Navigating Bias in Streaming

A Qualitatively Driven Mixed-Methods Study of User Experience and Agency on Netflix

by
Burak Kiraz
No. 16

Characters: 140.029
Pages: 58,3

Abstract

This thesis investigates the impact of algorithmic and user-driven biases in Netflix's recommendation system on the user experience and autonomy in content exploration. With a pragmatic approach and a qualitatively driven mixed-methods study, the study combines three methods: bias mapping of Netflix's UI, a 10-day diary study, and semi-structured interviews conducted before and after the diary study period.

The thesis identifies algorithmic biases, such as popularity, engagement, and positioning bias, which structurally favor a specific type of content, thereby limiting users' ability to discover new content. At the same time, cognitive and behavioral user biases, such as confirmation bias, choice overload, and trust in algorithms, contribute to reinforcing patterns in Netflix's decision-making processes. These biases interact with the UI's design choices, creating a user experience that strikes a balance between comfort and control.

The results show that users tend to passively navigate the interface, which autoplay-feature and visual hierarchies contribute to, as it prioritizes the platform's productions. Although users appreciate the system's personalization logic, many feel a lack of autonomy and variety. The diary study and interviews reveal that emotional and situational factors significantly influence these user patterns.

In conclusion, designing recommendation systems should focus on transparency, user control, and diversity in content presentation. Recommendation systems should support higher levels of user involvement and personalization without compromising freedom of discovery and critical reflection.

Table of Contents

1. Introduction

Streaming video-on-demand platforms (SVODs) are the leading way to watch films and TV shows. Many platforms, such as Netflix, Disney+, and MAX, offer a wide range of content, often presented through personalized recommendations. Netflix is one of the biggest streaming services available today and utilizes a recommendation system designed to interpret its users' diverse wishes and preferences. Recommendation systems generate personalized suggestions based on user behavior, interactions, and interests. Although such systems are designed to save users' time and enhance the user experience of these platforms, a growing body of research indicates that they also come with challenges and biases (Abdollahpouri et al., 2019; Bourreau & Gaudin, 2021; Schaffner et al., 2023; Sun et al., 2023).

Algorithmic and user-related biases, such as popularity bias, promotional bias, and confirmation bias, can expose users to similar content, potentially limiting their exploration of new and diverse content (e.g., Abdollahpouri et al., 2019; Schaffner et al., 2023). These biases can reinforce homogeneity in recommendations, giving users the illusion of choice. For example, the popularity bias reinforces titles that are far more popular, while the engagement bias encourages passive consumption, such as the autoplay feature. These patterns of passive interaction, exemplified by users accepting top-listed recommendations or relying on autoplay, can reduce user agency over time (Schaffner et al., 2023; Ahn & Lin, 2024). This is harmful because exposure to diverse content is limited, thereby weakening informed decision-making and making individuals overly reliant on platform logic rather than their intent.

This problem gains relevance as algorithmic decisions have increasingly been guiding media consumption in ways we cannot comprehend. Research on user interfaces and personalization strategies reveals how visual hierarchies and default settings can influence user behavior (Sun et al., 2023).

Therefore, it is essential to examine how recommendation systems influence the user experience and how interface design can either enhance or limit user autonomy.

To examine this, I have formulated the following problem statement:
How do algorithmic and user-driven biases in Netflix's recommendation system affect content discovery, and how might design solutions improve user exploration and agency?

To ensure this thesis answers the overall problem statement, the following four research questions will guide the research process:

1. What types of biases are embedded in Netflix's recommendation logic, and how are they manifested through its user interface?
2. How do users engage with and interpret Netflix's recommendations in everyday use, and how do these behaviors contribute to the reinforcement of bias?
3. To what extent does Netflix's interface support or hinder users' sense of autonomy and agency in content discovery?
4. How can insights on this problem be used for designing more inclusive and transparent interfaces?

This study will employ a qualitatively driven mixed-methods approach, combining a diary study, semi-structured interviews, and an analytical overview of Netflix's user interface. The purpose is to highlight how biases emerge and are experienced in practice by identifying potential ways to improve the recommendation system in both design and functionality, with a focus on users' perspectives.

1.1 Delimitation & Limitations

This thesis aims to understand how recommendation systems on streaming platforms influence user experiences of choice and content discovery. By focusing on Netflix, we can conduct an in-depth analysis of the complex interplay between algorithmic bias, user behavior, and interface design. While this approach limits broader comparisons across multiple platforms, it allows for a more nuanced understanding of the dynamics within a single, widely used system.

Moreover, as this thesis focuses on users' experiences with the system, it will not include technical aspects or the precise functions of algorithms. The analysis builds upon participant data, visual structures, and theoretical perspectives on bias and user agency. This approach means that the results are to be understood as experience-based and interpretive-driven rather than technically anchored. Although external influential sources, such as social media or recommendations from inner circles, are mentioned, these factors will not be analyzed in depth despite their potential role in users' experiences with SVOD platforms. The diary study's 10-day duration provides insights into short-term viewing patterns. However, it may not capture longer-term behavioral changes or the evolution of participants' preferences over time. While purposeful sampling allows for rich contextual insights, the small sample size of six means that the findings are not statistically generalizable to all Netflix users. Instead, it offers valuable perspectives on how a select group of people experiences the platform. Furthermore, as the diary study relies on self-reported data, which introduces the risk of self-reporting bias (e.g., participants may present themselves in a certain way or forget to log their watched content), several measures were taken into consideration to prevent these effects. These measures include providing clear and consistent instructions, ensuring participants' anonymity, and emphasizing the importance of honest reporting. Moreover, as the researcher, I acknowledge my lack of consistency in conducting follow-ups with participants, which resulted in minor variations in the dataset, as elaborated in the analysis section of this thesis.

2. Literature Review

This section presents a comprehensive literature review that explores biases in Netflix's recommendation system and their impact on content discovery. The review synthesizes relevant studies to establish a foundational understanding of systematic biases in subscription video-on-demand (SVOD) platforms. The literature review will be structured into two main areas: algorithmic biases in Netflix's recommendation system and user-centered biases, as well as their influence on engagement. By reviewing existing literature, this section highlights the key challenges in recommendation systems, justifying the present study.

2.1 Foundation for Review

Randolph emphasizes that a literature review must create a robust theoretical and empirical foundation by systematically exploring existing research (2009, p. 2). However, this thesis does not adhere to the formal criteria of a systematic review, but instead employs a critical and selective review approach. This is to engage with existing literature that addresses personalization, bias, and user interaction to support the exploration of user experience.
This study employs a selective critical review approach to examine how algorithmic biases impact Netflix's recommendation system. This literature review initiates content discovery research, where personalization and diversity considerations are in tension, by addressing gaps in the existing literature. This literature review follows the approach outlined by Cronin, Ryan, and Coughlan (2008), structuring the analysis as a critical framework for engaging with the existing research. Traditional reviews enable a comprehensive scrutiny of the literature, analyzing central issues, methods, and outcomes discussed in the materials

analyzed (Cronin et al., 2008, p. 4). This review was constructed to summarize the key findings and knowledge gaps in the literature, with a selective focus on content personalization, recommendation system bias, and user interaction analysis. The insights gathered here will inform both the theoretical framework and the methodological design discussed in later chapters.

2.2 Approach

To ensure the relevance and credibility of the reviewed literature, articles were found from academic databases, including Google Scholar, ACM Digital Library, and SpringerLink. The inclusion of ACM Digital Library is particularly relevant given its focus on recommender systems and human-computer interaction.

The interdisciplinary nature of the research, spanning recommender system biases, user behavior, and SVOD, necessitated a search strategy that bridged fields such as information science, media studies, and qualitative methodology. The following list of keywords and phrases was used in the search process:

- "Algorithmic bias in Netflix recommendations"
- "Biases in streaming platforms"
- "User behavior and filter bubbles"
- "Confirmation bias in media consumption"
- "Choice overload and recommender systems"
- “Streaming video on demand (SVOD)”

These keywords and phrases served as flexible guidelines in the search. They were refined and adapted throughout the process to ensure comprehensive and relevant results. In addition to keyword searches, citation chasing and backward referencing were used to locate relevant foundational works, such as Schaffner et al.'s (2023) study. Multiple search engines were consulted to ensure comprehensive coverage, and the inclusion criteria focused on peer-reviewed studies relevant to Netflix or comparable platforms.

2.3 Synthesis of Related Work

This section organizes prior research within the problem field into key themes related to biases in video-on-demand platforms’ recommendation systems. To ensure transparency, I will provide relevant examples of the literature search from each overall theme.

2.3.1 Algorithmic Biases in Netflix's Algorithm

For this theme, the literature search focused on systemic biases, primarily on Netflix or similar platforms. The following is one example of a keyword-based search session conducted in Google Scholar:

SVOD and challenges AND "recommendation system"
→ 286 results
→ SVOD and challenges AND "recommendation system" AND user engagement
→ 61 results
"User engagement" and challenges AND "recommendation systems" AND "streaming video on demand"
→ 12 results

From the 12 results in the final search, one highly relevant article was *Don't Let Netflix Drive the Bus: User's Sense of Agency Over Time and Content Choice on Netflix* (Schaffner et al., 2023). Other articles mentioned in this section, such as Bourreau & Gaudin (2021), Mansoury et al. (2023), and Van Es (2024), were identified through forward and backward citation chasing from this key source, as well as follow-up searches using new combinations of search terms.

Netflix's recommendation system exists to increase user engagement; however, it simultaneously introduces issues that affect users' opportunities to explore content (Bourreau & Gaudin, 2021, p. 25). One example of this is popularity bias, where the algorithm or system prioritizes popular content over lesser-known content. Two studies argue that this creates a "rich-get-richer" effect, where popular movies and series continue to be seen, while niche and underrepresented content are overlooked (Abdollahpouri et al., 2019; Mansoury et al., 2023).

Khoo (2022) and Anwar et al. (2024) both highlight another challenge within Netflix's recommendation system: the strategic placement of their studio project at the top of the recommendations rows, which reduces the visibility of third-party content. Such a challenge becomes a positioning bias. Furthermore, the engagement bias and autoplay feature play a crucial role in reinforcing recommendation loops (Anwar et al., 2024, p. 2). Schaffner et al. (2023) demonstrate how Netflix's autoplay feature tends to decrease users' active search for content, instead prompting users to turn to pre-picked recommendations (pp. 1-2). According to the researchers, the autoplay feature promotes passive consumption habits, which support the platform's engagement-driven model rather than stimulating broader and more diverse content exploration (Schaffner et al., 2023, pp. 20-21).

Other research shows that Netflix's user interface and systematic promotion of specific rows, such as "Top 10" or "Because you watched," forces users to find themselves in a loop of the same type of recommendations (Sun et al., 2023, p. 7). Following this, Van Es argues that Netflix advertises itself as allowing and enforcing personalization; however, its hidden strategy is that its recommendation system favors Originals and/or commercially attractive titles over meeting every user's preference (2024, pp. 4-5). This aligns with what Bourreau and Gaudin (2021) describe as a form of self-preferencing, where vertically integrated platforms tend to prioritize their content (Netflix Originals), often limiting the visibility of third-party titles (p. 25). This practice is understood as a promotional bias, as the system favors platform-owned productions despite individual user preferences.

One aspect of Netflix's interface that seems to be overlooked is the search feature. The scope of studies examining this is limited, perhaps because it is sometimes perceived as a user-controlled space for discovery. A targeted search on Google Scholar was employed:

"SVOD" and "search engine" AND Netflix AND recommendation

→ 353 results

"SVOD" and "search engine" AND Netflix AND "recommendation algorithm"

→ 30 results

Of the 30 results, only one article directly discussed the search feature. Olma, Rizun, and Strzelecki (2024) emphasize that users who understand the underlying mechanisms of Netflix's search and recommendation algorithms are better equipped to formulate effective queries and experience a greater sense of satisfaction and control (pp. 194–195).

Building on this foundation, two additional peer-reviewed articles were identified through citation chasing: Lamkhede & Das (2019) and Lamkhede & Kofler (2021). These studies investigate the technical and design-level integration of recommendation logic within Netflix's search system. They reveal how search results are not purely lexical but algorithmically curated to reflect user behavior, platform objectives, and engagement metrics.

This blending of recommendations within search constitutes a form of search bias, where the results shown to users are shaped not only by their input but by platform-driven prioritization strategies. As such, even active content discovery is subject to algorithmic steering, challenging the assumption that search guarantees user autonomy.

Based on these studies, Netflix exhibits several systematic biases that, in practice, work against its intended personalization. The factors include popularity, promotion, positioning, engagement, and search biases that collectively steer users towards a narrow band of content. However, it is essential to note that many of these findings are based on external observations and interface analysis, as Netflix's closed policies regarding its internal algorithms and user data limit access to internal information. Therefore, while such studies are valuable, their conclusions should be considered critical yet interpretative.

To summarize the findings in this section, Table 1 presents an overview of the algorithmic biases discussed in the literature and the studies that address them. This overview forms the basis for the empirical analysis in Chapter 5, where the same bias types will be explored through interviews and diary data.

Bias Type	Definition / Effect	Discussed In
Popularity Bias	Overpromotion of popular content, rich-get-richer cycle	Abdollahpouri et al. (2019); Mansoury et al. (2023)
Positioning Bias	Prominent placement of certain content rows (e.g., Originals)	Khoo (2022); Anwar et al. (2024)
Engagement Bias	Design choices (e.g., autoplay) that increase passive consumption	Anwar et al. (2024); Schaffner et al. (2023)
Promotional Bias	Limited visibility of third-party or niche content	Bourreau & Gaudin (2021); Van Es (2024)
Search Bias	Search results are structured to favor trending or sponsored content	Lamkhede & Das (2019); Lamkhede & Kofler (2021); Olma et al. (2024)

Table 1: Overview of the algorithmic biases based on related work

2.3.2 User-Centered Biases in Content Consumption

Now shifting the focus from the systemic biases to users and their potential biases when interacting with recommendation systems led to interesting articles, where one of these was found in the following search stream:

> "choice overload" AND "recommender system" AND Netflix
> → 258 results
>> → "choice overload" AND "recommender system" AND Netflix AND qualitative
>> → 110 results
>>> "streaming platforms" AND "user behavior" AND "choice overload" AND qualitative
>>> → 47 results

From the final search, the article "User's Dilemma: A Qualitative Study on the Influence of Netflix Recommender Systems on Choice Overload" by Romero Meza and D'Urso (2024) was identified as highly relevant.

Beyond algorithmic biases, user behavior also contributes to reinforcing content recommendation loops. Cognitive tendencies shape how users interact with Netflix's system, further exacerbating certain biases (Ahn & Lin, 2024; Mansoury et al., 2023). One of the most prevalent cognitive biases is confirmation bias, where users tend to prefer content that aligns with their past choices, thereby reinforcing filter bubbles. Research by Ahn & Lin indicates that Netflix's algorithm capitalizes on this tendency, frequently recommending similar genres and themes, which reduces opportunities for diverse content exploration (2024, p. 17).

Another key issue is choice overload bias, which occurs when users are presented with an overwhelming number of content options. This leads them to rely on Netflix's recommendations instead of actively searching for content. Romero Meza and D'Urso found that excessive content options can lead to decision fatigue, ultimately prompting users to opt for algorithm-driven selections rather than manual exploration (2024, p. 364).

Iordache et al. (2024) highlight that users' interaction with the interface, including thumbnails, visual hierarchies, and autoplay, creates a cognitive map of what users assume as "recommended" or "expected" (pp. 586–587). Furthermore, their study highlights how users rarely break patterns, indicating that personal preferences are strengthened rather than challenged (Iordache et al., 2024, p. 598). While mentioning the thumbnails, Sun et al. documented a bias in users' behavior, where they tended to choose the first thing they saw on the platform, which reinforces the algorithmic first choices (2023, p. 6). This is supported by Schaffner et al., who conducted an experimental study demonstrating that turning off the autoplay feature resulted in a reduction in daily watch time and shorter session lengths (2025, p. 13). Furthermore, they also found that users often did not realize the influence of autoplay until it was removed, pointing to the invisible design choice that can shape behavior (Schaffner et al., 2025, p. 15). Habit and perceived usefulness can significantly impact binge-watching behavior. Bastos et al.'s study reveals that even when users experience regret, it does not always entail reducing the behavior, especially when their habits have already been established (2024, p. 7). A related tendency is the recency bias, where users tend to engage with newly released or recommended content. This can reinforce shortsighted interaction patterns and limit the possibility of exploring more diverse content (Sun et al., 2023; Bastos et al., 2024). User-centered biases, therefore, seem to be part of a larger cycle of behavior reinforced by interface design, affective experience, and perceived utility (Bastos et al., 2024, pp. 6–7).

The literature suggests that users' cognitive tendencies and trust in the system actively contribute to reinforcing the influence of the recommendation system. Confirmation bias, recency bias, choice overload, and trusting the algorithm can lead users to accept the first recommendation instead of exploring the catalogue. These behavioral patterns reinforce habits and interface structures such as autoplay and visual hierarchy. Table 2 presents a clear overview of the discussed user-centered biases and articles that address them. The table will serve as a foundation for the analysis in Chapter 5, where biases are examined regarding both the data and the Netflix UI.

Bias Type	**Definition / Effect**	**Discussed In**
Confirmation Bias	Users prefer content that matches existing preferences, reinforcing filter bubbles and limiting diversity.	Ahn & Lin (2023)

Choice Overload Bias	Too many options lead to decision fatigue and defaulting to algorithm-suggested content.	Romero Meza & D'Urso (2024)
Trust in Algorithm Bias	Users assume the algorithm knows best, leading to passive acceptance of recommendations.	Mansoury et al. (2023)
Recency Bias	Tendency to choose newly released or recently recommended content, limiting exploration of older or diverse titles.	Sun et al. (2023); Bastos et al. (2024)

Table 2: Overview of the algorithmic biases based on related work

2.5 Interim Conclusion

Based on the related work, it appears that Netflix's recommendation system may be influenced by multiple biases, which can impact the user experience of exploring content. The biases are divided into two categories: algorithmic and user-centered biases. Algorithmic biases concern popularity, promotional, positioning, engagement, and search bias. They can limit users' exposure to different content by favoring popular, new, or platform-owned content, and by interface designs that encourage passive use of the platform. The user-centered biases included confirmation bias, choice overload, recency bias, and trust in algorithms. These tendencies typically mean that the user accepts the algorithm's initial recommendations without further exploration. The interface and habits reinforce these behavioral patterns.

The literature presents a complex interplay between the system's overall design and the user's own choices, which provides the foundation for this study's empirical research.

3. Methodology & PoS

The following section presents the methodology employed in this thesis project and outlines the research approach. The overall methodology will consist of interviews, a diary study, and thematic analysis to both collect and analyze the data.

