MASTER'S THESIS

Emotional Engagement in Narrative Medical Visualization: An Electrodermal Activity and Eye Tracking Study.

BEATRICE BUDICH

Department of Simulation and Graphics
Faculty of Computer Science
Otto von Guericke University Magdeburg

Supervisors:

Prof. Bernhard Preim, *University of Magdeburg, Germany*
Dr. habil. Monique Meuschke, *University of Magdeburg, Germany*
Assoc. Prof. Laura A. Garrison, *University of Bergen, Norway*

November 2025

MASTER'S THESIS

Emotional Engagement in Narrative Medical Visualization: An Electrodermal Activity and Eye Tracking Study.

BEATRICE BUDICH

November 2025

Otto von Guericke University Magdeburg

ABSTRACT

Data visualization graphically represents information, while Narrative Visualization integrates these representations into engaging narratives, making complex medical information accessible. This thesis explores how incorporating an individual character affects emotional engagement and comprehension in Narrative Medical Visualization. The study uses biometric measures, primarily electrodermal activity (EDA), to evaluate emotional responses during storytelling. Eye tracking and questionnaires supplement these measures to assess emotional and cognitive engagement. The findings suggest that incorporating individual characters in medical data stories enhances emotional engagement. The EDA analysis showed large effect sizes in maximum peak features. However, self-reported engagement often diverged from objective measures. The results revealed greater emotional engagement among women, indicating gender differences and the importance of participant balance. Participants experienced varied emotions and frequently reported curiosity and negative empathetic emotions. Character illustrations significantly affected emotional responses, supporting the hypothesis that they enhance engagement. Additionally, emotional priming effects were found in the within-subject dynamics. The order of the narratives influenced participants' levels of emotional engagement. Viewing a personalized perspective first tended to increase subsequent engagement, while viewing a neutral version first tended to decrease it in the subsequent story. Overall, there was no significant difference in recall between the character-oriented and neutral story. However, the results differed when the question was more closely related to the character's experience. This study underscores the potential of physiological measures combined with traditional methods to shape future research in Narrative Medical Visualization and emphasizes the need for greater emotional context in health-related storytelling.

ZUSAMMENFASSUNG

Datenvisualisierung stellt Informationen grafisch dar, während Narrative Visualisierung diese Darstellungen in ansprechende Erzählungen integriert und so komplexe medizinische Informationen zugänglich macht. Diese Thesis untersucht, wie sich die Einbindung eines individuellen Charakters auf das emotionale Engagement und das Verständnis in der Narrativen Medizinischen Visualisierung auswirkt. Die Studie verwendet biometrische Messungen, primär elektrodermale Aktivität (EDA), um emotionale Reaktionen während des Lesens von Datengeschichten zu evaluieren. Eye-Tracking und Fragebögen ergänzen diese Messungen, um das emotionale und kognitive Engagement zu interpretieren. Die Ergebnisse deuten darauf hin, dass die Einbindung individueller Charaktere in medizinische Datengeschichten das emotionale Engagement verstärkt. Die EDA-Analyse ergab große Effektstärken bei maximalen Spitzenmerkmalen. Das selbst berichtete Engagement wich jedoch häufig von den objektiven Messungen ab. Die Ergebnisse zeigten ein signifikant größeres emotionales Engagement bei Frauen, was auf geschlechtsspezifische Unterschiede und die Bedeutung einer ausgewogenen Teilnehmerzusammensetzung hinweist. Die Teilnehmer erlebten unterschiedliche Emotionen und berichteten häufig von Neugier und negativen empathischen Emotionen. Charakterillustrationen hatten einen signifikanten Einfluss auf emotionale Reaktionen, was die Hypothese stützt, dass sie das Engagement fördern. Darüber hinaus wurden emotionale Priming-Effekte in der Dynamik innerhalb der Probanden festgestellt. Die Reihenfolge der Geschichten beeinflusste das emotionale Engagement der Studienteilnehmer. Das Betrachten einer personalisierten Perspektive zu Beginn führte tendenziell zu einem höheren Engagement im weiteren Verlauf, während das Betrachten einer neutralen Version zu Beginn das Engagement in der nachfolgenden Geschichte tendenziell verringerte. Insgesamt gab es keinen signifikanten Unterschied in der Erinnerung zwischen der charakterorientierten und neutralen Geschichte. Die Ergebnisse unterschieden sich jedoch, wenn die Frage enger mit der Erfahrung des Charakters zusammenhing. Diese Studie unterstreicht das Potenzial physiologischer Messungen in Kombination mit traditionellen Methoden für die Gestaltung zukünftiger Forschung im Bereich der Narrativen Medizinischen Visualisierung und betont die Notwendigkeit eines stärkeren emotionalen Kontexts im gesundheitsbezogenen Storytelling.

ABSTRAKT

Datavisualisering representerer informasjon grafisk, mens Narrativ Visualisering integrerer disse representasjonene i engasjerende fortellinger, noe som gjør kompleks medisinsk informasjon tilgjengelig. Denne studien undersøker hvordan inkludering av en individuell karakter påvirker det emosjonelle engasjementet og forståelsen i Narrativ Medisinsk Visualisering. Studien bruker biometriske målinger, primært elektrodermal aktivitet (EDA), for å evaluere emosjonelle responser under fortelling. Øyesporing og spørreskjemaer supplerer disse målingene for å vurdere emosjonelt og kognitivt engasjement. Funnene indikerer at innlemming av individuelle karakterer i medisinske datahistorier øker det emosjonelle engasjementet. EDA-analysen viste store effektstørrelser i maksimale toppegenskaper. Imidlertid avveket selvrapportert engasjement ofte fra objektive målinger. Resultatene avdekket større emosjonelt engasjement blant kvinner, noe som indikerer kjønnsforskjeller og viktigheten av deltakerbalanse. Deltakerne opplevde varierte emosjoner og rapporterte ofte nysgjerrighet og negative empatiske følelser. Karakterillustrasjoner påvirket emosjonelle reaksjoner signifikant, noe som støtter hypotesen om at de øker engasjementet. I tillegg ble det funnet emosjonelle priming-effekter i dynamikken innenfor samme subjekt. Rekkefølgen på fortellingene påvirket deltakernes nivå av emosjonelt engasjement. Å se et personlig perspektiv først hadde en tendens til å øke det påfølgende engasjementet, mens å se en nøytral versjon først hadde en tendens til å redusere det i den påfølgende historien. Samlet sett var det ingen signifikant forskjell i hukommelsen mellom den karakterorienterte og den nøytrale historien. Resultatene var imidlertid forskjellige når spørsmålet var nærmere knyttet til karakterens opplevelse. Denne studien understreker potensialet i fysiologiske målinger kombinert med tradisjonelle metoder for å forme fremtidig forskning innen Narrativ Medisinsk Visualisering, og poengterer behovet for større emosjonell kontekst i helserelatert historiefortelling.

Although most providers now recognize HG as a serious condition with a biological basis, some may need to be better educated to understand that patients' [quality of life] can improve dramatically with adequate treatment, care, and understanding.

— Marlena Fejzo

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to Professor Bernhard Preim and Dr. Monique Meuschke for supervising this thesis and supporting me over the last three and a half years. Their close supervision of my visualization research projects has encouraged me to further develop my research skills.

I would also like to thank my supervisors at the University of Bergen. Professor Laura Ann Garrison inspired me to choose my thesis topic and provided close support throughout the development of the story design, preliminary evaluations, and conceptualization of the study design. I would like to thank the Folkehelseinstituttet for providing the MoBa dataset. Professor Marc Vaudel was available to answer my questions about the data and the application topic. He also enriched the data visualization concept with critical discussions about the appropriate metaphor. I would also like to thank Professors Helwig Hauser, Eduard Gröller, Jone Trovik, and Noeska Natasja Smit for their valuable feedback on the prototypes.

This study would not have been possible without the help of Professors Myra Spiliopoulou and Ernesto De Luca, who provided me with the study equipment and software. Additionally, I would like to thank Professor Myra Spiliopoulou for her advice on the study design to support the interpretation of the EDA data. I would also like to thank Anne Rother for sharing her experiences with research projects involving electrodermal activity technology.

I would like to thank Professor Bernhard A. Sabel and Dr. Marcel Martsch for sharing their expertise and insights on the psychological aspects of this study.

Finally, I would like to thank the former hyperemesis gravidarum patients for sharing their personal stories, as well as all the participants of this study who volunteered and provided valuable feedback.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

EDA Electrodermal Activity

HG Hyperemesis Gravidarum

1 INTRODUCTION

Data visualization focuses on representing data graphically and accurately, Narrative Visualization emphasizes the need to embed these data representations in a narrative to engage the general public.

Narrative Medical Visualization aims to make complex medical information more accessible, engaging and memorable for patients and broad audiences. By clearly presenting the impact of diseases and the benefits of certain treatment opportunities, Narrative Visualizations can motivate individuals to make more informed decisions about their health [72, 73].

Characters play a pivotal role in engaging audiences emotionally with complex medical information and influencing positive behavioral and lifestyle changes [75]. Characters are one of the most important components of a medical story among content, conflict and narrative framework [73]. Meuschke et al. (2022) argued that a data-driven disease story must contain at least one character which could be a human character or objects such as organs or diseases [72]. In a narrative, these characters can take different roles such as protagonist, antagonist, or supporting roles. A study has shown that personalizing a story by a patient character results in more focused attention and involvement with the story compared to not using a character [75].

Emotions are indispensable aspect in narrative experience. Previous research on narrative engagement indicates that emotional engagement is the most salient dimension for predicting effects of stories [10]. Segel & Heer (2010) have early motivated to investigate the effects of Narrative Visualizations on the reader's cognitive and emotional experience [97].

The main goal of this thesis is to investigate whether a data story with an individual human character elicits more emotional arousal. Another goal is to explore the potential of physiological measures to improve understanding of emotional engagement in Narrative Medical Visualization. Understanding how a user feels during experiencing a story would help to craft more engaging stories. The study results might influence the decision on which story version should be published on a support website for the education about a rare but severe pregnancy disease.

1.1 PROBLEM STATEMENT

Hyperemesis gravidarum

In this case study, I used methods of Narrative Medical Visualization to create a data story that should educate the general public about Hyperemesis Gravidarum (HG), a rare pregnancy condition characterized by severe, persistent nausea and vomiting. It has been reported to occur ranging from 0.3–10.8 % in different countries or ethnics [34]. Affected women often require hospitalization to receive hydration, nutrients, and medication. Due to the significant mental side effects, many women also require psychological support. This life-threatening condition due to disturbed metabolism and electrolyte balance [7, 30, 32] is often mistaken for normal pregnancy nausea accompanied by existing psychological problems [47]. Due to a lack of knowledge about the biological and genetic causes of the disease, and a failure to take advantage of treatment options, the health of both women and their unborn children is at risk [31, 33].

Understanding the biological and genetic causes of the disease, as well as feeling emotionally connected with affected women, could lead to increased support from family members, relatives, and friends. Marlena Fejzo, a former HG patient who experienced a lack of understanding for her condition, initiated the research on the causes of hyperemesis gravidarum. Because this research is relatively new, some doctors may still be unaware of the mechanisms and side symptoms of HG that affect the whole body and mind. Therefore, it is important to incorporate scientific findings into the story to increase credibility. Storytelling elements, on the other hand, can help to evoke empathy or sympathy. These feelings are often considered as a motivator for prosocial behavior [14].

Emotional engagement

Since emotional engagement is a precursor to sympathy or empathy, the main goal of the story is to evoke an emotional connection to the topic and those affected by it. Mittenentzwei et al. (2023) made a first attempt to investigate the emotional role of individual characters by comparing three stories where only the use of a character differed. The authors found that the patient character was more engaging. However, both the patient-centered story and the base condition story without a human character resulted in a very similar positive emotional response [75]. Therefore, the role of an individual character in medical data stories is not entirely clear from an emotional engagement perspective.

Evaluation of Narrative Visualization on a user's level of emotional engagement is a complex task which only a few previous works tried to deal with [69]. Most studies in the field of Narrative Visualization use survey methods collecting quantitative and qualitative feedback from participants' self-reports [2, 14, 38, 72, 75]. These studies have two major limitations: First, self-reports are highly subjective and lack reliability. People vary in their emotional awareness and ability to articulate or remember their feelings, leading to inconsistencies in self-reports [22]. Secondly, stories are designed along a tension arc, which consequently alters emotional engagement [37]. However, an evaluation via self-reporting during reading would disrupt emotional flow [96]. Most studies therefore only evaluate emotional responses or engagement for the entire story afterwards. The aspect of continuous change in emotional engagement throughout a narrative has not yet been investigated in Narrative Visualization.

1.2 Research Gaps

There is no study in Narrative Visualization that deeply focus on emotional engagement [1]. Furthermore, the role of characters in medical data-driven narratives is still underexplored.

The potential of objective measures such as psychophysiological data to quantify emotional engagement has not been explored in Narrative Visualization. There is also a gap in investigating the relationship between emotional responses and particular story elements. Additionally, the role of emotional priming has not been studied in the context of Narrative Visualization.

1.3 Research Questions

This study was driven by the following research questions:

RQ1: *"Does a fictitious, individual character arouse greater emotional engagement in the story's viewers for a medical condition compared to a general story with no individual human protagonist?"*

RQ2: *"Does experiencing a story in first-person perspective increase or decrease emotional engagement when the identical story is subsequently encountered in third-person perspective (and vice versa)?"*

RQ3: *"Does a fictitious, individual character increases understanding and memory in the story's viewers for a medical condition compared to a general story?"*

The first research question also explores how emotional responses are distributed throughout the story, and which visual elements (e. g., text, illustrations, and data visualization) elicit these responses.

1.4 Proposed Approach

With this thesis, I want to explore the potential of using biometric measures to directly access emotional responses depending on character usage. The proposed approach is an attempt to understand the effect of an individual character by comparing a personalized story from first-person perspective of a former patient to a more general story about hyperemsis gravidarum. The story concepts were implemented as web-based interactive slideshow prototypes.

To measure the different character approaches on emotional engagement, I propose a mixed method approach.

- The study's core method is Electrodermal Activity (EDA) where emotional responses are physically measured by change in skin conductance level over the entire course of the stories.
- Due to this method's limitation in measuring valence or specific types of emotions, I used a selection-based, multiple-choice questionnaire to obtain more granular results regarding emotional experiences.
- Eye-tracking was utilized to help better understand which elements in the story evoked emotional responses.
- A multiple-choice questionnaire was used to assess understanding and memory for the story's content. Additionally, an user engagement questionnaire and qualitative feedback were recorded to provide context for the data analysis. For example, the questionnaire asked which story the reader liked more and why.
- All participants viewed both stories in alternating order to explore how emotional priming with one story influences the other.

1.5 CONTRIBUTION

This study explores the emotional impact of characters in data stories using a novel methodology that measures emotional engagement through electrodermal activity, eye tracking, and a selection-based questionnaire. It enables the examination of emotional dynamics over time without disrupting participants' reading experience. The case study on hyperemesis gravidarum aims to foster empathy for those needing support and suggests that this evaluation method could be useful in other narratives requiring high emotional engagement. Furthermore, it connects emotional effects with understanding, memory, and content repetition. This thesis discusses emotional engagement from a psychological perspective and concludes with a definition of emotional engagement in the context of Narrative Medical Visualization. The reader will find details about the story development process and the purpose of the *"Martini Glass Structure for Character Design"* approach, based on the study's findings.

1.6 THESIS STRUCTURE

After introducing the topic, research problem, gaps, and questions, as well as providing an overview about the approach of this work and its contribution in Chapter 1, the thesis is divided into five parts:

PART 1: RELATED WORK The theoretical section positions my work within the existing body of knowledge and provides definitions relevant to my work. More specifically, it discusses how prior research on Narrative Visualization (Chapter 2), Narrative Medical Visualization (Chapter 3), and Psychology (Chapter 4) informed my work. In Chapter 5, I will describe various methods of measuring and evaluating emotional engagement, ranging from self-reports to physiological measures of electrodermal activity and eye tracking.

PART 2: METHODOLOGY First, I will present the prototype concept and the methods used to develop the story (Chapter 6), as well as the reasons behind the design choices. Second, I will provide details about the evaluation tools (Chapter 7), and thirdly, I will describe the sample, explain the experimental setup (Chapter 8), and the workflow for the data analysis (Chapter 9).

PART 3: RESULTS & DISCUSSION Chapter 10 presents the study's findings and evaluates the hypotheses. In Chapter 11, the results, practical implication, and limitations of this study are discussed.

