narrative extraction; narrative evolution; narrative visualization; YouTube; social media.

I. Introduction

Videos have become a dominant form of online content, shaping how people acquire knowledge and engage with information. Among platforms, YouTube stands out as the leading medium for video sharing and consumption, offering collaboration, diverse content, and valuable insights. Users watch 1 billion hours of video daily and upload 500 hours every minute [1], fueling exponential data growth. However, this growth also fosters divergent narratives, as some exploit YouTube to spread propaganda and misinformation. This is alarming given that nearly 70% of Americans rely on YouTube as an information source [2]. Understanding the evolution of such narratives has become increasingly complex. Most existing studies focus on metadata analysis [3], which overlooks the nuanced narratives embedded within video content. There remains a gap in computational methods that can extract and analyze these narratives for situational awareness and misinformation countermeasures.

To address this, our study employs the Generative Pre-Trained Transformer (GPT-5), an instruction-based model commonly used for text generation but adaptable for narrative extraction. A survey [4] found that zero-shot GPT-5 performs well in abstractive summarization—a task akin to narrative extraction. GPT's extensive training on large text corpora enables it to effectively process natural language [5] and grasp contextual meaning across segments [6], making it suitable for analyzing video narratives reliant on spoken language.

We apply this approach to the South China Sea dispute, a complex geopolitical conflict involving China and neighboring countries—the Philippines, Vietnam, Malaysia, Brunei, and Taiwan—over resource-rich maritime territories [7]. Covering videos from November 2014 to May 2023, our analysis examines narrative evolution and their impact on viewers' attitudes and values on YouTube.

We address the following research issues in this study:

- We explore how Generative AI, especially the Pre-Trained model (GPT-5), can be utilized for analyzing narratives in video content.
- We investigate the trend of narratives surrounding the South China Sea dispute and examine how they shift over time.

This study makes the following contributions:

- We analyze the actual video content using GPT-5 to extract narratives from YouTube videos by enhancing the understanding of embedded information.
- Our analysis uncovers narratives that intersect with other global issues, like the Russia-Ukraine conflict and COVID-19, showing how narratives within one geopolitical area can be shaped by or connected to broader global events.
- We highlight a significant five-fold rise in anti-Western, misinformation-driven narratives in 2022, that could sway public perception of the South China Sea by amplifying targeted narratives.

The paper is organized as follows. Section II presents a brief review of the literature related to narrative extraction. Section III explains the methodology. The research findings are presented in Section IV. Finally, we draw conclusions from the paper, identify its limitations, and suggest future research directions in Section V.

II. Literature Review

Numerous researchers have explored diverse approaches to computational narrative extraction over the years. The aim of *computational narratology* is to examine narratives from a computational and information-processing perspective, emphasizing the algorithmic processes related to narrative creation and interpretation. It involves the modeling of narrative structure using formal and computable representations, as outlined by Mani [8]. Narratives can be analyzed from various perspectives, including morality, toxicity, sentiment, emotion, plot, and story accomplishment. Initially, computational linguistics focused on shorter texts like fables and folk stories but has shifted to analyzing full-length novels and movie scripts. According to [9], the study of narratives in natural language texts can be condensed into five main stages:

pre-processing and parsing, identification and extraction of narrative components, linking components, representation of narratives, and evaluation. For example, in a survey paper, researchers combined blog posts and named entities to generate topics through network topic modeling. Parameters of the LDA model were tuned to determine the number of topics, and Natural Language Processing (NLP) techniques were used to extract sentences and identify noun and verb phrases. Grammar rules captured patterns to generate narratives, which were ranked using Term Frequency-Inverse Document Frequency (TF-IDF) scoring.

A. Narrative Extraction and Analysis

Gurung et al. [10] emphasize the narrative spread of content sharing the invocation of emotionally charged themes, triggering cognitive biases, and reinforcing existing beliefs, especially those tied to high-stakes issues like the South China Sea. Ranade et al. [11] utilize a comprehensive overview of the computational methods used to extract and analyze narratives across fragmented, multi-source datasets typical of social media. In his work, he identified the use of NLP techniques to detect central themes, tracking progressions and shifts in narrative emphasis, which are essential in understanding how South China Sea narratives evolve across different YouTube channels and over time. This, however, leads to a computational framework that reveals the extraction of narrative elements from YouTube videos and their comments that are related to territorial claims or regional power dynamics. Framing patterns and narrative arcs across different geopolitical actors were included. Kuenzler [12] applied the Narrative Policy Framework (NPF) to provide insights into how institutions and states constructed narratives manage public opinion, assert authority, and deflect criticism. The NPF examines how narratives are strategically constructed to either enhance an organization's reputation or achieve specific policy objectives. The NPF, therefore, provides a lens for analyzing how these entities utilize storytelling elements to influence public opinion on YouTube; this framework was also employed to examine how narratives evolved in response to geopolitical tensions and the role of digital media in shaping these dynamics. Sandberg [13] harnesses a methodological toolkit that could be adapted to study YouTube narratives in conflict zones. In his work, he outlines thematic, structural, performative, and dialogical approaches to narrative analysis. This thematic analysis focuses on identifying key themes such as sovereignty, security, and regional stability, while the structural analysis reveals the narratives used by different YouTube channels, showing recurring patterns in the representation of actors and adversaries in these narratives.

B. Challenges in Misinformation and Narrative Manipulation

Hussain et al. [14] explore the more troubling aspects of narrative-building on digital platforms, examining how various players manipulate social media to spread misinformation, influence public agendas, and create disinformation. They contend that YouTube's vast audience and limited content restrictions allow state-driven narratives to gain traction, which can influence public perception and contribute to a biased narrative landscape. This highlights the importance of vigilant monitoring and analytical tools capable of distinguishing between natural narratives and those crafted to influence opinions on complex geopolitical issues. Santana et al. [9] review the intricate process of extracting narratives from unstructured data, a frequent feature of digital media. Their research highlights the technical and methodological hurdles involved in parsing, annotating, and connecting narrative elements within vast amounts of text and video content, as commonly seen with YouTube data. The challenges of ensuring temporal consistency, handling multiple languages, and adapting NLP techniques for multimedia analysis emphasize the need for advanced computational tools specifically designed for the unique structure of YouTube narratives.

However, the emergence of pre-trained large language models has revolutionized these processes. According to a study by Stambach et al. [15], GPT has the ability to discern key characteristics and fulfill various roles across different domains, such as newspapers and political speeches, without the need for any additional training data beyond a prompt. Rather than relying on predefined rules, these large language models prioritize parameter tuning. In the approach proposed by Liu et al. [16], trainable continuous prompt embeddings were employed to enhance the accuracy of models like GPT and BERT by an impressive 80%. Recent studies [17] have put forward methods for comprehending figurative language, including sarcasm, metaphor, and idioms, in discriminative and generative tasks, effectively narrowing the gap between model performance and human understanding. Agarwal [18] demonstrated that GPT-4-based analysis can further reveal such biases between the text and the narratives extracted from the videos.

Narrative evolution provides insights into how stories, perspectives, and themes change over time. Understanding the evolution of narratives helps us comprehend societal shifts, cultural changes, online users' behaviors [19], and the dynamics of public opinion. For instance, authors in [20] analyzed blogs during the European migrant crisis to study narrative shifts related to refugees or migrants, using named-entity extraction. Similarly, authors in [21]–[23] identified and analyzed the influence of topics in blogs.

III. Data and Methodology

We collected YouTube videos focusing on the South China sea dispute. The data collection process began with the utilization of YouTube API keys, and these videos were gathered using specific keywords selected by subject matter experts. Some of these curated keywords are *Security*, *Chinese militia*, *U.S. interference*, *Philippine Coast Guard*, *Philippine Navy*, *People's Liberation Army*, and *Chinese Coast Guard/CCG*. Through this approach, we successfully amassed a dataset comprising 14,000 YouTube videos. The collected data encompassed metadata, including video ID, publication date, title, video description, and location. For transcription, we follow

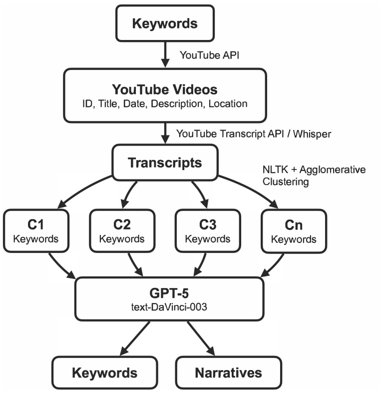


Figure 1. This figure illustrates the methodology for extracting keywords and narratives using GPT-5.

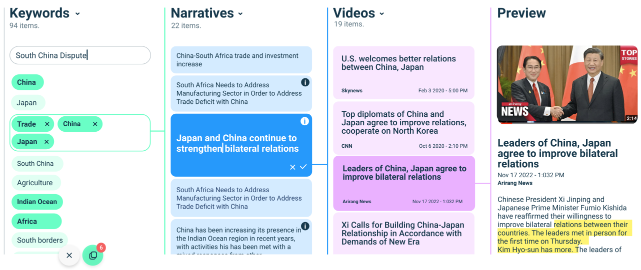


Figure 2. A drill-down style narrative visualization tool that delves into different facets associated with the South China Sea region.

the methodology suggested in [24]. There, the authors leverage parallel computing and the Python multiprocessing library to improve the speed of transcript collection from YouTube. This utilizes YouTube's Transcript API to extract YouTube-generated transcripts, and also uses OpenAI's Whisper model to generate transcriptions on videos without native YouTube transcriptions. Furthermore, the Google Translation API was employed to translate transcriptions from non-English videos. The initial number of collected videos was 14,000. However, after the transcription process, we were able to retrieve data from only 9,000 videos. This reduction in the number of videos was due to various factors, such as videos being taken down or removed during the transcription process, as well as some videos being live streams that did not have transcriptions available. After completing the transcription process, we utilized the NLTK library to eliminate punctuation and basic English stopwords. This helps remove unnecessary characters that could impact further analysis. It is important to consider that GPT-5 has a token limitation of 4096, which indicates the maximum number of tokens that can be processed in a single API call. To overcome this limitation, we implemented agglomerative clustering, as described in [25], to divide the transcripts into multiple chunks for further processing. This methodology is demonstrated in Figure 1.