3.1 The Pragmatic Approach

This thesis is anchored in the pragmatic research paradigm, where epistemology is created through experience, action, and reflection (Kelly & Cordeiro, 2020, p. 3; Morgan, 2014, p. 1047). Pragmatism rejects traditional dualisms, such as objectivity versus subjectivity, and emphasizes the usability and practical consequences of knowledge (Morgan, 2014, p. 1048). Drawing from John Dewey's pragmatic epistemology, research is not a passive pursuit of truth, but rather an active and ongoing process where experiences are evaluated and understanding is adjusted. Through continuous interaction and reflection, Dewey rejects the notion of constant truths (Dewey, 1938, pp. 3-5). For this study, it means that the goal is not to reach a final understanding of Netflix's recommendation system, but rather to generate knowledge on experiences, patterns, and pain points that can lead to valuable design implications. Therefore, the research design and methods employed are crucial for understanding complex issues, where technology, behavior, and experiences are interconnected and often constitute a "wicked problem" (Kaushik & Walsh, 2019, p. 6).

Moreover, Patton highlights how the methodology employed depends on the context and problem statements, as pragmatism allows for methodical flexibility rather than being bound by a specific philosophical tradition (Patton, 2015, pp. 154-157). Here, Patton refers to *generic qualitative inquiry,* where qualitative methods are employed without being explicitly grounded in a specific theoretical framework (Patton, 2015, p. 247). This is a pragmatic principle, focusing on practice-oriented questions, methods based on context, and analysis focused on action, which supports this thesis mixed-method perspective (Patton, 2015, p. 249).

3.1.1 Mixed-Method Considerations

This thesis is methodologically grounded in qualitative research, but incorporates quantitative elements that support an in-depth, interpretive analysis, as described by Creswell as a *qualitatively driven mixed-methods* study (Creswell, 2014, p. 101). Combining semi-structured interviews with a diary study enables insights into participants' reflections and patterns over a short period. Bryman highlights how such a combination can strengthen both the validity and reliability of the results (2012, p. 633). Thus, it aligns well in a pragmatic context where experience, practice, and theory are integrated.

Creswell and Plano Clark emphasize further that mixed-methods approaches are not only about utilizing different methods, but also using them in the best way possible, relating to the research questions and context (2011, pp. 69-70). This approach, with pragmatic anchoring, makes it possible for nuanced research into complex user experiences, where the context and its applicability drive methodological choices. Lastly, a concise and comprehensive research design mapping the overall structure of the study will be introduced in the next section.

3.2 Research Design & Approach

A research design is important to ensure a thought-through study. According to Flick, the research design for a qualitatively driven mixed-methods approach should be determined, flexible, and iterative, allowing data collection and analysis to inform each other. This means that there is room for adjustments following the process, allowing new insights to emerge (Flick, 2009, p. 92).

3.2.1 Research Strategy

To understand the complex nature of users' experiences in the meeting with Netflix's user interface, the following research design is mapped to have a clear overview of the following parts that entail gathering insights and knowledge to answer the research questions:

- Bias mapping & UI analysis: to understand which of the biases appear where and how Netflix reinforces these on a surface level.
- Semi-structured interviews; both pre- and post-diary study to ensure in-depth insights.
- Diary study: structured and contextual data collection method to see participants' day-to-day interaction with Netflix's recommendation system.
- Descriptive & Contextual analysis: to explain and understand the specific patterns gathered from the diary study.
- Thematic analysis: to analyze and interpret the qualitative material, focusing on the participants' statements.

The remaining sections for this chapter will focus on presenting all the methodology utilized in this study.

3.3 Purposive Sampling

Bryman describes purposive or purposeful sampling as a non-probability sampling method, where the researcher purposefully selects participants based on specific characteristics or criteria relevant to the research field and purpose (2012, p. 418). This method of sampling is suitable for qualitative research, where the focus is on achieving an in-depth understanding of phenomena, rather than generalizing results to a majority (Bryman, 2012, p. 418). Furthermore, the sampling method is used to select participants who are most likely to yield relevant and valuable information (Campbell et al., 2020, p. 653), making it essential in context-specific studies.

Stratton explains that the sampling method can be divided into two types: *random* and *subjective sampling* (2024, p. 121). Random sampling is used when participants are selected randomly from all potential participants, whereas subjective sampling does not use randomization in selecting participants. This study employs subjective purposive sampling to gather divergent and competing information from a specific group of participants. Considering the aim of this study is to explore personal experiences with recommendation systems, it is crucial to recruit participants who regularly engage with Netflix. The individuals selected for this study are based on their self-reported use of patterns during the pre-study interviews, rather than demographic variables.

In choosing such a sampling method, it is essential to acknowledge that the researcher selects participants, thereby creating a potential for bias in determining who is deemed relevant or available. Transparency surrounding these reflections is important and is intended to strengthen the study's ethical credibility.

3.4 Bias Mapping & UI Analysis

Bias mapping as a methodological approach will examine how Netflix's user interface (UI) reinforces biases in its recommendation system. Bias mapping in this context refers to the structured analysis of how UI elements, content positioning, and search feature contribute to systemic biases in content discovery.

Bias mapping is conducted through a systematic overview of Netflix's interface, aligning observed UI patterns with bias categories identified in the literature review (Chapter 2). This method adheres to the principles of systematic qualitative content analysis, as described by Flick (2009, p. 318) as a structured approach to identifying recurring patterns within a dataset.

To examine how biases are embedded in Netflix's UI, this study follows a three-step approach:

1. Systematic UI Review – Netflix's homepage, search bar, and recommendation rows were reviewed to identify patterns that reflect bias categories.
2. Bias Classification – Each UI element was classified based on biases documented in academic literature, including popularity bias, positioning bias, and search bias.
3. Integration with Systematic Overview – The findings were structured into a bias table, linking Netflix's interface elements to specific bias mechanisms described in existing research.

Observations will be documented through screenshots, categorized based on their associated bias type, and analyzed qualitatively to determine their alignment with literature-based bias types.

3.5 Semi-Structured Interview

As mentioned, this study conducts semi-structured interviews before and after the diary study period. Bryman places this method between the structured and unstructured interviews, which combines predetermined questions with the ability to allow for spontaneous elaboration during the conversation (2012, p. 471). DiCicco-Bloom and Crabtree emphasize Bryman's description as having open-ended questions, but not being led only by them (2006, p. 315). This means that having an interview guide will support the conversation. However, it should not serve as a checklist of questions to follow, and it is essential to respond to interviewees' statements and reflections to gather deeper insights (Gill et al., 2008; Ruslin et al., 2022, p. 24). As the method allows for a balanced combination of consistency and flexibility, it is also important to avoid leading questions and ensure that participants' perspectives stand out, rather than the researcher's assumptions being confirmed (Bryman, 2012, p. 474). Flick, therefore, emphasizes allowing participants to speak freely, providing room for reflection, and refraining from interrupting to lead the conversation, while striving to remain as neutral as possible during the conversations (Flick, 2009, p. 174).

The pre-interviews will uncover participants' reflections on their patterns and behavior regarding Netflix's recommendation system. In contrast, the post-interviews will focus on the same areas but investigate changes in behavior, experiences, or new insights they might have encountered during the diary study. These semi-structured interviews can therefore capture both pre-understanding and follow-up reflections that contribute to valuable insights into user behavior about meeting algorithmic recommendations.

3.6 Diary Study Method

The diary study is a generally qualitative method that collects data on users' experiences, behavior, and decision-making processes in their authentic situations over a specific period. Bartlett and Milligan present that the method is capable of "capturing life as it is lived" and allows access to both natural and non-disturbing experiences (2015, p.8). The overall idea of the process is to have participants log their experiences and reflections close to they happen, which reduces short-term memory bias and increases the quality of the data (Bolger et al., 2003, p. 581, Chun, 2016, p. 407–408; Consolvo et al., 2017, p. 72). Moreover, it also minimizes observer effects, which means that participants can reflect without other influences in their contexts (Carter & Mankoff, 2005, p. 899)

3.6.1 Choice, Implementation & Relevance

As the method provides detailed and authentic insights into users' interactions with Netflix's recommendation system, it is chosen as one of the methods for this thesis. The researcher limited their influence in the data collection, ensuring a stronger dataset and reducing the risk of bias (Ross et al., 1994, pp. 422–423; Lobato et al., 2024, pp. 1333–1334).

For this thesis, participants will log their daily experiences on Netflix over a 10-day period, inspired by studies that suggest this interval is sufficient to gather insights on both habits and variations in viewing behavior (Janssens et al., 2018). As there is no specific guideline to how a diary study is to be designed or conducted, other than some research suggesting digital diary methods as opposed to paper-based alternatives, as digital tools can increase the precision of data collection and lower the participant burden (Janssens et al., 2018, s. 9). The diary study will be using Microsoft Forms as the data collection tool, where participants will fill out one entry per viewed content, securing detailed documentation of each specific decision-making process, use of recommendation and its experienced relevance. The survey format will combine closed-ended answers (e.g., checkboxes and Likert scales) and open-ended answers to balance both the qualitative and quantitative aspects of this study. Although the method appears to have numerous benefits, potential challenges must also be acknowledged. Some studies suggest that using this method carries a risk of *entry fatigue*, where logging daily entries can reduce participants' motivation and, consequently, impact data quality (Olorunfemi, 2023, p. 420). Therefore, it is essential to ensure that the design of the diary forms is user-friendly (Bartlett & Milligan, 2015, pp. 93–94). This will be ensured through a short and manageable diary study.

Lastly, the data collected via Microsoft Forms generally consists of categorical variables such as content type, discovery methods, and satisfaction levels. These data will be analyzed and visualized through Tableau, as it allows the creation of interactive graphs and diagrams, making it an excellent tool for this study and its data properties.

3.7 Thematic Analysis

Thematic analysis (TA) is a flexible and well-established qualitative method used to identify, analyze, and highlight patterns in qualitative data (Braun & Clarke, 2006, p. 77). According to Flick, TA contributes to structuring data in meaningful themes that can strengthen interpretations (2009, p. 318). For this thesis, the method is a central analytical approach to uncover patterns and reflections from the participants based on the semi-structured interviews.

With a pragmatic perspective, where the aim is to gain knowledge about users' experiences, the analysis is predominantly deductive. It is therefore based on existing theories on algorithmic bias, user behavior, and interface design. Furthermore, it is inspired by the codebook-based thematic analysis, that combines structured coding with room for reflection and theoretical anchoring (Wolgemuth et al., 2025, s. 27). Specifically, after the interviews have been transcribed, a thorough read through highlighting key statements from the participants that are relevant to the study will be the beginning steps toward coding the interviews for the thematic analysis. Nowell et al. (2017, p. 2) explain the importance of transparency in the coding process details to ensure transparency.

This coding process is further supported by Naeem et al. (2023), who propose a structured six-step model for thematic analysis that guides researchers from initial coding to conceptual integration. Their model enhances analytical reliability and provides a valuable framework for maintaining coherence between data, theory, and emerging insights (Naeem et al., 2023, pp. 4–5) and using direct quotes from the participants, as it allows to emphasize the importance of each theme further (Nowell et al., 2017, p. 4). Therefore, in the respective analysis sections of the TA for both semi-structured interviews, I will explain the specific coding process, providing examples of quotes and subcodes that fit under each deductively driven theme.

Furthermore, Braun and Clarke highlight that doing thematic analysis means the researcher is actively creating the themes, and subjectivity is not seen as a mistake, but rather as terms that should be reflected and integrated within the analysis (2020, p. 3). As the analysis is primarily deductive, it also allows for the potential inclusion of inductive elements, especially given the post-interviews, where new themes may emerge. This combination of deductive and inductive logic reflects the thesis's pragmatic orientation and ensures openness to unexpected, yet relevant, insights within the data.

3.8 Ethical Considerations, Validity & Limitations

Before concluding this methodological framework, it is essential to highlight the importance of ethical considerations, validity, and limitations in the context of this study. When dealing with personal user data, ethical rigor is essential in the field of qualitative research. Flick stresses the importance of obtaining participants' consent, data anonymization, and participant confidentiality (2009, p. 40). This study adheres to these essential guidelines by anonymizing the participants in this thesis, obtaining their consent to record the interview sessions for transcription purposes, and ensuring the security of their data by keeping it within the confines of this thesis project.

Furthermore, when it comes to qualitative research, Flick argues that it is often evaluated through the lens of transparency, triangulation, and reflexivity (2009, p. 387). By triangulating diary studies, interviews, and UI analysis, I enhance the validity of this study. Flick (2004) highlights that triangulation is not only about validating data, but also about creating a more nuanced picture of the specific phenomenon being studied by comparing and utilizing different perspectives and methods (p. 179). However, some issues might arise in conducting any research. It is worth mentioning that this study has its limitations, such as participant self-reporting bias, which means that by having participants fill out the diary study and even being interviewed, there is a chance that they could present themselves in a positive light, as they think the researcher might want to hear. The time limit of this study also affects the data collection period, where a more extended diary study period could have been more beneficial in trying to understand the participants.

Lastly, it is essential to note that this study does not aim to understand the technological implications of algorithms on streaming platforms. Consequently, the knowledge emerging from this study is based on the user interface and user experiences, rather than on how the specific back-end system functions. This section aims to demonstrate that this study methodologically focuses on transparency, ethical responsibility, and triangulation. However, it is essential to recognize that all qualitative research is situated or limited by its context.

3.9 Interim Conclusion

Based on the presentations in this chapter, the methodology is employed to achieve a nuanced and context-based understanding of users' experiences with Netflix's recommendation system. By combining methods such as interviews, bias mapping, and diary studies, it creates a strong foundation for examining behavioral patterns and reflexive experiences. The pragmatic anchored approach enables the flexible application of methods and analysis, where both deductive and inductive elements are combined.

Not only will this methodological structure serve to answer the research questions, but it will also lay the foundation for understanding how algorithmic systems can shape user experiences. To analyze these aspects, it is essential to have a theoretical framework that sheds light on terms such as algorithmic bias, user agency, and interface design. The following chapter presents the theoretical framework, which informs the analysis and helps contextualize the empirical findings.

4. Theoretical Framework

To analyze and interpret the collected data, establishing a theoretical framework is fundamental. Theory serves as an analytical tool that enables the understanding and contextualization of complex phenomena. According to Shneiderman and Plaisant, theory contributes to creating transferable knowledge and identifying patterns (2005, p. 85).

The following chapter presents the central theoretical framing, which is used to examine Netflix's recommendation system and its influence on users' experiences. The theories encompass four main areas: algorithmic bias and personalization, filter bubbles, user agency and interface design, and choice overload and decision-making biases. These perspectives form a collective framework for interpreting the data collected during the analysis further.

4.1 Algorithmic Bias & Personalization

There exist multiple definitions of what an algorithm is. In specific public settings, the term is used broadly, sometimes applied to anything from dating applications to recommendation systems; however, two main types are referred to when it is related to the research field: mathematical structures and the specific implementation in software or systems (Mittelstadt et al., 2016, p. 3). When discussing bias or ethics, it is essential to understand algorithms as configured systems, which means they are shaped by the data they process and the social environments in which they operate (Mittelstadt et al., 2016, p. 3). When discussing algorithms and ethics, it is not just about the code, but also about how algorithms function in real life, including their design, decision-making processes, and the impact on users' choices and possibilities (Mittelstadt et al., 2016, p. 3). This context-based understanding is essential in this study as it guides in understanding how individual recommendations not only create comfort, but also bias.

Mittelstadt et al. (2016, p. 1) explain that recommender algorithms and AI decision systems often reflect human biases embedded in the interface or the training data. Furthermore, algorithms are described as "inescapably value-laden," which means the choices and input data can privilege specific values or interests over others (Mittelstadt et al., 2016, p. 1). An example of this is if an algorithm is trained on data to favor specific genres of movies, it may privilege that type of content over others. Algorithmic systems are shaped not only by training data but also by human decisions in their design and deployment (Mittelstadt et al., 2016, p. 3). Moreover, Mittelstadt et al. (2016) mapped the ethics of algorithms and identified bias as a central concern, thereby urging the continuation of research on how personalization may privilege specific information and impact individuals, environments, and society (pp. 14–15).

Therefore, it is essential to recognize that recommender systems are not neutral, as they can incorporate biases from both design and data, while personalization can both benefit and harm specific stakeholders.

4.2 Filter Bubbles

As mentioned in the previous section, personalization is essential to understand that although algorithms can tailor content to individual preferences, they can also limit users' exposure to diverse content through *filter bubbles* (Figà-Talamanca & Arfini, 2022, p. 1). It points to users being isolated in information environments that largely reinforce existing beliefs. This section will also include Alex Bruns' work on the phenomena, and it is worth mentioning that his work focuses on social and informational platforms. However, this study extends the theoretical lens to SVOD platforms, such as Netflix.

Eli Pariser introduced the term filter bubble, which was used as a metaphor for how algorithms risk creating "ideologically safe" information environments, where individuals are rarely confronted with opposing perspectives (Figà-Talamanca & Arfini, 2022, p. 1). However, more attention is being paid to the fact that such phenomena are not necessarily as ever-present or deterministic as the early concepts suggested. Bruns (2019) highlights that many studies have overstated the influence of algorithms and the active role of users in content curation. Bruns emphasizes that users' selective behavior and preferences largely shape the information environments they engage with, meaning that it is not merely algorithms that do so, but also users' habits that contribute to informational homogenization (Bruns, 2019, pp. 22-23). Bruns argues that, although media and algorithmic platforms can contribute to ideological homogenization, data suggest that many users are exposed to polarizing perspectives (Bruns, 2019, pp. 31-34). This is supported by Flaxman et al. (2016), who showcase that users are continuously confronted with diverse content, emphasizing that algorithmic platforms are part of the issue, but not solely (pp. 20-21).

This does not mean that the filter bubble effect is irrelevant; however, it does suggest that the phenomenon is more complex than initially thought. It should be understood in the context of an interplay between systemic design and users' preferences. Figà-Talamanca and Arfini (2022) highlight that it is an exaggeration to blame only systems and their algorithms for the existence of filter bubbles. They warn against technological determinism and point to humane cognitive tendencies, such as the wish for approval and comfort, that affect how users interact with recommendations (Figà-Talamanca & Arfini, 2022, pp. 10-12).

Moreover, Bruns also highlights the need to shift the focus from debating whether filter bubbles exist to examining how personalization in practice affects users' ability to engage in meaningful navigation and exploration within digital systems (2019, pp. 8-9). This is highly relevant to this study, given its focus on user interfaces and experience bias.

Lastly, Bruns argues that even in cases of ideological polarization online, it does not necessarily mean that users are isolated from alternative viewpoints. Instead, he showcases that users connect across digital spaces, even when their content consumption diverges. Polarization and diversity coexist, which means that the social structure of online interaction is far more complex than the notion of closed information loops suggests (Bruns, 2019, p. 181). Instead of assuming users solely live in isolated bubbles, the analysis should focus on how recommendation systems can promote or hinder active exploration and content diversity in practice.