PART 4: CONCLUSION & FUTURE WORK The final section (Chapter 12) provides a summary of the study and highlights its contribution to Narrative Visualization. The outlook section (Chapter 13) summarizes ideas for continuing research about the objective and other opportunities to evaluate the dataset. It also includes ideas that did not fit into the scope of this thesis.

PART 5: APPENDIX The questionnaires are listed in the appendix.

Part I

RELATED WORK

The purpose of this part is to provide the context of my work and to demonstrate how various research fields inspired my proposed approach. Chapter 2 introduces key concepts of Narrative Visualization. Chapter 3 presents the current state of research on Narrative Medical Visualization, which my work was mainly informed by. I also discuss the most significant studies in the context of character-driven medical data storytelling. Chapter 4 is a definition chapter about relevant terms of engagement, which is intended to establish a consistent vocabulary, particularly when discussing methods used to measure emotional engagement. This chapter also explains the concept of emotional priming. Chapter 5 discusses methods that have been used to examine emotional engagement in Narrative Medical Visualization and Narrative Visualization, as well as the insights and limitations of the presented studies. The final section refers to how emotional engagement is measured outside of visualization. It provides the reader with background information on electrodermal activity and eye tracking, as well as how these physiological measures are used to examine emotions.

The central questions guiding this part were:

1. Why is engagement important in the field of ***Narrative Visualization***?
2. What was found on a ***character's role*** in affecting emotional engagement?
3. What is the definition of ***engagement***?
4. What are the different ***categories*** of engagement?
5. How is emotional engagement ***measured***?
6. What are the ***challenges*** and ***limitations*** of previous studies?

2 NARRATIVE VISUALIZATION

Segel & Heer (2010) systematically analyzed practices in Data Journalism. The authors laid the groundwork for research on Narrative Visualization and in turn influenced data journalism. In their influential paper, *"Narrative Visualization: Telling Stories with Data"*, they explored storytelling principles in 58 existing data stories from online journalism. By classifying these examples into generalizable concepts, they provided a framework for analyzing narrative visualizations. For example, they identified seven ***genres*** of Narrative Visualization, including interactive slideshows and annotated charts. Furthermore, they classified design patterns in their design space consisting of ***genre***, ***visual narrative strategies*** (visual structuring, highlighting, and transition guidance), and ***narrative structure tactics*** (ordering, interactivity, and messaging).

Using this systematic approach, Segel & Heer proposed design strategies to improve data storytelling. For instance, they discovered that ***introductory tutorials*** are rarely used in data stories and they emphasized the advantage of this design strategy for introducing viewers to how to read the presented data visualization. They also found that effective narrative visualizations balance an ***author-driven*** and ***reader-driven*** approach by incorporating interaction and often employ ***multimodal cues***, such as voice-overs or annotations [97]. However, their analysis does not cover user studies to measure engagement or comprehension impacts.

The work by **Hullman & Diakopoulos (2011)** is also highly relevant to Narrative Visualization. In a similar manner as Segel & Heer (2010), they conducted a systematic qualitative analysis of 51 professionally produced narrative visualizations, primarily from international news outlets such as The New York Times. The authors intended to make visualization designers aware of the potential impact of their design choices on readers by highlighting certain risks, such as bias or manipulation. This ethical perspective is particularly important in the context of healthcare.

A key finding of the study was that visualizations are not purely objective. To promote ethical practices in data journalism, the authors classified design choices as ***visualization rhetoric techniques*** and explained how these techniques can influence users' perception and interpretation of visual information [49]. Limitations of their analysis is the lack of exchange with visualization creators, which could have provided insight into designers' choices. Furthermore, they did not compare users' interpretations with designers' intent in a controlled experiment.

Lee et al. (2015) is highly relevant to the field of Narrative Visualization. It addresses the need to define a visual data story and distinguish it from data visualization. They also provide insight into the visual data storytelling process, including the professions, tools, and channels involved. The authors aimed to investigate how raw data can be transformed into a coherent narrative structure to guide viewers' understanding. Their method involved a qualitative analysis of practices in data-driven journalism to develop a model for creating *"visually shared stories."*

A key contribution is the definition of a ***visual data story*** as a *"set of story pieces – that is, specific facts backed up by data."* [57] The authors introduced the term ***story piece*** to refer to a visualization that frames a single or few intended messages. They emphasized the role of narration, i. e., presenting these story pieces in a meaningful sequence or connection, highlighting or emphasizing the message to clearly communicate the intent.

Another contribution is their ***iterative storytelling process model*** with three steps that are influenced by external factors, such as the needs of the target audience and the context. These three steps are: (1) data exploration done by data analysts, (2) drafting story ideas and story pieces that make up a plot, usually done by scripters, and (3) creating and editing the story material into a presentation format, typically involving data visualization engineers and communication designers. The shared story is the final story made accessible to the audience. The authors suggested that a feedback loop from the audience would be promising [57].

The study provides an excellent overview of the terms, processes, and roles involved in creating visual data stories, demonstrating the interdisciplinary nature of Narrative Visualization. However, it does not offer comprehensive guidance on measuring the effectiveness of shared visual stories in terms of viewer understanding, engagement, or behavior change.

3 NARRATIVE MEDICAL VISUALIZATION

Communicating complex medical data to a non-expert audience is a relevant and challenging task in healthcare communication. Visualization techniques, such as infographics and interactive dashboards, are widely studied for their potential to improve comprehension by taking advantage of human capacities for processing visual information [9]. In the following, I will explore how Narrative Visualization enhances medical data communication.

Meuschke et al. (2021) introduced the field of Narrative Medical Visualization by their paper *"Towards Narrative Medical Visualization."* Through this foundational work, they addressed the challenge of conveying complex scientific and medical data to non-experts, such as patients and their relatives, via engaging storytelling. The authors' primary goal was to establish Narrative Medical Visualization as a novel approach by integrating narrative techniques into explanatory and exploratory visualization approaches.

Their methodology included a literature review, providing an overview of the challenges in Narrative Visualization, and a design-based study. To develop their data storytelling approach and provide a template for creating effective, data-driven disease stories, the authors employed a design-based research approach. They chose three diseases as use cases – liver cancer, brain aneurysm, and pelvic fracture – due to their relevance, e.g., a growing prevalence, broad occurrence across different age groups, high risk to overall human health and well-being, as well as a public interest in learning about these diseases.

The content was informed primarily by analyses of blogs and data from patients' ultrasound, MRI, or CT images. The authors presented a ***seven-stage template*** sequenced in a ***tension arc*** consisting of the following steps: (1) defining the disease, (2) explaining healthy anatomy, (3) listing and explaining typical symptoms, (4) diagnosing, (5) outline treatment options, (6) showing the disease prognosis, and (7) emphasizing prevention by communicating risk factors.

The key contributions of this work are revealing the potential of using Narrative Visualization in healthcare, the development of design guidelines, and suggesting future work directions [73]. A major limitation of this conceptual work is that the authors relied on expert feedback and did not conduct any empirical user studies.

Building on their previous work, **Meuschke et al. (2022)** examined the effectiveness of visual storytelling in improving ***understanding***, ***memorability***, and ***engagement*** with disease-related information. They conducted a questionnaire-based ***user study*** with 90 participants from diverse backgrounds. One half of the study participants (focus group) viewed an interactive, web-based story based on their seven-stage template about liver cancer. The other half (control group) read the same content in a traditional blog format. Given the increasing use of smartphones, the stories were offered in desktop and smartphone display formats.

Key findings indicate that the data story approach help users better comprehend and retain medical information. The authors investigated ***affective involvement*** with three questions. The results showed that participants were motivated to continue reading the story after the patient's introduction. However, the authors stated that users could feel more involved if the patient case was incorporated more deeply over the course of the story, thus establishing a stronger ***emotional connection*** through a ***relatable patient character***.

Participants stated that the interactive 3D visualizations were particularly engaging. The authors also highlighted the importance of concrete ***interaction instructions***, avoiding ***technical terms***, and ***personalizing*** stories. For example, offering reader-driven techniques could help prevent boredom when content is already known or does not meet personal interests. The authors mentioned that displaying their data story on a smartphone led to accessibility issues in terms of readability for some users. They argued that different devices and screen resolutions must be carefully considered during the design process [72].

One limitation of the study is that it could not determine which narrative elements or combinations of techniques improve memorability. Additionally, the focus and control groups were not compared with regard to engagement or willingness to change their lifestyle. This makes it difficult to reveal the clear difference between the data storytelling approach and the traditional presentation format. Other open question are which emotions were triggered and how story engagement develops over time.

Meuschke et al. (2022) stated that medical data stories require at least one ***character***. They discussed how a more patient-centered approach to storytelling could enhance ***sympathy*** and ***emotional engagement*** [72]. However, although their story included a character, they did not evaluate the effects of a character on story experience.

Mittenentzwei et al. (2023), on the other hand, focused specifically on investigating the ***role of a human character*** in medical data stories. In their paper, *"Do Disease Stories Need a Hero? Effects of Human Protagonists on a Narrative Visualization about Cerebral Small Vessel Disease,"* the authors examined the influences of a character on ***user experience***, such as engagement, emotional response, and self-referencing.

The method involved creating three interactive data story versions about Cerebral Small Vessel Disease. The content and overall style remained consistent; only the presence of a character varied.
In *version 1*, the ***patient*** was shown as the protagonist, while *version 2* told the story from the perspective of a ***physician***. Both story versions based on ***Joseph Campbell's Hero's Journey***. *Version 3* functioned as a basis version that did not incorporate a human protagonist.

These stories were evaluated in an online study in which 30 participants self-reported their level of engagement. To estimate ***emotional response***, participants were asked after completing the story to rate their "feeling" (valence) using a five-point scale ranging from extremely negative to extremely positive (see Chapter 4 for definitions of emotion and feeling). They also investigated perceived credibility and whether presenting disease risk factors leads to a motivation of preventive behavior. For this longer lasting engagement, they used the term "emotional flow."

The results showed higher engagement with stories that include a human protagonist where the patient-centered story outperformed the physician-centered version. On average, participants reported a positive emotional responses to both the patient-centered and the general story. Their emotional response to the physician-centered story was neutral. The authors could not identify a correlation between motivation to change lifestyle behaviors and character design. Self-identification and perceived credibility were slightly higher for the story that included a physician. In addition, participants were more willing to spend time with a story when it was told from the perspective of a human protagonist [75].

One limitation of this study is that participants, who were mostly students, did not represent the main target group. This means the average age was significantly below the peak incidence of the disease. This could explain the low identification score with the patient character.

Due to study termination the three participant groups were unbalanced, i.e., 23 % (patient story), 33 % (physician story), and 43 % (base condition story) [75]. This imbalance and a small sample size potentially leads to bias and reduces the study's statistical significance. Additionally, the authors noted that feedback from participants who dropped out of the study could have been valuable. Another limitation is that the study did not relate engagement to memory and understanding outcomes based on character usage. Furthermore, emotional response was evaluated by only one question for the entire story, which was neutral for both the patient-centered and the general story. This method might not provide detailed insights into emotional engagement. This work is the closest to my own study. However, my study will compare the concept of a patient story with that of a base condition story. It will also address several of the listed limitations by employing a more granular evaluation methodology.

Although characters are essential for humanizing complex medical data, enhancing audience engagement, and promoting behavioral changes – the traditional design process demands significant time and artistic skills. Transforming anonymous, epidemiological patient data from cohort studies, such as SHIP (Study of Health in Pomerenia [109]), into visual representations of authentic characters poses a challenge. While Meuschke et al. (2022) and Mittenentzwei et al. (2023) used static photographs to present their characters, Budich et al. (2023) proposed a novel, ***AI-generated character*** design approach [16] that was evaluated by Mittenentzwei et al. (2024) [74].

Figure 3: An AI-generated image of a patient character for a medical data story about fatty liver (Budich et al., 2023).

In *"Reflections on AI-Assisted Character Design for Data-Driven Medical Stories,"* **Budich et al. (2023)** aimed to explore the potential of AI tools to assist in character design and inspire further AI integration in storytelling. The method included involved creating an ***iterative, AI-assisted pipeline*** that uses ***Stable Diffusion*** – a text-to-image deep learning model – to help design informative, relatable, and engaging characters. This pipeline uses prompt engineering to create images and videos of characters that reflect demographic, disease-related, and lifestyle attributes, as well as emotions. ***Image-to-Image Translation*** is used to illustrate how characters evolve over time based on their data, facilitating translation that aligns with the progression of the narrative. Based on data from the SHIP cohort study [109] and to demonstrate the pipeline's output, the patient character *"Emma Winter"*, who has been diagnosed with fatty liver was created (Figure 3). One limitation the study encountered was the imperfect control over visual outputs of AI-generation. For example, using Image-to-Image Translation can result in inconsistent character variations. The paper discussed the importance of understanding audience perceptions and emotions in relation to AI-generated media, yet it did not evaluate the character design approach in a controlled study [16].

In their paper, *"AI-Assisted Character Design in Medical Storytelling with Stable Diffusion,"* **Mittenentzwei et al. (2024)** discussed and adapted the prompt scheme, including the use of ***keyword weights***. They also included the ***inpaint approach*** to better control over which parts of an image are altered. For instance, highlighting only the face to change facial expression improves character identity.

A key contribution of the paper is the authors' quantitative, crowd-sourced evaluation of the AI characters created by the semi-automated pipeline. The authors analyzed *authenticity* (recognizability of data attributes), *resemblance* (consistency of character identity), and investigated participants' *preferences* for static versus animated outputs. For the user study, the authors created two characters based on data from the epidemiological cohort study (SHIP) [109].

Key findings indicate that participants successfully identified many data aspects in the generated characters, proving their informativity. However, identifying emotions seemed to be more challenging. While some attributes, such as body weight, were easier to identify, others, such as occupation, were less clear. Study participants stated that they associate eye bags, wrinkles, and unhealthy skin color with visual signs of poor health, particularly connected to a poor diet.

Regarding resemblance, the authors found that character identity maintains best when only one attribute is changed, e.g., emotion. However, changing more than one attribute can lead to difficulties in generating consistent characters. Depending on how the AI model is trained, some attributes, e.g., overweight, can be linked with particular facial features such as an upturned nose, which can change the character's appearance unintentionally.

The authors also compared static AI images with AI characters in video format, examining emotional connection and engagement. Participants stated that they could relate better to the character by static images. On the other hand, participants also mentioned that they gained a better sense of the characters' physical characteristics via video. When asked for additional feedback, some participants reported that the animated characters felt more „dynamic" and „lifelike." However, static images were generally preferred over animations, despite the latter's potential. Reasons for disliking AI-generative videos included unnatural movements, an unrealistic appearance, and a monotonous voice.

AI-generated animations can fall into the so-called ***"Uncanny valley"***. This refers to the unsettling, even repulsive feeling that people experience when they encounter humanoid robots or animations that are very similar to humans but not completely perfect [74]. Given the rapid evolution of AI technology and improved video quality, this format could become more popular in the near future. In addition, as people become more familiar with AI-generated content in their daily lives, they also may accept this kind of imagery as just another form of visual media, such as paintings or photographs.

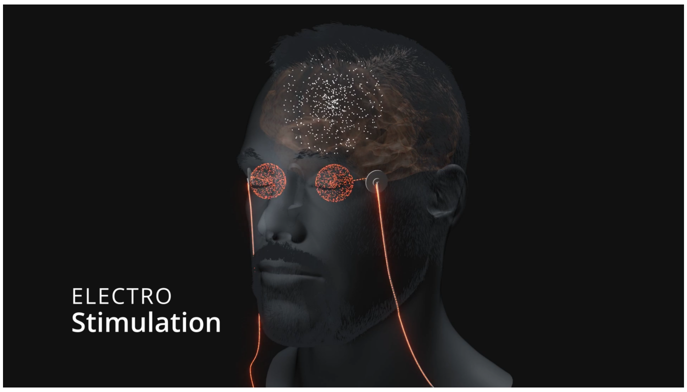


Figure 3.1: Character design for an animation explaining electrostimulation therapy for low vision [*created by the author*].

Budich et al. (2025) introduced a different ***character design*** approach to Narrative Medical Visualization which has a more personal accent. This approach combines 3D modeling, animation, and video footage of real patients telling their story in an interview. An example of the character design is shown in Figure 3.1. The study encompassed the development of a patient-centered medical communication video to visualize a complex therapy concept for vision restoration to improve patient understanding, adherence, and emotional engagement in the treatment process. A preliminary user study with patients showed a tendency for identification with the character. However, the sample size was too small to draw clear conclusions [17].