A. Narrative Extraction

In our Narrative extraction process, we utilized the GPT-5 model, specifically the *text-DaVinci-003* engine. To prioritize accuracy, we set the temperature parameter to 0, ensuring precise, factual results. For the max-tokens parameter, striking the right balance was crucial for providing both a concise narrative and relevant details [26]. For keyword extraction, we set the value to 8, aiming for an average of 5 top keywords. For the macro-narrative generation, we used a value of 35 for an average of 2 to 3 sentences. One of the important parameters is prompt crafting, as a well-designed prompt is crucial when using GPT-5. In our project, we employed the prompt "Given the transcript, provide keyword" for keyword extraction. Similarly, for narrative generation, we used a prompt instructing the model to generate a narrative based on the provided keywords. These prompts were used in [4] for news summarization. Each video's title, description, and transcript were considered; these narratives were then analyzed to gain deeper insights into their content. Furthermore, to assess the credibility of our narratives, we conducted a comparative analysis with the narratives generated by [14]. Their framework leveraged natural language processing techniques, including Part-Of-Speech (POS) tagging, chunking, and grammar rules, to extract actors and actions from the text and uncover associated narratives. We employed the Universal Sentence Encoder (USE) to compare both sets of narratives. The results indicated a significant average semantic textual similarity score of 0.7, indicating a strong alignment between the narratives.

B. Narrative Visualization

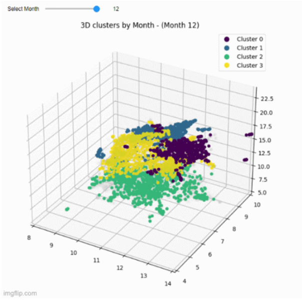
Our research introduces a custom-built narrative visualization tool as demonstrated in Figure 2, integrated into our web application called Vtracker [3], for analyzing videos. It utilizes a structured four-column format for analyzing videos. The tool displays influential entities, associated narratives, and detailed video information. It enables efficient exploration of video content in a visually appealing and user-friendly manner. Furthermore, users can edit narratives in the tool, allowing for user feedback and continuous improvement of the narrative structure. This iterative process helps refine and enhance the tool based on user preferences and feedback.

IV. Research Findings

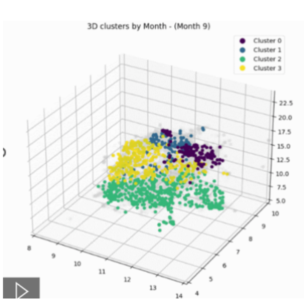
In this section, we discuss our research findings on the evolution of macro-narratives, and sentiment, emotion, morality, and toxicity analysis of the narratives.

A. Macro Narrative Evolution

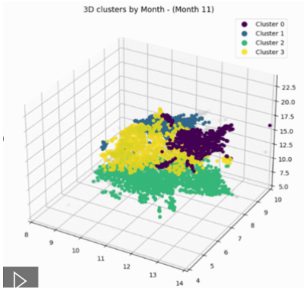
From the macro-narratives displayed in Figure 3, Cluster 0 examines how China's geopolitical strategy is affected by the crisis between Russia and Ukraine. Particular micro-narratives explore how China is changing its position and tactics in reaction to the Russia-Ukraine war each month from January to December. This covers issues including China's militarization of SCS islands, its position on sanctions, and Chinese-Russian joint military drills.



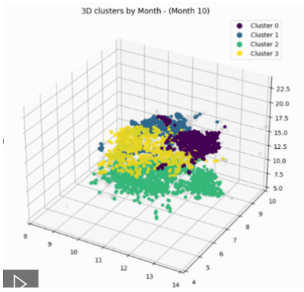
(a) This figure shows the 3D clusters (Clusters 0–3) of narrative by month.



(b) This figure shows the 3D clusters (Clusters 0–3) of narrative by month.



(c) This figure shows the 3D clusters (Clusters 0–3) of narrative by month.



(d) This figure shows the 3D clusters (Clusters 0–3) of narrative by month.

Figure 3. The figure shows the 3D clusters of narrative by month for Cluster 0, Cluster 1, Cluster 2, and Cluster 3.

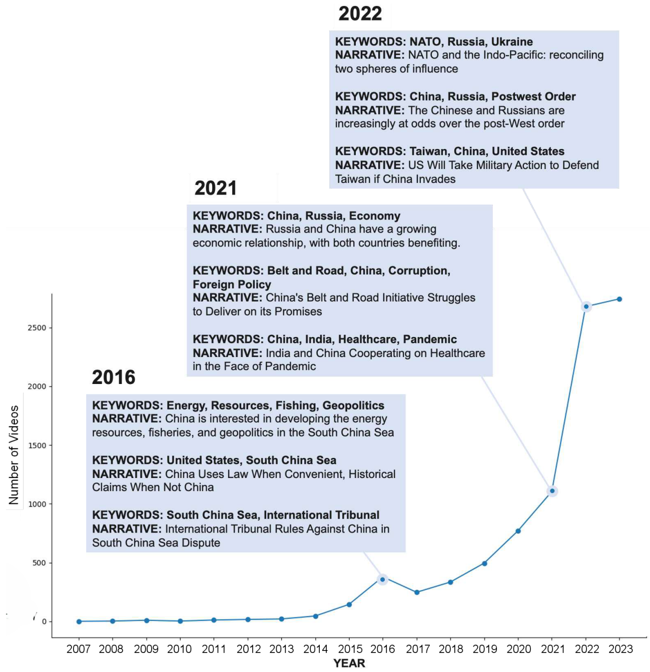


Figure 4. Posting trend of the South China Sea dispute from 2007 to 2023.

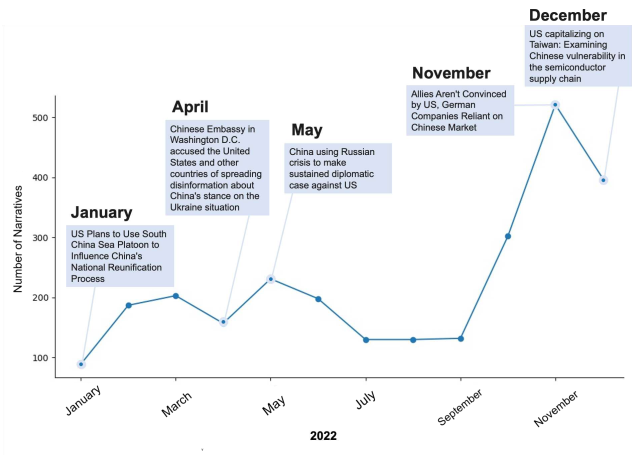


Figure 5. Narrative evolution surrounding the relationship between China and the United States.

China's hostility in the South China Sea is highlighted in Cluster 1. China's geopolitical initiatives, such as diplomatic protests, military drills, and strategic expansions, have heightened tensions with the United States and its neighbors. These developments are detailed in monthly micro-narratives. Cluster 2 highlights US security spending in the Philippines as a component of its larger Asia policy, which is entwined with a number of political and economic problems in the country. Cluster 3 macro-narratives critique President Biden's policies, about social inequality and economic difficulties in the United States, as well as unpopular diplomatic endeavors with Saudi Arabia.

Furthermore, we delve into the identified keywords and narratives associated with the dispute, thoroughly analyzing their alignment with real-life events. We draw parallels between our findings and the dynamic trends and shifts. The earliest video related to the South China Sea dispute dates back to 2007. From that period until 2014, there was limited attention given to the conflict. However, from late 2014 to 2016, there was a slight increase in discussions surrounding the matter. One prominent narrative that persisted during this time was illegal fishing, with videos highlighting fishermen's rights and shedding light on the issue. Notably, Indonesia took action by seizing 153 fishing vessels from neighboring countries, including 50 from Vietnam, 43 from the Philippines, and 1 from China [14].

From 2016 onward, discussions on "infrastructure" and "development" grew following the U.S. release of satellite images showing weapon construction on disputed South China Sea islands—a narrative that remains active. China asserted sovereignty based on historical fishing, while others argued this violated the Law of the Sea. Discourse on the dispute gradually expanded between 2016 and 2021, then surged by 108.30% from 2021 to 2022, as shown in Figure 4.

As demonstrated in Figure 5, the narratives in January and December 2022 exhibit a likely bias against the US, portraying it as attempting to exploit the South China Sea dispute. These narratives, often misleading, have been significantly magnified, being amplified fivefold. To identify various narratives, we applied K-means clustering to the 2022 narratives, resulting in the identification of 10 distinct clusters.

Among these clusters, we identified Clusters 2 and 5 as dominant narratives. These clusters are represented in Figure 6. To gain further insights, we conducted Latent Dirichlet

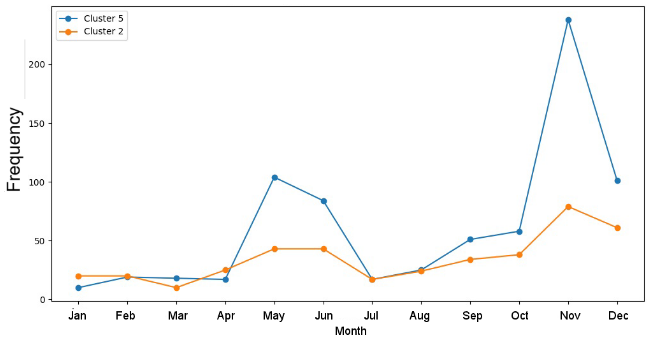


Figure 6. Trend of narratives for Clusters 2 and 5 throughout the year 2022.

Allocation (LDA) analysis on these top clusters to identify prevailing topics within the narratives.

Topics in Cluster 5:

Topic 1: [‘Ukraine’, ‘Russia’, ‘Russian’, ‘military’, ‘Trump’]
Topic 2: [‘foreign’, ‘policy’, ‘Russia’, ‘climate’, ‘alien’]
Topic 3: [‘vaccine’, ‘global’, ‘war’, ‘COVID-19’, ‘Philippines’]
Topic 4: [‘COVID-19’, ‘global’, ‘health’, ‘Taiwan’, ‘pandemic’]
Topic 5: [‘foreign’, ‘Asia’, ‘human’, ‘alien’, ‘United’]

In Cluster 5, May and November 2022 saw a significant number of posts focusing on the geopolitical dynamics between Ukraine and Russia. In May 2022, narratives centered around the COVID-19 pandemic and the Global COVID-19 Summit hosted by the United States, with Indonesia as a co-host. In November 2022, discussions included China’s position on the Ukraine peace talks, highlighted by Chinese leader Xi Jinping during the G20 Summit in Bali. The Philippines also discussed high unemployment rates and COVID-19 deaths in comparison to other Asian countries, particularly India.