4.3 User Agency and Interface Design in Algorithmic Systems

When it comes to recommendation systems, especially platforms like Netflix, it is essential to consider both user agency and interface design, as well as their impact on users' experiences and opportunities to explore diverse content. In comparison to traditional recommender systems, newer Human-Computer Interaction (HCI) and user-centered design that the user be a participant actor, not only a passive receiver (Konstan & Terveen, 2021, s. 32). By doing so, one can emphasize on giving the actual users of the interfaces more control, transparency and opportunity for feedback within the recommendation process. When referring to user agency, it describes the ability to affect and control decisions that algorithms make, which contrasts with earlier approaches where the accuracy of systems was prioritized over users' experiences and inclusion (Konstan & Terveen, 2021, p. 35). Moving on, I will present two theoretical frameworks by Don Norman (2013) and Shneiderman & Plaisant (2005), which offer a strong foundation for understanding how interface design can either support or hinder users' sense of agency.

According to Don Norman, Human-Centered Design (HCD) is not just a method, but a design paradigm that prioritizes users' needs, competencies, and behavior, and then designs around these (Norman, 2013, p. 210). This approach requires designers to understand not only what users want to achieve but also how they interpret, act, and feel within digital environments. Giacomin supplements this perspective with an important insight, where HCD is not merely seen as a set of techniques, but as a design mindset and a value-driven philosophy. He defines HCD as a practice that communicates, interacts with, and understands the individuals involved, and from there, develops solutions that are both cognitive and emotionally intuitive (Giacomin, 2014, p. 2). The user interface's ability to support users' understanding and possible action is not only a question of functionality, but also about trust, decision-making, and design ethics.

Norman also highlights the importance of both *feedback* and *feedforward*, meaning that the system should communicate not only what happens but also what is possible to do, which strengthens users' feelings of agency and control (2013, pp. 41, 83). In the context of a recommendation interface like Netflix, this becomes especially relevant, as users' sense of control is shaped by how the interface communicates the reasons behind its suggestions and the available options for response or customization.

Norman's concepts of *affordances*, *signifiers,* and *constraints* help analyze how users interact with a digital interface such as Netflix. Affordances are used to describe the perceived and actual properties of an object that determine its use, such as a button that affords interaction, like clicking (Norman, 2013, pp. 27-28). Concerning affordances, the cues that signal to the user where an action is possible, like arrows or highlighted sections within an interface, are signifiers (Norman, 2013, pp. 29-30). Norman then describes constraints as limiting a user's interaction with the system or platform. It can limit user agency by limiting their freedom of interaction. This, however, does not necessarily mean that constraints are only negative, as they can also guide users toward specific actions and away from others due to support or safety (Norman, 2013, p. 43). Together, these design principles create a framework for evaluating how interface elements afford or inhibit user agency. An interface such as Netflix relies heavily on affordances (e.g., play buttons, scroll rows), signifiers (e.g., "Top 10" labels, genre tags), and constraints (e.g., limited visibility of niche content, autoplay features). These visual and interactive elements are not neutral—they shape users' paths through the system, guiding attention and decision-making in ways that may support or suppress exploration and control.

Norman's framework can, therefore, support the analysis in understanding whether these elements either support or hinder user agency, not just in terms of ease of use, but also regarding how empowered users feel in navigating, resisting, or customizing the algorithmic logic presented to them. This analysis will help critically assess how design affects the balance between passive consumption and active engagement.

Sometimes, when trying to translate research results into design elements, having a guideline to follow can be helpful. Ben Shneiderman and Catherine Plaisant (2005) present a set of eight golden rules to follow when designing new interfaces (pp. 74-75). Although all Shneiderman and Plaisant's eight rules are valuable, this thesis will primarily focus on the seventh rule: supporting an *internal locus of control* (2005, p. 74). This rule is relevant to this study's aim of understanding how users experience bias, which is also related to users' feelings of autonomy. An internal locus of control refers to users' sense that they have control over their interaction with a system. An alternative understanding of this is that it is the user navigating the platform, not the platform navigating them (Shneiderman & Plaisant, 2005, p. 74). Furthermore, by focusing on this rule, the thesis highlights the importance of user agency, enabling users to make conscious and informed choices.

Another perspective on the relationship between users' control and the interface's design can be argued to differentiate between system-initiated personalization (SIP) and user-initiated customization (UIC) (Sundar & Marathe, 2010). Sundar and Marathe's study shows how "power users" prefer user-defined interfaces, as they strengthen control and agency, while less experienced users wish for an automated personalization, as it is more comfortable and requires less cognitive energy (2010, pp. 310-311).
Their theoretical "agency model of customization" emphasizes that when users have the opportunity to act as information gatekeepers, it increases their engagement, satisfaction, and feeling of ownership - which is why they argue for recommendation systems that support users' active participation (Sundar & Marathe, 2010, pp. 303-305).

Furthermore, Knijnenburg et al. (2012) present a user-centered evaluation framework for recommendation systems, focusing not only on algorithmic accuracy but also on how users subjectively experience the quality, variation, and transparency of recommendations (pp. 443-444). Users' experience of usefulness, usability, and control is often a stronger indicator of satisfaction than the actual performance of the algorithm. This approach supports a broader understanding of UX in recommender systems, where the interface's ability to provide an overview, affordances for action, and meaning shapes the overall experience — a particularly relevant focus in the context of Netflix's visual recommendation structure (Knijnenburg et al., 2012, p. 479).

Konstan and Terveen also support the focus on human-centered recommendation systems, arguing for the need for feedback mechanisms, the ability to tailor preferences, and transparency in how such platforms generate their recommendations (2021, p. 38). Furthermore, studies conducted by Jannach et al. (2017) also demonstrate that such control and autonomy given to users can enhance users' trust and sense of agency.

HCD principles are important in analyzing Netflix's UI and including them allows this study to evaluate whether the interface empowers users or constrains them through specific design choices. As recommendation systems become embedded in everyday media use, the ability for users to influence, understand, and navigate these digital platforms is both a design concern and an ethical imperative. Therefore, HCD functions as both an analytical and normative lens in this study, supporting the idea that elements such as transparency, feedback, and perceived control shape user agency.

4.4 Choice Overload & Decision-Making Biases

When giving users more control, it is essential not to overwhelm them with an excessive number of choices. Theoretical works on decision-making warn that too many choices can lead to *choice overload*, a situation where the user feels overwhelmed and has difficulty making a decision. Barry Schwartz's *Paradox of Choice* (2004) posits that at a certain point, having more choices can reduce satisfaction and even increase an individual's anxiety (pp. 3-4). Moreover, Schwartz demonstrates how individuals, especially those who strive for a "perfect choice," become even more dissatisfied when confronted with too many options, as they tend to compare and doubt whether their decision was the right one (2004, p. 85). He even goes on to claim that as the number of choices increases, the freedom of choice becomes the tyranny of choice (Schwartz, 2004, p. 235).

The phenomenon of having too many choices has been studied in the context of search engines and digital interfaces. Oulasvirta et al. exemplify how users who were given six search results had a higher feeling of satisfaction and confidence with their choices than those participants who had to choose between 24 choices (2009, p. 521). A mass of choices does not necessarily mean a better experience when it comes to digital design. Furthermore, Schwartz also references Herbert Simon's argument that a wave of information or choices can also hinder attention (2004, p. 79). This can be a genuine issue within streaming services, where personal feeds compete with the user's cognitive resources and capacity.

Scheibehenne et al. conducted a meta-analytical review of choice overload and found that the phenomenon is not universal. Instead, they suggest that it very much depends on contextual factors such as the complexity of the task, the individual's engagement, and how choices are presented (Scheibehenne et al., 2010, pp. 410–411).

This means that it is essential to note that too many choices do not always lead to lower satisfaction; instead, it can occur under specific contexts, which will be interesting to examine among the participants of this study.

4.5 Interim Conclusion

Altogether, this chapter presents a multifaceted theoretical framework that serves as the foundation for analyzing Netflix's recommendation system and its influence on users' experiences of variation, control, and navigation. The four perspectives – algorithmic bias, filter bubbles, user agency, and choice overload – address both algorithmic, behavioral, and design-based factors. Through Mittelstadt et al. (2016), algorithmic recommendations are value-laden and non-neutral. Furthermore, Bruns (2019) and Figà-Talamanca and Arfini (2022) highlight the interplay between algorithmic structure and users' preferences, which is especially relevant for SVOD platforms like Netflix.

Moreover, the theories also encompass user agency and interface design, as explored by Norman (2013), Shneiderman and Plaisant (2005), and Sundar and Marathe (2010), regarding how user interfaces support or limit users' feelings of control, understanding, and personalization. Here, terms such as feedback/feedforward, affordances, and internal locus of control become central analysis tools. Finally, Schwartz (2004) and Scheibehenne et al. (2010) provide insight into how complex decision-making and user interface can lead to meaningful variation, choice overload, or decision fatigue.

This theoretical framework provides a strong foundation for analyzing how Netflix's design choices influence both the content users encounter and the way it makes decisions. Therefore, this framework enables a structural and critical examination of the balance between comfort and control, automation and agency, and how users experience these factors in practice.

5. Analysis

This chapter analyzes the empirical material and studies how algorithmic and user-related biases shape the user experience with Netflix's recommendation system. The analysis is structured in five parts:

Firstly, Netflix's UI and its visual structures are analyzed through bias mapping, where algorithmic and design-based biases are visualized. Before the diary study, a thematic analysis of the pre-interviews was conducted to gain an understanding of the participants' habits, experiences, and opinions. The third section presents an analysis of the diary study data, providing insight into the viewing patterns of the participants. Then, a thematic post-interview analysis will identify behavioral changes, overall reflections, and new perspectives on users' interaction with the system. Lastly, all these findings will be synthesized in a comparative section.

This chapter is grounded in the theoretical framework presented in Chapter 4 and aims to address the research questions through a descriptive, thematic, and interpretive approach.

5.1 Analysis of Bias Mapping & Netflix's UI

At first glance, Netflix's user interface appears quite intuitive, effective, and personalized; however, certain aspects of its UI raise concerns about its design and the logic behind how recommendations are generated. These logics are not neutral. In the following section, the systemic biases identified in the literature review will be used to analyze how Netflix's UI amplifies both systemic and user-related biases. This analysis builds upon the theories presented in Chapter 4, focusing on algorithmic bias, user agency, and interface design. A systematic overview of the identified biases has been provided and will be referenced throughout this analysis. All screenshots provided in Appendix 2 were taken using a Safari browser on a MacBook.

5.1.1 Algorithmic Biases & Interface

Netflix's user interface is not just a visual layer but an algorithmic structure with carefully selected design elements that align with the platform's business intentions. As Mittelstadt et al. (2016) explain, algorithmic systems are far from neutral. This becomes evident with positioning bias, where the top placements on the homepage are usually dominated by popular or promoted content (see Appendix 2.3).

The screenshot that exemplifies positioning bias (Appendix 2.3, p. 6-7) illustrates the top section of the homepage, where the series "The Leopard" occupies the most visually dominant area. As users open the app, their attention is captured. Prioritizing 'featured' or 'original' content is aesthetic and strategic – the user is guided toward Netflix's commercial goals. Furthermore, UI elements such as the autoplay feature and minimal skip options reinforce nudging by reducing friction between exposure and engagement. Therefore, the interface curates what is most visible and what Netflix wants the user to watch, reducing user agency while sustaining an illusion of free choice. Through positioning bias, Netflix transforms its interface into a commercial amplifier, where placement becomes a form of power.

Another element is the autoplay feature, which activates whenever the user finishes an episode or remains idle (Appendix 2.3, p. 3-4). This reinforces engagement bias, where the system prioritizes continued viewership over deliberate choice. This may challenge the user's internal locus of control, a concept discussed by Shneiderman and Plaisant (2005), who emphasize the importance of trust and autonomy in UI design. When autoplay initiates content without clear user intent, it can reduce perceived control, shifting the decision-making power to the system.

Furthermore, promotional bias is also evident. Although Netflix Originals are no longer confined to a dedicated row, they are still widely embedded across multiple genre categories and rows (Appendix 2.3, p. 4). When titles are highlighted in this way, it can challenge transparency and the feeling of control. According to Norman, Shneiderman, and Plaisant, feedback and an internal locus of control are crucial for users' experience of agency (2013, p. 43; 2005, p. 74). This design choice can risk contributing to passive acceptance instead of active choices.

Finally, the search bias is notable (Appendix 2.3, pp. 4-5). The search bar uses autocomplete and trending suggestions that favor popular titles over relevance. While this does not definitively prove that Netflix forces users in a specific direction, it is reasonable to argue that the autocomplete is optimized for engagement metrics rather than purely user intent. These UI signifiers subtly guide the user's decision-making.

5.1.2 User-Related Biases & Cognition

However, not only does Netflix's UI amplify systemic biases; it also triggers and reinforces user-related or user-driven biases such as confirmation bias, choice overload bias, and recency bias (see Appendix 2.2, p. 2-3). According to Schwartz (2004), users' satisfaction drops when faced with too many decisions. This is reflected in Netflix's interface, which presents users with many sections, rows, and genres (Appendix 2.3). Although this may not overwhelm all users, the "endless scroll" and continuous flow of recommendations may lead to choice overload bias, where users default to what is immediately visible or familiar. Oulasvirta et al. (2023) demonstrate that fewer structured choices increase satisfaction, supporting the idea that effective design must strike a balance between variety and clarity.

Confirmation bias is also relevant, as users tend to repeat past behaviors by selecting familiar genres, actors, or formats. This, in turn, reinforces the algorithm's predictive logic, leading to repetitive suggestions. As Pariser's filter bubble theory suggests, such cycles limit the discovery of new information. However, as Figà-Talamanca and Arfini (2022) argue, this repetition is also user-driven, reflecting personal habits as much as algorithmic reinforcement.

Another significant user-driven bias is trust in the algorithm. Many users accept what is presented first or what autoplay selects as the "right" or "best" option. While this bias is difficult to verify empirically, it remains essential for ethical and design discussions. Norman and scholars like Konstan & Terveen (2022) emphasize the importance of agency and meaningful choice for trustworthy recommendation systems.

5.1.3 Critical Reflection

The bias mapping reveals clear patterns in how Netflix's interface is structured and how it interacts with users. However, bias is complex – it arises at the intersection of user behavior, interface design, and algorithmic logic. It is not enough to identify the presence of bias. This analysis shows that technical design (e.g., autoplay, layout, visual hierarchy) and user habits actively co-create the digital experience.

As Sunar and Marathe (2010) show, users' perception of control and agency over a system significantly affects their satisfaction and trust. The UI promotes a more passive form of personalization that appeals to users seeking convenience. This emphasizes the need to rethink the role of users in recommendation systems. Users are not passive – they have agency – but they need design features that support feedback, transparency, and informed exploration.

The analysis of Netflix's interface and bias mapping shows how the platform amplifies algorithmic and user-driven biases. These findings suggest a need to rethink design logic, not only to optimize engagement but also to promote content diversity, transparency, and user autonomy.

5.2 Thematic Analysis of Pre-Interviews

This section presents the results of the thematic analysis of the pre-interviews conducted prior to the diary study. As previously mentioned, the analysis employs a deductive coding approach, drawing inspiration from codebook-based thematic analysis (Wolgemuth et al., 2025). Accordingly, the main themes and subcodes were predefined based on the theoretical framework, including concepts such as algorithmic bias, user agency, and interface design.

Following transcription of the pre-interviews, relevant quotes were organized and coded in Miro. Using color-coded sticky notes, each quote was linked to a subcode that aligned with its content. Screenshots of this coding process in Miro are presented under each theme. For each theme, the associated subcodes are introduced, followed by supporting participant quotes, which are then interpreted about relevant theoretical concepts. All quotes are referenced in Appendix 5.

The three identified themes are: *Navigating the Interface*, *Emotional and Situational Viewing*, and *Trust and Frustration with Recommendations*.

5.2.1 Navigating the Interface

A pattern across all pre-interviews is participants' tendency to navigate through the most visually appealing elements on Netflix's homepage. Many describe a superficial navigation behavior that focuses on content already visible upon logging on to the platform. These tendencies are divided into four subcodes, which are elaborated upon below.

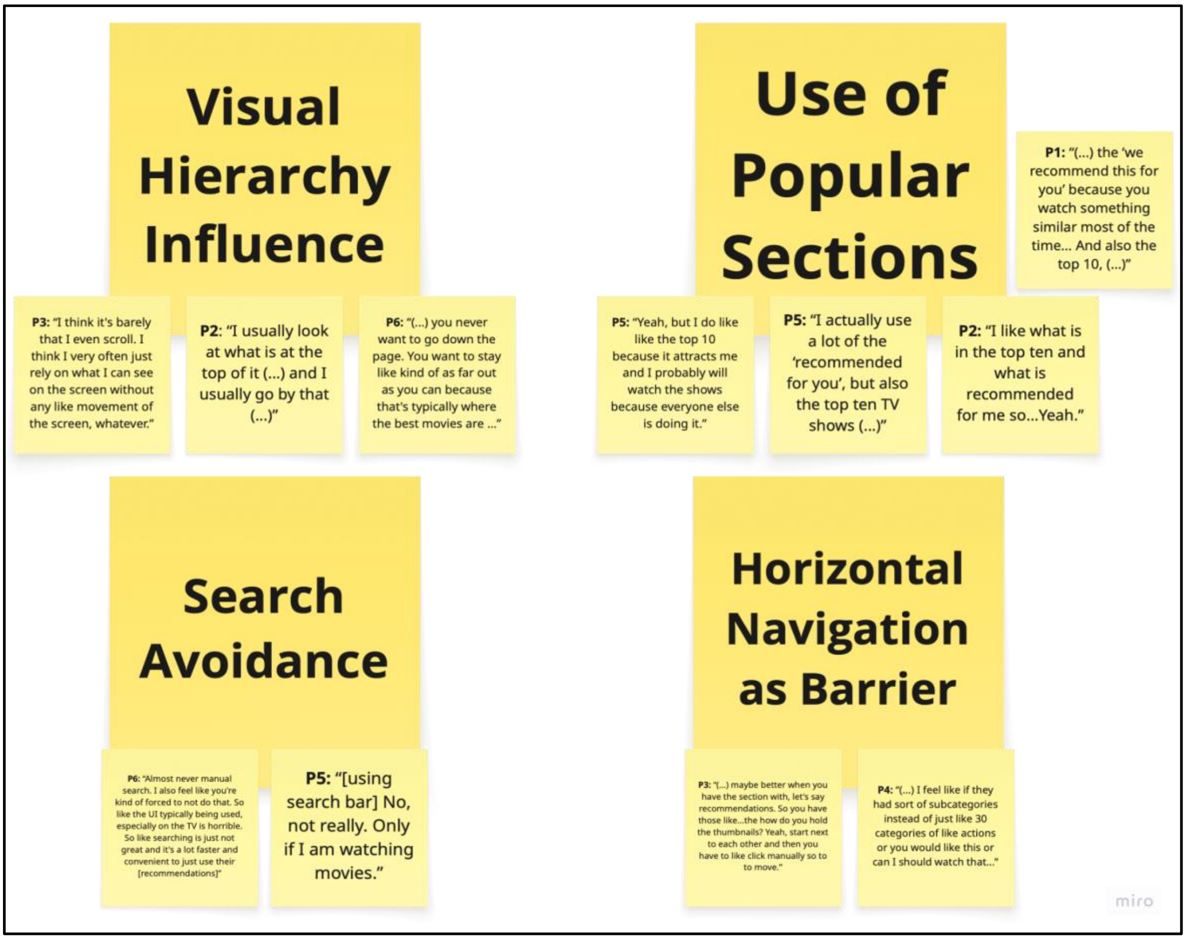

Figure 1: Thematic coding of Navigating the Interface via Miro

The subcode *"Visual Hierarchy Influence"* discusses how the layout and visual structure of Netflix's homepage guide users' attention. Many participants mention how they usually go for whatever first appears:

> "I usually look at what is at the top of it (...) and I usually go by that (...)" (P2, Appendix 5.1, p. 83).
>
> "I think it's barely that I even scroll. I think I very often just rely on what I can see on the screen without any like movement of the screen, whatever." (P3,

Appendix 5.1, p. 83).