The presented articles reveal a demand for analyzing the role of human characters in data stories in more depth. Previous studies did not focus on the emotional aspect, only used self-reported, unsynchronized methods, and neglected the change of emotional engagement over time. In the following, I will describe emotional engagement in the context of learning and healthcare narratives.

4 ENGAGEMENT

The previous chapter showed that there are a variety of terms related to the topics of affect and engagement. In this chapter, I will define the key terms and use these consistently throughout the thesis.

4.1 SIGNIFICANCE OF ENGAGEMENT

Cavanagh et al. (2019) stated that engagement is the *"necessary first step for learning."* [20] Several authors discussed the effects that occur during the engagement phase, thereby emphasizing the importance of creating engaging visualizations:

- Liem et al. (2020) found that engagement improves the ***effectiveness of communicating*** information [63].
- Engagement leads to ***pleasurable emotions*** [2]. These emotions are directly connected to ***user satisfaction*** and the intent to use a product [98].
- Busselle & Bilandzic (2009) stated that the outcomes of engagement are ***enjoyment***, ***persuasion***, and ***social reality construction***. Emotional perspective-taking processes foster ***identification*** or ***empathy*** [19].
- Liao & Wang (2023) found that healthcare narratives from literature and visual arts cultivate ***empathetic connections to patients*** and ***reflective thinking*** skills of health professional students [59].
- According to Boy et al. (2015), engagement is associated with the ***empowering*** users to gain knowledge from the presented material and discover their own insights [13].
- Studies have shown that emotional engagement significantly impacts ***motivation to learn*** a subject [78, 80].
- In health interventions and disease prevention, it is crucial to achieve maintenance of engagement after using an informative application. Yardley et al. (2018) stated that long-term engagement is *"a process of achieving **positive cognitive**, **emotional**, **behavioral**, and **physiologic change**."* [113]

In conclusion, engagement fosters motivation, learning, persuasion, enjoyment, user satisfaction, attitudinal changes, behavioral actions or changes, social reality construction, and other desired outcomes.

4.2 DEFINING ENGAGEMENT

Engagement can be used to describe and measure the experience, effectiveness, and impact of a data-driven story [70, 79]. O'Brien & Toms (2009) emphasized the importance of defining engagement in order to be able to measure, understand, and improve user experience [79]. However, engagement is a complex, interdisciplinary topic, making it difficult to identify a single definition [70].

The most abstract dictionary definition of engagement is as follows: *"the fact of being involved with something"* [25]

Another dictionary provides synonyms or closely similar words for feeling ***‚engaged'*** that are *‚involved'*, *‚interested'*, *‚fascinated'*, *‚occupied'*, *‚intrigued,'* and *‚immersed.'* Antonyms are *‚bored'* and *‚wearied.'* [71]

Since engagement is used in many different contexts, such as marriage, occupation or military [25], we must first define the context. In Narrative Visualization, data is presented within a story, blending traditional storytelling with interactive data visualizations to create an engaging and informative experience. In this context engagement relates to *user experience* and *user engagement* – concepts frequently used in Information Visualization and Human-computer Interaction – as well as *narrative engagement*.

- **Bilandzic et al. (2019)** defined ***narrative engagement*** as the *"experience of being deeply immersed in a story and connecting to its plot and characters."* [10] According to the authors, this experience is a central mechanism for narrative persuasion. Busselle & Bilandzic (2008) developed a scale to measure narrative engagement, including understanding, emotional engagement, attentional focus, and narrative presence [18].

User experience and user engagement can be considered as distinct concepts – which is more common in the industry, including product design, marketing, and digital platforms.

- ***User experience*** is considered to occur at the micro-level of practical user interaction and reflects the quality of interaction, such as intuitive, efficient usability and positive emotional experiences [46].
- ***User engagement***, on the other hand, is characterized by an established emotional connection (***affection***) between the user and a system, based on repeated interactions and high involvement with the system [115].

In contrast, in the literature ***user engagement*** is defined as a ***category of user experience*** that consists of an emotional, cognitive, and behavioral dimension that occurs while using a system [78].

Attfield et al. (2011) added a time dimension to the definition by stating that engagement is an *"emotional, cognitive and behavioral connection that exists, at any point in time, and possibly over time, between a user and a resource."* [3] This concept suggests that user engagement can probably occur at any level, ranging from a single or first session (micro-level) to long-term relationship across multiple sessions or a lasting effect (macro-level). This ***macro-level engagement*** is crucial for medical and health education where the viewer, including patients, should adapt their habits towards a healthier lifestyle [113].

O'Brien & Toms (2008) indicated four distinct stages that can occur in ***microlevel engagement***: (1) point of engagement, (2) period of sustained engagement, (3) disengagement, and (4) reengagement. The absence of engagement is called nonengagement. Common reasons why users do not engage with a system include technical issues, fatal interruptions, delays, cognitive overload due to overwhelming displays or tasks [78].

(1) ***Point of engagement***: A user can be intrinsically motivated to engage when they have a particular goal in mind or for social reasons. For instance, a user may become interested due to a friend's recommendation. Engagement can also be extrinsically motivated, i.e., initiated by the medium itself when something resonates with user's interests. The authors found that users can also be drawn in by visual aesthetics, stories, the personality of a person or character, and humor.

(2) ***Period of sustained engagement***: At this stage, the prerequisites are focused attention and interest in the task or application itself. The main reasons for sustained engagement are the novelty of the experience, the level of interest, an appropriate level of challenge, feedback, and user control. Low task urgency can lead to disengagement, while high urgency can maintain engagement.

    The authors stated that this phase is typically accompanied by positive affect, such as enjoyment, fun, as well as physiological arousal. However, negative emotions can also foster sustained engagement. For example, a reader may experience negative emotions while deeply immersed in a sad story.

(3) ***Disengagement***: There are internal and external factors that cause distraction from an application or task. Internal causes may include thoughts about the urgency of other tasks, mind wandering, or long-lasting tasks that result into loss of attention. External factors include technical and usability issues that interrupt

engagement, as well as the lack of or excessive challenge. The authors stated that this phase can be accompanied by a negative affect, such as confusion, overload, frustration, or boredom.

(4) ***Re-engagement***: Users return to an application at a later point in time because they had positive experiences with it previously. Re-engagement motivations include the experience of having fun, feeling rewarded, receiving incentives, and learning or discovering something new [78].

In summary, creating engaging narrative visualizations is a complex task that combines user experience, user engagement with narrative engagement. Disengagement can occur when technical, medium, or story-related aspects distract or confuse the viewer. To avoid disengagement, all levels of engagement must be considered. Therefore, creating reliable and engaging digital visual narratives requires a structured approach [74] or templates [72] as well as collaboration between various professions [57], including data analysts, data visualization engineers, journalists, communication designers, psychologists, and experts on the subject of the story.

4.3 DEFINING EMOTIONAL ENGAGEMENT

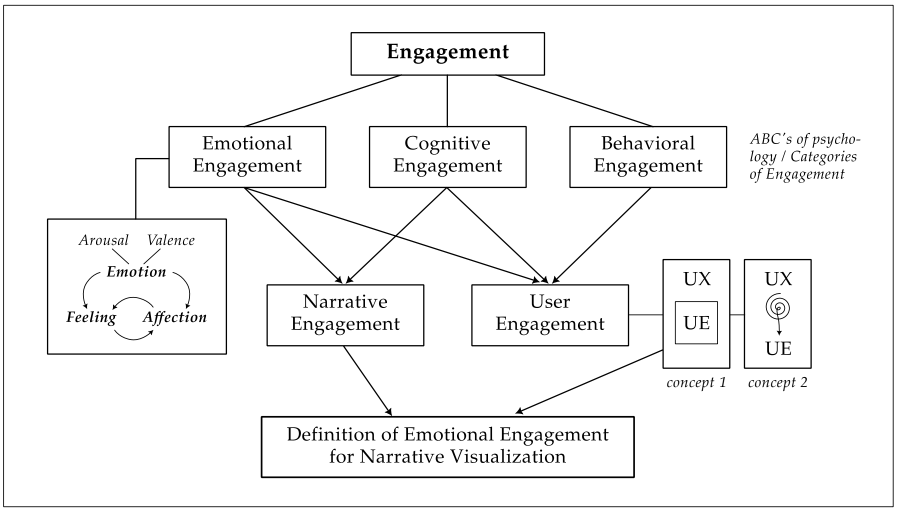


Figure 4.1: Connecting different concepts towards a definition of emotional engagement for Narrative Visualization [*created by the author*].

Broader definitions of engagement refer to the ***‚ABCs of psychology,'*** i. e., emotions, behavior, and cognition. These components have been described as fundamental and interconnected aspects of human experience [58]. According to ***Psychological Engagement Theory***, psychological engagement is defined as *"a state where individuals are emotionally and cognitively connected to a task, activity, or relationship [...] for achieving optimal outcomes."* [91]

In an educational context, emotional engagement is studied for its significant impact on learning processes, such as cognitive and behavioral engagement [65]. While emotional engagement is commonly defined as the *"experience of positively valenced and energy-mobilizing emotions (e. g., interest) during a learning activity"* [93], this definition does not represent the wide range of emotional experiences with narratives, including negatively valenced emotions [19, 55].

Busselle & Bilandzic (2009) proposed the following definition in a narrative context: ***"Emotional engagement*** *concerns emotions viewers have with respect to characters, either feeling the characters' emotions (empathy), or feeling for them (sympathy) [. . .], and appears specific to the emotional arousal component of narrative engagement."* [19]

While ***affection*** means a *"feeling of attachment, e. g., fondness, tenderness, and liking,"* [87] ***affect*** (adj. *affective*), on the other hand, is *"any experience of feeling or emotion, ranging from suffering (most negative state) to elation (most positive state), from the simplest to the most complex sensations of feeling."* [86] According to this definition, affect includes both feelings and emotions. However, it is important to distinguish between the two terms.

Feelings are purely mental and the subjective evaluation of a raw, instinctive emotion [89]. They occur when the brain's cognitive areas, e. g., the prefrontal cortex, process and label an emotion. These interpretations depend on memories, context, and personal experiences, leading to potentially different interpretations of the same emotion among individuals. Furthermore, people vary in their awareness of internal affective processes and their ability to articulate their feelings, leading to inconsistencies in self-reports [22, 96].

Therefore, feelings are difficult to evaluate via a study. ***Emotions***, on the other hand, *"are designed to engage with the world"* [88] – making them more accessible to evaluation. Emotions are described as a *"complex reaction pattern, involving experiential, behavioral, and physiological elements, by which an individual attempts to deal with a personally significant matter or event."* [88] These reactions, or physiological responses to a stimulus, are driven by the limbic system of the brain and are therefore instinctive and often automatic. Emotions can be visible as facial expression [28], or measured physically by biometric responses, such as sweating or increased heart rate [56]. The differences between emotions and feelings are briefly summarized in Table 4.1.

	EMOTIONS	FEELINGS
Main difference	Physiological, unconscious response to a stimulus	Personal, conscious interpretation of an emotional response to a stimulus
Origin	**Physiological** *(limbic system, physical changes in the body, e.g., increased heart rate, sweating, adrenaline release)*	**Cognitive** *(prefrontal cortex, influenced by memories, context, and individual perspectives)*
Consciousness	**Instinctive** *(often automatic, arise unconsciously, out of the person's control, rooted in biology and evolution)*	**Conscious** *(require awareness and mental reflection)*
Duration	**Short-lived** *(seconds to minutes)*	**Longer-lasting** *(hours, days, or longer)*
Universality	**Universal** *(basic emotions: happiness, sadness, fear, anger, surprise, disgust, are considered across cultural differences)*	**Individualized** *(vary greatly between individuals and cultures making them more individual and varied)*

Table 4.1: Differences between emotions and feelings [28, 85].

Lang et al. (1993) proposed a two-dimensional model to categorize emotions. Figure 4.2 depicts these dimensions in a Cartesian diagram. The x-axis represents ***emotional arousal***, which ranges from low (lethargic / calming) to high (alert / arousing). The y-axis refers to ***emotional valence*** ranging from negative (unpleasant) to positive (pleasant) with different levels. The emotion *fine* is defined as neutral and of low intensity [23, 56].

The twenty emotion types used in the evaluation are placed along these two dimensions. As shown in Figure 4.2, there is a denser distribution along a triangular shape encompassing uplifting and fulfilling emotions on the top, while the lower triangle contains negative emotions with increasing discomfort and distress.

However, emotions cannot be divided into distinct categories; rather, there are gradations between emotions that are closely related. In addition, emotions can have different degrees of intensity within the dimensions. For example, the emotion of depression is often attributed a low intensity. However, in combination with high psychological stress this emotion can reach a moderate level of arousal. Similarly, the emotion of misery lies between moderate and high arousal depending on the person's experience. Therefore, the figure should only provide an overview of the model and the emotional space used in the study.

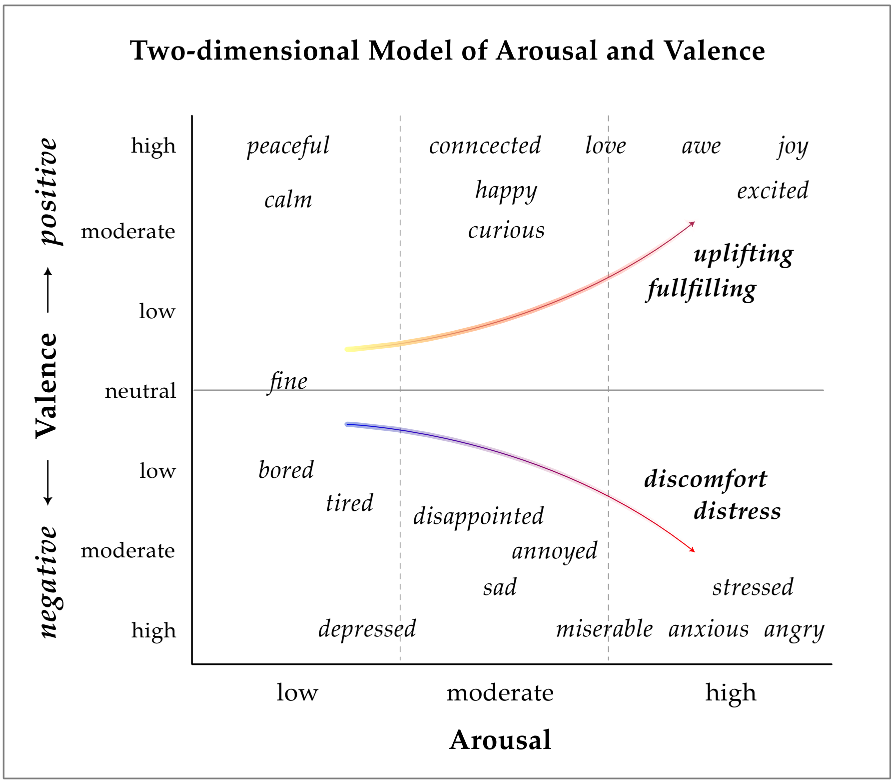


Figure 4.2: The two-dimensional model of arousal and valence based on Lang (1993), containing all 20 emotion types used in this study [*created by the author*].

The following is an attempt to define emotional engagement in the context of Narrative Medical Visualization, based on the aforementioned concepts and terms of engagement.

In the context of Narrative Medical Visualization, **emotional engagement** refers to the degree to which a viewer experiences an emotional response, connection, or investment with a medical story, including its plot, characters, story pieces, visualizations, and visual elements. This involves cognitive processes (understanding) that can result in feelings and affection, such as empathy (sharing the characters' emotions), sympathy (feeling for the character), curiosity, and motivation. These feelings make the content personally relevant, memorable, and impactful. They deepen understanding of a condition or stimulate behavioral responses, such as positive lifestyle changes.

4.4 EMOTIONAL PRIMING

Since the third research question addresses the within-subject dynamics of emotional engagement, this section will briefly describe the theoretical background of emotional priming.

Emotional or affective priming is a psychological concept that describes the emotional interaction between previous and subsequent stimuli. Priming occurs when the perception of a current stimulus is influenced by a previous one. Experiments on affective priming have found that participants who see faces expressing negative emotions (e. g., sadness or anger) are more likely to interpret a subsequent presentation of a neutral face as negative. The same phenomenon was found with happy faces, which caused participants to interpret a neutral face as positive. The priming effect occurs when there is a relationship between the prime and target stimuli and works most effectively when the two stimuli are in the same modality, e. g., photographs [105].