Topics in Cluster 2:

Topic 1: [‘China’, ‘security’, ‘trade’, ‘infrastructure’, ‘Africa’]
Topic 2: [‘deterrence’, ‘China’, ‘United’, ‘Taiwan’, ‘defence’]
Topic 3: [‘China’, ‘trade’, ‘armed’, ‘politics’, ‘deal’]
Topic 4: [‘Russia’, ‘Ukraine’, ‘China’, ‘NATO’, ‘Putin’]
Topic 5: [‘politics’, ‘human’, ‘China’, ‘work’, ‘social’]

Cluster 2 focuses on China’s engagement in security, trade, and infrastructure, as well as its relationships with Africa. Topic 1 gained prominence in March 2022 and was influenced by Japanese Foreign Minister Yoshimasa Hayashi, expressing concerns about challenges faced by African nations due to the pandemic and the conflict in Ukraine [22]. Topic 2 delves into the delicate position of the United States regarding Taiwan. In this cluster, Russia and Ukraine also feature but in the context of defense and NATO’s assistance to Ukraine. Disinformation surrounding Australia’s nuclear submarines investment is addressed, along with concerns about China’s military expansion in the South Pacific. The Quad’s focus has evolved to provide public goods, but Southeast Asian countries express mixed responses, cautious about potential tensions with China.

These findings demonstrate the use of computational toosl to detect narrative evolution and dynamics. Despite 2022 having the highest frequency of content related to the dispute, the top two narratives are focused on varied topics rather than the actual dispute itself. Socio-computational methods presented in this study enable detection of such narrative shifts that can help policy/decision makers.

V. Conclusions and Future Work

This paper utilized GPT-5 for extracting narratives from YouTube videos discussing the South China Sea dispute. We also developed a narrative visualization tool to visualize these narratives along with their corresponding keywords. Our analysis covered the posting trend from 2007 to 2023, with a significant surge in postings related to the dispute observed between 2021 and 2022. In order to understand this surge, we conducted a thorough analysis of the year 2022. Our findings revealed a five-fold increase in anti-West narratives from January to December 2022. Additionally, we analyzed the top narrative clusters for 2022, which highlighted a noticeable divergence. This work can aid analysts in delving deeper into the narratives embedded within videos and inform strategic actions based on their findings.

This study demonstrates that combining LLMs with narrative visualization enables scalable and systematic analysis of narrative dynamics within video-based information ecosystems. By extracting and clustering narratives from YouTube content over time, the approach advances computational social science beyond text-centric analyses and provides a practical framework for detecting narrative inflection points, divergence, and escalation. The identification of a sharp rise in anti-West narratives during 2021–2022 illustrates the utility of AI-assisted methods for revealing temporal shifts that are difficult to observe through manual or episodic analysis, while visualization supports human-in-the-loop sensemaking and interpretability in high-volume, multimodal data.

Beyond methodological contributions, the findings have broader implications for international relations, media studies, and public policy. The observed narrative surge and divergence surrounding the South China Sea dispute offer empirical insight into information competition and strategic narrative formation in contested geopolitical contexts. More broadly, this work provides a practical, ethically grounded analytical framework to monitor digital narratives across domains, supporting early warning, evidence-based strategic communication, and policy evaluation. It contributes to interdisciplinary efforts to understand and respond to evolving information environments using transparent and responsible AI tools.

The current method lacks accounting for replicated videos and off-topic content, posing complexities. We manually disregarded irrelevant clusters, but we need a computational approach to remove such content accurately. In the future, we will evaluate the accuracy of our results with more existing methodologies. Furthermore, testing unexplored parameters of

GPT-5 could enhance our results. Additional content analysis, such as sentiment, emotion, toxicity, and morality analysis, could reveal deeper insights explaining which narratives resonated with the target audience and how the audience reacted. Finally, extensive testing for the narrative visualization tool is necessary to further validate its effectiveness and usability.

ACKNOWLEDGMENTS

This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920), Army Research Office (W911NF-23-1-0011, W911NF-24-1-0078, W911NF-25-1-0147), Office of Naval Research (N00014-21-1-2121, N00014-21-1-2765, N00014-22-1-2318), Air Force Office of Scientific Research (FA9550-22-1-0332), Air Force Research Laboratory, Defense Advanced Research Projects Agency, the Australian Department of Defense Strategic Policy Grants Program, Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment, and the Donaghey Foundation at the University of Arkansas at Little Rock. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

REFERENCES

[1] Statista Research Department, *Hours of video uploaded to youtube every minute*, Statista, statistic no. 259477, Retrieved January, 2026, 2023. [Online]. Available: https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/.

[2] B. Auxier and M. Anderson, "Social media use in 2021," Pew Research Center, Tech. Rep., 2021, Retrieved January, 2026, pp. 1–4. [Online]. Available: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.

[3] T. Marcoux, O. Adeliyi, D. S. Banjo, M. I. Gurung, and N. Agarwal, "Exploring online video narratives and networks using vtracker," in *Cyber Security and Social Media Applications*, S. T. Özyer and B. Kaya, Eds., Springer, 2023. DOI: 10.1007/978-3-031-33065-0_6.

[4] T. Goyal, J. Li, and G. Durrett, "News summarization and evaluation in the era of gpt-3," *arXiv preprint arXiv:2209.12356*, vol. 1, pp. 1–4, 2022, Retrieved January, 2026. [Online]. Available: https://arxiv.org/abs/2209.12356.

[5] K. Ethayarajh, "How contextual are contextualized word representations? comparing the geometry of bert, elmo, and gpt-2 embeddings," *arXiv preprint arXiv:1909.00512*, 2022, Retrieved January, 2026. [Online]. Available: https://arxiv.org/abs/1909.00512.

[6] R. Dale, "Gpt-3: What's it good for?" *Natural Language Engineering*, vol. 27, no. 1, pp. 113–118, 2021. DOI: 10.1017/S1351324921000107.

[7] M. Blaine, "The geography of territorial disputes in the south china sea," *Artificial Intelligence Review*, pp. 287–303, 2022. DOI: 10.1007/s10462-021-10000-x.

[8] I. Mani, *Computational modeling of narrative*. Morgan & Claypool Publishers, 2012, vol. 5, pp. 1–142. DOI: 10.2200/S00465ED1V01Y201204HLT016.

[9] B. Santana *et al.*, "A survey on narrative extraction from textual data," *Artificial Intelligence Review*, vol. 56, pp. 8393–8435, 2023. DOI: 10.1007/s10462-022-10338-7.

[10] M. I. Gurung, N. Agarwal, and A. Al-Taweel, "Are narratives contagious? modeling narrative diffusion using epidemiological theories," in *FOSINT-SI, ASONAM 2024*, 2024.

[11] P. Ranade, S. Dey, A. Joshi, and T. Finin, "Computational understanding of narratives: A survey," *IEEE Access*, vol. 10, pp. 101 575–101 587, 2022. DOI: 10.1109/ACCESS.2022.3205314.

[12] J. Kuenzler, "From zero to villain: Applying narrative analysis in research on organizational reputation," *European Policy Analysis*, vol. 7, no. S2, pp. 405–424, 2021. DOI: 10.1002/epa2.1123.

[13] S. Sandberg, "Narrative analysis in criminology," *Journal of Criminal Justice Education*, vol. 33, no. 2, pp. 212–229, 2022. DOI: 10.1080/10511253.2022.2027479.

[14] M. Hussain, H. A. Rubaye, K. Bandeli, and N. Agarwal, "Stories from blogs: Computational extraction and visualization of narratives," in *Text2Story@ ECIR*, 2021, pp. 33–40.

[15] D. Stammbach, M. Antoniak, and E. Ash, "Heroes, villains, and victims, and gpt-3: Automated extraction of character roles without training data," *arXiv preprint arXiv:2205.07557*, 2022, Retrieved January, 2026. [Online]. Available: https://arxiv.org/abs/2205.07557.

[16] X. Liu *et al.*, "Gpt understands, too," *arXiv preprint arXiv:2103.10385*, 2021, Retrieved January, 2026. [Online]. Available: https://arxiv.org/abs/2103.10385.

[17] T. Chakrabarty, Y. Choi, and V. Shwartz, "It's not rocket science: Interpreting figurative language in narratives," *Transactions of the Association for Computational Linguistics*, vol. 10, pp. 589–606, 2022. DOI: 10.1162/tacl_a_00485.

[18] N. Agarwal, "From metadata to meaning: Gpt-4 reveals bias trends in youtube," in *Proceedings of the First IARIA International Conference on AI-based Media Innovation (AIMEDIA 2025)*, Venice, Italy: IARIA, Jul. 2025, pp. 83–88, ISBN: 978-1-68558-330-9.

[19] S. Shajari, N. Agarwal, and M. Alassad, "Commenter behavior characterization on youtube channels," in *eKNOW 2023, Fifteenth International Conference on Information, Process, and Knowledge Management*, Apr. 2023, pp. 59–64.

[20] M. Hussain, K. Bandeli, S. Al-Khateeb, and N. Agarwal, "Analyzing shift in narratives regarding migrants in europe via blogosphere," in *Text2Story@ ECIR*, vol. 585, 2018, pp. 33–40.

[21] N. Agarwal, H. Liu, L. Tang, and P. S. Yu, "Identifying influential bloggers in a community," in *Proceedings of the 1st ACM International Conference on Web Search and Data Mining (WSDM'08)*, Stanford, California, USA: ACM, Feb. 2008, pp. 207–217. DOI: 10.1145/1341531.1341569.

[22] M. Shaik, Z. Stine, and N. Agarwal, "Developing situational awareness from blogosphere: An australian case study," in *Proceedings of the Eleventh International Conference on Social Media Technologies, Communication, and Informatics (SOTICS 2021)*, Barcelona, Spain: IARIA, Oct. 2021.

[23] R. Strating and J. Wallis, "Maritime sovereignty and territorialisation: Comparing the pacific islands and south china sea," *Marine Policy*, p. 105 110, 2022. DOI: 10.1016/j.marpol.2022.105110.

[24] M. C. Cakmak and N. Agarwal, "High-speed transcript collection on multimedia platforms," in *Proceedings of the 8th IEEE Workshop on Parallel and Distributed Processing for Computational Social Systems (ParSocial 2024)*, California, USA: IEEE, May 2024. DOI: 10.1109/IPDPSW62448.2024.00123.

[25] E. Tokuda, C. Comin, and L. Costa, "Revisiting agglomerative clustering," *Physica A: Statistical Mechanics and its Applications*, vol. 585, p. 126 433, 2022. DOI: 10.1016/j.physa.2021.126433.

[26] L. Floridi and M. Chiriatti, "Gpt-3: Its nature, scope, limits, and consequences," *Minds and Machines*, vol. 30, pp. 681–694, 2020. DOI: 10.1007/s11023-020-09548-1.