"(...) you never want to go down the page. You want to stay as far out as you can because that is typically where the best movies are …" (P6, Appendix 5.1, p. 84).

Although the interface allows users to scroll, participants explain that they tend to stay at the top. According to Norman, visibility promotes user understanding and decision-making, meaning that what users see will be used (2013, p. 22). Specifically, Participant 6 highlights that the best movies are at the top of the page, which supports the positioning bias as the interface strategically places their preferred content, influencing participants' decision-making (Anwar et al., 2024).

To this, the subcode *Use of popular sections* suggests a preference for shortcuts such as "Top 10" and "Recommended for you". Emphasizing their previous preference to stay at the top of the page, some participants explain their use of these shortcuts in choosing their content:

"...the 'we recommend this for you' because you watch something similar most of the time… And also the top 10..." (P1, Appendix 5.1, p. 83).

"I like what is in the top ten and what is recommended for me, so…Yeah." (P2, Appendix 5.1, p. 83).

"Yeah, but I do like like the top 10 because it attracts me and I probably will watch the shows because everyone else is doing it." (P5, Appendix 5.1, p. 84).

"I actually use a lot of the 'recommended for you', but also the top ten TV shows (...)" (P5, Appendix 5.1, p. 84).

Therefore, for some participants, there is a bias towards engagement and popularity, which means that the system seems to place popular content in specific positions, which means that when users use these placements to find content, the algorithm prioritizes this type of content over other, which creates a self-reinforcing feedback loops (Mittelstadt et al., 2016, p. 5).

Supporting the preference of using the first couple of rows on Netflix is the subcode *Search avoidance*. Here, some participants speak of how they rarely use the search feature and instead focus on the guidance of the main interface:

"Almost never manual search..." (P6, Appendix 5.1, p. 84).

"[using search bar] No, not really. Only if I am watching movies." (P5, Appendix 5.1, p. 84).

Specifically, Participant 6 states that they only use the search feature if they must watch a movie, perhaps indicating that if they know what they want to watch, they will search for it: otherwise, not. Moreover, this behavior suggests that the search feature is either unnecessary or overly complicated to use, which in turn influences users' motivation to search for specific content or actively use relevant search terms. It weakens users' sense of agency, as the structure actively limits users' possibilities for expressing their preferences directly and instead guides them to stay through algorithmically curated content.

Lastly, *Horizontal Navigation as Barrier* describes how some participants feel that the specific choices in the UI, such as horizontally scrolling through the content rows, can diminish the navigation flow:

> "(...)Yeah, start next to each other and then you have to like click manually so to to move." (P3, Appendix 5.1, p. 84).
>
> "(...) I feel like if they had sort of subcategories instead of just like 30 categories of like actions or you would like this or can I should watch that..." (P4, Appendix 5.1, p. 84).

Participant 3 states that they find the requirement to manually click to move through the horizontal rows annoying, while Participant 4 feels that there are too many rows. These pain points indicate a need for deeper exploration and can either lead to decisions based on superficiality or a return to content they are already familiar with. Just like Schwartz argues, too many choices can lead to an overload that overwhelms the user (2004, p. 85). In contrast, Scheibehenne et al. argue that choice overload is not a universal issue. Instead, it depends on the complexity of the issue at hand, meaning the user's level of involvement and presentation of options (Scheibehenne, 2010, pp. 410–411). In the context of Netflix, it is worth considering whether the issue lies in the number of options or the way they are presented. However, it is also possible that both are complementing each other and causing issues.

All subcodes suggest participants primarily navigate through visual content on the homepage, reflecting both positioning and popularity bias. The underutilization of the search function reinforces these biases, which can lead to reduced user agency. Moreover, while participants who manage to scroll a bit seem to be overwhelmed by choices, their friction with the horizontal navigation inhibits their curiosity about other content.

5.2.2 Emotional & Situational Viewing

The second theme addresses the emotional situations, such as daily routines and mental capacity, that shape participants' use of Netflix. Instead of using the platform for curiosity and viewing new content, participants tend to use it as a source of comfort, mental regulation, or background noise. The following section will delve into the four subcodes under this theme.

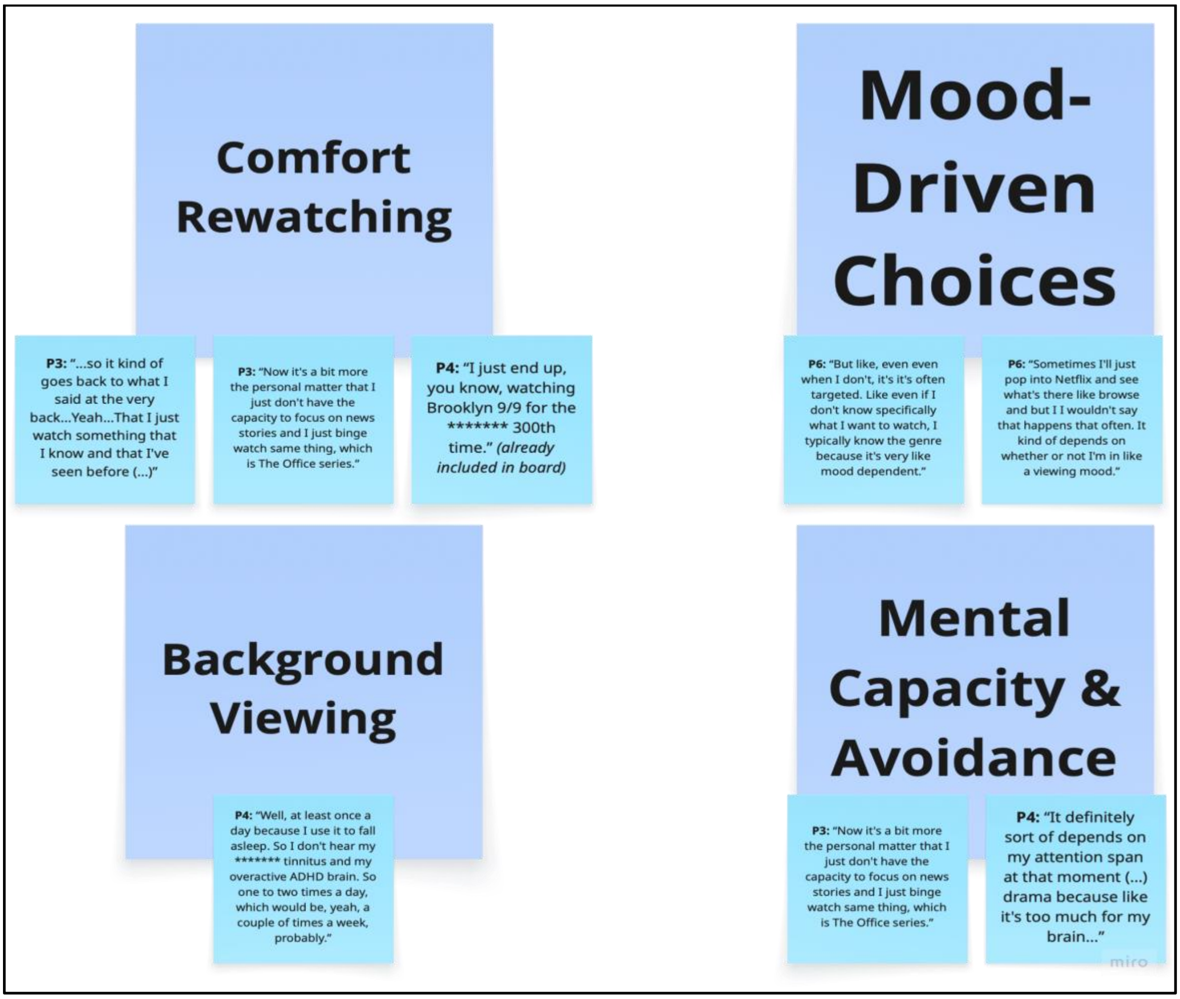

Figure 2: Thematic coding of Emotional & Situational Viewing via Miro

The first subcode is *Comfort Rewatching,* which encapsulates how rewatching familiar content provides participants with emotional comfort. Many users express how they intentionally prefer rewatching the duplicate content:

> "I just end up, you know, watching Brooklyn 9/9 for the ******* 300th time." (P4, Appendix 5.2, p. 84)
>
> "...so it kind of goes back to what I said at the very back…Yeah…That I just

watch something that I know and that I've seen before (...)" (P3, Appendix 5.2, p. 84)

Participant 3 states that, regardless of the reason for using Netflix, they somehow end up rewatching the same TV show continuously, which Participant 4 also emphasizes. This suggests that users' personalization on Netflix is rooted in their behavioral rewatching sessions and may not be presented with new content. This supports Figà-Talamanca and Arfini's idea of user-driven filter bubbles, where the user is the cause of the options by repeatedly choosing the same (2021, p. 1).

Moving along, participants 3 and 4 also speak into this further under the subcode *Mental Capacity and Avoidance*. This subcode speaks volumes about the fact that for some users, the mental state tends to control the decision-making:

> "Now it's a bit more the personal matter that I just don't have the capacity to focus on new stories and I just binge watch the same thing, which is The Office." (P3, Appendix 5.2, p. 84)
>
> "It definitely sort of depends on my attention span at that moment (...) because like it's too much for my brain..." (P4, Appendix 5.2, p. 84)

Above, it shows that although users rewatch for comfort, it might not always be the algorithm's fault, as users are not mentally available to receive the recommendations. This again emphasizes Talamanca and Arfini's belief that algorithms are not always to blame.

Mood-Driven Choices showcase further how specific moods can influence the specific kind of decisions made by participants:

> "Even when I don't [know what I want], it's often targeted. Like even if I don't know specifically what I want to watch, I typically know the genre because it's very like mood dependent." (P6, Appendix 5.2, p. 84)
>
> "Sometimes I'll just pop into Netflix and see what's there like browse and but I I wouldn't say that happens that often. It kind of depends on whether or not I'm in like a viewing mood." (P6, Appendix 5.2, p.84)

These statements by Participant 6 show that, at times, logging on to the platform is targeted with a specific mood, leading to different choices. Netflix becomes an emotional regulation mechanism. Instead of navigating based on content type, they specifically choose genres that align with the specific mood they are in. Emphasizing Knijnenburg et al. (2012), it is essential to recognize that watching Netflix and utilizing recommendations are dynamic and context-based rather than static (p. 1027).

Lastly, the subcode *Background Viewing* showcases how some participants use Netflix as background noise, which serves a therapeutic purpose:

> "I use it to fall asleep. So I don't hear my ******* tinnitus and my overactive ADHD brain." (P4, Appendix 5.2, p. 84)

Meaning that not every time users use the platform, they are actively searching for something to focus on, but more something that can serve as an emotional buffer—traditional understandings of what "engagement" means that Netflix plays a role in users' self-regulation. The platform becomes a functional element in everyday life, rather than just an entertainment source.

These subcodes suggest that users' emotional and situational conditions influence their experiences with Netflix, particularly in terms of whether they receive personalized recommendations. It suggests that algorithmic personalization is not solely about technical precision, but situational relevance and users' mental and emotional availability in the moment. Just as Knijnenburg et al. argue, systems should also measure how well they meet users' needs in a specific situation (2012, p. 1028).

5.2.3 Trust & Frustration with Recommendations

The third and last theme for the pre-interviews relates to the tension between participants' trust in Netflix's recommendation system and moments of frustration, doubt, and resignation. Many participants describe how the platform is helpful and relevant to their overall needs. However, at times, they are skeptical due to repetitiveness, mismatched recommendations, and the algorithmic logic that is not always transparent. This ambivalence reflects key issues relating to algorithmic bias, filter bubbles, and user agency.

Figure 3: Thematic coding of Trust & Frustration with Recommendations via Miro

First, we have the subcode *Trust During Passive Interaction*, which encompasses situations where the recommendation system is intuitive, especially when users are relaxing or not directly evaluating the system:

> "No, I think as of right now, I think the system works because I feel like. When I'm watching Netflix, I'm like in the zone. You know, I just watch and then something pops up like ohh that, That's nice. I'll watch that..." (P1, Appendix 5.3, p. 85)
>
> "(...)if it's like a show or movie, I'm like watch it to the end where it comes up with the like. If you 'if you like this or not', I'm like, you know what I did like this a lot" (P1, Appendix 5.3, p. 85)

Participant 1 describes being in a specific state of mind when watching and how they are more receptive to Netflix recommendations, finding that the content seems to fit these moments well. They emphasize this by liking the content that Netflix recommends. As Norman points out, well-designed systems can create a sense of ease and intuitiveness, where users proceed through actions with minimal cognitive effort and friction (Norman, 2013, pp. 41, 83). However, at the same time, such seamless interaction may mask underlying biases and limit users' critical reflection, which challenges the perception of the system's reliability.

The subcode *Overwhelming Options* speaks into the many options that can lead users to confusion or difficulty in making decisions:

> "I must say I feel overwhelmed sometimes. It's a lot of information on a lot of different series and just having to choose and figure out if this one is better than the other." (P2, Appendix 5.3, p. 85)
>
> "(...) it becomes like this bottomless pit. But there's a lot of crap…that's overwhelming (…)" (P6, Appendix 5.3, p. 85)

Both participants state that they experience Netflix as having much information that they need to manage. Participant 6 even goes as far as calling it a "bottomless pit," emphasizing their view that the interface has too much content. These statements reflect a case of choice overload, which, for a recommendation system, can cause users to feel less satisfied with the experience and even limit their user agency (Schwartz, 2004, pp. 3-4).

Ignored Feedback reflects a specific situation where the algorithm does not seem to react to users' past actions:

> "Usually the algorithm is like, hey, you really like this thing? Why don't you watch this thing? And I'm like, I already did watch that thing." (P4, Appendix 5.3, p. 85)
>
> "Sometimes I'm thinking, why has this come up? But it's not something I have reacted on." (P2, Appendix 5.3, p. 85)
>
> "I…I don't think like for years and years I haven't watched any reality shows on Netflix at all. Yeah. So when it's sort of recommending Love is Blind or whatever, I'm just like, seriously?!" (P4, Appendix 5.3, p. 85)

These specific instances indicate a break in the feedforward and feedback loops, where the system fails to adapt to users' actual preferences, undermining their trust and control (Norman, 2013, p. 29). One interesting aspect is how it seems to recommend content that has already been watched or even suggest genres that the user claims they never watch. The statement recommending *Love is Blind* to Participant 4 can indicate a case of promotional bias, as the show is a Netflix Original, which is why it was recommended to the participant.

Speaking of wrong genre recommendations, *Genre Mismatch* reflects how users experience recommendations that are out of touch with their preferred genres:

> "sometimes I'm kind of like baffled by why you would recommend Paw Patrol when I just watched John Wick." (P6, Appendix 5.3, p. 85)
>
> "(…)if I watch Gossip Girl, I think it will like recommend other kind of girly school kind of thing and I find like the options that they suggest, not very. It's not something that I would watch or else maybe I watch one episode, but it's not really for me (...)" (P5, Appendix 5.3, p. 85)
>
> "it's annoying sometimes when you watch something (…) It will recommend the new biggest thing on Netflix because we watched this one property. But these properties have nothing to do with each other (...)" (P6, Appendix 5.3, p. 85)

Flaxman et al. demonstrate in their quantitative analyses that users, despite algorithmic personalization, are not necessarily confined to closed information environments; they are often exposed to diverse content (2016, pp. 20–21). This supports the idea that repetitive recommendations and mismatched genres are not just signs of system flaws but may also reflect a broader information strategy that attempts to balance diversity and predictability. However, based on participants' frustration, it seems that this balance does not always feel meaningful or effective from the user's perspective. Such mismatches in recommendations can be random and unserious, as Participant 6 states that they randomly receive suggestions, such as Paw Patrol, despite not watching kids' shows.

However, it is not only mismatches that seem to be an issue; some participants, along with mismatches, experience repetitive recommendations at times, which *Repetition Fatigue* covers:

> "It's only like the same shows, so sometimes it's just the same stuff that keeps coming up..." (P5, Appendix 5.3, p. 85)
>
> "Sometimes it's just like, because you watched this show, and it's not the same show at all. It has a little bit of similarities, but it's not really..." (P5, Appendix 5.3, p. 85)
>
> "(...) I think I will just stick to what I know so could be that maybe they are a bit repetitive, but again I never really give it like… I never gave it a second thought." (P3, Appendix 5.3, p. 85)

It suggests a superficial personalization logic based on genre similarities rather than more profound semantic relevance. Here, unlike previous reflections, it seems that this indication of filter bubbles is not due to comfort rewatching, but instead the algorithm's logic, which does not allow the user to reach out and receive a new set of content.

Across all five subcodes, users' experience of the recommendation system fluctuates between trust and frustration, depending on the situation and its results. Trust tends to be high when the system aligns with users' expectations and allows for effortless navigation, but it declines when users encounter confusion, repetition, or mismatched recommendations. These reflections highlight the importance of developing transparent, context-sensitive, and user-centered recommendation systems that facilitate understanding and informed decision-making, rather than relying on passive acceptance.

