RecGaze: The First Eye Tracking and User Interaction Dataset for Carousel Interfaces (Extended Abstract)*

Santiago de Leon-Martinez†
Brno University of Technology
santiago.deleon@kinit.sk

Jingwei Kang
University of Amsterdam
j.kang@uva.nl

Robert Moro
Kempelen Institute of Intelligent Technologies
robert.moro@kinit.sk

Maarten de Rijke
University of Amsterdam
m.derijke@uva.nl

Branislav Kveton
Adobe Research
kveton@adobe.com

Harrie Oosterhuis
Radboud University
harrie.oosterhuis@ru.nl

Maria Bielikova
Kempelen Institute of Intelligent Technologies
maria.bielikova@kinit.sk

1 Introduction

Understanding where users look and what items are observed when presented with results is vital for many off-policy estimation (OPE) methods [6, 8, 11, 14, 16]. In traditional single-ranked-list interfaces user browsing behavior tends to follow unidirectional top-down browsing [9]. Different domains and scenarios may, however, require different interfaces with more complex browsing behaviors that complicate determining observation propensities. For instance, in web image search results are typically displayed in a 2D grid (i.e., single 2D ranked list), allowing users to inspect the results in a vertical as well as a horizontal direction [18]. *Carousel* and *multi-list* interfaces have become a popular way for e-commerce and streaming services to display items (see Figure 3 in Appendix for an example of a carousel interface). These interfaces organize items into topics, where a topic consists of items that share certain characteristics. Multiple topics are shown below each other in a list that can be scrolled vertically and each topic is shown as a row that can be scrolled horizontally [2].

While carousels are an effective means of displaying recommendations through their use and success in practice (e.g., Netlifx or Spotify), there is little research on why carousels may be better than other interfaces. Moreover, there is little research in how users interact with them, which may greatly impact how systems are designed in practice and how they should be evaluated, especially for methods like OPE that rely on observations. In contrast to the hundreds of publications on simpler interfaces like single 1D ranked lists and single 2D ranked lists (cf. [3]), we have only been able to locate fewer than 30 publications on carousels and multi-lists. We believe that this lack of research is due to:

(1) a lack of publicly available datasets of interaction data with carousels and multi-lists, and
(2) a lack of empirical studies of users' browsing behaviors.

Our goal in this paper is to address the first problem. We address the second in a follow-up work [6], an extensive analysis of the gaze data revealing how users browse carousel interfaces and providing initial values for propensity estimation. We hope that the release of a dataset of interaction data with carousels enables researchers to advance the field of carousel and multi-list interfaces.

This work extends beyond just carousel/multi-list interfaces and presents the first publicly available recommender dataset with eye tracking. Eye tracking technology has been steadily developing over the years. Examples of this can be seen in the continued development and popularity of virtual reality (VR) and augmented reality (AR) devices (e.g., Apple Vision Pro [1] or Meta Quest 3 [15]), which point to a near future where a VR/AR headset or glasses replace today's devices. Gaze will soon be an accessible online data stream for recommender system designers that can revolutionize recommenders by providing item observations, item dwell times, and gaze interaction sequences before a click. The latter can greatly help in determining what leads to a click or purchase.

This work seeks to achieve the following objectives:

(1) Provide the first recommendation dataset with eye tracking data (along with cursor and selection explanations), enabling researchers to empirically understand how users browse carousel interfaces and advance the field of gaze-based recommenders;
(2) Provide the first click feedback dataset in carousels and the second in multi-lists with more than 9 times as many interactions (movie selections) than prior work, enabling research in the underdeveloped field of carousel/multi-list [10]; and
(3) Present the first gaze visualizations in carousel interfaces supporting golden triangle or F-pattern browsing behavior.

2 User Study Methodology

To better understand how users browse carousel interfaces, we designed an eye tracked user study of 40 screens (see Figure 3 in Appendix for a sample screen), where participants picked one movie found on a carousel page (or exited without selection). The user study was designed with the following goals in mind:

(1) Determine any general browsing behaviors across users;
(2) Determine the impact of carousel genre preference on browsing;
(3) Examine the differences in browsing behavior across 3 browsing tasks: free-browsing, semi-free-browsing, direct search;
(4) Survey users to address open questions in carousel design [13];
(5) Provide an extensive mouse, click, and gaze dataset along with user explanations for future research.