Modeling the Winkum Game

Istar Abdullahi
Computer Science (Grinnell College)
Grinnell, USA
email: abdullah2@grinnell.edu

Erin Jarvis
Computer Science (Grinnell College)
Grinnell, USA
email: jarviser2@grinnell.edu

Patrick Sales Garcia
Computer Science (Grinnell College)
Grinnell, USA
email: salesgar2@grinnell.edu

Benjamin Cerrato
Computer Science (Grinnell College)
Grinnell, USA
email: cerratob@grinnell.edu

Shayak Nandi
Computer Science (Grinnell College)
Grinnell, USA
email: nandisha@grinnell.edu

Fernanda M. Eliott
Computer Science (Grinnell College)
Grinnell, USA
email: eliottfe@grinnell.edu

***Abstract*—This research began from a reflection: are there any out-of-the-box, unexplored ways to investigate: 1. Group-influenced alliances and strategic decision-making, as well as 2. Reasons for people, institutions, or organizations to decide to commit to certain agreements but ultimately fail to follow through on them? And finally, 3. Can an observer get a hunch on others' intentions as the group dynamics and decision-making unfold and evolve? That reflection guided us to the Winkum Game, which, perhaps due to its complexity and lack of a winner in the traditional sense, has not attracted much interest. We are currently mapping the game's variables (not only its rules but also bridging it to group dynamics) for future computational modeling and experimentation. This could take the form of using either hand-designed agents to simulate specific human-grounded scenarios, or learning agents (applying Reinforcement Learning techniques) to investigate agent-agent dynamics and outcomes. The Winkum Game unfolds interesting power dynamics analogies: since winkers select and launch exchanges (picking sitters), they seem to be better positioned in terms of power, and we expect intriguing scenarios to emerge from flipping roles within the game, helping simulate real-world unpredictability and power dynamics.**

***Keywords*-Computational Modeling; Cooperation; Coordination; Decision-making; Group Dynamics; Winkum Game.**

I. INTRODUCTION

What decision-making insights can a game without a winner in the traditional sense reveal? Curious about the Winkum Game's resourcefulness in providing analogies for real-world applications, we decided to investigate ways of modeling it computationally. Given the game's intricate group dynamics, we anticipated that identifying which variables to model would be challenging. Although more difficult than anticipated, the design process has been insightful for decision-making reflection and enlightening for real-world analogies. This is a work in progress, but we are particularly excited to share it with a multi-disciplinary community. More importantly, we invite researchers across areas to partner with us as we computationally model the game, and we hope to attract more interest in the Winkum Game.

In the Winkum Game, there is a circle of chairs, and all but one chair is occupied by a group of players – the "sitters". Another group of players stands behind each chair – the "winkers. Therefore, there should be n players in the game, where $n = (2 \times \text{number of chairs}) - 1$, and we would suggest having at least $n = 7$ for richer analysis. The winker behind the empty chair must wink at a sitter, who should try to escape from their current chair to sit on the empty chair. "Try to escape" because a sitter can only move to the empty chair if: they have been winked at, and are able to avoid, in the standing process, being tapped by the player behind them.

Figure 1 illustrates the game, Figure 2 summarizes the winker's and sitter's decisions within the game, and we provide additional details in Section III. Up until now, we introduced the game using "players" to facilitate understanding. However, from now on, we use "agents" instead. We define an agent as an autonomous figure that undertakes decision-making (which could be either a human or a learning agent).

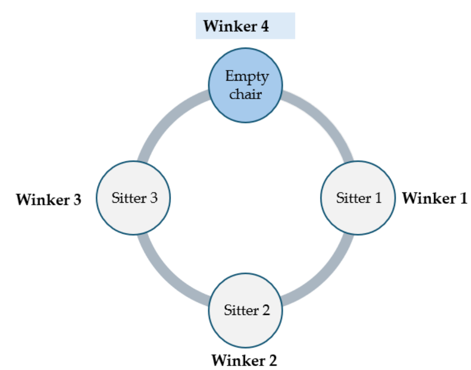


Figure 1. Winkum Game for $n = 7$: four chairs (circles), three sitters, and four winkers. Empty chair color-coded in blue, highlighting who will wink. Players are numbered to facilitate visualization, but sitters dynamically change positions.

Our contributions include spotlighting the game's multi-disciplinary applications and its potential for computational modeling. As this project evolves, we will use it as an analogy to 1. simulate real-world scenarios, e.g., agreements between countries or financial institutions, 2. analyze human behavior, and 3. embed the game and our modeling into a web application to enable people to interact with it. For example, if we refer to the original motivation of the game, which is for people to play and have fun, an instructor could use the application to set the game's variables, simulate their classroom, and brainstorm how well the game could work for their students (by visualizing *fun* levels and potential student *isolation* outcomes).

This work is organized as follows: in Section II, we briefly contextualize what makes the game interesting to model; in Section III, we detail the game; and in Section V, we discuss and conclude.

II. Winkum Game and Networked Emotions

Given the game's strong group dynamics flavor likely impacting humans' internal variables, we saw interesting linkages with *Networked Emotions* – which refer to the view of "emotions as multi-layered processes in which intraindividual processes are tightly coupled and often cannot be separated from interindividual processes" [1]. Thus, in this work, by *layered feeding processes across variables*, we specifically refer to networked emotions as we seek to account for the social nature of emotions and the messy layers of emotion and emotion regulation in strategic decision-making.

Layered feeding processes across variables make the Winkum Game particularly insightful for analyzing group dynamics. For instance, there are three main variables' umbrellas: 1. overall game, 2. winkers, and 3. sitters. What we call *kindness* is among them, a variable that can favor winkers' cooperation (the *Morality-as-Cooperation* Theory [2] inspires our modeling mindset). Specifically, winkers and sitters privately rank each other (e.g., using three categories: not interested, neutral, or interested). If a sitter does not receive enough interest, isolation can emerge (when a sitter is never or rarely winked at, resulting in sitting in the same chair for a long period and being "exposed", since others can easily see the "exclusion or isolation"). Thus, high *kindness* levels make a winker more willing to wink at isolated sitters or even prevent that from happening – therefore enabling a winker to suppress self-interests and cooperate with others [3] [4] (in other words, wink at a "not interesting" sitter).

For now, we named this variable "kindness" instead of "empathy", seeking to ground the modeling on an act's external observation and refrain from assuming an agent's internal drivers; although we do acknowledge that "kindness" has a strong positive connotation, which points us to changing the terminology in the future (for example, a winker may be seen as "kind" but be driven by strategy only).

As we continue to find multidisciplinary pathways to model and interpret the game, our current inspiration sources include: *reputation* [5] as a possible guide for designing the agents' interests, and [6] for kindness, moral reputation, and related considerations. We use as inspiration the "morality-as-cooperation" theory [2], and finally [7] [8] for a hybrid take on Reinforcement Learning (RL) techniques (dry and wet) and the emergence of cooperation in utility-based computational approaches.

III. Modeling the Winkum Game

The design process has been more difficult than anticipated, and it is currently in its early stages. In this section, we present our reflections on designing and subsequently computationally simulating the Winkum Game. Seeking to put the game "on the radar" of various communities, we focus on a high-level presentation and leave additional technical details for future work. Our next steps are to finish modeling, implementing, and testing the game, as well as to develop a software application (e.g., a Shiny app [9]) to enable people to set the game's variables and simulate different group dynamics scenarios.

We envision this software application providing two main options: enabling people to 1) run different predefined scenarios and 2) set the variables to run customized scenarios. Both options will generate data graphics for users to investigate outcomes and reflect on group-influenced decision-making. Possible visualizations are:

- Agent "winkability" and "tappability";
- How the "fun" variable changes across agents over time (see "fun levels" below);
- How dynamic is the game;
- Heat maps of who is winked at (including isolation);
- Escape rate *vs* tap rate.

Also, as we elaborate on our project in future work, we will explore further linkages between our approach and InvestESG, a multi-agent reinforcement learning benchmark for studying climate investment as a social dilemma [10], as well as possible uses within generative AI systems, such as the GOVernance of the Commons SIMulation (GOVSIM), a generative simulation platform for investigating interactions and cooperative decision-making in large language models [11]. Within that direction, [12] offers interesting reflections on the modeling of opponent shaping.

Below, we summarize the game's components identified so far, and we bridge them to realistic settings whenever possible. Finally, we will start modeling using a Gaussian distribution to set each agent's variables, gaining a sense of the design and game dynamics before implementing RL techniques.

Game's goal. Since there is no winner, what is the goal? Looking at it holistically, we can see the game as a way for agents to show and evaluate each other's interests as it evolves, a sort of "intelligence gathering". Through the game, winkers can reveal their interest in specific sitters (stakeholders) or pretend to be interested to deceive others; in other words, we can use it as an analogy for organizations or politicians who show interest in launching commercial, ecological, or political relationships with specific organizations or political agendas, but do not necessarily hold such interests.

Sitters can get a glance of their "popularity" across winkers through how many times they have been winked at in comparison to the others, and show their interest by vigorously trying to move to a winker's chair if winked at, or making less effort to leave a chair if winked at by a less interesting winker (or "commercial partner"). Overall, agents should have "fun" (which could be translated, for example, to "fostering commercial relations or advancing a political agenda"). Still, the game should be dynamic, enabling winkers to wink at different sitters and sitters to explore different chairs; therefore, there should be a balance between "showing interest" and

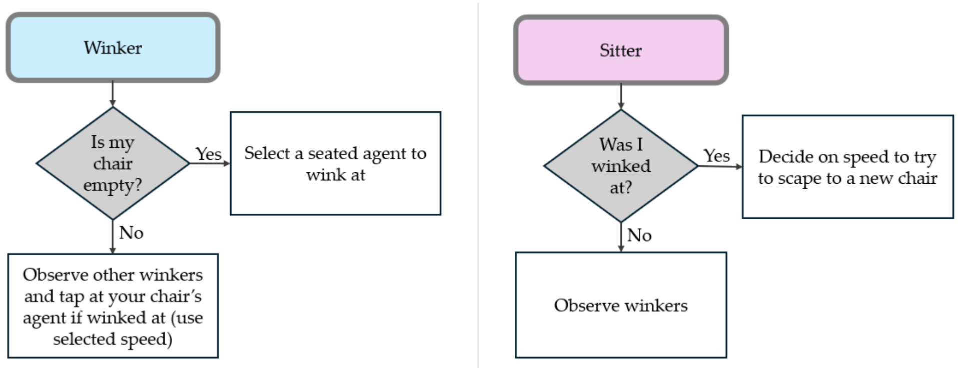


Figure 2. Summarizing decisions *per* agent's role: winkers at left and sitters at right. Note that other strategies are possible; for instance, a winker could, instead of observing other winkers, focus on observing their sitter and tap if noticing any movement, and a sitter could pretend to have been winked at, trying to move to a different chair.

"exploring different options" (offering interesting insights to RL's exploration-exploitation dilemma [13]).