The emotional priming effect is characterized by its unconscious nature. Furthermore, emotional priming influences cognitive processes, such as semantic processing, recall, judgment, and decision-making [100]. Studies found that emotional cues attract more attention than neutral stimuli and that these emotionally charged stimuli positively impact attention and recall [82, 83]. However, the extent of this effect can differ depending on the type of emotional stimulation. While positive emotional stimuli are more likely to enhance memory performance, Windmann & Kutas (2001) demonstrated that negative emotions can lower the threshold for recognizing relevant information and distort memories [112].

Emotional priming have also been researched in the context of storytelling. Rao et al. (2019) conducted a within-subjects study to investigate the effects of emotional priming on active storytelling. Before participants were instructed to create a story, they were emotionally primed by two: first, participants viewed an emotion displayed on a screen; second, participants were instructed to express an emotion through their facial expressions. These conditions were then compared to a condition without a prime. They found that both emotional priming conditions led to richer storytelling, as measured by the variety of ideas and language used [92]. Storytellers can use emotional priming to influence their audience's emotional state. For instance, displaying emotional facial expressions at the beginning of the story can generate more emotionally rich narratives.

5 MEASURING EMOTIONAL ENGAGEMENT

5.1 METHODS IN NARRATIVE MEDICAL VISUALIZATION

Self-report methods, User Engagement Scales

Studies in the field of Narrative Medical Visualization commonly using self-report methods, in particular questionnaire-based evaluations with Likert scales. The questionnaires contain items from user engagement scales [2, 79].

Likert scales of affective involvement

Based on Amini et al. (2018), **Meuschke et al. (2022)** evaluated affective involvement, including enjoyment with the following questions on 5-point Likert scales [2, 72]:

- *"The beginning of the story motivated me to continue."*
- *"The story triggered emotions in me."*
- *"I felt actively involved in the story."*
- *"The story was entertaining."*
- *"The story was boring."*
- *"I would recommend my friends to watch the story."*

Mittenentzwei et al. (2023) used the scale developed by O'Brien & Toms (2009) [79] to evaluate involvement with respect to different character roles:

- *"I felt involved in the story."*
- *"Viewing the story was fun."*

The authors also asked study participants to rate their "emotional response" to the story from *"extremely negative"* to *"extremely positive"* on a 5-point Likert scale [75].

To compare different character representations, **Mittenentzwei et al. (2024)** employed specific questions:

Single choice questions

- *"Which format made you feel more emotionally connected to the character?"*

Likert scales of emotional states

- *"What are your feelings about the person in the video?"* providing a range from, e. g., *"unemotional"* to *"hair-rising"* [74].

Results of these studies are discussed in more detail in Chapter 3.

5.2 METHODS IN NARRATIVE VISUALIZATION

Airaldi et al. (2025) recently conducted a literature review of 116 studies about evaluation methods in Narrative Visualization and best practices for designing data visualizations. They found that narrative visualizations are challenging to evaluate due to their human-centered nature, and their aim to access at the interaction between cognition and emotions. Traditional methods with quantitative metrics, i. e., interaction patterns, number of interactions, or time spent [13], often fail to evaluate emotional responses to narrative elements or narrative clarity. Qualitative methods on the other hand, such as "think-aloud" protocols, "walk-throughs," and interviews give more insights into users' comprehension and experience, but are also time-consuming to conduct and analyze. Therefore, most user-centric studies on Narrative Visualization use questionnaires with Likert scales [1, 75].

The following highlights works in Narrative Visualization that have evaluated affect and the methods used for this purpose.

Garretón et al. (2024) examined attitudinal changes through data stories. The authors experimentally studied three effect dimensions of data visualization: persuasiveness, emotions, and memory. They related persuasion and emotional engagement by stating that persuasion can result from analytical reasoning or emotional appeal. They also refer to the Elaboration Likelihood Model [84], which posits that the personal relevance of the topic influences persuasion.

In their case study, the authors developed a story about potential causes of the Chilean water crisis that are not discussed in the mainstream media. They developed three versions of the story with different visual representations. Version 1 uses only illustrations, version 2 uses only data visualizations, and version 3 combines illustrations and data visualizations from versions 1 and 2.

Positive and Negative Affect Schedule (PANAS), 5-point Likert scales

To measure whether the emotional impact is greater for stories with data visualizations, Garretón et al. used a short version of the original Positive and Negative Affect Schedule (PANAS [111]). The I-PANAS-SF scale was chosen for its brevity, containing ten items, i. e., ten emotions (*upset, hostile, alert, ashamed, inspired, nervous, determined, attentive, afraid*, and *active*).

Measuring emotional states at different time points

Participants were asked to rate their emotional state with respect to the topic before and after reading the story, as well as one week later. The results showed that the differences in affect between the time points were greatest with the story that only uses data visualization [38]. This method allows to estimate changes in affect through data storytelling, though it does not consider emotional changes during the storytelling experience, and it has limitations due to delay and self-reporting. Garretón et al. also conducted a qualitative analysis of free-form responses regarding the usefulness of the visual representations. They found that participants rarely answered from an emotional perspective (7 out of 133). However, more subjects responded emotionally to versions containing data visualizations than to the illustrative version. Participants expressed feelings about the topic, such as *"powerlessness," "shocking," "devastating,"* and *"frustration"*. This result shows that free-form responses allow for a wider range of emotions, as the stated feelings were not provided in the questionnaire. Furthermore, the small response size revealed that, when investigating emotions, participants must be explicitly asked to express their emotions in order to acquire a sufficient amount of data [38].

Free-form questions

Boy et al. (2017) investigated whether including human-related visuals in data storytelling evoke more ***empathy***, ***distress***, and increase the willingness to donate – compared to standard charts. They conducted seven experiments investigating different implementations of ***Anthropomorphism*** – specifically human-like pictographs. Each experiment consists of two story versions about human rights topics. The focus group stories used the Anthropomorphism concept to present data and the control group stories were designed with pie charts.

To examine the effects on emotions, the authors divided their set of questions into two categories: empathic concern and personal distress. Participants were asked to rate their feelings using 7-point Likert scales according to empathy-related adjectives, i. e., *"sympathetic, moved, compassionate, tender, warm,* and *softhearted;"* and distress-related adjectives, i. e., how much they felt *"alarmed, grieved, upset, worried, disturbed, perturbed, distressed,* and *troubled."* Qualitative feedback was collected via free-form responses.

7-point Likert scales of emotion adjectives

Qualitative feedback

The results showed that there was no significant difference in empathy or distress between the two representations. Boy et al. raised the question of whether it is not anthropomorphism itself that has no effect, but rather, whether the abstract style of pictograms evokes significantly fewer emotions and empathy. The authors suggested that the lack of life-like representations may have led to the indifference between the groups. They proposed further ideas for humanizing data visualization, such as using more detailed drawn illustrations of human faces, visual metaphors, photographs, animations, cinematic effects (zooming), and voice-overs. These techniques could lead to greater empathy and identification with the narrative character [14].

Nowak et al. (2018) chose a purely qualitative approach to examine emotional affect and engagement by conducting introspective interviews with users that viewed narrative visualizations. For the experiment, two interactive data stories with high emotional content were chosen. The first group of participants viewed the "U.S. Gun Deaths" visualization by Periscopic and the second group viewed "Plane Truth" from Information is Beautiful.

Elicitation interview

Nowak et al. used the elicitation interview technique that was designed to re-create the experience during an activity. They adapted the evaluation method by adding a breathing exercise between the data storytelling session and the interview. This techniques should help participants to focus on their internal stages and relax. The interview session was divided into iterative steps. First, the interviewer led the attention to the particular experience of viewing the data story with the question: *"Could you go back to the moment when you sat down to view the data visualization? What do you see? What do you smell? What is around you? How do you feel?"* Secondly, to access the narrative in a temporal dimension the interviewee was asked: *"What happens next?"* Thirdly, to delve deeper into the understanding of the stated experience, the interviewer asks about certain aspects of the experience. For example, *"If you agree, I would like to go back to your experience of X. How did you feel at this time?"*

Interviewees often used affective terms to describe their experience and gave meaning to it. For example, the animation of the bullets in the "Gun Deaths" Visualization was connected to emotional states such as *sadness* or *shock*. Nowak et al. found that this interview technique enables rich insights into emotional engagement and sensory experience. It can be useful to understand which elements of a data story evoke emotions. However, the method relies on participants' memory, their ability to go into deep introspection (without rationalizing), and their willingness to share their deep affective experiences [77]. Other limitations arised from a small number of participants and the time-consuming nature of the method.

Next, I discuss the paper *"Negative Emotions, Positive Outcomes? Exploring the Communication of Negativity in Serious Data Stories"* by **Lan et al. (2022)** because it focuses on negative emotions, which play a crucial role in designing serious data stories. The paradox that negative content or sensation evokes positive emotions is called *"positive negative experience," "tragic entertainment,"* or *"valance transformation."*

The authors conducted two studies. First, they carried out a workshop analyzing a corpus of serious data stories with data story experts and identified 19 design strategies for serious data stories through discussion.

Expert Workshop

In addition, the authors identified three benefits for evoking negative emotions as a result of the discussion with data storytelling experts:

(1) *"Lead to contemplative experiences"* (i. e., encourage thinking, reflection, and action-taking)

(2) *"Enhance user engagement"* (i. e., attract attention and interest)

(3) *"Leave a long-term impression"* (i. e., negative emotions can last longer than entertainment)

Secondly, the effectiveness of these techniques was evaluated in a lab study involving 35 participants. The authors selected thirteen data stories about the Coronavirus pandemic from the workshop and showed each participant three of them. After viewing each story, participants were asked to complete a questionnaire. The same questionnaire was given to them directly after each data story.

Lab study

First, the authors used 48 emotions from the ***Emotion Annotation and Representation Language Model*** (EARL) [99], including 24 negative and 24 positive emotion types. Participants were asked to rate each emotion on a 9-point Likert scale, where 0 represented "did not feel the slightest bit" and 8 represented "the most extreme ones experienced in life". The EARL scale classifies emotions into groups such as passive (e. g., sad), agitation (e. g., stress), caring (e. g., empathy), reactive (e. g., interest), and positive and lively (e. g., amusement). Secondly, to correlate emotions with outcomes, the following indicators were asked about using a 7-point Likert scale: contemplative experience, user engagement, and meta-emotion.

48 EARL emotions, 9-point Likert scales

During the first study session, participants were interviewed to obtain qualitative feedback on emotional experiences and the effectiveness of the 19 design methods. To investigate long-term memory, the same participants were interviewed again after two or three weeks. The follow-up interview questions focused not only on content-related memories, but also on which emotions were recalled and what had evoked them.

Interviews

Lan et al. found that negative emotions were most strongly correlated with ***contemplative experience*** among all the indicators. Although these data stories were designed to primarily elicit negative emotions, participants also experienced positive emotions. Additionally, these were positively correlated with contemplative experience. However, the results were not compared to stories designed to elicit positive emotions.

While some participants agreed that negative emotions encouraged them to think about and reflect on the presented topic, others stated that strong negative emotions prevented them from thinking rationally or staying engaged. The authors discussed ***emotional intelligence*** as a possible factor in how well a person can handle negative emotions. They found no significant correlation between negative emotions and the likeability of a data story. The results showed that the most effective design methods for serious data stories were *tough task*, *emotive language*, *role-play gaming*, *details behind data*, *self-run animation*, and *picture with negative semantics*. A long-term memory test showed that participants could recall the overall topic, with negative emotions being the most memorable. However, most of the emotions previously reported faded, as did concrete data values and visual details. This study effectively combined designers' intentions with user outcomes. Their deep focus on emotions and variety of methods provided valuable insights into emotional engagement [55].

In summary, previous studies in Narrative Visualization have evaluated emotional engagement using self-report methods, primarily questionnaires with 5–9-point Likert scales to measure emotions and engagement. Some researchers use interviews as their main method or as an additional method. Workshops are useful for capturing the intentions of visualization engineer and designers.

I noticed a high degree of variability in the depth of emotional investigation, ranging from asking one single question to interactive interview settings or questionnaires with large sets of emotion types. However, most studies do not consider potential changes throughout the story, nor do they explicitly distinguish between arousal (the strength of the emotional response) and valence (the inherent pleasantness or unpleasantness of an emotion) in their data analysis.

In addition, limited work has been done to understand a character's role in serious data stories from the perspective of emotional engagement. Furthermore, I could not identify a single publication that investigated emotional engagement in Narrative Visualization with Electrodermal Activity. The next section will provide a conscious overview about this electrical phenomena, its working principle, and a selection of articles that study emotions using Electrodermal Activity.

5.3 ELECTRODERMAL ACTIVITY

As discussed in Chapter 3 and Chapter 5, previous work in Narrative Visualization heavily rely on subjective measures. On the other hand, physiological studies rarely focus on narrative contexts. This study addresses this gap by investigating how character-driven medical data storytelling influences emotional engagement, using electrodermal activity and eye tracking as primary measurements to provide objective insights. Emotional facial expressions may seem like a straightforward physiological measurement. However, studies have shown that there is a negative correlation between emotional arousal and facial expressiveness. This was attributed to increased effort to suppress expressions. Buck et al. (1744) assumed that socially learned inhibition of overt affective responses inhibits facial expressiveness [15]. Emotional experiences can be measured in terms of autonomic nervous system activation, which is indicated by increased cardiovascular output, i. e., an increase in heart rate or blood pressure. Heart rate is also useful to study the valence dimension of emotions. For example, sadness is accompanied with a deceleration in heart rate [12, 54]. However, EDA is more sensitive to measure emotional reactions to visual stimuli. This section will focus on electrodermal activity as core method to study emotional arousal. I aim at to provide the reader with the necessary background information about how Electrodermal Activity (EDA) is measured and how the data can be interpreted. I also will give an overview in which fields EDA is used, which objectives are investigated, and the limitations of this method. The last subsection will present works that combine EDA and eye tracking.

Physiological measurements

5.3.1 *Basics of Electrodermal Activity*

Research on Electrodermal Activity started in the 1880s in Germany and is mainly used in psychophysiology, clinical and applied psychology, and medical disciplines. However, EDA is also used to test Virtual Reality environments [53, 62, 64], Human-Computer Interaction [23, 48], and is also used in narrative contexts [37, 94, 95]. The main objectives in these studies are measuring emotional engagement, learning, and cognitive load, such as task difficulty.

Application areas

Objectives

Electrodermal Activity (EDA) encompasses the electrical phenomena of the skin. In a nutshell, when an emotionally relevant stimulus is perceived, the autonomic nervous system is activated, resulting in sweat secretion. This increases skin conductance, allowing electricity to flow more easily which can be measured by the current flow between two electrodes. While the sympathetic nervous system activates sweat glands, an inhibitory effect by the parasympathetic system is discussed [12].

Definition

Physiological background

Studies have shown that electrodermal activity responses are correlated with physical activity, as well as the intensity and psychological significance of a stimulus [12]. Changes in electrical properties are related to the functions of the autonomic nervous system, which is controlled by the sympathetic nervous system. Key brain regions involved in the sympathetic nervous system include the hypothalamus and the amygdala. These brain regions are associated with fight-or-flight reactions and the regulation of emotions and stress. When an emotionally significant event occurs, the sympathetic nervous system is activated, triggering the release of stress hormones such as adrenaline and cortisol. These hormones then signal the sweat glands to release sweat, which is an evolutionary adaptation designed to cool the body in the event of an immediate threat. The eccrine sweat glands on the palmar and plantar surfaces have been identified as significant origins of sweat production during emotional responses with a high density of sweat glands. The production of sweat on the palms of the hands and soles of the feet increases adhesion to surfaces, which can be advantageous in flight situations [12].

The Signal & Parameters

EDA consists of a tonic and phasic component. The *tonic signal*, also called the ***skin conductance level*** or baseline, represents the slowly changing component of the signal associated with a person's general arousal or mood. In healthy humans a normal skin conductance level ranges between 1 and 20 microsiemens. The *phasic component* is characterized by a relatively fast increase in the signal. A single peak in the phasic component is called ***skin conductance response***. This component represents the emotional response or reaction [12].

When an emotional relevant stimulus is presented, there is a ***latency*** to the onset of response. This latency depends on the external environment where low temperature causes longer latencies. Edelberg (1972) has shown that 1.8 seconds are a characteristic value for latency in ambient temperature [27]. A ***skin conductance response*** starts with the so called ***response onset*** and then rises relatively steep until it reaches its maximum. This time window is called ***rise time***. The maximum point marks the ***amplitude*** (height) of the emotional response and is followed by the ***recovery time*** where the signal drops to its normal level. As the signal decay is usually flatter than the signal rise, the rise time is shorter than the recovery time. In highly arousing situations, skin conductance responses can superimpose each other. Another term for stimulus-related emotional response is ***electrodermal reactivity***.