The study consisted of three parts: (i) a pre-survey, (ii) movie selection tasks, and (iii) a post-survey. The survey sections gather user information, particularly their genre preferences, ideas on

*Extended abstract of the SIGIR '25 resource paper with the same title [5].
†Also with Kempelen Institute of Intelligent Technologies.

Table 1: Task summaries of interaction feedback, the number of movies selected (users can exit without selecting) and average screen time (until selection or exit).

Task	Users	Screens	Interactions	Movie Selections	Avg. screen time
Free-browsing	87	30	2,607	2,432	56.99s
Semi-free	87	5	435	383	43.85s
Direct Search	87	5	435	424	11.57s
All Tasks	87	40	3,477	3,239	49.67s

carousel interfaces, and feedback on the interface. The main part of the study are 40 screens of movie selection tasks. Of these 40 screens, 30 are free-browsing (pick any movie on the screen), 5 are semi-free browsing (pick a movie from your favorite genre), and 5 are direct search tasks (find a specific movie). The same screens and tasks were used for every participant. We used a larger number of free-browsing screens as this scenario is the closest to real-world interactions and while examining behavior in other scenarios may be insightful, we did not want to make the study too long.

Additionally, after the free-browsing and semi-free browsing search tasks, participants were asked to give feedback on their movie selection: familiarity with the movie selected and why they selected it. Along with selection explanations, we gathered mouse movements, clicks, gaze, and screen recordings throughout the movie selection tasks to provide the most complete recommender system dataset that is publicly available. A summary of the tasks and screens with the number of interactions (screens where feedback data was gathered) can be found in Table 1.

3 Dataset Features

We provide two datasets for the community to use due to the sensitive nature of raw gaze and cursor data that could be used to potentially link participants between datasets. The first is a public dataset that can be found on Zenodo[1] with summarized fixation and mouse information with already timestamped labeling of Areas of Interest (AOIs) for cursor, clicks, and fixations. The supplementary material (stimuli, all screen layouts, etc.) are available in a GitHub repository.[2] The public dataset is designed to be an already preprocessed dataset that can be used out of the box for research in recommender systems, information retrieval and human computer interaction. It is organized in 4 dataframes in csv format. For the list of provided user features, item features, and summary feedback dataframes see Table 2 in the Appendix. We additionally provide a fourth dataframe that is a simple click feedback dataset with only the final movie selection click per screen per user (in total 3,239 selections) and also contains the selection explanation information.

The second is a non-public dataset that contains sensitive information that can be provided by reaching out to the authors. It is better suited for researchers that are more interested in eye/mouse tracking research or would like to do the gaze preprocessing (e.g., AOI linking and fixations) themselves.

[1] https://zenodo.org/records/15270518
[2] https://github.com/santideleon/RecGaze_Dataset

4 Use Cases for RecGaze Dataset

The RecGaze dataset is the first recommender dataset to provide a wide range of feedback data (eye tracking, cursor, clicks, and selection explanations). For this reason, it can lend itself to many applications. First and foremost, it allows one to comprehensively analyze and understand possible user behaviors within the carousel interfaces by combining all available modalities as illustrated in Section 5. Secondly, it can be used to fine-tune designs of these interfaces and their underlying models. The following are additional possible use cases for which the dataset may serve:

Click models for carousel interfaces. Designing click models based on observed browsing behaviors like the F-pattern (see Figure 1) and using the data to determine parameters [12].

Propensity Estimation and OPE. Gaze data linked to item AOIs provides observation propensity and can be used as initial propensity weights for similar carousel interfaces [6, 11, 16]. Moreover, the behaviors observed in gaze data (and other modalities) across users and screens can be used for OPE or offline simulations [14].

Whole page optimization. As an extension of click models or with general insights found in the data, one can better understand how different carousels interact with each other, particularly taking into account preferences, and design better page layouts [7, 17].

Impressions. The dataset can also be used for impressions, as they can be extracted and compared to the "ground truth" gaze data.

5 Visualizing Browsing of Carousels

In Figures 1 and 2 , we present the first visualizations of gaze data in carousel interfaces. For Figures 1 and 2a, fixation data with (x, y) pixel position value were used (non-public dataset), and Figures 2b and 2c used position value for clicks and cursor respectively (non-public dataset) rather than AOI allocation (public dataset).