When does the game end? Although the game lacks a clear ending mark, it could be set to end once the agents' average "fun" level drops below a certain threshold. At that point, sitters and winkers could switch places and continue playing: a winker could randomly be chosen to remain as winker, and sitters and winkers could be randomly assigned in the first round. Alternatively, new agents could be added, or simply end the game. We assume agents' interests align with making the game last. Therefore, winkers may wink at "isolated" sitters out of genuine kindness, to deceive others, or to prevent the game from ending (by keeping all agents at least minimally engaged); either way, from an external perspective, we simply call it "kindness".

Kindness. Before the first game begins, agents have their roles defined (winker or sitter) and privately rank each other (not interested, neutral, or interested). If a sitter is never or rarely winked at, isolation can emerge, and agents' ranks can be dynamically updated as they learn more about each other's interests as the game runs. For instance, sitters may have less fun if they identify peers as being isolated (and, of course, isolated sitters should have their fun levels drop, as well as winkers who are stuck with a sitter, or, on the contrary, have a hard time attracting or keeping sitters on their chair). Therefore, different ranking combinations will likely create distinct group dynamics, influencing fun levels, game duration, and agents' *participation*.

Actions and participation. Winkers' actions are to select who to wink at (wink selection) and the tapping speed to keep a sitter. If a winker assumes the corresponding sitter is winked at, they will decide how effective the tapping should be; for example, if a winker has been stuck with a sitter for a long time, they can pretend to try to keep a sitter but facilitate their escape. Game participation comprises observing agents and the winker in charge of winking at each round, and winking. The sitters' actions are to observe the current winker in charge of the winking and escaping speed: if the sitter believes they have been winked at, they will select the escaping speed based on their interest in moving to that specific winker's chair. They can also pretend an understanding to have been winked at, in the interest of moving to a different chair. Whereas game participation involves observing agents, receiving winks, and trying to escape.

Fun levels. To allow an agent's internal goals flexibility, it makes sense to set an overall "game's fun" variable in addition to individual agents' "fun" levels. The overall "fun" could either be an average of fun or designed upon something else, such as based on agents *participation*.

Engagement. Individual engagement levels increase if an agent participates in the game (and decrease otherwise). Once the engagement is depleted, the agent's fun decreases. The *engagement* variable defines how long an agent can stay engaged without participating before losing interest.

Escapability and tapability represent an agent's overall ability (or interest) to escape or successfully tap their opponent, encompassing factors like speed and reaction time.

Consistency indicates how reliably an agent performs at their best during a game.

Environment. For now, we consider as *environment* the number of chairs, game's rules and agents. Note that we will refine the environment in future work, as we agree with [14] that the environment is an important entity at the application level and that aspects conceptually distinct from the agents themselves should *not* be assigned or hosted within agents.

Simulation. A simulation can be defined as letting the game run until it reaches some halt condition (such as having agents switch roles up to a certain number of times and switches being triggered by a drop in average "fun").

Reinforcement function. We are still examining the game to define state and state transitions, agent architecture, and suitable RL algorithms. For the reinforcement function, we envision potential within variables, such as: participation, time spent with an agent according to their category, fun levels, winking success (for winkers), and escaping success (sitters).

Finally, additional variables should be considered if we seek to model the game as close as possible to real settings. For instance, one should also consider "visibility", which is determined by the sitter's seating position relative to the winker, and a sitter seated directly across the winker should be easier to wink at – offering analogies to the ease of making partnerships with "easily reachable partners" even if they are not as "interesting" as others.

IV. Discussion

The Winkum Game offers a flexible framework for studying strategic decision-making and testing group engagement and dynamics theories. We claim that the Winkum Game can be leveraged for broader research into group interactions, providing insights into how factors, such as active participation, kindness, tolerance for inactivity, and relationships among individual agents shape group performance and morale. For instance, things become particularly interesting once one analyzes questions, such as: How do we measure the authenticity of the fun factor in human settings? Is the fun displayed sincere or performed? Likewise, what about participants who may experience fun even without participating too much in the game? How do we account for the impact of "isolated" participants in the group dynamics?

Once setting up a simulation, to what degree should sitters and winkers be interested in interacting with each other? What is the relationship between overall "fun" and the interest level among participants? In what ways should the group dynamics influence individual ranks? (For example, by identifying a winker's group of interest, a sitter may lose interest in that winker). What does resilience mean in this game? How can an isolated sitter proactively respond to isolation? How to treat isolation and kindness?

Similar to [15], who seeks to model the evolution of cooperative behavior in realistic systems (by applying an evolutionary game in which lack of cooperation likely causes a cascading failure effect), we are particularly interested in the Winkum Game as we envision unexplored potential in investigating real-life group dynamics. As mentioned earlier, the computational modeling requires thorough reflection and multidisciplinary investigation as the agents' internal variables and the duration of the game feed from each other (intraindividual and interindividual processes [1]).

Finally, the Winkum Game offers a rich way to investigate group dynamics and decision-making. In the game, there is no traditional winner, and it focuses on agents having fun as they dynamically change pairs (which could be interpreted as dynamic agreements that unfold over time). At the same time, it is a game that enables agents to model other agents' mental models and strategies. While it may seem "naive" to read too much into this game, it enables us to simulate agents who get strategic insights into other agents' (or opponents') preferences. The game can be used as a metaphor to model different stakeholders' decision-making and simulate various scenarios, power dynamics, and preferences. For example, Commons research [16]–[18] can enlighten our modeling; whereas [19] offers insightful avenues for reflecting on how stakeholders govern themselves and make decisions.

V. Conclusion and Future Work

Various studies can benefit from computational simulation (e.g., designing simulations of group behavior, enhancing data-collection methods, and providing insights for analysis and visualization to reveal hidden patterns and relationships). Our work aligns with this direction by introducing our initial scaffolding to model the Winkum Game and investigate group dynamics and decision-making. In our approach, we focus on agents' perceptions of fun, individual levels of engagement, and the popularity/likability of agents, all of which reflect the agents' commitment and willingness to keep playing the game. If we assume winkers are better positioned in terms of power, we can simulate real-world unpredictability by flipping power dynamics (e.g., by flipping roles within the game).

We suggest at least three applications of the Winkum Game: 1. Using it as an analogy to simulate real-world scenarios, e.g., agreements between countries, financial institutions, or politicians and political agendas; 2. Analyzing decision-making and group dynamics; 3. Enabling an instructor to set the game's variables, simulate their classroom, and brainstorm how well it would work for their students, which may enlighten considerations of how heterogeneous groups (e.g., age, gender, diversity) can improve group dynamics.

As the use of generative AI systems increases and hybrid spaces emerge, the Winkum Game may offer insights into these new forms of interaction. To conclude, we hope this work will spark multidisciplinary collaborations and attract greater interest in using the Winkum Game as an analogy for investigating group dynamics.

References

[1] A. Kappas, "Social regulation of emotion: messy layers," *Frontiers in psychology*, vol. 4, p. 51, 2013.

[2] O. S. Curry, D. A. Mullins, and H. Whitehouse, "Is it good to cooperate? testing the theory of morality-as-cooperation in 60 societies," *Current anthropology*, vol. 60, no. 1, pp. 47–69, 2019.

[3] M. Tomasello, "Human culture in evolutionary perspective," *Advances in culture and psychology*, vol. 1, pp. 5–51, 2011.

[4] M. Tomasello and A. Vaish, "Origins of human cooperation and morality," *Annual review of psychology*, vol. 64, pp. 231–255, 2013.

[5] E. Brigatti, "Consequence of reputation in an open-ended naming game," *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 78, no. 4, p. 046108, 2008.

[6] D. Sperber and N. Baumard, "Moral reputation: An evolutionary and cognitive perspective," *Mind & Language*, vol. 27, no. 5, pp. 495–518, 2012.

[7] F. Eliott and C. H. C. Ribeiro, "Emergence of cooperation through simulation of moral behavior," in *Hybrid Artificial Intelligent Systems*. Springer International Publishing, 2015.

[8] ——, "Moral behavior and empathy modeling through the premise of reciprocity," *HUSO 2015*, p. 42, 2015.

[9] W. Chang et al., *shiny: Web Application Framework for R*, 2026, r package version 1.12.1.9000.

[10] X. Hou, J. Yuan, J. Z. Leibo, and N. Jaques, "Investesg: A multi-agent reinforcement learning benchmark for studying climate investment as a social dilemma," *arXiv preprint arXiv:2411.09856*, 2024.

[11] G. Piatti, Z. Jin, M. Kleiman-Weiner, B. Schölkopf, M. Sachan, and R. Mihalcea, "Cooperate or collapse: Emergence of sustainable cooperation in a society of llm agents," *Advances in Neural Information Processing Systems*, vol. 37, pp. 111 715–111 759, 2024.

[12] J. Duque et al., "Advantage alignment algorithms," *arXiv preprint arXiv:2406.14662*, 2024.

[13] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction (2nd ed.)*. MIT press, 2018 [1998].

[14] D. Weyns, G. Vizzari, and T. Holvoet, “Environments for situated multi-agent systems: Beyond infrastructure,” in *Environments for Multi-Agent Systems II: Second International Workshop, E4MAS 2005, Utrecht, The Netherlands, July 25, 2005, Selected Revised and Invited Papers 2*. Springer, 2006, pp. 1–17.
[15] W.-X. Wang, Y.-C. Lai, and D. Armbruster, “Cascading failures and the emergence of cooperation in evolutionary-game based models of social and economical networks,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 21, no. 3, p. 033112, 08 2011.
[16] E. Ostrom, *Governing the commons: The evolution of institutions for collective action*. Cambridge university press, 1990.
[17] E. Ostrom, J. Burger, C. B. Field, R. B. Norgaard, and D. Policansky, “Revisiting the commons: local lessons, global challenges,” *science*, vol. 284, no. 5412, pp. 278–282, 1999.
[18] A. Agrawal, J. Erbaugh, and N. Pradhan, “The commons,” *Annual Review of Environment and Resources*, vol. 48, no. 1, pp. 531–558, 2023.
[19] R. C. Ellickson, *Order without law: How neighbors settle disputes*. Harvard University Press, 1991.