Measuring Methods

There are three methods of measuring EDA: The endosomatic method and two exosomatic methods, which involve applying either direct current (DC) or alternating current (AC) via electrodes on the skin. The sensor device (Section 7.1) used in this study employs the ***exosomatic measurement method*** [29] and applies a consistent direct current (DC) voltage of, e. g., 0.5 V to the skin to measure skin conductance. The sensor receives a single-channel signal from the current flow between two silver–silver chloride (Ag/AgCl) electrodes. Conductance level is the reciprocal of skin resistance and can be calculated using an alternative formula of Ohm's law: $I = E \times G$, where the current flow through the electrodes (I) is the fixed voltage value (E) multiplied by the conductance of the participant's skin (G) [29].

For a comprehensive summary of the fundamentals, measuring methods, and applications of EDA, the reader is referred to *"Electrodermal Activity"* by **Boucsein et al. (2012)** [12].

5.3.2 *External & Internal Influences on EDA*

External influences

Some aspects affect EDA measurements and must be taken into account when conducting an experiment. For example, EDA can vary with ambient temperature. Cool temperatures have a decremental effect, while heat has an incremental effect [12]. Seasons also have a clear influences on EDA measurements where skin conductance level is lower during winter months, intermediate in early summer, and high during heat spell. Venables & Christie (1973) suggest that hormonal changes between winter and summer cause these effects [108]. Therefore, it is important to control the room temperature around 23°C. Additionally, it is recommended that the experiment be conducted within one season. Movement produces artifacts that must be addressed during data preprocessing. Poor electrode placement can also lead to a noisy signal. When the electrodes become loose due to hand movement or excessive sweating, they lose contact with the skin, resulting in a low signal [81].

Individuals have physiological differences, such as different skin thicknesses and densities of sweat glands, as well as different levels of hydration (dryness of the skin) and medication. These differences can increase or decrease sweating [12].

Internal influences

When measuring emotional engagement with EDA, it is also important to be aware of internal physiological or psychological sources that may influence sweat gland activity. Physiological responses vary based on individual and cultural differences. ***Individual differences***, such as personality, health, and emotional regulation [12, 28]. For example, autonomic nervous system diseases (e. g., Parkinson's disease or multiple system atrophy), and neuropathies (e. g., diabetic neuropathy or peripheral nerve damage) are correlated with ***disturbed sweat regulation***. Some psychological disorders and their variation among individuals can influence the signal. People with an anxious personality may have heightened sympathetic nervous system responses. On the other hand, poor electrodermal conditionability in terms of electrodermal ***hypoactivity*** (low skin conductance levels) and ***hyporeactivity*** (low skin conductance responses) can also be a product of psychological disorders, such as depression, psychopathy, and schizophrenia [12].

There are also lateralization effects. Emotional arousal is mapped differently to body parts in left-handed individuals than in right-handed people. For example, the approach motivation of right-handers is linked to the left hemisphere. However, in left-handers, it appears to be processed by the right hemisphere [12].

Several studies found that skin temperature is positively correlated with skin conductance responses and negatively correlated with latency [66, 110]. Cold hands as a symptom of, e. g., Raynaud's syndrome, produces low quality data due to poor blood circulation in the hands. Cold hands can also be an issue in healthy persons, particularly in the winter months. Participants should warm up their hands before the experiment starts. In addition, the specific temperature at the hands should be recorded during the experiment, a feature that the used EDA sensor in this study offers. Regarding small body temperature fluctuations during the day, it would be optimal to evaluate all participants during the same day time. In addition, using a baseline or separating the overall mood from emotional arousal can be helpful to isolate stimulus-related responses from other internal states.

Demographic properties that can have an influence on EDA are age and gender. A decrease in the number of active eccrine sweat glands and reduced sweat gland function occurs commonly with individuals over 65. Physiological and psychological changes during aging must be considered as a factor for reduced skin conductance level and skin conductance response. However, other studies showed that some elderly participants produced higher amplitudes with emotionally relevant stimuli, such as photographs of children [12]. In general, females tend to have higher skin conductance levels, while males show higher arousal when stimulated.

The suitability of electrodermal activity as a measurement tool should be tested before starting an experiment by ***monitoring*** the person's baseline values and skin temperature via a real-time application, to ensure a reliable data collection.

5.3.3 *Studying Emotions with EDA*

As described by **Boucsein et al. (2012)**, EDA is used to investigate psychophysiological states and processes, such as information processing, memory, decision-making, and orientational habits. In addition, EDA is used for assessing personality traits and psychological disorders, e. g., anxiety disorders, depressions, psychopathy, and schizophrenia.

One relevant insight from previous studies is that an ***inverted-U relationship*** exists between arousal and performance. This means that low to moderate levels of arousal lead to better performance in learning and understanding tasks, whereas very low and very high arousal states are associated with poorer performances [6]. Emotions and stress are closely related because negative stress (distress) as well as positive stress (eustress) come with strong emotional responses. Emotional excitement and motivational processes occur as a loop-like neuronal activity within the limbic system (Papez-circuit). Within this circuit sensory information is being checked by the thalamus and compared with memory via the hippocampus. The hypothalamus has the function to activate the autonomic nervous system in the case of a significant event and elicits concomitants of various emotional states [12]. Studying psychophysiological patterns for different emotions is a challenging problem and seems to vary widely across individuals. However, a study using EDA showed that *anger* results in a high frequency of non-stimulus related skin conductance responses, while *anxiety* increases the overall arousal level. However, Malmo (1962) could also show that large individual differences exist [5]. It has been found that *pleasure* raises the skin temperature of the hand. There is an inverted U-shaped relationship between pleasantness and skin conductance responses where high pleasantness and high unpleasantness both result in strong skin conductance responses [103]. The type of medium can affect emotional engagement: A study compared short videos to single static images taken from the videos. The skin conductance response amplitudes were higher for the video format, but the valence differed less between the two stimuli [101].

Emotional patterns

EDA has frequently been used as an indicator of emotions and stress when presenting all kinds of emotion-inducing and stressful stimuli [12]. Affective computing uses machine learning methods to computationally detect and interpret emotions. **D'Amelio et al. (2025)** conducted a systematic review and meta-analysis of affective computing on EDA data. The data is processed through feature extraction (e. g., amplitude, latency, and frequency of conductance changes) and fed into machine learning models that primarily employ a two-dimensional framework of arousal and valence. The surveyed studies predominantly use classification algorithms to categorize emotions. The authors refer to theoretical frameworks of affective science which emphasize the non-distinct nature of emotions, and therefore suggested that models of affective computing should rather treat emotions as continuous gradients than discrete categories. In this regard, ***regression models*** for emotion recognition are underexplored. The authors also concluded from their meta-analysis results that EDA data consistently demonstrates stronger predictive power for arousal than valence, showing significantly higher accuracy in arousal models. This aligns with EDA's physiological basis in sympathetic activation,

Affective Computing

providing evidence that it reliably captures emotional intensity (e. g., in high-arousal states like stress or excitement) but less effectively distinguishes positive from negative valence [24]. This aligns with Boucsein (2012) who stated that *"psychophysiological investigation of emotional states requires not only physiological measures but also subjective reports."* [12] The authors highly recommend the use of a multivariate methodology in psychophysiological research of emotional states.

Multivariate methodology

In the following, I will present experimental studies that are more closely related to the topic of my thesis, i. e., Human-Computer Interaction, Storytelling/Narrative Engagement, and Education. The focus is on how the authors employed EDA as their primary method and the insights that the studies provided. ⟶

HCI & user experience

The ***advantages*** of EDA for evaluating user experience include its non-invasive nature and interpretability in measuring emotional states, engagement, stress cues, and discomfort during interaction [23, 81]. From a practical standpoint, modern wearable sensors [41] are affordable and allow for continuous long-term recording during motion, e. g., in daily activities and exercise, as well as the collection of real-time data during interactions with technology. For instance, humanoid robots in healthcare could use EDA data to respond empathetically and responsibly to human emotions. Another area that could benefit from real-time EDA data is the learning environment. This data could provide insights into students' emotional engagement and motivation, allowing for the real-time adaptation of educational settings [23, 81].

While **Lang et al. (1993)** found that self-reported emotional arousal and valence were consistent with EDA [56], Richardson et al. (2018) observed the opposite. These studies used different types of stimuli. Lang et al. (1993) showed participants colored photographs of varying emotional relevance (from neutral to arousing).

Storytelling

Richardson et al. (2018), on the other hand, compared auditory stories to a video format of the same narrative. For recording self-reports on narrative engagement from 102 participants, the scale developed by Busselle and Bilandzic (2009) [19] was adapted. While study participants rated higher for the video format than for the audiobook with respect to ***attention***, ***narrative***, ***character***, and ***presence***, all physiological measurements were higher for the audio book which suggests that emotional and cognitive engagement was more intense for the auditory format. In addition to ***electrodermal activity***, the authors recorded ***heart rate*** and ***skin temperature***, and analyzed the data using Bayesian mixed models.

The authors assumed that the audio format might impose a greater cognitive load during the co-creation process than videos, which are already rich in stimuli. The video format may also have a more relaxing effect, which could explain lower arousal and the preference for it. They concluded that more research is needed to distinguish between emotional and cognitive engagement [94].

Education

Stuhldreher & Brouwer (2025) used wearable EDA sensors (EdaMove 4 [41]) in a real-life setting of a lecture. The study aimed to investigate whether engagement with a lecture differs when the audience has or does not have a personal relationship with the speaker. Half of the audience had a personal relationship with the lecturer, while the other half had a professional one. Attentional engagement was also assessed using a questionnaire handed out after the lecture. Participants were asked to rate their engagement on a 10-point Likert scale for each predefined event in the lecture.

The authors found no significant correlation between self-reported attentional engagement and EDA data. They discussed the challenges and limitations of their mixed method. Firstly, they stated that participants might find it difficult to rate their engagement, particularly when asked with a time delay after the stimuli were presented – in this case, events within a 50-minute festive inaugural lecture. Secondly, asking participants to report their level of engagement immediately after each event would interfere with their overall engagement. Thirdly, participants might be hesitant to report low engagement due to social expectations and acceptability, particularly when it comes to an event that is supposed to be engaging, such as a festive lecture.

EDA, on the other hand, offered a continuous, objective signal over the course of the lecture that can help to identify moments when participants are engaged. However, the authors mentioned that they encountered low signal issues during EDA recording. They used a threshold of 0.3 microsiemens and labelled all values under this threshold as 'NaN'. If missing data made up more than 10 % of the recording, the participant's data was excluded from the analysis [104].

In the following section, I will focus on studies that employed EDA and eye tracking as their core methods, and what insights were gained using this methods. ⟶

5.4 EDA & EYE TRACKING

Storytelling

The study by **Frey et al. (2020)**, titled *"Physiologically Driven Storytelling: Concept and Software Tool,"* describes a method to support real-time adaptation of the narrative using physiological signals, including electrodermal activity and eye tracking. The authors developed a tool, called Physiological Interactive Fiction (PIF), which is an open-source system for text-based interactive storytelling. It processes signals like EDA, breathing, and eye tracking in real-time to infer the reader's emotional and cognitive states, such as arousal, valence, and reading difficulty, and is able to dynamically branch the story accordingly. EDA was used to detect arousal and emotional responses, and played a key role in their concept. For example, peaks in the EDA data during exciting or fearful moments (e. g., deceiving police or encountering spiders) can trigger branches that escalate intensity, adjust character reactions to match the reader's inferred state, or simplify complex sections if high arousal signals cognitive overload. Their lab study showed that EDA contributes to high classification accuracy (92.9 % for arousal), enabling adaptations like empathy-building branches where characters mirror the reader's fear [37].

Rühlemann (2020) conducted a pilot study on emotional engagement involving two serious stories designed to elicit negative emotions, such as sadness and distress, respectively. The storyteller read each story aloud to two participants. All participants wore EDA sensors and eye tracking glasses. The sessions were recorded via video, and the timestamps from the video were used to label the EDA and eye tracking data according to story parts, such as the climax. Using these methods, Rühlemann found that the narrator influences the emotional response of the audience. Audience arousal levels corresponded to the storyteller's arousal intensity, particularly when they had direct eye contact (mirroring effect). Other identified factors include narrative relevance and recipient susceptibility [95].

These insights are useful for Narrative Visualization because they suggest that stories with individual characters could take advantage of the mirroring effect to elicit empathy in audiences. Designers should ensure that characters look in the direction of the audience, simulating eye contact. However, factors such as the recipient's sensibility and the narrative's relevance to the individual are partially beyond the designer's control. However, emphasizing the relevance of the topic and narrative intent at the beginning of the story would be promising.

HCI & Education

Liberman & Dubovi (2022) compared narration techniques to evaluate the modality principle in VR-based learning environments. The modality principle states that people learn better from visual stimuli when combined with audio narration instead of on-screen text. To test this theory, the authors developed two desktop VR applications (low immersion, no head-mounted display) in which only the narrative technique differed. One used on-screen text, while the other employed spoken narration. The study was conducted with nursing students and focused on investigating cognitive experiences while learning medication administration procedures.

While, the theory of the modality principle states that the distribution is more effectively distributed between visual and auditory processing, the authors found more electrodermal activity peaks in the spoken version. They suggested that it requires more cognitive effort from learners due to the allocation of more brain areas to process multisensory information. Additionally, the eye tracking data showed shorter fixation durations for the orally narrated version, suggesting less focused attention. The post-knowledge test showed no significant difference between the different design approaches [62]. Asking participants about their motivation to learn with the application, for example, could have provided additional context and helped interpret the higher arousal level in the EDA data.

In a subsequent study, **Liberman et al. (2023)**, compared high-immersion VR (using a head-mounted display) to low-immersion desktop VR. The objective of the investigation focused on both attentional engagement and cognitive load, specifically the effects of immersion techniques on declarative (factual) and procedural (step-by-step) knowledge. Similar to the previous study, these applications were designed to teach medication administration procedures and foster declarative and procedural knowledge. Eye tracking, either integrated into the head-mounted display or remote for the desktop version, was employed to capture real-time gaze data, including fixations, saccades, and dwell times on predefined areas of interest (AOIs), such as the medication cabinet or electronic patient records. EDA and eye tracking were supplemented with self-reported scales (e. g., NASA-TLX for subjective load) and pre/post-knowledge tests to evaluate knowledge gain. This methodology enabled a time-sensitive examination of cognitive dynamics without relying solely on retrospective reports. The results revealed different effects for low versus high immersion. While the procedural knowledge test results were higher for low-immersion VR, the declarative knowledge gain was higher in high-immersion VR.

EDA and eye tracking data provided context for these results. For high-immersion VR, higher and more frequent arousal amplitudes were measured, suggesting that this environment is associated with greater cognitive load and emotional arousal, likely due to sensory richness and spatial navigation. Additionally, eye fixations were shorter, reflecting fragmented attention and increased scanning effort. The occurrence of eye blinks was also reduced, which is linked to cognitive load. However, high-immersion VR led to longer dwell times on relevant areas of interest, implying deeper exploration and encoding through spatial contextualization. Conversely, desktop VR promoted a more sustained focus on areas of interest, reducing off-task glances by up to 15 %. Based on these results, the authors concluded that the benefits of immersion are context-dependent, and that low-immersion designs can be as effective as high-immersion VR [61].

Ma et al. (2023) integrated EDA, eye tracking, and facial expression recognition as components of multimodal learning analytics (MMLA) to objectively capture real-time physiological and behavioral indicators of learning engagement. They utilized an online learning platform with a statistics course that incorporates various modalities and interaction elements, including text, images, animations, audio, quizzes, exercises, and kinesthetic interactions. In a small user study, the authors investigated MMLA's potential to analyze cognitive and emotional engagement during learning activities.

Electrodermal activity was used to measure stress levels and arousal. Facial expression analysis should provide information about emotional valence. However, facial expressions can be less informative due to individual differences and the suppression of emotional expression learned through socialization. Eye tracking detected saccades and fixations, which were visualized using heat maps. Emotions were reported via a questionnaire considering the six emotions: joy, surprise, contempt, sadness, disgust, and anger. Data analysis revealed no significant correlation between interaction elements and electrodermal activity peaks or self-reported emotions. Nevertheless, participants reported feeling more emotionally engaged during the active parts of the course, such as quizzes and exercises. Animations evoked more emotional engagement and cognitive activity than other interaction types. The study results did not include learning outcomes, which would have made the analysis more meaningful. The number of emotion types was limited. It is also important to choose an appropriate set of emotions; for example, disgust is probably rarely experienced when learning statistics. Overall, the emotional valence of the learning experience was negative, implying that improvements to the learning environment are needed [67].