In Figure 1, fixation data from all tasks across every user was combined to generate an aggregated heat map on scale to the stimuli (also included in GitHub). The heat map also takes into account fixation duration. Therefore, the heat is representative of both how long and how often users were fixating on a location. The heat map points to a strong top-down browsing behavior, left-right browsing behavior (at least on page 1), and a bias towards initially presented movies (unswiped). These preliminary results support the golden triangle browsing behavior or F-shaped pattern seen in single 2D ranked lists [19], which supports our original hypothesis that this behavior would also be seen in carousels [4].

However, on pages 2 and 3 more fixations can be seen on the farthest right movies which go against the general trend of F-pattern seen, particularly left-right browsing. We hypothesize that the increase in fixations on the last presented movie is due to swiping behavior. This was confirmed in a follow-up analysis of the gaze data [6], showing that users almost exclusively swipe from the right side after viewing the last item and after the swipe users re-examine the same (last) movie position. Users thus exhibit a right-left browsing behavior after a right swipe resulting in the flipped F-pattern seen in Figure 1.

For Figures 2a to 2c, we selected one participant and one free-browsing screen to visualize the types of data available. They present scan paths showing the points where data were registered

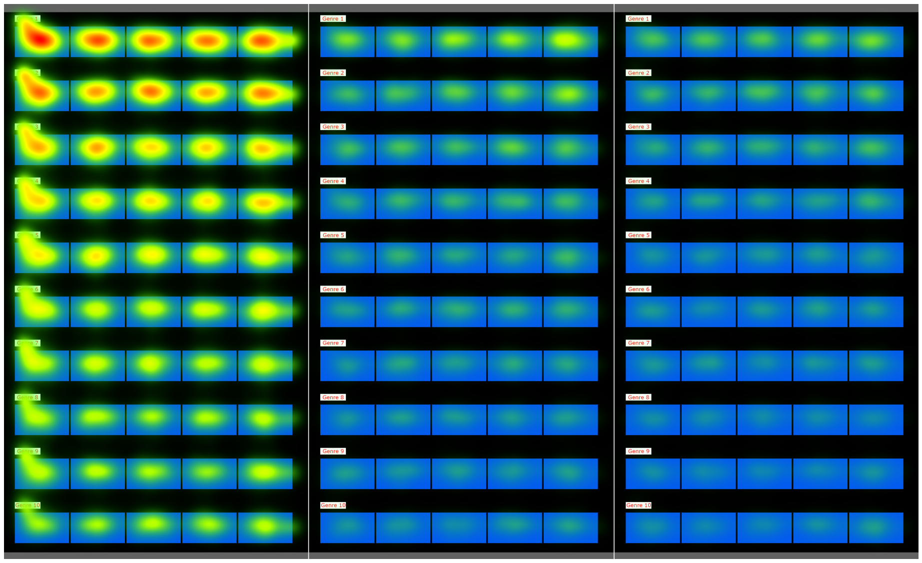
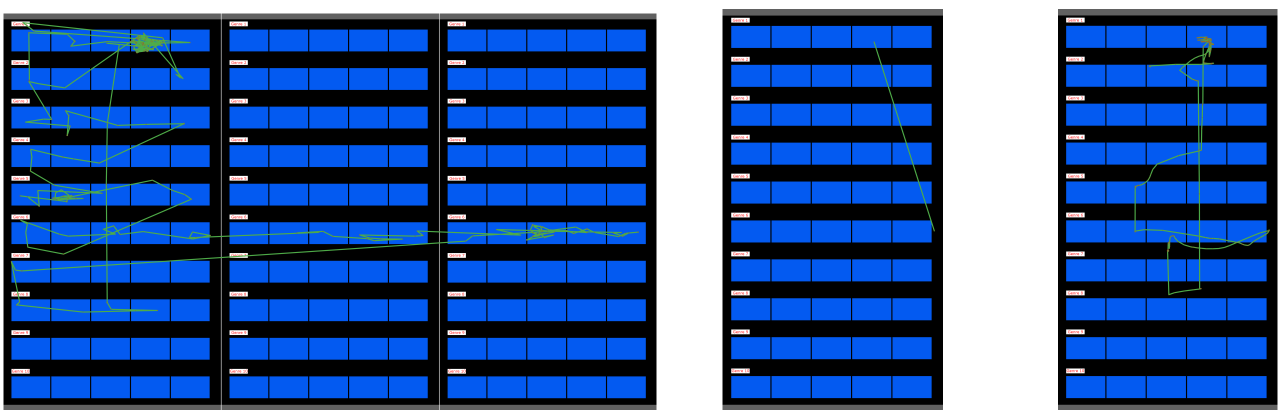


Figure 1: Aggregate fixation heat map for every user on all 30 free-browsing screens with duration weighting shown on a to-scale background stimuli of the movie screens (with movie posters shown as blue and genre text as white boxes). It includes horizontal displacement, so initial 5 movies can be distinguished from second and third set from swiping.