An Investigation into Why Horse Racing Experts Chose Not to Enter Their Horses in Other Similar-Distance Races

Yasuhiko Watanabe
Ryukoku University
Seta, Otsu, Shiga, Japan
email: watanabe@rins.ryukoku.ac.jp

Hideaki Nakanishi
Ryukoku University
Seta, Otsu, Shiga, Japan
email: t170517@mail.ryukoku.ac.jp

Yoshihiro Okada
Ryukoku University
Seta, Otsu, Shiga, Japan
email: okada@rins.ryukoku.ac.jp

***Abstract*—In this study, we focus on horse racing experts, such as racing horse owners and trainers, and investigate why they chose not to enter their horses in other similar-distance races. Using sire line, distance of races, and order of finish as clues, we analyze the 36922 horses registered with Japan Racing Association from 2010 to 2017 statistically. The results of the statistical analysis showed that many horse racing experts chose not to enter their horses in other races of any distance when their horses did not perform well in the first races or the two or three consecutive races of that distance. This result also suggested that many horse racing experts did not consider sire lines when choosing not to enter their horses in other similar-distance races as they do when selecting race distance.**

***Keywords*-decision making; expert; Thoroughbred horse; sire line; race distance.**

I. INTRODUCTION

It is often said that finishing something is harder than starting it. Take horse racing experts, such as racing horse owners and trainers, for example. The first thing they have to do with their horses is to choose which races to enter them in. On the other hand, the last thing they have to do with their horses is to choose not to enter them in any more races. If the above saying is correct, it is harder for horse racing experts to choose not to enter their horses in any more races than it is to choose which races to enter them in. In this study, we focus on horse racing experts and investigate why they chose not to enter their horses in other similar-distance races.

In order to investigate why horse racing experts chose not to enter their horses in other similar-distance races, we focus on sire line, distance of races, and order of finish. A sire line is a term that refers to the paternal lineage or ancestry of a horse, especially a racehorse. Many horse racing experts often say that a sire line can indicate the potential abilities or characteristics of a horse, such as which distance races they are good at [1] [2]. In this study, we analyze horses classified into one sire line type statistically and report why horse racing experts chose not to enter them in other similar-distance races.

The rest of this paper is organized as follows: In Section II, we survey the related works. In Section III, we survey information about racehorses and show how to collect it. In Section IV, we show how to analyze racehorse information statistically and discuss why horse racing experts chose not to enter their horses in other similar-distance races. Finally, in Section V, we present our conclusions.

II. RELATED WORK

Thoroughbred horses originated from a small number of Arab, Barb, and Turk stallions and native British mares approximately 300 years ago [3]–[5]. Since then, they have been selectively bred to improve speed and stamina, and are consequently superior competitive racehorses. Wade et al. reported a high-quality draft sequence of the genome of the horse and suggested that the horse was domesticated from a relatively large number of females, but few males [6]. McGivney et al. reported that centuries of selection for favourable athletic traits among Thoroughbreds acts on genes with functions in behavior, musculoskeletal conformation, and metabolism [7]. Recently, some genomic regions were identified as a candidate region influencing racing performance in racehorses [8]. Many researchers applied statistical models to evaluate various parameters on racing performance in Thoroughbred horses [9]. Martin, Strand and Kearney reported that the most influential parameter was distance raced [10]. Cheetham et al. investigated whether both race earnings and number of race starts were associated with horse signalment (age, sex, and breed), gait, and race surface [11]. Wells, Randle and Williams investigated how temporal, behavior, and loading related factors associated with the period before the start of the race influences racehorse performance [12].

Statistical researches are conducted not only in horse racing but also in other sports, such as baseball. In recent years, statistical researches brought a new batting approach that batters should aim for big fly balls rather than grounders [13]. Kato and Yanai reported that Shohei Otani, the Japanese superstar slugger in Major League Baseball (MLB), always aims for hitting fly balls [14]. This new batting approach, the so-called "fly-ball revolution", shows that even experts may make mistakes. As a result, it is important to discuss how experts made decisions and why they made mistakes. Yerkes and Dodson studied the relationship between arousal and performance and showed that a little stress can help we perform a task, however, too much stress degrades our performance [15]. Weinschenk used Alex Rodriguez as an example to show that even experts can make mistakes when the stakes are high [16]. Aircraft pilots are under a great deal of mental stress when they are flying their planes. Shappell and Wiegmann focused on preventing errors in aviation, including decision errors, and propose a framework for analyzing and

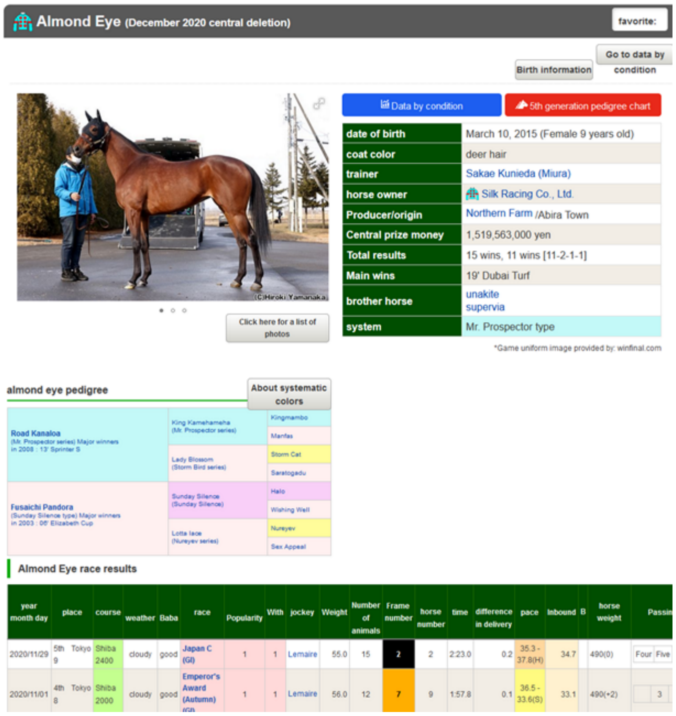


Figure 1. An example of horse information provided by Keiba Lab.

classifying human errors [17]. Kang and Yoon studied the types of errors that both younger and older adults make when learning how to use new technologies [18]. They found that older adults used different strategies than younger adults. However, they did not report how experts made mistakes.

III. A COLLECTION OF RACEHORSE INFORMATION

Keiba Lab [19] is one of the most popular horse racing information sites in Japan. This site records various information about all racehorses registered with Japan Racing Association (JRA) and registered users can freely access it. Figure 1 shows an example of horse information provided by Keiba Lab. As shown in Figure 1, the horse information provided by Keiba Lab consists of personal information and race results. Personal data consists of name, date of birth, age, sex, coat color, breeder, birth place, owner, trainer, ancestors up to three generations ago, sire line, career statistics, career prize money, and so on. Race results consist of venue, event date, distance, weather, racetrack, surface, race name, favourite, order of finish, jockey, weight, horse number, frame number, time, and so on. In order to discuss why horse racing experts chose not to enter their horses in other similar-distance races, we collected information about 36922 horses registered with JRA from 2010 to 2017 from Keiba Lab. Table I shows the number of horses registered with JRA from 2010 to 2017.

On Keiba Lab, various sire lines are used to classify horses. We surveyed how racehorse sire lines diverged and grouped them into Native Dancer Line, Nearctic Line, Royal Charger Line, and others. For example, Figure 1 shows that the sire line

TABLE I
THE NUMBER OF HORSES REGISTERED WITH JRA FROM 2010 TO 2017.

year	number of registered horses
2010	4470
2011	4524
2012	4505
2013	4595
2014	4672
2015	4663
2016	4738
2017	4755
Total	36922

TABLE II
THE NUMBER OF HORSES CLASSIFIED INTO THE THREE MAIN SIRE LINE TYPES.

sire line	number of horses
Native Dancer Line	8799
Nearctic Line	6383
Royal Charger Line	18104
others	3636
Total	36922

of *Almond Eye* was Mr. Prospector Line. It branched out from Native Dancer Line. As a result, in this study, we determined that the sire line of *Almond Eye* was Native Dancer Line. Then, we classified 36922 horses registered with JRA from 2010 to 2017 into these four types. Table II shows the number of horses classified into these four sire line types. As shown in Table II, 90 percent of the 36922 horses were classified into the three main sire lines: Native Dancer Line, Nearctic Line, and Royal Charger Line.

36922 horses had competed in races of various distances. We grouped the race distances into five types: 1000 – 1399m, 1400 – 1799m, 1800 – 2199m, 2200 – 2799m, and more than 2800m. Then, we investigated which distance races and how many times the 36922 horses had competed in. For example, *Almond Eye* had competed in one 1000–1399m race, six 1400–1799m races, four 1800–2199m races, and four 2200 – 2799m races. Table III shows the distance and number of races the 36922 horses had competed in. Table IV shows the number of times the 36922 horses of four sire lines had competed in races of various distances. Table V shows the number of horses of four sire lines had competed in races of various distances. The point to note is that horse owners get prize money when their horses place in the top five in races held by JRA.

IV. ANALYSIS OF DECISION MAKING OF HORSE RACING EXPERTS

In this section, we investigate why horse racing experts chose not to enter their horses in other similar-distance races. Furthermore, we associate the result of the investigation with experts' race selection.

A. Basic idea

It is widely recognized that inherited variation in physical and physiological characteristics is responsible for variation in individual aptitude for race distance. Many horse racing

TABLE III
THE DISTANCE AND NUMBER OF RACES THE 36922 HORSES HAD COMPETED IN.

	race distance					
	1000-1399m	1400-1799m	1800-2199m	2200-2799m	2800m-	Total
number of races	5144	7433	7247	2083	1244	23151

TABLE IV
THE NUMBER OF TIMES THE 36922 HORSES OF FOUR SIRE LINES HAD COMPETED IN RACES OF VARIOUS DISTANCES.

	race distance					
sire line	1000-1399m	1400-1799m	1800-2199m	2200-2799m	2800m-	Total
Native Dancer	27008	31619	28568	4173	2511	93879
Nearctic	18710	22444	20072	2838	1647	65711
Royal Charger	42525	67514	71758	13181	5848	200826
others	9879	12058	10780	1817	876	35410
Total	98122	133635	131178	22009	10882	395826

experts would agree that if the best race distance of ancestors is known, the offspring's best race distance is most likely to take after them. As a result, we focus on three factors of racehorses: sire line, race distance, and order of finish. In this section, we first investigate whether horse racing experts entered their horses of certain sire lines in races of certain distances too many times or too few times. The result of this investigation shows which distance races the experts thought were favorable or unfavorable for racehorses of a certain sire line. Next, we investigate whether horse racing experts entered their horses into races of a certain distance too many times. The result of this investigation shows experts' judgements of horses' performance. By using experts' judgements of horses' performance, we investigate why horse racing experts chose not to enter their horses in other similar-distance races. Finally, we associate the result with experts' race selection.

B. Detection of race distance and sire line combinations that horse racing experts selected too many times or too few times

In order to detect cases where horse racing experts entered their horses of certain sire lines into races of certain distances too many times or too few times, we conduct the statistical analysis by using Hypothesis *ES*.