5.2.4 Takeaways

The three themes demonstrate how participants' experience of Netflix's recommendation system involves a combination of passive navigation, emotional use, and fluctuating trust. While the user interface contributes to comfort and visual dominance, rather than facilitating clear exploration, the decision-making processes are highly influenced by mood and mental capacity, rather than curiosity. The participants experience the system as both helpful and frustrating, depending on how it meets their needs. Collectively, it indicates that personalization is not only about technical precision, but also a dynamic and situational relationship between the user and the system. The following section will delve into the analysis of the data collected through the diary study.

5.3 Diary Study Analysis

This section presents the qualitative and quantitative analysis of the data collected from the diary study, where six participants recorded their interactions with Netflix's recommendation system. The goal of the analysis is to identify tendencies in the participants' user behavior, feelings toward the relevance of the recommendations, and possible pain points with the UI. The data is structured and processed to ensure consistency, transparency, and analytical applicability, which I will explain in the following subsection.

5.3.1 Data Processing & Methodological Reflection

The original dataset is exported from Microsoft Forms and includes 56 entries (Appendix 6, p. 86). Every entry equaled a daily log and could include up to five content items (meaning a film, one TV episode, or multiple episodes from the same show). By integrating the

branching logic into the survey form, the number of questions each Participant had to answer depended on the number of items they watched.
To ensure the applicability of the data, I will utilize Tableau for visualization, so I processed it from a wide format (due to branching) to a long format. Each row in the processed dataset represents a single item of content per day, including various information such as its title, how it was found, and whether it was finished watching.

The collected dataset finishes with 52 unique diary days. This number is lower than the 56 entries because some participants submitted multiple answers on the same day, primarily due to forgetting to complete the survey the previous day. Instead of cleaning out these minor issues, I left them in the dataset as is. This was to ensure a level of authenticity in the data collection by also acknowledging human mistakes and the kinds of deviations that can occur in real-life and practice-based studies. Furthermore, two participants logged the same show more than once during the same day, likely because they watched multiple episodes of the same show. As the study is designed to define a single content item as a movie, one episode, or multiple episodes of the same show within a single session, I considered these entries redundant and removed them during processing.
To ensure transparency, the specific number of diary study days for each participant is as follows:

- Participant 1: 10 Days
- Participant 2: 9 Days
- Participant 3: 9 Days
- Participant 4: 7 Days
- Participant 5: 10 Days
- Participant 6: 7 Days

With the dataset processed, I proceed to analyze the diary study. The following section will examine both descriptive patterns (e.g., types of content watched, completion rates, and discovery methods) and more nuanced trends related to repetition, perceived relevance, and user behavior. These observations are supported by visualizations created in Tableau and informed by the theoretical framework established earlier in the thesis.

5.3.2 Analysis of the Diary Data

This section presents a descriptive analysis and contextualizes the diary data. Through descriptive statistics and visualizations in Tableau, I examine the types of content users typically watch, how they discover it, and whether they finish it. Furthermore, the analysis will shed light on participants' satisfaction with the recommendations and how they might differ across content type and recommendation source.

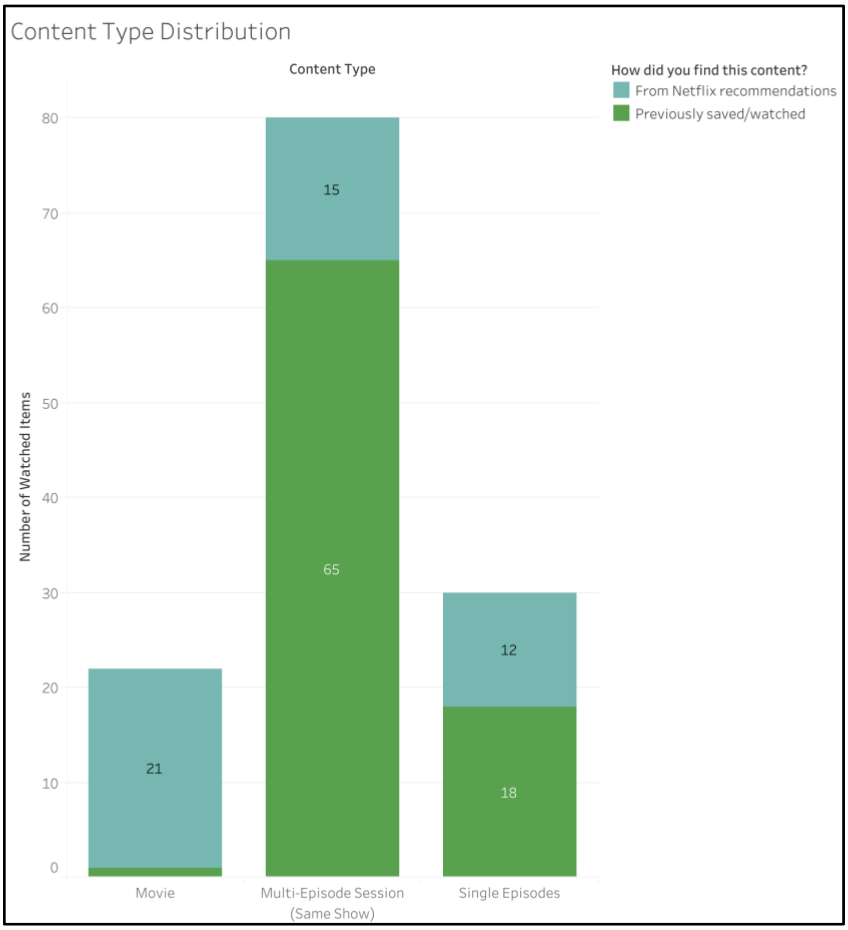

Figure 4: Content Type Distribution

Starting the descriptive analysis, Figure 4 presents the distribution of content types and how they were discovered, specifically whether they were recommended by Netflix or selected independently by the participant (e.g., previously saved or manually searched for).

The figure reveals that series, particularly "multi-episode session (same show)", dominate among the three content types. This category accounts for 80 out of 132 logged items, with the majority (65 out of 80) being content that the participants had either already saved or watched before. This pattern is also reflected in the participants' free-text responses. For instance, Participant 2 noted, "Same as yesterday – one of the first suggestions and it showed a funny short clip of it before choosing" (Appendix 6.1, p. 84), highlighting how autoplay previews may nudge users toward repetitive selections. Likewise, Participant 3 explained, "It started automatically (autoplay)" (Appendix 6.1, p. 84), reinforcing the idea that not all choices result from active exploration. Furthermore, it relates to how Netflix's interface supports and promotes a specific type of content. Just as Mittelstadt et al. (2016) describe algorithms as value-laden, coded with specific values.

In contrast, movies and single episodes are more evenly distributed. However, in the case of single episodes, the difference between self-initiated and recommended discovery is relatively small. Notably, most of the movie-type content appears to have been surfaced primarily through Netflix's recommendation features, rather than through prior user intent. This suggests that shorter, standalone content may rely more heavily on algorithmic visibility to capture attention, unlike series, which often benefit from continuity or previous familiarity.

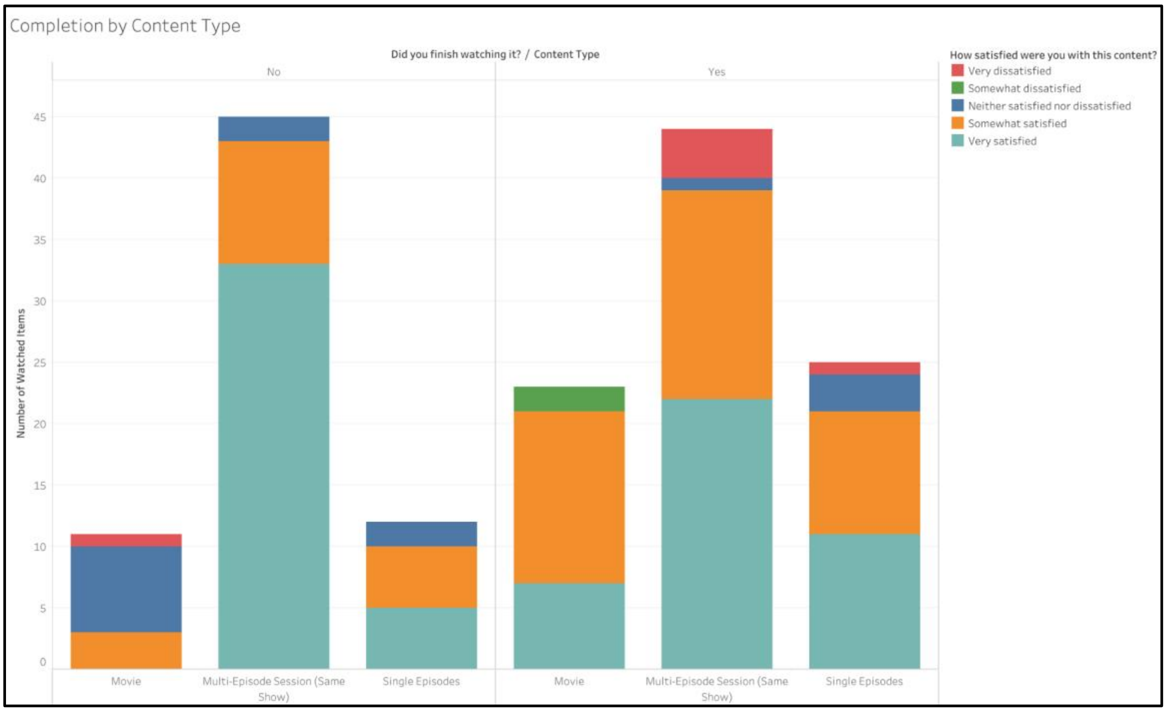

Figure 5: Completion Rates by Content Type

Figure 5 illustrates the relationship between content type, completion rates, and satisfaction. It shows that series (multiple-episode sessions) have the highest completion rate, with 44 out of 89 total viewed series completed. In contrast, movies and single episodes are less frequently completed, though the difference is modest. This suggests that while content type influences completion, the variation is not drastic. It would be relevant here to consider whether the content was recommended or not, which is addressed further in Figure 6.

Moreover, the figure illustrates the correlation between satisfaction and completion. Participants generally rated multiple-episode sessions as the most satisfying, regardless of whether they completed them. However, some of the free-form answers indicate that satisfaction may not always stem from novelty or enthusiasm. Participant 3 explained, "It's this comfort series that I will just start playing (...) sometimes I will not even pay attention to what's happening” (Appendix 6.1, p. 84). This aligns with the notion that emotional comfort and habitual viewing can influence both completion and perceived satisfaction, even in the absence of active engagement.

This suggests that completion does not necessarily reflect high engagement or enjoyment. Some participants might complete a series out of habit, a desire for continuity, or due to a lack of alternatives, rather than out of active enthusiasm. In other words, completion can be driven by contextual or emotional momentum rather than genuine satisfaction.

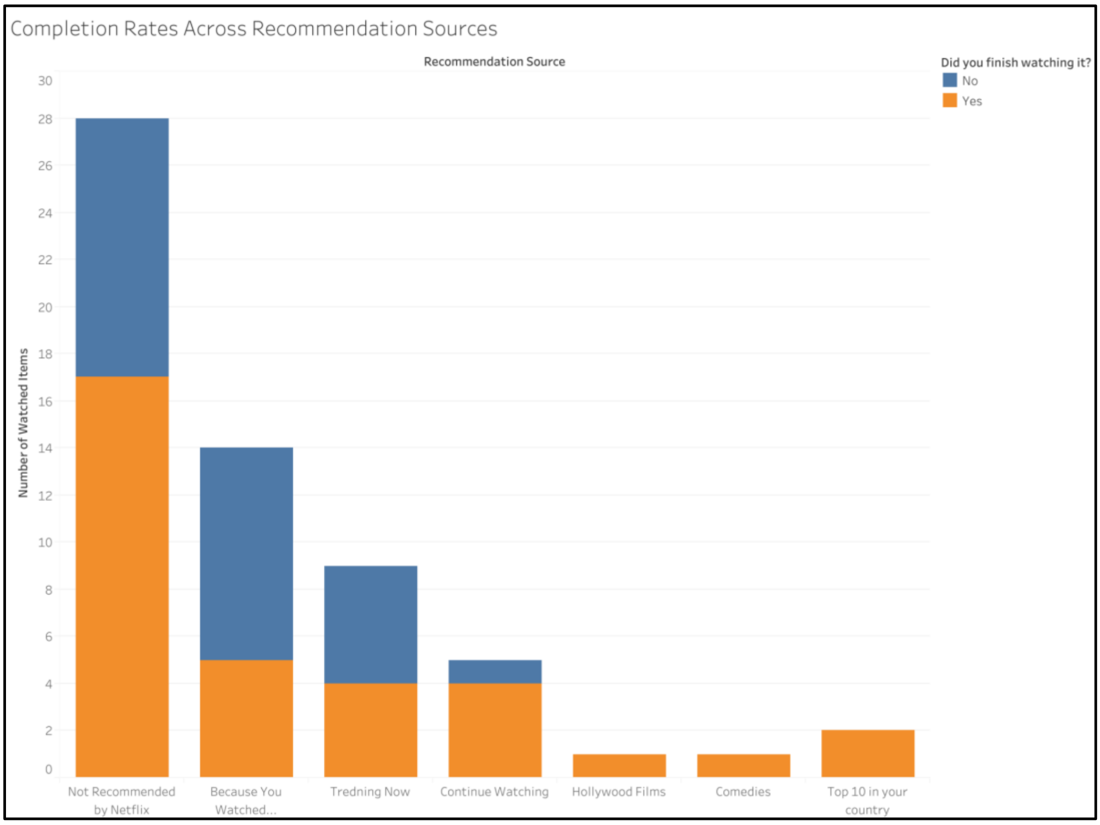

Figure 6: Completion Rates Across Recommendation Sources

Figure 6 illustrates how completion rates differ across various recommendation sources, such as *“Because you watched”*, *“Trending Now”*, and *“Not Recommended by Netflix”*. The category *“Not Recommended by Netflix”*—referring to content found manually or outside of Netflix’s algorithmic interface—shows the highest number of completed items (17 out of 28), corresponding to a completion rate of around 61%.

In comparison, Netflix-recommended categories collectively account for 17 completed and 15 not completed titles, totaling 32 items. These include “Because You Watched” (5/14), “Trending Now” (4/9), “Continue Watching” (4/5), and a few niche categories like “Hollywood Films”, “Comedies”, and “Top 10 in Your Country”, each with perfect but very small completion counts.

While both groups resulted in the same number of completed titles (17 each), the *completion rate* was slightly higher for content not recommended by Netflix (17 out of 28, ~61%) compared to Netflix-recommended content (17 out of 32, ~53%). This modest difference may suggest that user-driven discovery — even if it requires more effort — tends to lead to more committed or satisfying viewing experiences. However, it is worth noting that this primarily reflects the number of watched items.

This could suggest that user-driven discovery, rather than passive acceptance of algorithmic nudges, leads to more meaningful engagement. This observation is reinforced by Participant 5, who described having to scroll down the homepage to find something they wanted to watch: “It was placed 8–9 rows down. Therefore, I scouted the Netflix homepage before I chose that TV show” (Appendix 6.1, p. 84). This indicates a conscious act of exploration beyond default recommendations. In contrast, Participant 5 also noted another case where the content was not visible: “My boyfriend searched for it, so it did not just appear on my homepage” (Appendix 6.1, p. 84). These examples demonstrate that algorithm-independent selection often requires effort but yields higher satisfaction and completion rates, as indicated by the numerical differences between sources.

While algorithmically recommended items are visible and easily accessible, they may not align as closely with users' actual preferences, resulting in lower completion rates despite higher exposure.

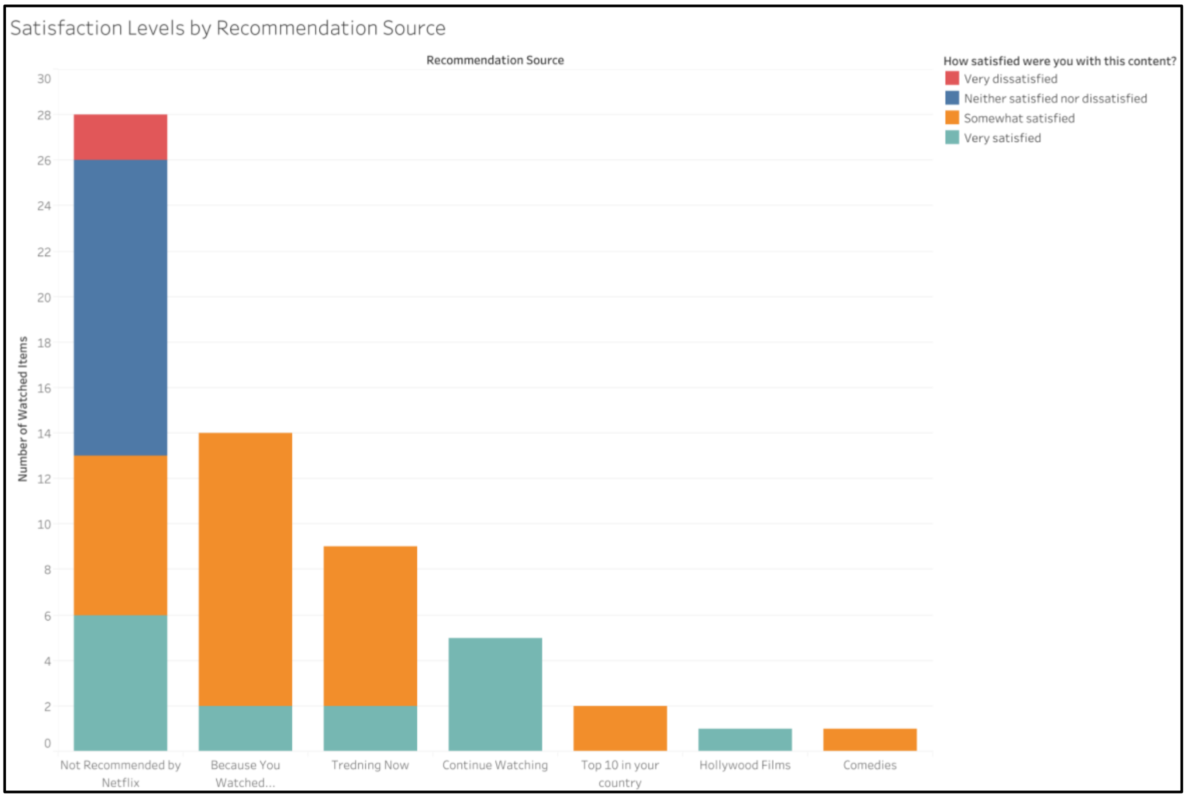

Figure 7: Satisfaction Levels by Recommendation Source

Figure 7 builds on the findings of Figure 6 by analyzing satisfaction levels about the source of the recommendation. The "Not Recommended by Netflix" category displays a more even spread across satisfaction levels, from "very satisfied" to "neither satisfied nor dissatisfied." This suggests a broader range of user expectations and more varied emotional responses when the content is selected independently. In contrast, content recommended by Netflix — especially from categories such as "Because you watched…" — shows a stronger clustering around "somewhat satisfied" and "very satisfied."