(a) Fixation scan path with swipe differences **(b) Click scan path** **(c) Cursor scan path**

Figure 2: Visualizations of fixations, clicks and mouse cursor movements for one selected user and one free-browsing screen.

(red circles) with radius proportional to the duration. For this screen, the participant began to fixate on the first, left-most movie of genre 2 and ended on the fourth movie of genre 1, as can be seen in the fixation data from Figure 2a. Figure 2b shows the clicks for two same location forward swipes on genre 6 and the final click to select the fourth movie of genre 1. Finally, the cursor path can be seen in Figure 2c from an initial cursor position hovering the third movie of genre 2. The scan paths provide a detailed summary of the participant's browsing/interaction sequence. For example, multiple fixations (Figure 2a) and mouse cursor movements (Figure 2c) on the selected fourth movie of genre 1 are indicative of the user reading its description (tracing the text with the cursor).

It is also possible to add user preference information to interpret these results. In Figure 2a, we can see that the participant often examined multiple movies of each genre that they scrolled to. Outside of genre 1, the majority of fixations and all swipes are found in genres 5 (Drama) or 6 (Animation), which are both user's preferred ones. On the other hand, genre 4 (Fantasy), which was marked as the top genre by the user, lacked fixations when compared to the other ones. We provide this as an example and further explore and confirm the impact of genres on browsing in a follow-up work [6].

6 Conclusion

Although carousel and multi-list interfaces are widely used in commercial services, thus impacting users daily, there is little research that is focused on this type of interface. In this work, we help address the barriers to research in carousels and encourage a return to eye tracking and empirical understanding of user behavior by providing the RecGaze dataset. We demonstrated the utility of the dataset for a range of use cases and provided the first gaze analysis of user behavior in carousel interfaces supporting the golden triangle or F-pattern browsing behavior. By providing this dataset, we hope to enable and advance research in carousel and multi-list interfaces as well as gaze-based recommendation systems.

References

[1] Apple 2025. Apple Vision Pro — apple.com. https://www.apple.com/apple-vision-pro/. [Accessed 03-02-2025].

[2] Walid Bendada, Guillaume Salha, and Théo Bontempelli. 2020. Carousel Personalization in Music Streaming Apps with Contextual Bandits. In *Proceedings of the 14th ACM Conference on Recommender Systems*. Association for Computing Machinery, 420–425.

[3] Aleksandr Chuklin, Ilya Markov, and Maarten de Rijke. 2015. *Click Models for Web Search*. Morgan & Claypool Publishers.

[4] Santiago de Leon-Martinez. 2024. Understanding User Behavior in Carousel Recommendation Systems for Click Modeling and Learning to Rank. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining* (Merida, Mexico) *(WSDM '24)*. Association for Computing Machinery, New York, NY, USA, 1139–1141. https://doi.org/10.1145/3616855.3635734

[5] Santiago de Leon-Martinez, Jingwei Kang, Robert Moro, Maarten de Rijke, Branislav Kveton, Harrie Oosterhuis, and Maria Bielikova. 2025. RecGaze: The First Eye Tracking and User Interaction Dataset for Carousel Interfaces. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Padua, Italy) *(SIGIR '25)*. Association for Computing Machinery, New York, NY, USA, 3702–3711. https://doi.org/10.1145/3726302.3730301

[6] Santiago de Leon-Martinez, Robert Moro, Branislav Kveton, and Maria Bielikova. 2025. Riding the Carousel: The First Extensive Eye Tracking Analysis of Browsing Behavior in Carousel Recommenders. arXiv:2507.10135 [cs.IR] https://arxiv.org/abs/2507.10135

[7] Weicong Ding, Dinesh Govindaraj, and S V N Vishwanathan. 2019. Whole Page Optimization with Global Constraints. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Anchorage, AK, USA) *(KDD '19)*. Association for Computing Machinery, New York, NY, USA, 3153–3161. https://doi.org/10.1145/3292500.3330675

[8] Shashank Gupta, Philipp Hager, Jin Huang, Ali Vardasbi, and Harrie Oosterhuis. 2024. Unbiased Learning to Rank: On Recent Advances and Practical Applications. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining* (Merida, Mexico) *(WSDM '24)*. Association for Computing Machinery, New York, NY, USA, 1118–1121. https://doi.org/10.1145/3616855.3636451

[9] Marti A. Hearst. 2009. *Search User Interfaces*. Cambridge University Press.