Hypothesis *ES* If experts did not enter too many times or too few times their racehorses of certain sire lines into races of certain distances, we would expect that experts entered their horses of sire line s_i into races of distance d_j at most $N_{ES}(s_i, d_j)$ times

$$N_{ES}(s_i, d_j) = P_{ES}(d_j) \times \sum_j N_{entry}(s_i, d_j) \qquad (1)$$

where d_j is the type of race distance. We classified race distances into five types:

TABLE V
THE NUMBER OF HORSES OF FOUR SIRE LINES HAD COMPETED IN RACES OF VARIOUS DISTANCES.

	race distance				
sire line	1000-1399m	1400-1799m	1800-2199m	2200-2799m	2800m-
Native Dancer	5045	7135	5599	1269	574
Nearctic	3641	5102	4053	1005	395
Royal Charger	8807	14666	13074	3794	1320
others	2056	2844	2312	561	207
Total	19549	29747	25038	6629	2496

d_1 1000 – 1399m
d_2 1400 – 1799m
d_3 1800 – 2199m
d_4 2200 – 2799m
d_5 2800m –

$N_{entry}(s_i, d_j)$ is the number of times horses of sire line s_i were entered into races of distance d_j, as a result, $\sum_j N_{entry}(s_i, d_j)$ is the total number of times horses of sire line s_i were entered into races. $P_{ES}(d_j)$ is the probability that an expert enters his/her horse into a race of distance d_j. $P_{ES}(d_j)$ is

$$P_{ES}(d_j) = \frac{\sum_i N_{entry}(s_i, d_j)}{\sum_i \sum_j N_{entry}(s_i, d_j)} \qquad (2)$$

where $\sum_i N_{entry}(s_i, d_j)$ is the total number of times horses were entered into races of distance d_j and $\sum_i \sum_j N_{entry}(s_i, d_j)$ is the total number of times horses were entered into races.

If this hypothesis is rejected by an two-sided binomial test [20], we determine that experts entered their horses of sire lines s_i into races of distance d_j too many times or too few times.

C. Detection of horses that horse racing experts judged to have performed well

If a horse perform well in a race of a certain distance, experts will try to enter the horse into another race of a similar distance. As a result, if horses are judged to have performed well in races of a certain distance, experts may enter them into other similar-distance races repeatedly. In order to detect cases where horse racing experts entered their horses into races of certain distances too many times or too few times, we conduct the statistical analysis by using Hypothesis *EJ*.

Hypothesis *EJ* If an expert did not enter too many times or too few times his/her racehorse of a certain sire line into races of a certain distance, we would expect that the expert entered horse h_k into races of distance d_j at most $M_{EJ}(h_k, d_j)$ times

$$M_{EJ}(h_k, d_j) = P_{EJ}(s_i, d_j) \times \sum_j M_{entry}(h_k, d_j) \qquad (3)$$

where s_i is the sire line of horse h_k and d_j is the type of race distance. We classified race distances into five types in the

TABLE VI
THE P-VALUES OF EXPERTS' RACE SELECTIONS FOR HORSES OF NATIVE DANCER LINE.

sire line	race distance				
	1000–1399m	1400–1799m	1800–2199m	2200–2799m	2800m–
Native Dancer	1.0000	0.3024	0.0000	0.0000	0.0825

same way that we did in Hypothesis *ES*. $\sum_j M_{entry}(h_k, d_j)$ is the number of times horse h_k were entered into races. $P_{EJ}(s_i, d_j)$ is the probability that an expert enters his/her horse of sire line s_i into a race of distance d_j. $P_{EJ}(s_i, d_j)$ is

$$P_{EJ}(s_i, d_j) = \frac{N_{entry}(s_i, d_j)}{\sum_j N_{entry}(s_i, d_j)} \quad (4)$$

where $N_{entry}(s_i, d_j)$ is the number of times horses of sire line s_i were entered into races of distance d_j. As a result, $\sum_i N_{entry}(s_i, d_j)$ is the total number of times horses were entered into races of distance d_j.

If this hypothesis is rejected by an two-sided binomial test [20], we determine that an expert entered horse h_k of sire lines s_i into races of distance d_j too many times or too few times.

D. *Results of the investigation*

In order to investigate why horse racing experts chose not to enter their horses in other similar-distance races, we apply Hypothesis *ES* and *EJ* tests on the 8799 horses of Native Dancer Line registered with JRA from 2010 to 2017, as shown in Table I. The significance levels for Hypothesis *ES* and *EJ* were 0.05. First, we calculated the p-values of experts' race selections and experts' judgements of horses' performance by applying Hypothesis *ES* and *EJ*, respectively. Table VI shows the p-values of experts' race selections for horses of Native Dancer Line. Table VII and Table VIII show the number of horses of Native Dancer Line determined by Hypothesis *EJ* to and not to have repeatedly competed in races of various distances, respectively, and also show the number of times the horses had competed in the races and the average number and the Standard Deviation (SD) of them.

First, we consider experts' expectations that their race selections suggested. Table VI, the results obtained by applying Hypothesis *ES*, shows that the p-value of race distance type d_1 (1000 – 1399m) was more than 0.975. As a result, experts entered horses of Native Dancer Line into races of distance type d_1 (1000 – 1399m) too many times. In other words, many experts strongly thought horses of Native Dancer Line were favorable to win in races of distance type d_1 (1000 – 1399m). On the other hand, the p-values of race distance type d_3 (1800 – 2199m) and d_4 (2200 – 2399m) were less than 0.025. As a result, many experts strongly thought horses of Native Dancer Line were unfavorable to win in races of distance type d_3 (1800 – 2199m) and d_4 (2200 – 2399m).

Next, we consider why experts chose not to enter their horses in other similar-distance races. There are many reasons

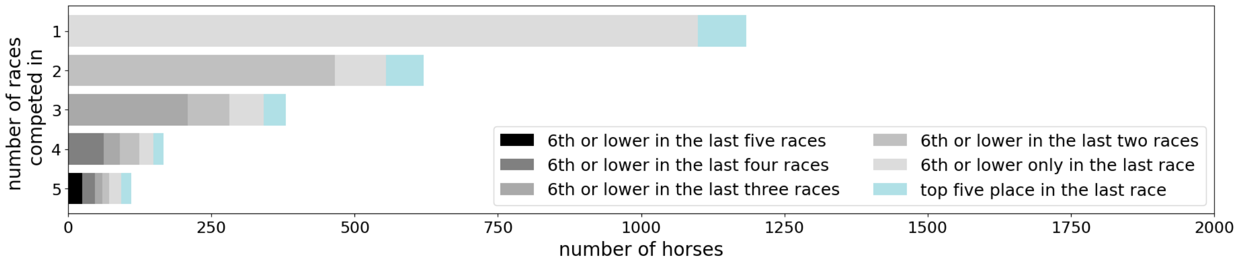
TABLE VII
THE NUMBER OF HORSES DETERMINED BY HYPOTHESIS *EJ* TO HAVE REPEATEDLY COMPETED IN RACES AND THE NUMBER OF RACES THEY HAD COMPETED IN.

	race distance				
	1000-1399m	1400-1799m	1800-2199m	2200-2799m	2800m-
horses competed repeatedly	2320	1575	1940	628	376
races competed in	20005	13491	18207	3308	2248
ave. of races competed in	8.6	8.6	9.4	5.3	6.0
SD of races competed in	7.7	7.4	7.6	5.2	5.0

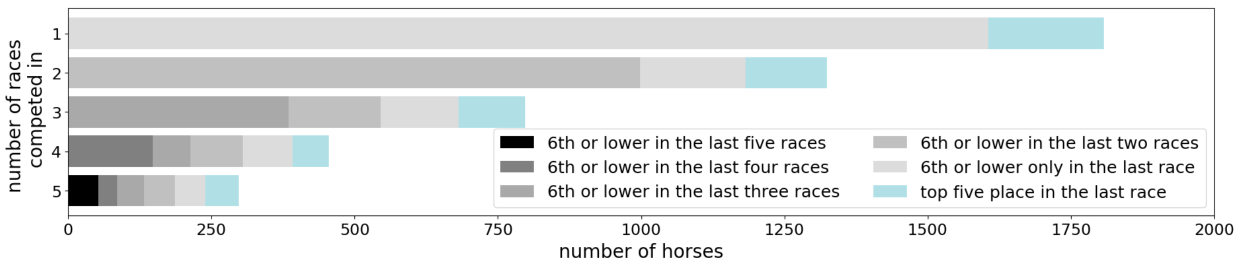
TABLE VIII
THE NUMBER OF HORSES DETERMINED BY HYPOTHESIS *EJ* NOT TO HAVE REPEATEDLY COMPETED IN RACES AND THE NUMBER OF RACES THEY HAD COMPETED IN.

	race distance				
	1000-1399m	1400-1799m	1800-2199m	2200-2799m	2800m-
horses competed not repeatedly	2725	5560	3659	641	198
races competed in	7003	18128	10361	865	263
ave. of races competed in	2.6	3.3	2.8	1.3	1.3
SD of races competed in	2.3	3.0	2.7	0.7	0.6

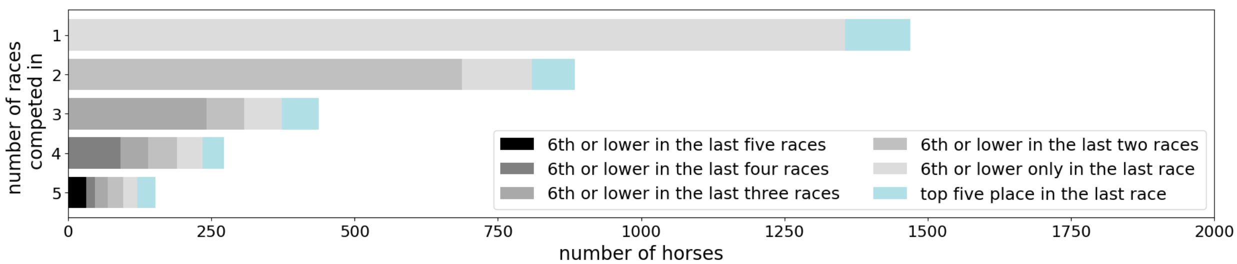
why experts chose not to do it. However, horses' performance was an important factor in experts' decisions, especially underperforming horses. As a result, we surveyed horses of Native Dancer Line determined by Hypothesis *EJ* not to have repeatedly competed in races of a certain distance and examined how many consecutive races they had finished in sixth place or lower, including their last races. The reason we focused on sixth place is because no prize money is offered when a horse finishes in sixth place or lower in races held by JRA. Figure 2 shows the number of consecutive races in which the horses competed in races of a certain distance five or fewer times finished in sixth place or lower, including their last races. Figure 2 shows that, in all race distances, there were not many horses that finished in fifth place or higher in their last races. We think that it is difficult for experts to choose not to enter their horses in other similar-distance races after they finished in fifth place or higher in the most recent races and won the prize money. On the other hand, the most common case was that horses that finished in sixth place or lower in their first races of a certain distance never competed in races of that distance again. This is the most common case even in races of distance type d_2 (1400 - 1799m) where the average number of races competed in was 3.3, as shown in Table VIII. As a result, many experts chose not to enter their horses in other races of certain distances when they finished in sixth place or lower in their first races of that distance. Next, we consider horses competed in races two or more times. Figure 2 shows that, for horses that had competed in races two or three times, the most common case was that they never finished in the top five in the races. In other words, many experts chose not to enter their horses in other races of a certain distance when their horses finished in sixth place or lower in the two or three consecutive races of that distance. For horses that had competed in races four or five times, the most common case