As the studies presented show, it is common practice and recommended to supplement EDA with other methods or measures, such as heart rate, skin temperature, eye tracking, and questionnaires. However, self-reported information should be viewed with caution due to social desirability bias and differences in how individuals assess their engagement.

Part II

METHODOLOGY

This part provides details about the story design, development, and preliminary user experience study. After the describing the project's background (Section 6.1) and the data and information (Section 6.2) used to craft the stories, it presents the concept (Section 6.3), narrative structure (Section 6.3.5) and narrative techniques (Section 6.3.6) for designing an emotionally engaging medical data story. Section 6.4 describes the prototyping techniques, and Section 6.5 presents the results of the pre-pilot study. In Section 6.6, I present the final design of both story versions. In the next chapter (Chapter 7), I describe the instruments – EDA, eye tracking, and the questionnaire designs – that have been used in the experiment to measure emotional engagement and its effects. Finally, Chapter 8, I will provide an overview of the composition of the study participants (Section 8.1) and the experimental procedures (Section 8.2).

This part is divided into 3 chapters:

- Story Design & Stimulus Development
- Measuring Instruments
- Experimental Procedures & Participants

6 STIMULUS DEVELOPMENT

The concept and design of the first story prototype (told from an individual perspective) were mainly developed during my exchange year at the University of Bergen, Norway. After the oral examination, refinements were made to the interface design and data visualizations of this prototype, and the second prototype, telling a more general story, was finalized, including all the illustrations. Therefore, the story concept and parts of its development is not part of this thesis. However, to provide a comprehensive overview of this project, I will include the story concept and summarize the design choices. Please note that all texts were written within the scope of my Master's thesis.

6.1 PROJECT BACKGROUND

The project started in August 2023 as a ***visualization project*** at the University of Bergen supervised by Asso. Prof. **Laura Garrison** and Prof. **Marc Vaudel**. We regularly met online with Dr. habil. **Monique Meuschke** to discuss the project's progress. Laura Garrison introduced me to the topic of a data-driven story about hyperemesis gravidarum (HG) with the ***research question*** on *"How can data stories inspire empathy?"* The project's ***intent*** can best be summarized with a quote from HG researcher Marlena Fejzo:

> *"Although most providers now recognize HG as a* ***serious condition*** *with a* ***biological basis****, some may need to be better* ***educated*** *to understand that* ***patients' [quality of life]*** *can improve dramatically with adequate* ***treatment****,* ***care****, and* ***understanding****."*
>
> — Marlena Fejzo, In: Nature reviews disease primers 2019 [34]

I also had the chance to talk and discuss my story prototypes with HG reseachers, Prof. **Jone Trovik**, **Kimber MacGibbon**, the head of the Her Foundation, a hyperemesis gravidarum support organization in the United States of America. Professors **Helwig Hauser**, **Eduard Gröller**, and **Noeska Smit** provided me with insightful feedback to improve my prototypes.

6.2 DATA

Cohort data

The data story was developed using the MoBa data set from ***"The Norwegian Mother, Father and Child Cohort Study"*** (*norweg. "Den norske mor, far og barn-undersøkelsen,"* MoBa) [35, 68]. This population-based pregnancy cohort study is one of the world's largest health studies, collecting data from 102,810 pregnant Norwegian women who provided information on their health during pregnancy in three extensive questionnaires. Participants also provided samples of DNA, RNA, blood, plasma, and urine, which are stored in a biobank for bio-genetics research [102]. With a consensus rate of 41 %, the dataset is a rich sample of the Norwegian population. The study's primary goal is to investigate health and medical factors that can impact the health of pregnant women and their children. The Norwegian Institute of Public Health recruited pregnant women from across Norway between 1999 and 2008. The study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. The data is subject to regulations by the Regional Committees for Medical and Health Research Ethics [68].

Diagnosing HG

Hyperemesis gravidarum (HG) is a rare but devastating condition that affects 0.3–10.8 % of pregnant women [34]. However, there is no single, universally accepted uniform definition of HG. In addition to the most prominent symptom of persistent nausea and vomiting, different definitions include other diagnostic criteria, such as dehydration, the presence of ketones in the urine, weight loss, and electrolyte imbalances. Other definitions incorporate the requirement of hospitalization. Nevertheless, there is still no internationally accepted agreement on the diagnostic criteria [52]. Ziman et al. (2023) encountered a similar challenge when developing an interactive visualization dashboard using MoBa data to assist experts in analyzing the outcomes of children of HG mothers. The authors defined HG as a condition manifested by at least one hospitalization. They identified women who probably experienced HG using both the hospitalization variable and the prolonged nausea and vomiting variable [114]. Since blood samples were collected between weeks 17-20, higher levels of the hormone GDF-15 that can predict HG with a high degree of accuracy might be promising. DNA samples were also collected which could be useful to identify women who are genetically more sensitive to GDF-15 variants that trigger persistent vomiting [31].

The challenges of the data analysis can be summarized as follows:

Challenges

- Dealing with the absence of a concrete variable in the dataset for the diagnosis of hyperemesis gravidarum.
- The definition of HG which includes at least one hospitalization could not be applied because the sub-dataset was aggregated and did not allow tracking of subjects over time. For example, if

women reported hospitalization in the fifth month, it was only possible to correlate other symptoms to these women in that same month.

- Patients in the group of interest are known to have higher attrition rates due to the severity of their condition. This can introduce additional bias and uncertainty to the data [114].
- Due to data privacy regulations, data is aggregated into ten-step bins, where values less than ten are defined as "<10." Regarding the relative group size of HG patients, this introduces significant uncertainty in the data.

Research methods

Marc Vaudel – a bioinformatics researcher and domain expert who conducts multidisciplinary research on large cohorts to understand metabolic health during pregnancy and early childhood – introduced me to the medical topic, dataset, and provided me with literature recommendations. I employed the following methods to gain a deeper understanding of the condition and how it affects women emotionally:

- First, I read the recommended literature about hyperemesis gravidarum, focusing on the condition's bio-genetic causes and its accompanying symptoms, e. g., [34].
- Second, I reviewed literature that considered the experiences of women affected by HG. For instance, Havnen et al. (2019) conducted interviews with affected women to understand the consequences of condition on daily life and how doctors and the social environment recognized their suffering [47].
- Third, I researched hyperemesis gravidarum support websites and blogs where women reported their personal stories [4, 36].
- Fourth, image research was conducted. In particular, an illustration in [34] inspired the creation of an animated illustration explaining the disease's physiological processes.

MoBa questionnaires

Next, I familiarized myself with the MoBa questionnaire documentation. The first questionnaire was sent to participants after the 15^{th} week of pregnancy where they were asked to report on symptoms such as nausea in four-weeks bins (weeks 0-4, 5-8, 9-12). The second questionnaire only addressed diet and was therefore less relevant to this project. The third questionnaire, which covered the second half of the pregnancy (weeks 13-16, 17-20, 21-24, 25-28, and 29+), was sent after the 30^{th} week. Some of the questions differed between the first and third questionnaires. For example, only the first questionnaire asked about sleeping problems, leading to missing values.

Approach

Based on the challenges I listed above, it was not possible to identify women with HG in the data with high confidence. To address this issue, I considered that nausea and vomiting (NV) during pregnancy typically ends between weeks 12 and 20, with many women feeling better by week 16. When these symptoms were reported after the fourth month, it was called prolonged nausea and vomiting (PNV), and this data was more likely to include women with HG. Based on a list of symptoms gathered from previous research, I found it interesting to investigate whether different groups of nausea and vomiting (NV), prolonged nausea and vomiting (PNV), nausea only, and no nausea are correlated differently with possible side symptoms. To do so, I related the symptoms to the variables in the MoBa documentation [90] and requested the data for each month.

Data variables

Through exploratory data analysis, I verified that most of the symptoms mentioned in the literature reflecting the physical and psychological effects of the disease also showed different distributions in the MoBa data when comparing the following groups: NV, PNV, nausea only, and no nausea.

The following variables were selected for the final data visualization:

- *Nausea and vomiting*
- *Long-term nausea and vomiting*
- *Nausea only*
- *Hospitalization*
- *Unusual fatigue and drowsiness*
- *Sleeping problems*
- *Headaches*
- *Depression*

Table 6.1 shows an example of the aggregated MoBa sub-dataset. Each variable has a separate code for the four-week time intervals. For example, the 'category' variable contains the variable 'CC390' (long-term nausea and vomiting in weeks 21–24), for which 2,700 women in total reported these symptoms. 'Level' expresses a binary value where '1' refers to agreement, and 'missing' does not refer to a missing value; it means logical 0 or disagreement. When the 'outcome' is labeled as 'any' this means that the frequency of the category variable in relation to the whole cohort is considered. Summing the numerical values of the first two rows gives the total number of participants, which is 102,810. When 'outcome' contains a variable, then the 'category' variable becomes an independent variable. Rows 3 and 4 show examples where long-term nausea and vomiting is correlated with dependent variables, e. g., CC462 (unusual fatigue/drowsiness in weeks 21–24) and CC678 (depression in weeks 21–24).

CATEGORY	LEVEL	OUTCOME	OUTCOME_LEVEL	NUMBER
CC390	1	any	any	2700
CC390	missing	any	any	100110
CC390	1	CC462	1	850
CC390	1	CC678	1	140

Table 6.1: Example of aggregated MoBa data in which an independent variable (category) is correlated with a dependent variable (outcome): CC390: Long-term nausea and vomiting in weeks 21–24, CC462: Unusual fatigue/drowsiness in weeks 21–24, CC678: Depression in weeks 21–24.

6.3 STORY CONCEPT & DESIGN

In this section, I will introduce the concept of the medical data stories. Both versions blend storytelling with data visualization to highlight the impact of hyperemesis gravidarum on women's health. The story is grounded in evidence-based insights, embedding cohort data from the MoBa study and scientific results to provide context.

Individual Story: The first approach is a character-driven disease data story that uses a relatable, individual human narrative. The story aims to make complex epidemiological data engaging, empathetic, and actionable by portraying a character who serves as a role model for seeking help. The story is told from a first-person point of view.

General Story: The second approach uses a more objective, neutral, and general point of view. This data-driven disease story focuses on communicating the impact and spread of the disease through the same verifiable data. The story starts with a third-person point of view. In the data visualization part, different women share their experience in first-person perspective.

6.3.1 *Narrative Intent & Target Audience*

The main goal is to raise awareness of hyperemesis gravidarum, a disease with severe symptoms that impact the physical and psychological health of pregnant women. It must be clearly distinguished from "normal" nausea and vomiting during pregnancy. Cohort data and scientific findings support this message.

The second narrative intent considers that women with this condition require medical and daily support; otherwise, the health of both the mother and her child is at risk. Therefore, another explicit goal is to communicate the importance of early detection and encourage affected women to seek medical and psychological help.

The third narrative intent is more implicit. However, it is very central to the story. It aims to promote sympathy and empathy for HG patients and encourage supportive behavior from their social environment, including partners, family members, and friends.

Therefore, the *target audience* encompasses broad audiences, including HG patients, their relatives, and friends. The story is not only for those who are directly involved, but it can also educate and prepare others for supportive behavior.

In the following, I will describe the story concept and design choices along the four story ingredients: 1) content, 2) narrative characters, 3) conflict, 4) narrative structure [72].

6.3.2 *Content*

During the conceptual phase, Marc Vaudel helped reviewing the content. For creating the story's content three different sources were used. Context information influenced the data analysis and choice of the character:

- *Data-driven:* As described in Section 6.2, content was directly derived from **cohort data**. For accompanying symptoms of HG, context research was conducted on research papers to match variables in the data.
- *Context-driven:* Domain knowledge from **research papers** on HG was used to explain the disease's background, i. e., its biological and genetic causes. Laura Garrison, Marc Vaudel and Jone Trovik advised me on the most effective metaphor and character design choices for conveying the severity of the disease, as well as how to sequence information within the story. For example, in narrative medical visualization, information on prevention or treatment options is usually presented later in the story [72, 75]. After speaking with the **experts**, I decided to place crucial information on medication opportunities earlier in the story. However, in some extreme cases, such as the one described in this true story, medication does not alleviate the symptoms. In these cases, women urgently need additional help and hospitalization to get through the illness. **Blogs** on HG support websites were a rich source of information, offering insights into the symptoms, how patients experienced this devastating time, and how they felt when their child was born.

- *Character-driven:* The **patient character** was inspired by a blog entry from an Australian HG support website where I found a story that gave detailed insights into the disease process and emotional impact. Originally, I wanted to tell the woman's story through this data story. However, due to difficulty contacting the **author of the blog entry**, I had to transform the real character into a fictional one. More details are described in Section 6.3.4.

6.3.3 *Conflict*

The story begins with the main character looking forward to the pregnancy. However, the conflict between the protagonist and the antagonist (in this case, the disease) emerges early in the narrative when the character realizes that her symptoms began just before the fourth week of pregnancy. The conflict escalates over dehydration and severe symptoms despite taking medication, until she finally declares that HG feels like she is dying and is admitted to the hospital.

6.3.4 *Character Design*

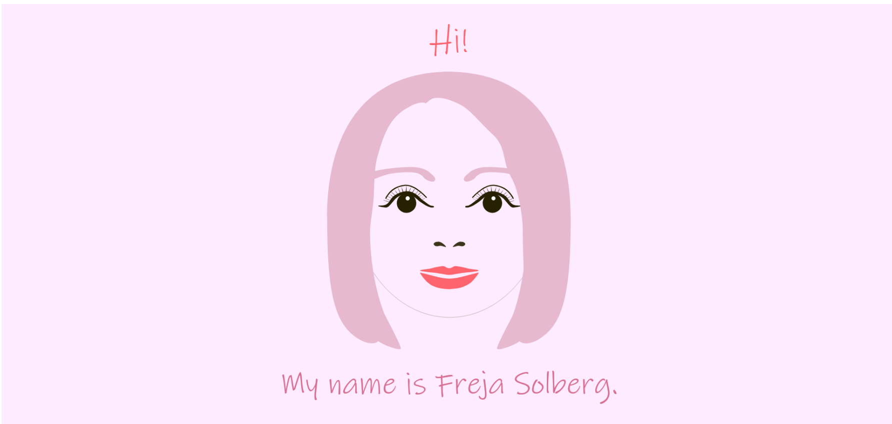


Figure 6.1: Character design for the protagonist introduction.

Character inspiration

Throughout the individual story, the patient protagonist is represented both visually, through illustrations or icons, and in the text. The character was developed based on a real HG patient who wrote a blog entry on an HG support website [60]. In this post, the woman described how her symptoms first appeared in the early stages of pregnancy and how they persisted throughout the entire pregnancy. She also described many accompanying symptoms, such as sleep problems, fainting due to drowsiness, and physical and psychological pain. These symptoms illustrate the severity of the disease. However, I chose this story mainly because of her emotional account of repeatedly reaching out to a hospital for support. In later stages, she was rejected by doctors who told her that her problems were psychological. As I discovered through my research, doctors often mistake HG for normal nausea and vomiting. I originally wanted to tell this woman's story because of her persistence. She would have been a great role model for other women. Unfortunately, I was unable to contact her to ask for her permission to use her story for potential publication, so I decided to change the character's name and appearance.

Character name

To transform the real character into a fictional one, first, I changed the character's name. Finding an appropriate name for the character went through a process, from a more American-sounding name *"Mila Baxter"* to a more Norwegian-sounding name *"Freja Solberg"*. Norwegian women stated that they could not relate to the name "Mila Baxter" because the letter X is not part of the Norwegian alphabet. One lesson learned is that a character's name influences the overall perception and identification with the character. Adjusting the name for usage in different cultural contexts could benefit character identification. The name *"Freja Solberg"* was chosen based on the following considerations: Norwegians are known for being proud of their cultural heritage. They are descendants of pagans who worshiped the Norse gods and some still feel connected to these beliefs today, expressing this through arts and music. In Norse mythology, *Freja (eng. Freya)* is the goddess of fertility and love, fitting the story's topic. The last name *"Solberg"* – a composed word of *"sun"* and *"rock"* or *"little mountain,"* which is a metaphor for the idea that, *"despite the difficulties of this time, there is light — the child being born."*

Figure 6.2: Character design for the protagonist, who is experiencing the effects of the disease.