[10] Dietmar Jannach, Mathias Jesse, Michael Jugovac, and Christoph Trattner. 2021. Exploring Multi-List User Interfaces for Similar-Item Recommendations. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, New York, NY, USA, 224–228.

[11] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. 2017. Unbiased learning-to-rank with biased feedback. In *Proceedings of the tenth ACM international conference on web search and data mining*. 781–789.

[12] Jingwei Kang, Maarten de Rijke, Santiago de Leon-Martinez, and Harrie Oosterhuis. 2025. Rethinking Click Models in Light of Carousel Interfaces: Theory-Based Categorization and Design of Click Models. In *Proceedings of the 2025 International ACM SIGIR Conference on Innovative Concepts and Theories in Information Retrieval (ICTIR)* (Padua, Italy) *(ICTIR '25)*. Association for Computing Machinery, New York, NY, USA, 44–55. https://doi.org/10.1145/3731120.3744585

[13] Benedikt Loepp. 2023. Multi-list Interfaces for Recommender Systems: Survey and Future Directions. *Frontiers in Big Data* 6 (2023), 1239705.

[14] Dan Luo, Lixin Zou, Qingyao Ai, Zhiyu Chen, Dawei Yin, and Brian D. Davison. 2023. Model-based Unbiased Learning to Rank. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining* (Singapore, Singapore) *(WSDM '23)*. Association for Computing Machinery, New York, NY, USA, 895–903. https://doi.org/10.1145/3539597.3570395

[15] Meta 2025. Meta Quest 3 — Meta.com. https://www.meta.com/quest/quest-3/. [Accessed 03-02-2025].

[16] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as Treatments: Debiasing Learning and Evaluation. In *Proceedings of The 33rd International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 48)*, Maria Florina Balcan and Kilian Q. Weinberger (Eds.). PMLR, New York, New York, USA, 1670–1679. https://proceedings.mlr.press/v48/schnabel16.html

[17] Yue Wang, Dawei Yin, Luo Jie, Pengyuan Wang, Makoto Yamada, Yi Chang, and Qiaozhu Mei. 2016. Beyond Ranking: Optimizing Whole-Page Presentation. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining* (San Francisco, California, USA) *(WSDM '16)*. Association for Computing Machinery, New York, NY, USA, 103–112. https://doi.org/10.1145/2835776.2835824

[18] Xiaohui Xie, Yiqun Liu, Maarten de Rijke, Jiyin He, Min Zhang, and Shaoping Ma. 2018. Why People Search for Images using Web Search Engines. In *WSDM 2018: 11th International Conference on Web Search and Data Mining*. ACM, 655–663.

[19] Qian Zhao, Shuo Chang, F. Maxwell Harper, and Joseph A. Konstan. 2016. Gaze Prediction for Recommender Systems. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) *(RecSys '16)*. Association for Computing Machinery, New York, NY, USA, 131–138. https://doi.org/10.1145/2959100.2959150

Appendix

We provide an example screen of the movie selection interface viewed by participants during the user study shown in Figure 3. Additionally, we provide a large table that describes the major features of the RecGaze dataset in detail with accompanying examples in Table 2.