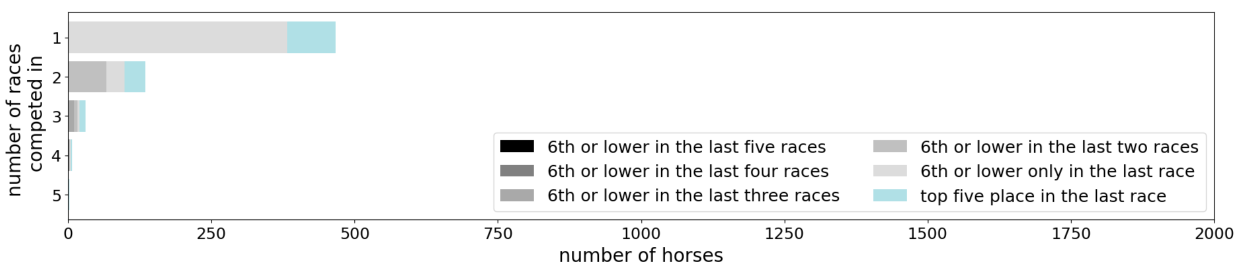
(a) race distance: 1000 – 1399m



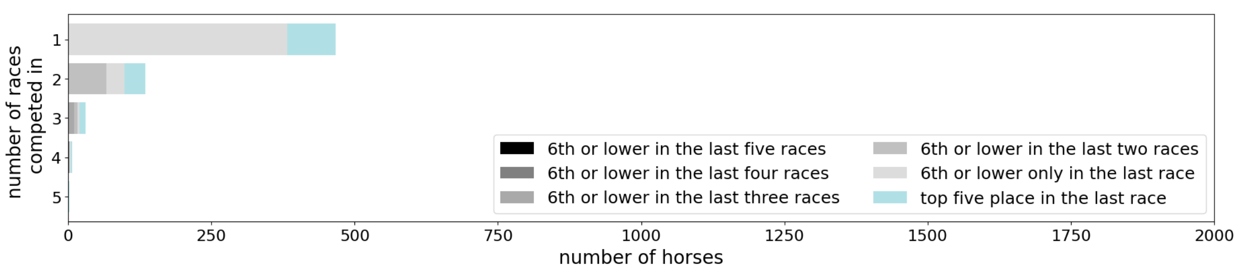
(b) race distance: 1400 – 1799m



(c) race distance: 1800 – 2199m



(d) race distance: 2200 – 2799m



(e) race distance: 2800m –

Figure 2. The number of consecutive races in which horses of Native Dancer Line that competed in races of a certain distance five or fewer times and were determined by Hypothesis *EJ* not to have repeatedly competed in races of that distance finished in sixth place or lower, including their last races.

 ISBN: 978-1-68558-353-8

was still that they never finished in the top five in the races. However, the other common case was that horses finished in sixth place or lower in the last two or three races, not every race. Furthermore, for horses that had competed in races six or more times, the most common case was that horses finished in sixth place or lower in the last two or three races, not every race. In summary,

- it is most likely that experts chose not to enter their horses in other races of a certain distance when their horses finished in sixth place or lower in their first races of that distance.
- it is also likely that experts chose not to enter their horses in other races of a certain distance when their horse finished in sixth place or lower in the last two or three consecutive races.

The point to note is that, even if the race distance was different, the timing of when many experts chose not to enter their horses in other similar-distance races was not much different. For example, Figure 2 shows that the timing of when many experts chose not to enter their horses of Native Dancer Line in other similar-distance races was not much different whether the type of race distance was d_1 (1000 – 1399m) or d_3 (1800 – 2199m). However, many experts strongly thought horses of Native Dancer Line were favorable to win in races of distance type d_1 (1000 – 1399m) while they were unfavorable to win in races of of distance type d_3 (1800 – 2199m). As a result, it is possible that many experts did not consider sire lines when choosing whether or not to enter their horses in other similar-distance races as they do when selecting race distance.

V. Conclusion and Future Work

In this paper, we investigated why horse racing experts chose not to enter their horses in other similar-distance races. We analyzed sire lines, race distances, and race results of the 36922 horses statistically and showed that many horse racing experts chose not to enter their horses in other races of any distance when their horses finished in sixth place or lower in the first races or the two or three consecutive races of that distance. The result of the investigation suggested that many horse racing experts did not consider sire lines when choosing whether to enter their horses in other similar-distance races as they do when selecting race distance. To generalize this finding, we intend to analyze race performance data from other time periods and compare the results with those obtained in this study.

References

[1] F. Tesio, *Breeding the Racehorse*. J. A. Allen, 1958.

[2] D. Epstein, *The Sports Gene: Inside the Science of Extraordinary Athletic Performance*. Penguin Publishing Group, 2013.

[3] M. A. Bower *et al.*, "The cosmopolitan maternal heritage of the Thoroughbred racehorse breed shows a significant contribution from British and Irish native mares," *Biology Letters*, vol. 7, no. 2, pp. 316–320, 2011. [Online]. Available: https://doi.org/10.1098/rsbl.2010.0800 [accessed: 2026-03-05]

[4] E. Cunningham, J. J. Dooley, R. Splan, and D. Bradley, "Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses," *Animal Genetics*, vol. 32, no. 6, pp. 360–364, 2001. [Online]. Available: https://doi.org/10.1046/j.1365-2052.2001.00785.x [accessed: 2026-03-05]

[5] E. Hill *et al.*, "History and integrity of thoroughbred dam lines revealed in equine mtDNA variation," *Animal Genetics*, vol. 33, no. 4, pp. 287–294, 2002. [Online]. Available: https://doi.org/10.1046/j.1365-2052.2002.00870.x [accessed: 2026-03-05]

[6] C. Wade *et al.*, "Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse," *Science*, vol. 326, no. 5954, pp. 865–867, 2009. [Online]. Available: https://doi.org/10.1126/science.1178158 [accessed: 2026-03-05]

[7] B. A. McGivney *et al.*, "Genomic inbreeding trends, influential sire lines and selection in the global Thoroughbred horse population," *Scientific Reports*, vol. 10, no. 1, p. 466, 2020. [Online]. Available: https://doi.org/10.1038/s41598-019-57389-5 [accessed: 2026-03-05]

[8] E. W. Hill, B. A. McGivney, J. Gu, R. Whiston, and D. E. MacHugh, "A genome-wide SNP-association study confirms a sequence variant (g. 66493737C>T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses," *BMC Genomics*, vol. 11, no. 1, pp. 1–10, 2010. [Online]. Available: https://doi.org/10.1186/1471-2164-11-552 [accessed: 2026-03-05]

[9] C. Wylie and J. Newton, "A systematic literature search to identify performance measure outcomes used in clinical studies of racehorses," *Equine Veterinary Journal*, vol. 50, no. 3, pp. 304–311, 2018. [Online]. Available: https://doi.org/10.1111/evj.12757 [accessed: 2026-03-05]

[10] G. Martin, E. Strand, and M. Kearney, "Use of statistical models to evaluate racing performance in thoroughbreds," *Journal of the American Veterinary Medical Association*, vol. 209, no. 11, pp. 1900–1906, 1996. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/8944806/ [accessed: 2026-03-05]

[11] J. Cheetham, A. Riordan, H. Mohammed, C. McIlwraith, and L. Fortier, "Relationships between race earnings and horse age, sex, gait, track surface and number of race starts for Thoroughbred and Standardbred racehorses in North America," *Equine Veterinary Journal*, vol. 42, no. 4, pp. 346–350, 2010. [Online]. Available: https://doi.org/10.1111/j.2042-3306.2010.00032.x [accessed: 2026-03-05]

[12] I. Wells, H. Randle, and J. M. Williams, "Does the start of flat races influence racehorse race performance?" *Applied Animal Behaviour Science*, vol. 253, p. 105682, 2022. [Online]. Available: https://doi.org/10.1016/j.applanim.2022.105682 [accessed: 2026-03-05]

[13] T. Sawchik, *Has the Fly-Ball Revolution Begun?*, FanGraphs Baseball, 2017. [Online]. Available: https://blogs.fangraphs.com/has-the-fly-ball-revolution-begun/ [accessed: 2026-03-05]

[14] M. Kato and T. Yanai, "Launch fly balls for better batting statistics: Applicability of "fly-ball revolution" to Japan's professional baseball league," *International Journal of Performance Analysis in Sport*, vol. 22, no. 3, pp. 437–453, 2022. [Online]. Available: https://doi.org/10.1080/24748668.2022.2075302 [accessed: 2026-03-05]

[15] R. M. Yerkes and J. D. Dodson, "The relation of strength of stimulus to rapidity of habit-formation," *Journal of Comparative Neurology and Psychology*, vol. 18, no. 5, pp. 459–482, 1908. [Online]. Available: https://doi.org/10.1002/cne.920180503 [accessed: 2026-03-05]

[16] S. Weinschenk, *100 Things Every Designer Needs to Know About People, 2nd Edition*. New Riders Publishing, Jun. 2020.

[17] S. A. Shappell and D. A. Wiegmann, *The Human Factors Analysis and Classification System–HFACS*, U.S. Department of Transportation Federal Aviation Administration, 2000. [Online]. Available: https://www.skybrary.aero/sites/default/files/bookshelf/1481.pdf [accessed: 2026-03-05]

[18] N. E. Kang and W. C. Yoon, "Age- and experience-related user behavior differences in the use of complicated electronic devices," *International Journal of Human-Computer Studies*, vol. 66, no. 6, pp. 425–437, 2008. [Online]. Available: https://doi.org/10.1016/j.ijhcs.2007.12.003 [accessed: 2026-03-05]

[19] *Keiba Lab*, Keiba Lab. [Online]. Available: https://www.keibalab.jp/ [accessed: 2026-03-05]

[20] M. Hollander and D. A. Wolfe, *Nonparametric Statistical Methods, 2nd Edition*. Wiley-Interscience, Jan. 1999.

 ISBN: 978-1-68558-353-8