Visual design

I opted for a visual design approach with a flat style in order to have as much freedom as possible in illustrating the effect's of the disease and emotions of the character through facial expressions. The first illustrations matched the appearance of the real woman who posted a photo of herself with her family. To transform her into a fictional character, I changed the hairstyle and hair color. Facial features were also slightly changed. According to the topic, the femininity of the character was emphasized in the facial appearance and by the use of a pastellish-reddish color palette in the introduction (Figure 6.1). As the story should transport the emotional experience the protagonist goes through, I filled approx. 70 % of the screen height with the head of the character. Similar to Mittenentzwei et al. (2024), who found that participants associated eye bags, wrinkles, and an unhealthy skin color with poor health [74], I progressively added these features to the character's face as her condition worsened. Examples are provided in Figure 6.2 and Figure 6.3. Green skin represents nausea. Wrinkles and cracked lips represent dehydration. Eye bags are associated with sleep problems and exhaustion. Sunken cheeks symbolize the unhealthy weight loss experienced by most women with HG. One of my supervisors proposed using a more diverse character, i. e., a colored person. However, it did not seem meaningful to use a character, representing a minority in Norway for a story based on Norwegian data. However, this approach could be taken into account for a more broad story, told through the lens of several women.

Figure 6.3: Character design for the protagonist, who is experiencing the effects of the disease II.

Scripting Large parts of the story were scripted based on the text from the blog post. The sequence of this blog story was mostly maintained. The first-person perspective was adopted for the individual story to tell an authentic story with an internal, emotional view of the experience.

6.3.5 *Narrative Structure*

First, I will describe the narrative structure of the individual story. In Section 6.3.7, I will explain how the story was adapted for the general story approach. Each story piece (slide) transports at least one message or provides context for subsequent messages, and is supported by data visualization or illustration. For the final prototypes, see Section 6.6.

Story Title: "Surviving Hyperemesis Gravidarum"

1. Introduction: In the first part, the protagonist introduces herself, saying, *"Hi! My name is Freja Solberg."* An illustration of her joyful face is shown. A reddish color palette is used to convey positivity and femininity. Then, the topic of pregnancy is introduced with a thought bubble depicting a woman about to give birth. This illustration was inspired by a video in which a woman showed her positive pregnancy test to the camera with tears of joy in her eyes.

Conflict

2. Spiral of Escalation: The second part establishes the conflict between the protagonist, Freja, and the antagonist, HG (the disease). The conflict begins at a moderate level, such as stating that *"it turned into a nightmare,"* showing her frightened face, and escalates to a higher, more severe level.

Disease definition

Treatment

This part conveys the most important information about HG from a relatable first-person perspective. For example, it defines the disease, describes the early onset of nausea and vomiting, and emphasizes the importance of seeking treatment, such as visiting a doctor, taking anti-nausea medication, and considering a hospital visit for fluids and medication. To visually emphasize the change in scene from the positive introduction to the escalation of the conflict, the background color shifts from pastel pink to dark violet, and finally to dark blue. Changes to the character's appearance illustrate her HG symptoms, including distress. These symptoms are supported by first-person text. Her face and lower lip turn green to symbolize nausea and vomiting. The same color is used in the data visualization for this symptom. Other health changes due to the symptoms are illustrated by an increasing number of wrinkles, tired-looking eyes from fatigue, cracked lips from dehydration, and a sunken face from weight loss. Finally, she is shown drowning in an ocean, symbolizing the lowest point woman may experience during this time. The text states that she is admitted to a hospital.

3. Explaining the Disease: The third part explains the disease based on the research of Fejzo and colleagues. Here, the story varies from the original blog post. This decision was made because, in the original story, the doctors were unaware of the disease. This story aims to address this lack of knowledge, so educational material was incorporated as a central aspect.

- The protagonist finally receives an education about the disease and gains an understanding of her situation. The visual style is lighter and clearer, providing a focus on studying the information rather than evoking strong emotions. The text briefly explains that the disease is caused by GDF15, a growth factor and hormone released by the growing placenta. GDF15 travels through the bloodstream to the brain, where it triggers receptors in the vomiting center of pregnant women.

High levels of this hormone were found in the blood of HG patients. To illustrate the elements and mechanism of the disease, an illustration in Fejzo et al. (2019) [34] was used as a template. This illustration was simplified by removing unnecessary details, to avoid overwhelming the viewer. The illustration was initially static, but was then refined with animation to guide the viewer through the process.

Animated medical illustration

• A second story piece communicates the message that *"Genetically, some women have a greater sensitivity to GDF15, which is linked to persistent nausea and vomiting."*

4. **Data-driven Insights:** This section begins with an informational slide that provides details about the MoBa data, e. g., name of the long-term cohort study, number of participants, and enrollment period.

Introduction data visualization concept

• **Data Visualization Introduction:** Here, the concept of the data visualization concept is introduced by the question *"How do women with HG feel?"* and the ***Iceberg Metaphor***. This metaphor shows the most prominent symptoms most commonly associated with HG above the horizon line (x-axis) and the less obvious symptoms below it. To emphasize the emotional impact, I repeated the ocean blue background from the second part and displayed Freja's head melting with the underwater part of the iceberg. All symptoms are depicted in a specific color used in the data visualizations and are annotated with their names. • Since the final data visualization might overwhelm the general public, I used the incremental construction technique to introduce elements of the visualization step by step. The left side of each page in this section is reserved for the visualization, while the right side provides explanatory text. • First, the x-axis is introduced, with each month illustrated by the stage of the developing embryo. • Next, the y-axis is displayed, annotated with the absolute number of women who took part in the cohort study.

Incremental construction

Pictogram-based annotation

• **Data Visualization Part:** On the third page, the ***area chart*** is applied, showing the data for the main variable, i. e., nausea with vomiting during pregnancy for each month. The chart's title, *"Nausea with Vomiting,"* remains stable throughout the data visualization part. Each area has a unique, distinguishable color chosen to represent the associated symptoms. Color and smoother lines were also used to make the data visualization more emotional. For instance, green can be associated with disgust and is therefore assigned to nausea and vomiting. Each data point is interactive – by hovering over a circle the number of percentage is displayed in a tooltip. • The fourth page considers women who were hospitalized due to nausea and vomiting. The area chart, shown in signal red (association with the red cross symbol), is barely visible due to the scale of 102,080 women. I used two techniques to illustrate the number of women likely affected by HG. First, I stated the total number of women (3,084) and percentage (3 %) in text narration. Second, I animated the data visualization with

Semantic color coding

Interactive data points

Zooming

a zoom-in function. Users can activate the animation by clicking the zoom-in button. • The fifth page emphasizes that nausea and vomiting usually subside by the fourth month, but can persist throughout pregnancy for HG patients. This key aspect is highlighted in the area chart. The question *"What is below the iceberg?"* introduces the first accompanying symptom, 'unusual fatigue and drowsiness,' below the x-axis as the dependent variable of 'nausea and vomiting,' and is being displayed in relative percentages. A dull, unsaturated color was chosen to match the feeling of drowsiness and extreme fatigue. The right side of the page, as well as an annotation on the y-axis, provides a less technical explanation of how to read the lower part of the chart and offers a numerical example. The symptom is also displayed above the x-axis relative to the total number of study participants. This area chart allows for a better comparison of the data points to the dependent variable and works well for the most common side-symptom. However, when side-symptoms are less common, the data points vanish against zero percent; therefore, the symptoms are shown below as 'relative percentages'. An icon with Freja's face appears in the lower area chart. A glowing effect indicates interactivity. When the user hovers over the icon, a tooltip appears, showing an illustration of the character's face that fits the symptom, as well as a quote associated with the symptom. • The seventh page introduces the side effect 'headache.' I chose a color close to red which is most commonly associated with pain but already used for hospitalization. • The next page displays 'sleeping problems' (night-blueish color) and 'depression' (black) together as psychological problems. I explain that data on sleeping problems are missing from the fourth month on and visualize this aspect with a 'fading out' effect after the last data point.

Engaging question

Explicit instructions

Humans behind data

Missing data

Comparative visualization

• **Comparative Visualization:** The next two slides juxtapose the focus group with women only reported nausea and neither nausea nor vomiting. • Only nausea is displayed with a less saturated greenish color. • The percentages of women who did not report neither nausea nor vomiting are displayed in pink and show milder symptoms.

5. Resolution: The resolution reveals the protagonist's outcomes. The text was inspired by the final portion of the HG patient's post, which highlights the relief and positive emotions she experienced when her child was born. The visual style is similar to that of the introduction to connect to the original positive atmosphere. This ending was chosen to end on a hopeful note. • **Call to Action**: Additionally, the main message was conveyed through a call to action, encouraging individuals to seek help and ask a doctor for medication.

7. Sources & Credits: The last three pages provide links to a hyperemesis gravidarum support website, thank the MoBa study and its participants for providing the data, and offer background information about the development of the story.

6.3.6 *Narrative Techniques*

Narrative techniques in data storytelling are methods used to craft a compelling and coherent story around data to engage, inform, and persuade an audience. These techniques bridge raw data and human understanding, making complex information accessible and memorable. The following narrative techniques, adapted from Segel & Heer (2010) and Lan et al. (2022), have been used to design the data stories [55, 97]. While Segel & Heer systematized techniques for data storytelling in general, Lan et al. (2022) focused on serious data storytelling techniques that can be used to elicit emotions in the viewers.

1. *Narrative Structure Techniques*

   - *Linear Ordering:* This author-driven technique was used to maintain a clear narrative structure.
   - *Consistent Visual Platform:* The layout concept was carefully structured to support orientation.
   - *Incremental Construction:* This technique was used to gradually reveal information towards the final more complex data visualization and ensure understanding of its arguments, set the pace and rhythm of the narrative flow.
   - *Explicit Instruction:* Instructions on how to read the data visualization are necessary for uncommon data visualizations, such as the one used in this story.
   - *Messaging Techniques,* such as the use of headlines, annotations, engaging questions, multi-messaging, message repetitions, and introductory texts.
   - *Highlighting Key Insights:* Critical data points are emphasized to clearly communicate messages. For example, using saturated versus unsaturated colors for parts of the area chart to set focus.
   - *Emotive Language:* Large parts of the text transports emotions through the protagonist experience or women's quotes.
   - *Hover highlighting / Details:* With this reader-driven technique, users can freely interact with and explore the data visualization.

2. *Visual Narrative Techniques*

   - *Visual Metaphors:* Pairing data or information with imagery that supports messages, e. g., the Iceberg metaphor, ocean metaphor.
   - *Humans behind Data*: This technique aims to build empathy in users by adding the protagonist's face and quotes to the data visualization.

- *Zooming/Animation:* A zoom-in interaction was used to highlight hidden aspects of the data, i. e., a small but meaningful portion of the data. Animation was found to have an emotional effect [55, 67].
- *Semantic Color Coding:* Using colors associated with semantics can convey meaning, emotions, and atmosphere.
- *Dark Background*: This technique can evoke negative emotions and leave the viewer feeling melancholic. For example, the background turns dark purple when the protagonist is confronted by the antagonist (the disease).
- *High-saturated Colors*: Signal red can arouse emotions, such as alarm and fear.
- *Pictures with Negative Semantics:* Many of the illustrations are designed to evoke negative emotions, e. g., by expressing the protagonist's suffering through facial expression and shapes.

6.3.7 *Design of the Story Versions*

To better understand the effect of individual characters, I created two versions of the story that contain the same messages and information. The only difference between the two stories is the character approach. The ***independent variable*** is defined by the use of an individual protagonist (individual story) versus the absence of an individual human character (general story). The personalized story is told from the first-person perspective of a former patient, while the second version tells a more general story about hyperemesis gravidarum.

The following changes were made to transform the individual story into a ***general story***: The text was reformulated into third-person perspective, changing as little as possible. For example, *"It was supposed to be the happiest time of* ***my*** *life"* became *"It was supposed to be the happiest time of* ***their*** *lives."* Instead of introducing an human protagonist, the disease itself is presented as the central character. An infographic depicting one hundred simple icons of women's heads showcases its prevalence. The icons are arranged to form a question mark, highlighting those who are proportionally affected by HG.

To illustrate the messages' objective, all illustrations of the human character were replaced with an icon-based style. This style is similar to that of the personalized story and maintains consistent size, format, color scheme, and design quality. Maintaining consistent design quality in both stories is crucial because uneven quality could create a positive bias for the more visually appealing story. These methods should ensure that the influence of character usage (the independent variable) can be isolated as much as possible.

6.4 PROTOTYPING TECHNIQUES

In this section, I will describe the techniques used to realize the story concept. Creativity techniques such as brainstorming, sketching, illustrative drafting, and exploratory data visualization were used to develop and create the story design. The development went through iterative feedback loops with data visualization experts, domain experts, and pilot users as described in Section 6.5. The cohort data was preprocessed using Python. The first charts were created using Python in Jupyter Notebook to get a first impression of the data. Then JavaScript and the D3.js library [11] were used to build the area charts, including labels, axes, interactive elements (hover and click elements, tooltips), and animation. To smooth the line segments of the area chart at each data points, a cardinal spline for controlling the tangents was implemented with a tension parameter $t = 0.6$ (tighter curve). Inkscape [50] was used to create the illustrations and animation images for the medical illustrations. The final animation for the medical illustration was created using GIMP [39]. All elements were integrated into a ***web-based, interactive slideshow*** prototype using HTML5 and CSS5.

6.5 PRE-PILOT STUDY

I evaluated both stories in a qualitative pre-pilot study in Bergen in Spring 2024. This preliminary study was part of the human-centered design approach to evaluate user experience with the aim to improve both prototypes for the final evaluation. I translated the individual story into Norwegian for one Norwegian participant to feel more comfortable. The evaluation was conducted in two parts. First, I evaluated the individual story with two participants, a Hyperemesis Gravidarum researcher and a former HG patient. The second part of the study was conducted after the individual prototype was refined and the general story in a preliminary state. I evaluated both version with four other participants from diverse background, including another former HG patient, a data visualization expert, and two other persons from broad audience. Since, all study participants were female, I planned to include more male participants in the final experiment.

The procedure was explained before participants started to read the stories on a laptop. They were asked to think-aloud to express immediate feedback, including thoughts and feelings. Afterwards, I conducted a short interview about their overall experience with the story and iteratively ask whether they encountered any problems during the exploration, regarding user experience or comprehension. The first two sessions were documented in writing, while the last four were audio recorded.

Key insights from the pre-pilot study (Part 1)

P1: Female, age: 63, HG researcher

- *mentioned that the story's title "Surviving Hyperemesis Gravidarum" is appropriate*
- *suggested to include information about medication earlier in the story*
- *character design does not reflect the extreme experience, suggested, e. g., a greener face, hair style less perfect*
- *had no problems to interpret the data visualization*
- Main insight: The character design had to be refined to better convey the HG experience.

P2: Female, age: 28, Former HG patient

- *felt connected to the story, reflected on her own experience*
- *had problems interpreting the data visualization but could follow when the visualization was explained step by step*
- Main insight: The data visualization needs more explicit instructions and incremental construction.