However, this clustering should not be taken as a clear sign of stronger engagement or deeper enjoyment. Instead, it may reflect the influence of interface design, where users tend to follow the most visible or familiar options. Such behavior may stem more from convenience or habit than from genuine interest. In this sense, the high satisfaction ratings indicate passive contentment rather than active approval.

Interestingly, the broader distribution of satisfaction for non-recommended content may point to more deliberate choices and authentic reactions, both positive and negative. Participant 3 reflected on this dynamic, stating: "Now it's a bit more the personal matter that I just don't have the capacity to focus on new stories and I just binge-watch the same thing"

(Appendix 5.2, p. 81). This illustrates how satisfaction is not always a direct result of algorithmic success but is somewhat shaped by emotional context and user readiness. In such cases, the alignment between the recommendation and the need is incidental, not intentional, highlighting the limitations of personalization that lack situational awareness.

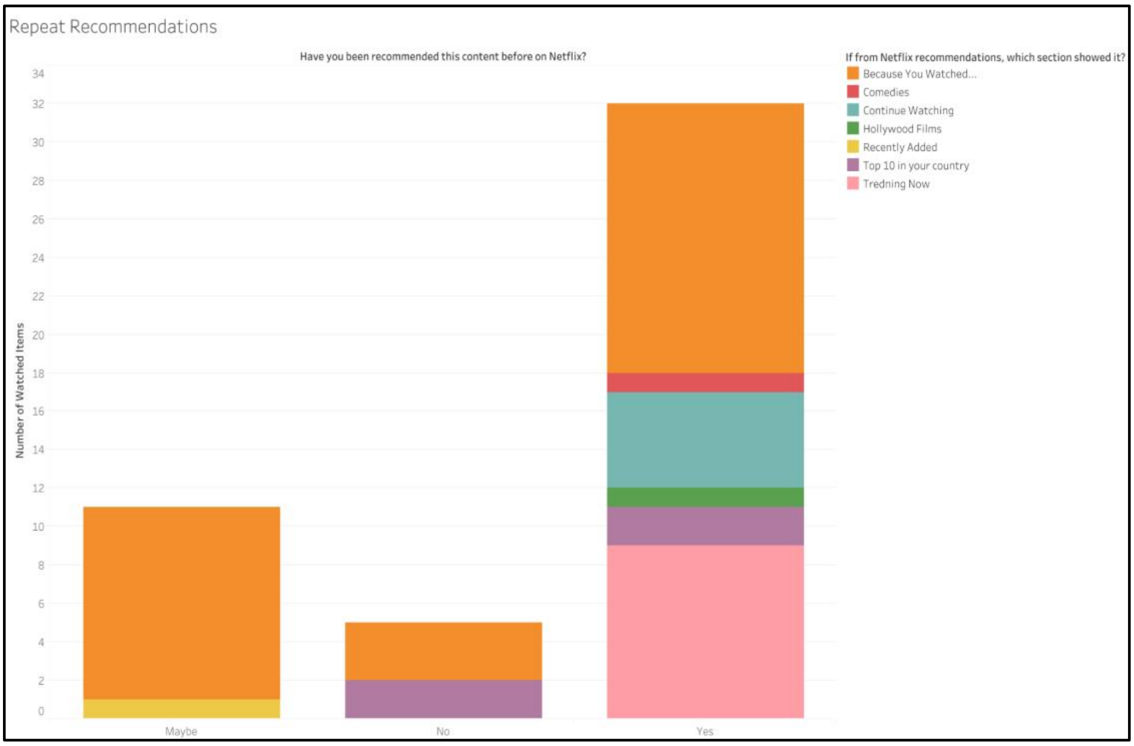

Figure 8: Repeat Recommendations

Looking at Figure 8 on Repeat Recommendations, we see a large group of viewed content have been recommended to the participants before - a total of 34 titles, and the colors indicate where this recommendation found place it seems that section such as “Trending Now” and “Because you watched…” repeats as a primary source for repeated recommendations. It suggests that Netflix’s recommendation system tends to recirculate the same type of content. Several participants noticed this repetition. When users continue to see recommendations for content they have already shown they dislike, it can erode their confidence in the system's ability to ‘know’ or adapt to their preferences. This pattern suggests that the algorithm is not genuinely personalizing suggestions but rather relying too heavily on past behavior, creating a loop that recycles the same kinds of content, even if it is not wanted. Such repetitive patterns suggest an algorithmic feedback loop, where previous behavior is reused for future recommendations. This can create a filter bubble, limiting the possibility of discovering new content. Just as Figà-Talamanca and Arfini describe, variation tends to become superficial - in other words, an illusion of diversity based on minimal differences (2021).

5.4 Thematic Analysis of Post-Interviews

This thematic analysis analyzes the post-interviews conducted after the end of the diary study. Like the pre-interviews, this analysis follows a deductive approach, where the coding is based on the same three overall themes. However, the reflexive nature of the diary study allows for inductive elements to emerge, as participants expressed new patterns and reflections.

To ensure analytical continuity, the three main themes are maintained: Navigating the Interface, Emotional and Situational Viewing, and Trust and Frustrations with Recommendations. Two new themes emerged during the coding: Study-Induced Behavior Shifts and Personalization Features and Suggestions. These themes encapsulate participants' awareness and reflection, as well as their specific evaluations and recommendations, which speak to how the study design has influenced their experience with Netflix's recommendation system.

All interviews are coded through Miro and organized into subcodes under each theme. The following sections present each theme, accompanied by connecting subcodes, quotes, and interpretations. All quotes will be referenced to Appendix 8 (p. 135).

5.4.1 Navigating the Interface

Post-interviews reveal that participants still experience some limitations in their navigation of the UI. Although the experiences are emphasized, they are also connected with more reflection and awareness, which is likely due to a combination of conducting the pre-interviews and the diary study.

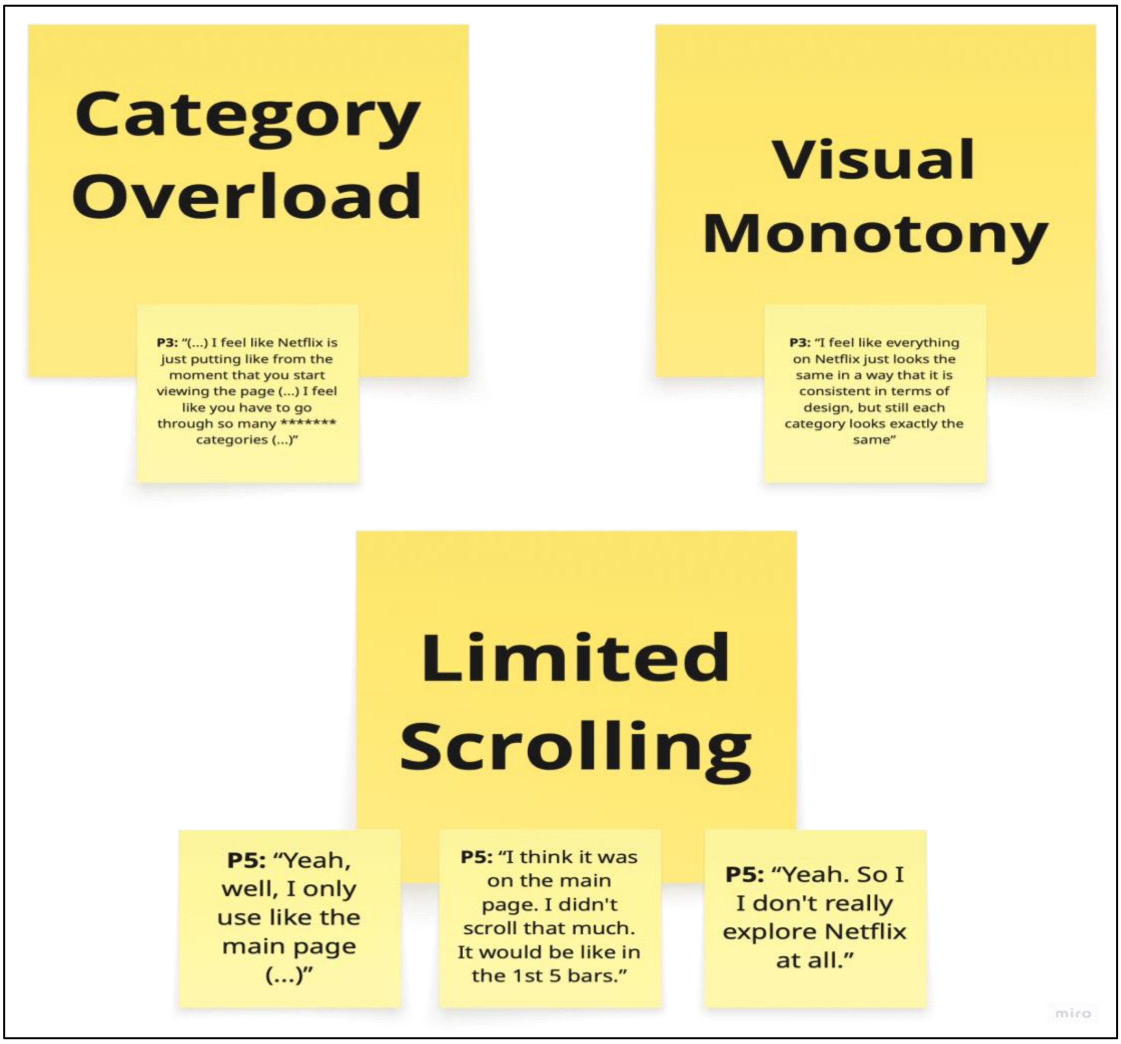

Figure 9: Thematic coding of Navigating the Interface via Miro

The first subcode is *Category Overload,* which, as a similar subcode from the pre-interviews, emphasizes how the horizontal rows and genres feel overstimulating or hindering for users' motivation:

> "I feel like Netflix is just putting like from the moment that you start viewing the page (...) you have to go through so many ******* categories." (P3, 8.1, p. 135)

Although variation and numerous choices can be an advantage, too many options can lead to frustration and difficulty in making decisions. As Schwartz emphasizes, it does not empower the user (2004, pp. 3-4). As this issue recurs, so does the behavior of staying at the top of the page.

Limited Scrolling reveals that many participants still do not scroll past the top part of the homepage, which may be due to the numerous options that overwhelm the individual:

"I didn't scroll that much. It would be like in the 1st 5 bars." (P5, Appendix 8.1, p. 135)

"Yeah, well, I only use like the main page (...)" (P5, Appendix 8.1, p. 135)

"Yeah. So I I don't really explore Netflix at all." (P5, Appendix 8.1, p. 135)

These main rows are typically algorithmically prioritized and serve as anchors for attention, where positioning and popularity bias emerge. Norman describes this as constraints, not suggesting technical limitations, but design choices that directly direct behavior (2013, p. 43). Even if the user technically has access to the entire catalogue of content, they remain at the top, reinforcing biases and limiting content diversity.

Lastly, *Visual Monotony* highlights how the interface's similar aesthetics create, for some, a perceptual exhaustion:

"I feel like everything on Netflix just looks the same in a way that it is consistent in terms of design, but still each category looks exactly the same" (P3, Appendix 8.1, p. 135)

Consistency in design is generally a good UI principle; however, in this context, the number of similar visuals can cause users to lose track of orientation. It can be challenging to differentiate between rows, and its visual uniformity counteracts the feeling of variation. The feeling of variation and transparency is just as important as algorithmic personalization, especially in terms of engagement and satisfaction (Knijnenburg et al., 2012, pp. 443-444). The post-interviews showcase similar and continuous experiences in navigation, however, with greater reflection and awareness. Overstimulating category rows and visual monotony can reduce engagement and decision-making processes. Participants remain at the top of the page, where algorithmic placements dominate, reinforcing positioning and popularity bias. These patterns of staying at the top are also a consequence of the design-based constraints (Norman, 2013, p. 43). Even though users do have access to the catalogue, the interface's visual hierarchy signals which actions they "expect" or categorize as "easy". Netflix's recommendations do not just become a content-based choice, but a design-based one, where affordances (such as the play button and visual highlights) and constraints (limited choices without scrolling) actively shape users' decisions.

5.4.2 Emotional & Situational Viewing

For the post-interviews, the emotional and situational aspects remain in participants' descriptions of their engagement with Netflix. The pre-interviews highlight habitual rewatching and mood-based decisions, and the post-interviews emphasize the same factors.

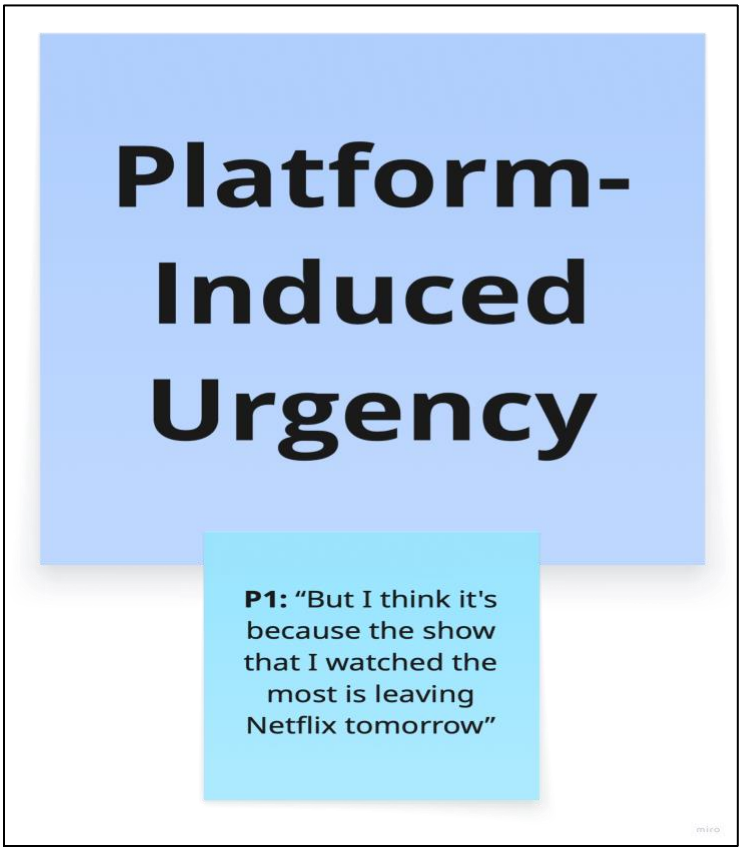

Figure 10: Thematic coding of Emotional & Situational Viewing via Miro

As this theme generally repeats many of the same points from the pre-interviews, these quotes are not included to avoid repetitive arguments. However, a quote reflecting *Platform-Induced Urgency* made it clear that Netflix can also create a feeling of rush:

> "…because the show that I watched the most is leaving Netflix tomorrow." (P1, appendix 8.2, p. 135)

Participant 1 states that the platform notifies them about limited availability, which not only informs them but also potentially changes the users' behavior. If the content being removed is content users watch for comfort, it may lead to a surge in viewing sessions due to the fear of no longer being able to enjoy the content. One could argue that this logic can subtly encourage users to watch the content and even prompt them to view something they might have saved or never intended to watch.

Emotional and habitual patterns are less apparent in the post-interviews, yet they still underscore that Netflix's interface and algorithmic signals can influence what users watch, when, and why. If someone's favorite TV show is canceled the next day, this can perhaps strengthen the emotional connection they had to the content and lead to faster and less reflective reactions. From a design and recommendation perspective, it raises questions about transparency and control within the system.

Furthermore, considering these user patterns, recommendation systems should take users' affective needs into account. Giacomin (2014) emphasizes that human-centered design is not only about being user-friendly but also about addressing emotional and practical aspects in users' day-to-day lives. In this case, Netflix is not just entertainment, but also a mental buffer, which the platform does not necessarily induce. It raises the question of the extent to which personalization should be more situational and tailored to users' conditions in the moment.

5.4.3 Trust & Frustration with Recommendations

Just as with the other main themes, the theme reflecting trust and frustrations with the recommendation system is also less apparent in the post-interviews. However, it is still emphasized, and although not new reflections, they have become more nuanced, which also highlights the complexity of the relationship between users and algorithms.

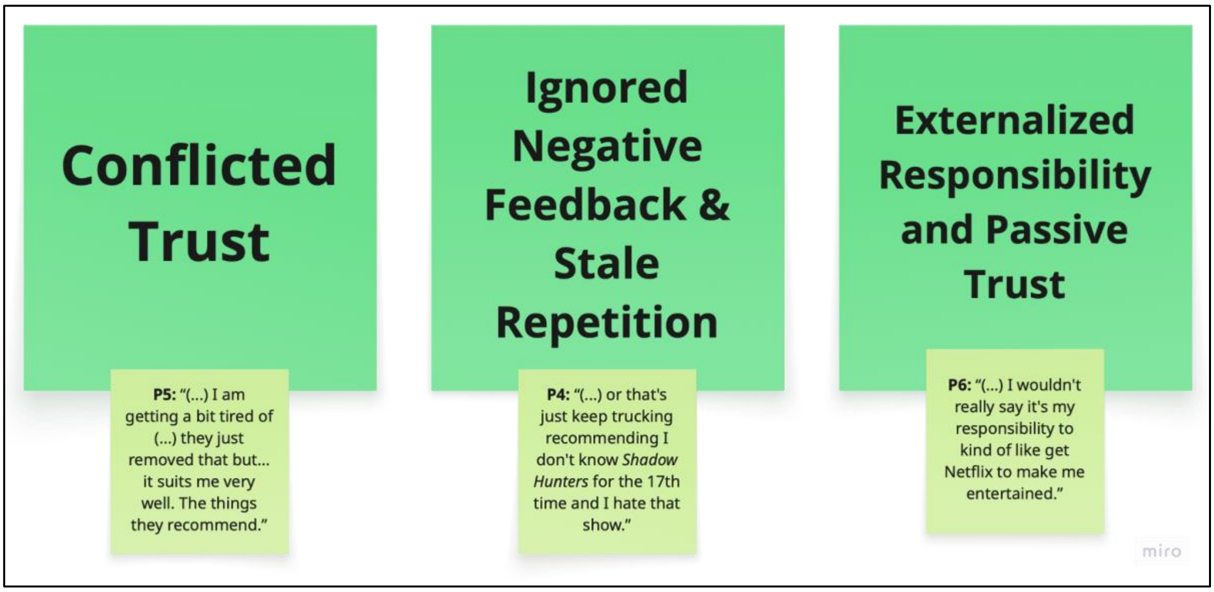

Figure 11: Thematic coding of Trust & Frustration with Recommendations via Miro

Ignored Negative Feedback and Stale Repetition encapsulates how the system still recommends content that the participants do not enjoy or care for:

"…keeps recommending Shadow Hunters for the 17th time and I hate that show." (P4, Appendix 8.3, p. 136)

Participant 4 states that they are continually being recommended the show *Shadow Hunters*, and it infuriates them deeply. It highlights the issues with Netflix's recommendation system in how it personalizes content. Just as Mittelstadt et al. argue that algorithms can be coded with specific interests or values, which can lead to certain content being favored over others (2016, p. 3).