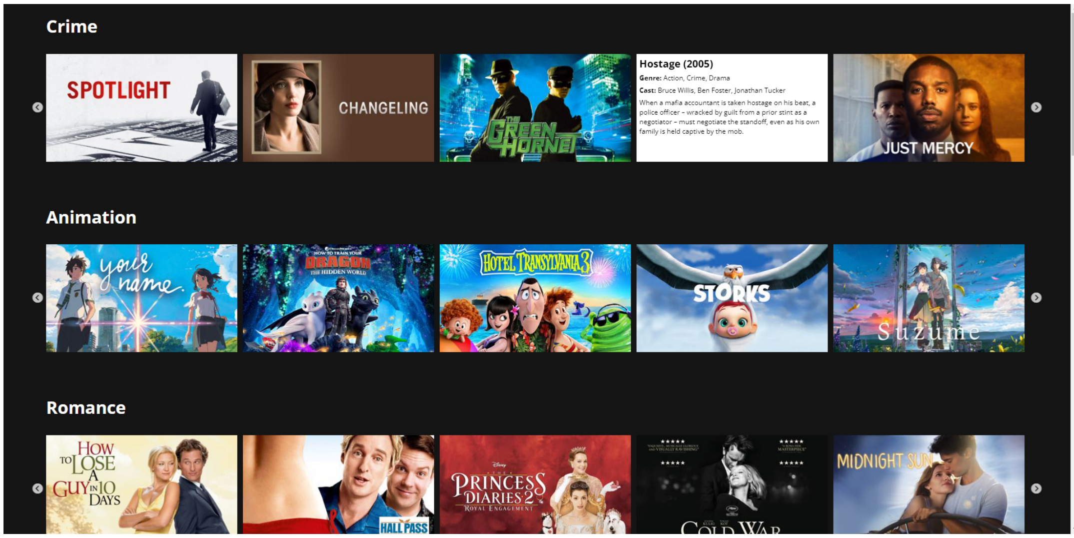


Figure 3: A capture of a free-browsing screen at initial presentation with the first 3 carousels shown and mouse-over details of a movie in the top row. Movies are ranked (left to right) by popularity based on IMDB number of votes. Movie posters and mouse-over details were shown (removed here due to copyright). See GitHub repository for a sample screen recording gif.

Table 2: Major features of the RecGaze Dataset. Features marked with * are part of the non-public dataset and square brackets [] refer to multiple responses. To be concise, we use shorthand names, combine similar features, and denote different columns using bold. For a detailed explanation of the features refer to the GitHub documentation.

Category	Feature name	Type	Example
User features	UserID	Multi	KInIT_18 (Location_ID)
	Age*	Range	30-39
	Gender*	Text	Man
	Genre preferences	Multi	**Q_Top:** Action; **Q_Preferred:** [Action, Comedy]; **Q_Ratings 1-5:** [Action_5, Animation_1...]
	Interface expertise	Multi	3-4 times per week
	Movie expertise	Multi	1-2 times per month
	Decision-making	Int 1-7	**Q_1:** 1; **Q_2:** 3; **Q_3:** 5; **Q_4:** 7; **Q_5:** 4; **Q_6:** 5;
	Genre/Movie #	Text	**Q_1:** Yes, I felt overwhelmed by # of Genres; **Q_2:** No, I didn't feel overwhelmed by # of movies
Item features	MovieID	Int	372058 (ID from TMDB)
	Screen position	Multi	**TaskID/Screen**: 7; **Carousel_position**: 1; **Genre**: Horror; **Movie_position**: 6
	Title & Year	Multi	**Title**: Hostage; **Year**: 2005
	Description	Text	"When a mafia accountant is taken host..."
	Poster	Img	http://{url}
Feedback & AOI label	Gaze (90 Hz)*	Multi	**Time:** 1.241s; **Position**: 120, 300 px;
	Fixation (90 Hz)*	Multi	**Time:** 1.241s; **Position**: 120, 300 px; **AOI:** Movie_372058; **Duration**: 0.891s
	Cursor (event)*	Multi	**Time:** 2.450s; **Position**: 100, 600 px; **AOI:** GenreText_1_Horror
	Click (event)*	Multi	**Time:** 10.903s; **Position**: 100, 142 px; **AOI:** BackSwipe_Genre_1_Horror
Summary feedback	Fixation sequence	Multi	**Time:** 1.241s; **AOI:** Movie_372058; **Duration**: 0.891s
	Cursor sequence	Multi	**Time:** 2.450s; **AOI:** GenreText_1_Horror; **Duration**: 1.902s
	Click sequence	Multi	**Time:** 10.903s; **AOI:** BackSwipe_Genre_1_Horror;
Selection explanation	Movie familiarity	Text	I have seen part of the movie
	Why selected?	Text	[Because of the poster, Because of the details, Other: "I like the director"]
Other	Screen recording*	Video	25 fps .avi video recording