Externalized Responsibility and Passive Trust create an interesting conversation, as some participants do not feel that any issues regarding feeling entertained are on anyone but Netflix's shoulders:

"I wouldn't really say it's my responsibility to get Netflix to make me entertained." (P6, Appendix 8.3, p. 136)

This comment suggests resignation, where the user is not actually seeking or expecting transparency or control, but instead adapting to the system's logic, which can still reinforce biases, as it almost actively allows the system to do as it pleases. However, on the other hand, it can be said that this is more of a call for the system to improve, as the participant is the consumer, and Netflix's recommendation system should function intuitively and effectively.

The post-interviews emphasize that users' experiences are a dynamic balance between trust, resignation, and skepticism. Instead of challenging, which can be difficult, users seem to prefer to adapt to the system's limitations. It suggests that user agency is not only about technical control, but also about how users emotionally and practically interact with a complex system.

5.4.4 Study-Induced Behavior Shifts

Moving forward to the new emerging themes in the post-interviews. The first theme is how the diary study may have influenced users' behavior during the 10-day process. Although this study did not intend to change users' behavior, many participants described a shift in their awareness, search patterns, and decision-making processes. It indicates that by participating, it became a brief interruption of their automated habits and platform-induced use of Netflix.

Figure 12: Thematic coding of Study-Induced Behavior Shifts via Miro

Raised Awareness of Habits shows how the logging process makes some users reflect on their habits more than usual:

> "It was OK actually it got me thinking about what I'm actually watching and perhaps what I should focus more (...)" (P2, Appendix 8.4, p. 136)
>
> "(...) it's kind of funny because I haven't really thought about how I sort of move around on the website or the streaming service (...)" (P4, Appendix 8.4, p. 136)

This is not an odd occurrence, as it indicates a temporary transition from automated to intended use. According to Norman, this occurs when users pause and actively evaluate their interactions, which is usually a result of an awareness-induced incident, such as the diary study (2013, p. 83).

Furthermore, the subcode *Intentional Timing of Viewing* speaks on how some participants adjusted their viewing habits, however, not because of the system's recommendations, but because the diary study provoked a metacognitive reflection:

> "I think, yeah, maybe because of the study for OK, maybe now it's a good time to to watch it..." (P3, Appendix 8.4, p. 136)

Here, Participant 3 mentions that a new show seemed fitting, as they are participating in the study. Bruns highlights how algorithms are not alone, and users are more selective if they are aware (2019, pp. 22-23). It indicates that the user can easily shift between automated platform logic and intentional actions, thus explaining the participants' choice to watch new content.

Prompted Active Search describes another intentional behavior induced by the diary study, where a participant actively searched for something: "I did search for something… perhaps subconsciously because of the study." (P2, Appendix 8.4, p. 136). Participant 2 reflects on their act of searching and is unsure whether it is an automated action or intentionally done, as indicated by the diary study. However, such a break from passive behavior can be interpreted as a positive one, where the user experiences a moment of empowerment, actively taking responsibility for filtering and making choices (Konstan & Terveen, 2021, p. 35).

The subcode' *Higher Algorithm Engagement"* describes how the study creates awareness and curiosity for Netflix's recommendations: "I was more inclined to go by recommendation..." (P6, Appendix 8.4, p. 136). Participant 6 states how they felt more inclined to use Netflix's recommendations. It indicates a reflexiveness between context, cognition, and the user interface, where the diary study creates a temporary openness to the system's logic that was not evident during the pre-interviews.

Not only does some participants use the recommendations, but some also states how they even watch more Netflix than usual as a response to taking part of the study, which reflects the subcode *Increase in Viewing Volume*: "I just feel like I watch more Netflix because of the study." (P5, Appendix 8.4, p. 136). Although Participant 5 watches Netflix most days of the week, the study makes them watch more than usual. The existence of a reflexive logging mechanism creates a sense of accountability, where it seems they feel a need to log something. This change of behavior suggests that users are not only shaped by preferences but also by situational and methodological factors.

Lastly, *Superficial Reflection & Ambivalence* cover the participants who did not necessarily see the potential biases in behavioral changes possibly caused by the study:

> "Every time I went on Netflix I was like, ohh this is good because I have to fill out the survey. But that's just in my head." (P1, Appendix 8.4, p. 136)
>
> "Yes, the same amount of hours wasted." (P1, Appendix 8.4, p. 136)

This speaks into Giacomin's surface-level engagement, where users engage in self-evaluation without internalizing or acting meaningfully on their reflection (2015, p. 2).

However, these findings should not be interpreted as a methodological bias, although being aware of it matters, but rather as a valuable analytical lens. It shows that Netflix's default instances are affected by automated interactions and even frictions, such as self-reporting, which can increase user agency. It supports the idea that user behavior is not only shaped by algorithms but by an interplay between design, context, and awareness. The study's ability to activate self-reflection emphasizes how user inclusivity and feedback mechanisms can function as critical breaks and self-reinforcing recommendation cycles.

5.4.5 Personalization Features & Suggestions

Even though the purpose of recommendation systems is to adapt to users' preferences, they also function as non-transparent gatekeepers, which raises questions about how much control users have—and whether they wish for it.

Figure 12: Thematic coding of Personalization Features & Suggestions via Miro

This theme has been touched upon across the interviews, focusing on participants' specific reflections on Netflix's recommendation system and personalization logic, as well as their satisfaction and suggestions for improvement. Their reflections cover control, autonomy, and trust with algorithmic systems.

Firstly, the subcode *Satisfaction with System* speaks on how participants generally feel satisfied with Netflix's recommendations:

> "My algorithm is made for me." (P5, Appendix 8.5, p.136)
>
> "I sort of already feel like I do [have control] by watching the things that I want to..." (P2, Appendix 8.5, p. 135)

Their statements indicate implicit trust and experiences of intuitive design, where the system reacts naturally to user behavior, which Norman highlights as a system should (2013, p. 210). However, as Mittelstadt et al. warn, satisfaction can also mask the reinforcement of biases, as the algorithm can confirm already existing preferences and keep users in a closed loop (2016, p. 6).

However, even with satisfaction, there is still room for improvement. Some participants speak on specific suggestions for improvements. *Better Feedback Timing* relates to a participant who reflects on one specific feedback mechanism:

> "If that function came before 'did you enjoy the movie?', I think that would actually help my algorithm." (P5, Appendix 8.5, p.136)

Above, Participant 5 requests that the question about whether the content watched was liked or not be displayed earlier. They usually do not stay through enough of the end credits and therefore miss it. A design should integrate feedback that serves as a proactive mechanism, strengthening users' feelings of control (Shneiderman & Plaisant, 2005, p. 74).

With feedback also comes feedforward, and *Persistent Preferences* speak into this logic. One participant wants the system to update preferences continuously: "When you first set up your Netflix account... I feel like you should still have that option [to update preferences]." (P4, Appendix 8.5, p. 136). Here, Participant 4 states a desire for a system that combines passive learning with active user input, which can create a balance where users can customize their changing behavior or moods, resulting in better feedback logic.

Aligned with customization, some also want to be able to remove specific content they do not enjoy: "I really don't want to see anything about horror movies... nice to have that feature actually." (P2, Appendix 8.5, p. 135). However, not all wish for the ability to customize their content: "If there was an option to do that, I probably wouldn't use it." (P6, Appendix 8.5, p. 136). This reflects Sundar and Marathe's distinction between user-initiated customization and system-initiated personalization, where power users tend to value more control over the interface. In contrast, others prefer low-effort, system-driven personalization that requires minimal cognitive input (2010, pp. 310-311).

Then, one participant suggested a more dynamic approach under the subcode *Mood-Based Suggestions*: “Maybe it can ask you what mood you’re in today?” (P3, Appendix 8.5, p. 135). This suggestion highlights a form of situational personalization, where the system adapts to emotional and contextual factors upon user login. Giacomin argues for human-centered design that addresses both cognitive and affective needs, which he believes can enhance the user’s sense of relevance and connection (2014, p. 2).

Collectively, this theme reveals that personalization must strike a balance between the system and users’ trust, control, and autonomy. Some participants are satisfied with the system’s current logic, while others suggest minor improvements that can strengthen their feeling of agency, which does not necessarily mean more features. Instead, these improvements should consider better timing and emotional states. Jannach et al. highlight that when users are offered features that allow for direct control and adjustment of recommendations, it increases their trust and experience of the system. It indicates that even small design changes, such as allowing feedback or removing unwanted content, can strengthen users’ feelings of agency and engagement.

5.4.6 Takeaways

The analysis of the post-interviews repeats the three main themes from the pre-interviews, where awareness and reflection are heightened. The participants recognize the limitations in navigation, the emotional dependence on recognized content, and repetitive recommendations, but describe experiences that are more nuanced, which can be an effect of the diary study. Two new themes emerge, providing insight into how users not only adapt to the system but also temporarily alter their behavior when they become aware of their patterns and habits. Multiple participants report changes in their viewing habits, an increase in awareness, and new possible improvements to how personalization should function.

6. Discussion

This chapter presents the overall findings, focusing on answering the research questions, comparing them to related work, examining the design implications these findings create, and, finally, reflecting on the overall methodology. Moreover, these main aspects will be discussed in their subsections below.

6.1 What Do These Findings Mean?

The analysis reveals that Netflix's recommendation system and user interface promote and reinforce specific biases (specifically positioning and popularity bias), primarily through factors such as the placement of specific content at the top and in recommendation rows (e.g., "Because you watched…"). This mechanism prioritizes both aesthetics and strategic promotion of its own and popular content. Participants experience confirmation bias (the tendency to choose what is familiar) and choice overload. These are reinforced by the interface's visual hierarchy and automated features, making it easier to choose the content that appears first. Even though Netflix's UI offers many options, the analysis reveals that participants typically make decisions at the top of the homepage, and it feels easier to repeat content than to explore new content. It creates an experience of control, though in practice it is limited by the system's own structure and design choices. These findings align with earlier literature on how algorithmic and interface-based affordances influence user navigation and perpetuate bias (Bourreau & Gaudin, 2021; Schaffner et al., 2023).

Firstly, analyzing Netflix's UI through bias mapping (Appendix 2.3, p. 3) answers the research question: *What types of biases are embedded in Netflix's recommendation logic, and how are they manifested through its user interface?* It shows how algorithmic and design-based decisions create and shape algorithmic biases. These biases manifest both visually and functionally through autoplay, positioning logic, and recommendation rows such as "Because you watched…". These structures represent both aesthetic priorities and strategic reinforcing mechanisms where Netflix promotes its own productions and popular content.
Furthermore, the analysis reveals how user-driven biases, such as confirmation bias and choice overload, are indirectly reinforced through interface design. The visual hierarchy keeps users in top placements, and autoplay reduces reflective decision-making, particularly when rewatching or selecting familiar content. These observations emphasize how Netflix's interface subtly channels users toward predefined pathways, supporting a recommendation logic that is both commercially motivated and normatively embedded. Relating to Romero Meza and D'Urso's findings, participants in this study express frustration with the repetitive nature of recommendations (2024). Confirmation bias and emotional comfort, as described by users, align with related work on how affective and habitual factors influence content selection (Bastos et al., 2024; Ahn & Lin, 2023).

Secondly, the interviews and diary study answer the research question: *How do users engage with and interpret Netflix's recommendations in everyday use, and how do these* *behaviors contribute to the reinforcement of bias?* Overall, the participants showcased both trust and resignation in the system. Pre- and post-interviews reveal that many users experience the recommendations as intuitive and meaningful, especially in relaxing situations. However, frustrations emerge with repetitive content or irrelevant recommendations. One example of these repetitive recommendations is "Shadow Hunters" (P4, Appendix 8.3, p. 135), where the user expresses a lack of interest in the show. Nevertheless, they continue to face it on their "personalized" feed.

The diary study also reveals that participants tend to rewatch a significant amount of content (especially multi-episode TV shows), which can be attributed to the UI structure and algorithmic logic being influenced by users' past preferences, creating a feedback loop. Such behavior creates algorithmically sustained behavioral passivity (Mittelstadt et al., 2016). Rather than encouraging exploration, Netflix becomes a mirror of users' past comfort actions, potentially creating filter bubbles rooted in emotional safety over novelty.

As the pre- and post-interviews reveal, many watch for comfort and describe how decisions on what to watch depend on their mood and emotional capacity. It suggests that emotional and mental readiness is key when engaging with recommendations. Knijnenburg et al. (2012) argue that users' satisfaction with personalized systems depends on situational relevance instead of static preferences. Furthermore, although the bias mapping showcased how Netflix's system reinforces biases, it is worth mentioning that the analysis reveals a more complex relationship between users and the interface. Multiple participants logged in to watch the same shows repeatedly, justifying these decisions based on their mood. Algorithms are not the only reason users tend to rewatch content, but they can surely reinforce it. As the diary study shows, the completion rates were moderately close across recommended and non-recommended content (non-recommended being slightly higher), indicating that users tend to engage less with Netflix's recommendations. The data support the argument that people tend to rewatch content, thus completing the same material repeatedly. All of which aligns with arguments on how visual hierarchy and UI play a central role in shaping users' interaction with recommendation systems (Iordache et al., 2024).

Reflecting on Bruns' (2019) critique of deterministic filter bubble narratives, although algorithmic structures matter in these conversations, users' selective behaviors are just as important. While Netflix's interface often reinforces familiar patterns, it occasionally includes features that nudge users toward exploration, such as feedback mechanisms or mood-based filters, which can help shift behavior beyond routine viewing. Still, as Van Es (2024) and Khoo (2022) argue, nudging mechanisms are often framed as "self-preferencing" systems that prioritize engagement over user autonomy, raising further ethical concerns at the boundary between personalization and manipulation.

This brings us to the third research question: *To what extent does Netflix's interface support or hinder users' sense of autonomy and agency in content discovery?* As it is essential to understand the complexity of the relationship between the user and the system, it appears that Netflix's UI offers numerous options. However, at times, they can seem more like an illusion. Here, Norman's (2013) terms of affordances and constraints can be used to interpret these "easy" choices that the interface offers. Although users can seek and navigate more broadly, the interviews and diary study still suggest that they do not take advantage of this freedom, which is not always due to users' lack of interest but sometimes to how the interface favors repetition and quick decisions. However, it raises the question of whether the experienced autonomy in engaging with the UI is legitimate or only a consequence of design choices. As many participants expressed, their decision is mostly made at the top of the homepage, and it usually reflects what they are familiar with or what is popular. It can suggest soft nudging, where the freedom of choice exists but is predefined by specific recommendation paths and the design's visual dominance. It is worth noticing how single users can showcase resistance against specific algorithmic recommendations, such as consciously searching for content.

Therefore, autonomy in the context of Netflix’s recommendation system is more about balancing the system’s effectiveness and users' ability to challenge that. Shneiderman and Plaisant (2005) emphasize the importance of systems that support trust and control through user-centered design. While these values are not inherently contradictory, the findings in this study suggest that Netflix’s interface tends to prioritize comfort in ways that can unintentionally limit exploration and critical reflection. It does not mean that the system is failing the user. However, it highlights a need for new thinking about how UI design can encourage meaningful participation and personalized influence without overwhelming the user with too many options. These findings raise ethical questions about the SVOD platform's responsibility to avoid manipulating design and support informed, autonomous user decisions. The boundaries between personalization and manipulation are worth studying for future reference.

This thesis, therefore, contributes to a growing body of literature that bridges system-level critique with practical user experience, emphasizing how emotional needs, structural design, and algorithmic personalization intersect in shaping user agency. As Knijnenburg et al. (2012) emphasize, user satisfaction arises from a contextual and affective fit – something that cannot be measured solely by click-through data. However, it requires a nuanced understanding of user motivations and limitations. Through the diary study and interviews, it becomes evident that users avoid deliberately searching for content, generally due to how "easy" it is to take the first presented options. The analysis reveals that, in practice, Netflix's interface restricts user agency by making automated options more accessible, faster, and visually prominent than their alternatives. At the same time, it should be acknowledged that users' non-usage of features, such as searching, is not only explainable by the interface's structure. Post-interviews show that it also depends on users' opinions, motivation, and awareness.

Therefore, on the surface, Netflix's interface appears open and user-controlled, but it does have subtle limitations and structures, which create a feeling of control without actual influence. Algorithms do not just limit users' autonomy; they also hinder it through interaction design, the absence of solid feedback mechanisms, and the emotional discomfort caused by repetition and unpredictability.

6.2 Design Implications

The final research question: *How can insights on this problem be used for designing more inclusive and transparent interfaces?* Focuses on how the findings from this study can contribute to making recommendation interfaces more inclusive and transparent. The thesis demonstrates that users' experiences with the system are shaped not only by algorithmic decisions but also by habits, visual dominance, and emotional needs. It all points to different design implications that can strengthen users' agency and feelings of control.

Firstly, Netflix should work to increase the visibility of the system's logic, clarifying why something is recommended. One specific example could be to add small texts under each recommendation, e.g. where users can be shown: “Recommended because you saw psychological thriller with female leads” or “People who saw the same show, also watched X”. Such clear explanations for each recommendation allow users to gain insights into why content is being recommended. Furthermore, it would allow them to evaluate and even reject the given recommendation. According to Vultureanu-Albişi & Bădică, explainable recommendations are crucial to user-centered design, as they strengthen users’ mental model of the system and enhance the experience of control, trust, and transparency (2022, p. 3).

Furthermore, it could be helpful in specific situations, such as when participants in this study express confusion about certain recommended content (e.g., Shadow Hunters, Paw

Patrol), which suggests a lack of effective feedback mechanisms. Another design, which presents the reason for the recommendation and directly allows users to say "no, thanks" to a specific genre, can enhance users' experiences of transparency and participation. Furthermore, the interview shows how features such as (liking or disliking) are easily missed or introduced too late. Participants request the ability to adjust their preferences actively. This would require a dynamic personalization design, where the platform gathers situational feedback based on mood, time, or genre. It can decrease the experience of automated repetition.

Secondly, the analysis reveals that the uniformity of the interface and its visual hierarchies lead users to engage in superficial navigation. Many people choose the first option they see or know, and rarely scroll through the options. The illusion of freedom of choice exists, but the alternative becomes invisible. Related studies also document that users experience frustration over not being able to customize their content rows, such as “Top 10,” and that a customizable layout can decrease choice overload and strengthen users' feelings of agency in the navigation process (Romero Meza & D’Urso, 2024; Iordache et al., 2024). A solution to this could be allowing a more flexible homepage with collapsible content rows that are user-defined.

Lastly, the recommendation system should strike a balance between comfort and challenge—the analysis reveals how users cope with comfortable options, especially in high-pressure situations. Therefore, systems should not necessarily pressure users to explore but mildly suggest mild and context-based nudges that invite variation—e.g., "Do you feel like watching something new today?" or the possibility to pause or edit repeated content temporarily.

This study demonstrates that more inclusive and transparent interfaces do not require the introduction of significant new features; instead, they necessitate improved timing, visibility, and user understanding in complex situations. By designing for both cognitive relief and emotional relevance, recommendation systems can support meaningful user journeys, where personalization is tailored to the individual, rather than predefined.

6.3 Methodological & Epistemological Reflections

Before concluding this thesis, it is essential to reflect on and discuss the findings regarding the overall methodology and epistemology. Moreover, it is essential to reiterate that this study does not aim to generalize these findings to all Netflix users, but rather to provide in-depth and contextual insights into how personalization can be implemented in practice. It is essential to acknowledge both the strengths and limitations of the approach used in this study.

Methodologically, this study's strengths lie in its triangulation of bias mapping, interviews, and a diary study, which give a multidimensional understanding of user experiences. The combination creates a reflexive progress, where participants' reflections and patterns are observed and interpreted. However, the diary study is dependent on users' self-deportation, which is fragile to selective memory or social desirability to be a certain way. For example, some users may have forgotten to log their viewing habits, which is unfortunate but not impossible. However, upon further reflection, having participants log the content they watched the day before is not the best solution. This logging should be closer to the viewed content for future reference, as on the same day. Since the diary study data is self-reported, some behavioral patterns can be underrepresented, especially if users forget to log shorter sessions (as the study highlights, watched content exceeds 50% completion). Future studies could combine the diary study method with a passive data collection method or a larger participating group to capture an even more nuanced picture. Even though daily reminders were utilized, the researcher's role as facilitator and data receiver may have affected the interaction and logging process - for example, as there was no subsequent follow-up on missing diary entries, which could have helped complete the data.

The pragmatic approach allowed for the combination of qualitative and quantitative data without committing to a single conceptual understanding of truth. Central to this approach is the perception that knowledge is experience-based, situated, and action-oriented, as something created in practice rather than existing independently (Morgan, 2014; Kaushik & Walsh, 2019). This means that the study aims not to uncover how Netflix works technically, but rather how the system is experienced, interpreted, and used in everyday media consumption. This epistemological position has enabled a user-centered approach to reading algorithmic systems. However, at the same time, it makes it difficult to state whether the user's experiences are due to actual algorithmic mechanisms or simply perception. For example, when a participant feels that the system repeats the same recommendation 17 times, it is analytically relevant, but epistemologically unclear whether this has happened. Thus, the validity of the thesis does not lie in reproducibility but in its ability to shed light on an experienced phenomenon through multiple simultaneous perspectives.

The study's focus on Netflix, without a comparative perspective on other platforms, has allowed for a deep dive into a single case. However, it also limits the extent to which the findings can be broadly interpreted. For example, some identified bias types or design choices could be general to the streaming industry. The same applies to users' emotional and cognitive navigation patterns – they are hardly unique to Netflix. However, it is precisely through the depth of the case and the practical focus that the thesis contributes important knowledge about experienced personalization and the user's interaction with algorithmic design.

Overall, the methodological design of the thesis has produced nuanced and reflective data. At the same time, the experiences from the fieldwork suggest the need for longer-term, more participatory methods, such as observation or user log analysis, to understand behavior over time with greater precision. The study shows that methodological reflection is not only about uncovering weaknesses but about understanding what one's approach allows one to see – and what it does not allow.

7. Conclusion

This thesis’s goal was to examine how algorithmic and user-driven biases in Netflix’s recommendation system affect content discovery, and how design solutions might improve user exploration and agency. In an era where digital media platforms are increasingly shaping our culture, it is essential to understand how personalization technologies not only offer convenience but also include structures that can limit content diversity and user influence. Through a qualitatively driven mixed-methods approach, this thesis sheds light on how a recommendation system’s visual structure, features, and algorithmic prioritizations contribute to specific biases. Positioning bias, popularity bias, promotional bias, and engagement bias play a crucial role in determining how content is perceived and experienced as attractive. These design choices are closely tied to users’ behavior, where habits, moods, mental capacity, and comfort-seeking influence how Netflix is used.

The results show that algorithms and users’ preferences create a reinforcing dynamic, where decisions tend to be repeated rather than challenged. The diary study reveals that most viewed content was repeated, typically series with multiple episodes from the same show, while new content was rarely sought out. The interviews confirm that users tend to stay at the top of the platform’s homepage and rarely use the search feature. However, it is not due to a lack of interest in new content, but the interface’s structure makes the passive path the easiest and most intuitive. Users are thus limited by design choices, creating an illusion of control and variation, without necessarily realizing this fully in practice.

Furthermore, it becomes clear that users seek emotional comfort through familiar and recognizable content. This is elaborated upon through pre- and post-interviews, which support the notion that content consumption is not only determined by rational preferences but also by emotional needs and situational factors. Netflix is therefore not just a source of entertainment, but a tool for emotional regulation – via background viewing, watching to relax, or mood-based choices. These situational factors are central in understanding why users tend to choose the first or easily accessible content instead of exploring something new.

However, although this thesis often focuses on algorithmic structures, it also demonstrates that user behavior can be equally responsible for reinforcing biased consumption patterns. Personalization is therefore not only something that is done to users, but can also emerge with them. This is a complex issue, not a static or absolute one, but a dynamic process shaped by mutual interplay between system design and user behavior. Recognizing this complexity enables us to move beyond simplistic critiques of filter bubbles or algorithmic manipulation and toward a more nuanced understanding of user behavior in this context.

Based on these insights, this thesis argues the need for more inclusive and transparent designs in recommendation systems. This does not imply significant technological changes, but rather relatively minor design adjustments, such as clarifying why content is recommended, enabling users to customize by removing disliked content, or allowing them to reevaluate their preferences continually. Moreover, other suggestions for small design interventions – such as proactive feedback timing, mood-based prompts, and collapsible recommendation rows – could help interrupt passive loops and foster more reflective viewing behavior. These ideas are more experiential rather than technical, but they aim to support users in making sense of and shaping their journeys.

Lastly, this thesis contributes to an emerging body of research. Designing a better recommendation system requires not only understanding how algorithms work, but also how users feel and decide in their everyday encounters with them. Future studies might expand on this work by exploring cross-platform comparisons or incorporating passive tracking data to complement self-reported patterns. Still, this thesis offers insights into how discovery and autonomy unfold within algorithmically curated platforms, revealing not only what shapes user experiences but also how design can create space for agency within them.

8. Reference List

Abdollahpouri, H., Mansoury, M., Burke, R., & Mobasher, B. (2019). The unfairness of popularity bias in recommendation. *RMSE Workshop Held in Conjunction with the 13th ACM Conference on Recommender Systems (RecSys)*, 1–7.

Ahn, Y., & Lin, Y.-R. (2024). Break out of a pigeonhole: A unified framework for examining miscalibration, bias, and stereotype in recommender systems. *ACM Transactions on Intelligent Systems and Technology, 15*(4), Article 73. https://doi.org/10.1145/3650044

Anwar, M. S., Schoenebeck, G., & Dhillon, P. S. (2024). Filter bubble or homogenization? Disentangling the long-term effects of recommendations on user consumption patterns. *Proceedings of the ACM Web Conference 2024 (WWW '24)*, 1–12.

Bartlett, R., & Milligan, C. (2015). *What is diary method?* Bloomsbury Academic. https://doi.org/10.5040/9781472572561

Bastos, M., Naranjo-Zolotov, M., & Aparício, M. (2024). Binge-watching uncovered: Examining the interplay of perceived usefulness, habit, and regret in continuous viewing. *Heliyon, 10*, e27848. https://doi.org/10.1016/j.heliyon.2024.e27848

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*(1),579-616. https://doi.org/10.1146/annurev.psych.54.101601.145030

Bourreau, M., & Gaudin, G. (2021). Streaming platform and strategic recommendation bias. *Journal of Economics & Management Strategy, 30*(1), 25–47. https://doi.org/10.1111/jems.12452

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2020). Conceptual and design thinking for thematic analysis. *Qualitative Psychology, 7*(1), 3–26. https://doi.org/10.1037/qup0000176

Bruns, A. (2019). *Are filter bubbles real?* Polity Press.

Bryman, A. (2012). *Social research methods* (pp. 384–430, 469–498, 565–583, 627–651). OUP Oxford.

Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., Bywaters, D., & Walker, K. (2020). Purposive sampling: Complex or simple? Research case examples. *Journal of Research in Nursing, 25*(8), 652–661. https://doi.org/10.1177/1744987120927206

Carter, S., & Mankoff, J. (2005). When participants do the capturing: The role of media in diary studies. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*, 899–908. https://doi.org/10.1145/1054972.1055098

Chun, C. A. (2016). The expression of posttraumatic stress symptoms in daily life: A review of experience sampling methodology and daily diary studies. *Journal of Psychopathology and Behavioral Assessment, 38*, 406–420. https://doi.org/10.1007/s10862-016-9540-3

Creswell, J. W. (2009). Chapter five: The introduction. In *Research design* (pp. 100–111). SAGE Publications.

Creswell, J. W., & Plano Clark, V. L. (2011). Core mixed methods designs. In *Designing and conducting mixed methods research* (2nd ed., pp. 51–100). SAGE Publications.

Cronin, P., Ryan, F., & Coughlan, M. (2008). Undertaking a literature review: A step-by-step approach. *British Journal of Nursing, 17*(1), 38–43. https://doi.org/10.12968/bjon.2008.17.1.28059

Consolvo, S., Harrison, B., Smith, I., & Chen, M. Y. (2017). Diary studies and experience sampling. In J. A. Rode, I. Smith, & A. K. Dey (Eds.), *Mobile user research* (pp. 71–84). Springer. https://doi.org/10.1007/978-3-031-02485-6_4

Dewey, J. (1938). *Logic: The theory of inquiry* (pp. 1–10, 139–158). Henry Holt & Company Int.

DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical Education, 40*, 314–321. https://doi.org/10.1111/j.1365-2929.2006.02418.x

Figà-Talamanca, G., & Arfini, S. (2022). Through the newsfeed glass: Rethinking filter bubbles and echo chambers. *Philosophy & Technology, 35*(1), 1–34.

Flaxman, S., Goel, S., & Rao, J. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly, 80*(S1), 1–26.

Flick, U. (2004). Triangulation in qualitative research. In U. Flick, E. von Kardorff, & I. Steinke (Eds.), *A companion to qualitative research* (pp. 143–190). SAGE Publications.

Flick, U. (2009). *An introduction to qualitative research*. SAGE Publications.

Giacomin, J. (2014). What is human centred design? *Design Journal, 17*(4), 606–623.

Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Methods of data collection in qualitative research: Interviews and focus groups. *British Dental Journal, 204*(6), 291–295. https://doi.org/10.1038/bdj.2008.192

Iordache, C., Loisen, J., & Van Audenhove, L. (2024). Integrating discoverability and prominence in video-on-demand consumption choices: A qualitative user study in Belgium. *The International Communication Gazette, 86*(7), 581–603. https://doi.org/10.1177/17480485231191898

Jannach, D., Naveed, S., & Jugovac, M. (2017). User control in recommender systems: Overview and interaction challenges. In *EC-Web 2016* (LNCS vol. 278, pp. 21–33). https://doi.org/10.1007/978-3-319-52196-9_2

Janssens, K. A. M., Bos, E. H., Rosmalen, J. G. M., Wichers, M. C., & Riese, H. (2018). A qualitative approach to guide choices for designing a diary study. *BMC Medical Research Methodology, 18*, 140. https://doi.org/10.1186/s12874-018-0579-6

Kaushik, V., & Walsh, C. A. (2019). Pragmatism as a research paradigm and its implications for social work research. *Social Sciences, 8*(9), 255. https://doi.org/10.3390/socsci8090255

Kelly, S., & Cordeiro, J. (2020). Three principles of pragmatism for research on organizational processes. *Methodological Innovations, 13*(2), 1–10. https://doi.org/10.1177/2059799120937242

Khoo, O. (2023). Picturing diversity: Netflix's inclusion strategy and the Netflix recommender algorithm (NRA). *Television & New Media, 24*(3), 281–297. https://doi.org/10.1177/15274764221102864

Knijnenburg, B. P., Bostandjiev, S., O'Donovan, J., & Kobsa, A. (2012). Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction, 22*(4–5), 441–504.

Konstan, J. A., & Terveen, L. G. (2021). Human-centered recommender systems: Origins, advances, challenges, and opportunities. *AI Magazine, 42*(3), 31–42. https://doi.org/10.1609/aimag.v42i3.2846

Lamkhede, S., & Das, S. (2019). Challenges in search on streaming services: Netflix case study. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1371–1374. https://doi.org/10.1145/3331184.3331440

Lamkhede, S. D., & Kofler, C. (2021). Recommendations and results organization in Netflix search. *Proceedings of the 15th ACM Conference on Recommender Systems*, 577–579. https://doi.org/10.1145/3460231.3474602

Lobato, R., Scarlata, A., & Wils, T. (2024). Video-on-demand catalog and interface analysis: The state of research methods. *Convergence, 30*(4), 1331–1347. https://doi.org/10.1177/13548565241261992

Mansoury, M., Duijvestijn, F., & Mourabet, I. (2023). Potential factors leading to popularity unfairness in recommender systems: A user-centered analysis. *BNAIC/BeNeLearn Joint International Scientific Conferences on AI and Machine Learning*, 1–16.

Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 1–21.

Morgan, D. L. (2014). Pragmatism as a paradigm for social research. *Qualitative Inquiry, 20*(8), 1045–1053. https://doi.org/10.1177/1077800413513733

Naeem, M., Ubaid, R., Saleem, H., & Faraz, N. A. (2023). A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. *The Qualitative Report, 28*(1), 276–293.

Norman, D. A. (2013). *The design of everyday things: Revised and expanded edition*. Basic Books.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1), 1–13. https://doi.org/10.1177/1609406917733847

Olma, K., Rizun, M., & Strzelecki, A. (2024). Exploring factors driving satisfaction from recommendation algorithms on Netflix. In S. Hassan, M. S. Khan, & W. W. Goh (Eds.), *Responsible AI and Analytics for an Ethical and Inclusive Digitized Society* (pp. 251–266). Springer. https://doi.org/10.1007/978-981-97-3699-7_13

Olorunfemi, D. (2023). Diary studies in research: More than a research method. *International Journal of Market Research, 66*(4), 410–427. https://doi.org/10.1177/14707853231222139

Oulasvirta, A., Hukkinen, J. P., & Schwartz, B. (2009). When more is less: The paradox of choice in search engine use. In *Proceedings of the 32nd international ACM SIGIR* *conference on Research and development in information retrieval* (pp. 516–523). ACM.

Patton, M. Q. (2015). *Qualitative research & evaluation methods* (4th ed., pp. 150–164, 243–251). SAGE Publications.

Randolph, J. J. (2009). A guide to writing the dissertation literature review. *Practical Assessment, Research & Evaluation, 14*(13), 1–13.

Romero Meza, L., & D'Urso, G. (2024). User's dilemma: A qualitative study on the influence of Netflix recommender systems on choice overload. *Psychological Studies, 69*(3), 349–367. https://doi.org/10.1007/s12646-024-00807-0

Ross, M. M., Rideout, E. M., & Carson, M. M. (1994). The use of the diary as a data collection technique. *Western Journal of Nursing Research, 16*(4), 414–425.

Ruslin, Mashuri, S., Rasak, M. S. A., Alhabsyi, F., & Syam, H. (2022). Semi-structured interview: A methodological reflection on the development of a qualitative research instrument in educational studies. *IOSR Journal of Research & Method in Education, 12*(1), 22–29. https://doi.org/10.9790/7388-1201052229

Schaffner, B., Stefanescu, A., Campili, O., & Chetty, M. (2023). Don't let Netflix drive the bus: User's sense of agency over time and content choice on Netflix. *Proceedings of the ACM on Human-Computer Interaction, 7*(CSCW1), Article 128, 1–32. https://doi.org/10.1145/3579604

Schaffner, B., Ulloa, Y., Sahni, R., Li, J., Kim Cohen, A., Messier, N., Gao, L., & Chetty, M. (2025). An experimental study of Netflix use and the effects of autoplay on watching behaviors. *Proceedings of the ACM on Human-Computer Interaction, 1*(CSCW Preprint), 1–22.

Shelton, S. A., & Coogler, C. H. (2025). Ma(r)king themes: Annotating as thematic analysis. In J. R. Wolgemuth, K. W. Guyotte, & S. A. Shelton (Eds.), *Expanding approaches to thematic analysis: Creative engagements with qualitative data*(pp. 27–48). Routledge. https://doi.org/10.4324/9781003389149

Shneiderman, B., & Plaisant, C. (2005). Guidelines, principles, and theories. In *Designing the user interface: Strategies for effective human-computer interaction* (pp. 60–106). Pearson/Addison Wesley.

Sun, R., Akella, A., Kong, R., Zhou, M., & Konstan, J. A. (2023). Interactive content diversity and user exploration in online movie recommenders: A field experiment. *arXiv preprint*, arXiv:2309.13296. https://doi.org/10.48550/arXiv.2309.13296

Sundar, S. S., & Marathe, S. S. (2010). Personalization versus customization: The importance of agency, privacy, and power usage. *Human Communication Research, 36*(3), 298–322.

Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research, 37*(3), 409–425.

Schwartz, B. (2004). *The paradox of choice: Why more is less*. HarperCollins. (pp. 1–5, 77–110, 235)

Stratton, S. J. (2024). Purposeful sampling: Advantages and pitfalls. *Prehospital and Disaster Medicine, 39*(2), 121–122. https://doi.org/10.1017/S1049023X24000281

van Es, K. (2024). Exploring Netflix myths: Towards more media industry studies and empirical research in studying video-on-demand. *Critical Studies in Television: The International Journal of Television Studies*, Advance online publication. https://doi.org/10.1177/17496020241297618

Vultureanu-Albişi, A., & Bădică, C. (2022). A survey on effects of adding explanations to recommender systems. *Concurrency and Computation: Practice and Experience, 34*(20), e6834. https://doi.org/10.1002/cpe.6834

Wolgemuth, J. R., Guyotte, K. W., & Shelton, S. A. (Eds.). (2025). *Expanding approaches to thematic analysis: Creative engagements with qualitative data* (pp. 11–47). Routledge. https://doi.org/10.4324/9781003389149