Key insights from the pre-pilot study (Part 2)

P3: Female, age: 28, Former HG patient, Student in nutritional science

- *felt well understood as a former patient*
- *wished that she had received informative support when she was affected, suggested the use of the story for prenatal care*
- *«Nausea is green for us» (comment on color usage)*
- *recognized the Iceberg metaphor and stated its appropriateness*
- *Incremental Chart Construction helped with interpretation*
- *suggested more annotation of the anatomical illustration*
- Main insights: absolute to relational percentages not clear, additional annotation for anatomical parts

P4: Female, age: 60, Librarian

- *had a little background knowledge about the disease*
- *had no problems interpreting the data visualization and the absolute to relational percentages*

P5: Female, age: 23, Design student

- *was not familiar with the disease*
- *expressed being emotionally moved and empathetic while reading the story and quotes*
- *mentioned that this information, in particular the percentage of women that suffer from the disease, should be more publicized*
- *exploratory behavior, liked quote interaction*
- *wished for more interactivity*
- Main insights: problems interpreting the relative percentages, Y-truncation was a little tricky

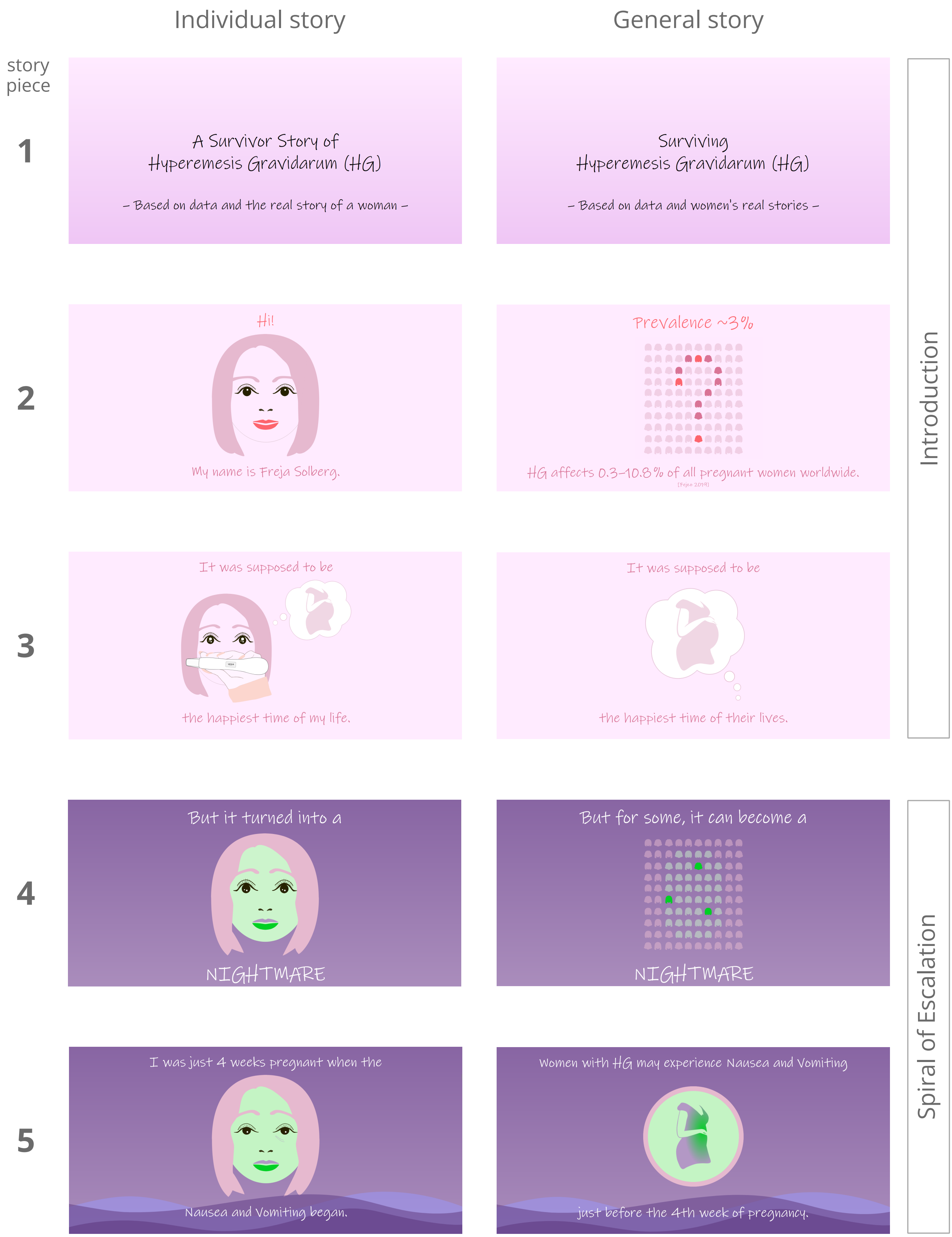
P6: Female, age: 25, PhD-student in visualization

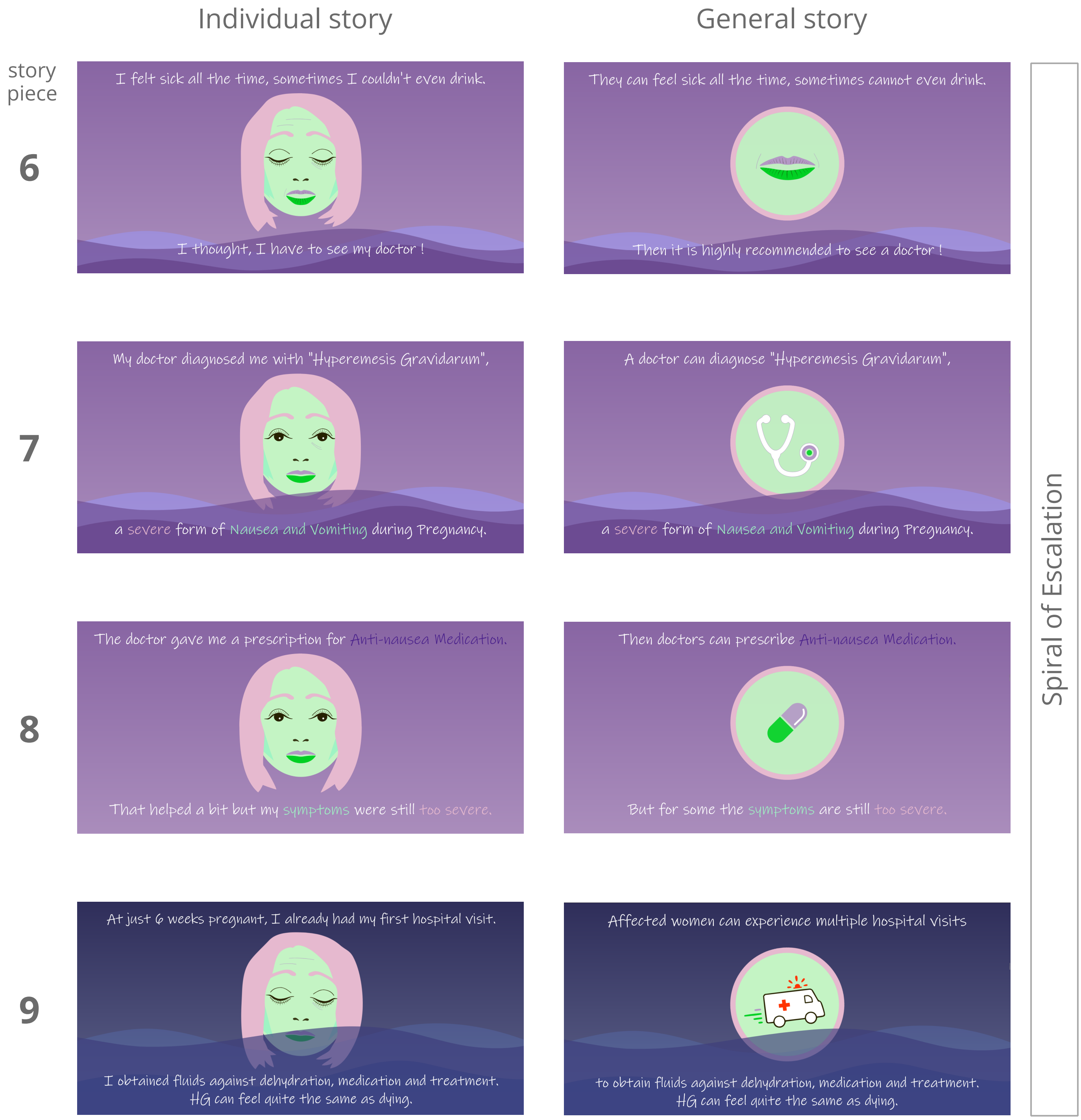
- *stated she stayed engaged throughout the story*
- *expressed surprise about the use of quotes and animation*
- *story felt not too long*
- *expressed being emotionally moved and empathetic while reading the story and quotes*
- Issues: mentioned problems with data visualization annotation, and reading the chart at the bottom

Participants reported feeling engaged by both versions of a study, particularly appreciating the use of quotes, data visualizations, and animations; however, they felt a stronger connection to the individual story. Pre-pilot study findings highlighted two main issues: unclear instructions for reading lower charts and significant differences between the individual and general stories. To address these, annotations were added to the charts, and a consistent layout and graphic design were implemented across both story types, ensuring only the character approach varied.

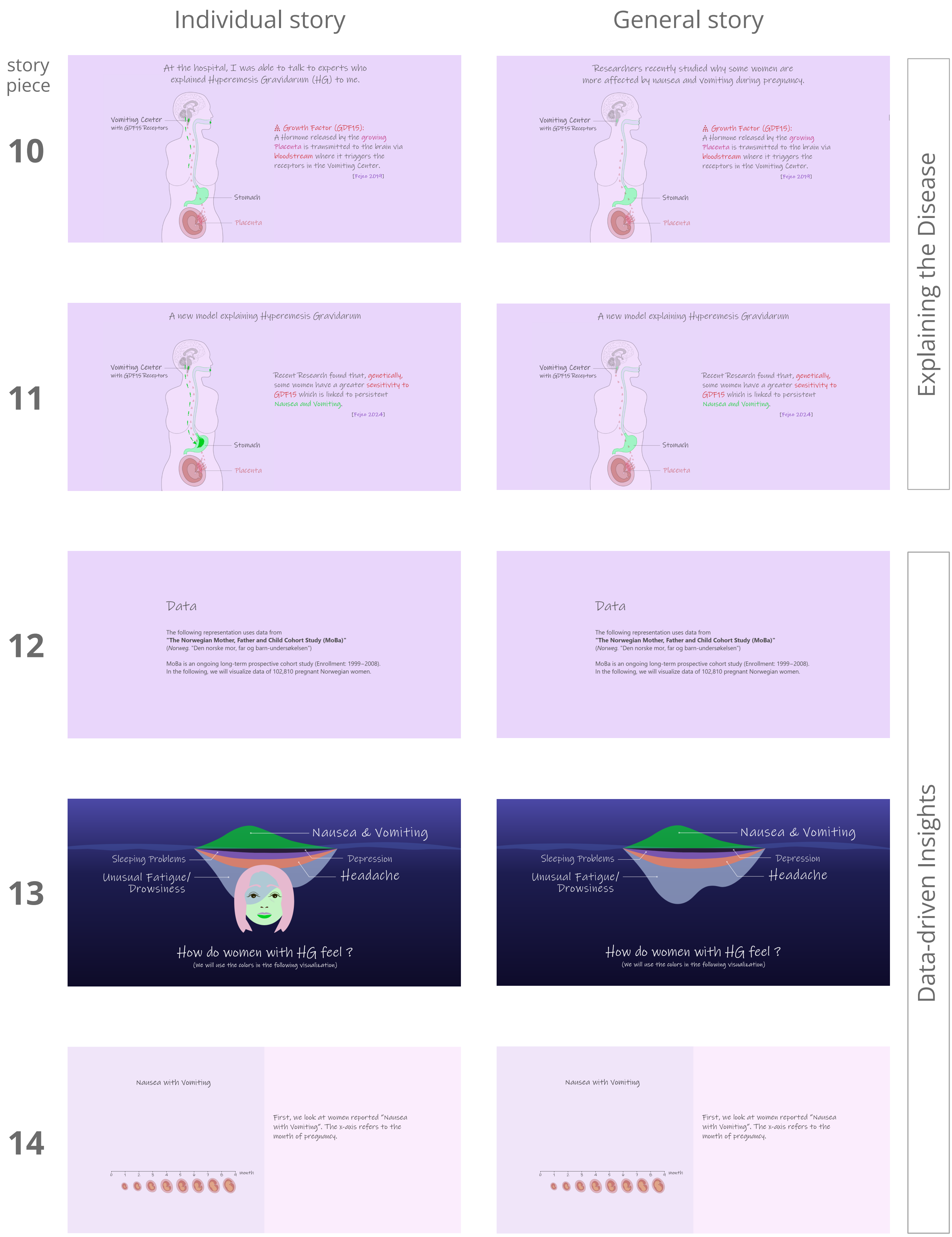
6.6 REALIZATION OF THE STORY VERSIONS

On the following pages, I will present the two revised versions of the story prototypes. The left-hand column includes all the story pieces of the individual story, which are juxtaposed with the corresponding story pieces of the general story in the right-hand column. All story pieces are labeled according to their respective part of the story. Please note that Slide 29 was not included in the subsequent analysis, as it only provides background information about the project. Some story pieces are shown twice if they include interactive elements.


Individual story
General story
story piece
1
A Survivor Story of
Hyperemesis Gravidarum (HG)
– Based on data and the real story of a woman –
Surviving
Hyperemesis Gravidarum (HG)
– Based on data and women's real stories –
2
Hi!
My name is Freja Solberg.
Prevalence ~3%
HG affects 0.3–10.8% of all pregnant women worldwide.
3
It was supposed to be
the happiest time of my life.
It was supposed to be
the happiest time of their lives.
Introduction
4
But it turned into a
NIGHTMARE
But for some, it can become a
NIGHTMARE
5
I was just 4 weeks pregnant when the
Nausea and Vomiting began.
Women with HG may experience Nausea and Vomiting
just before the 4th week of pregnancy.
Spiral of Escalation



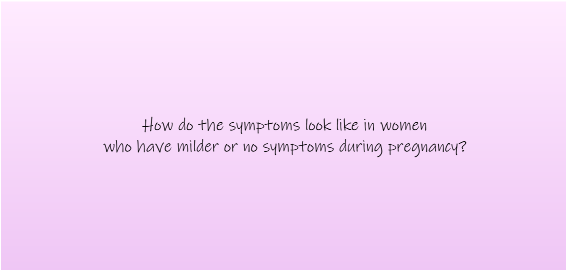
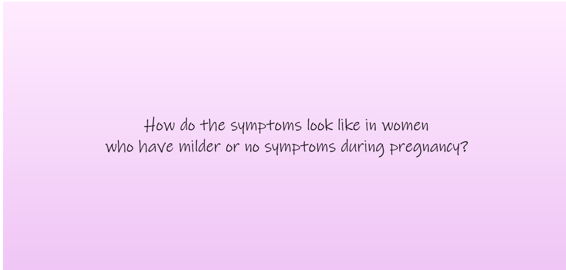
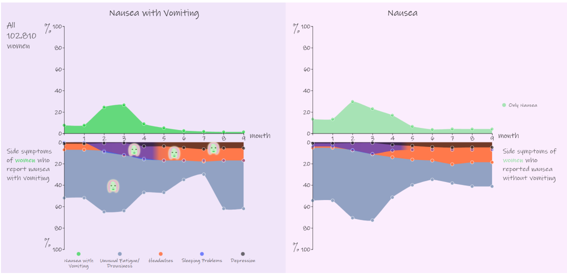
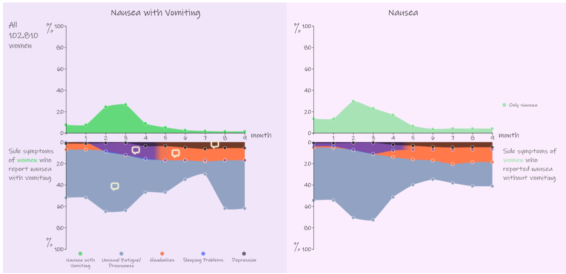
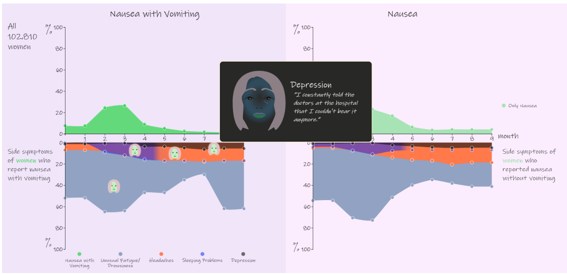
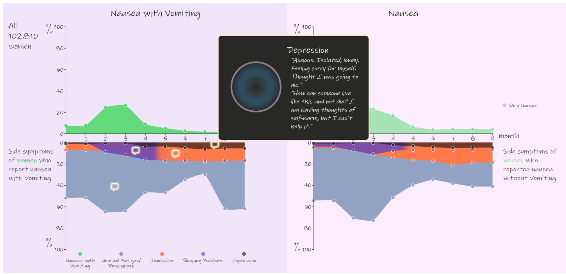
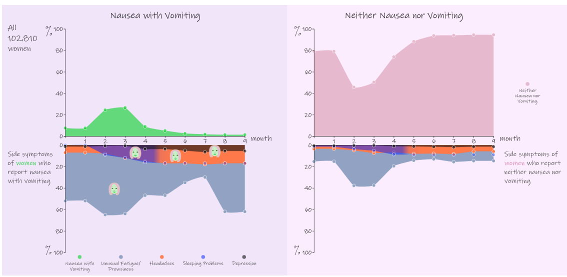
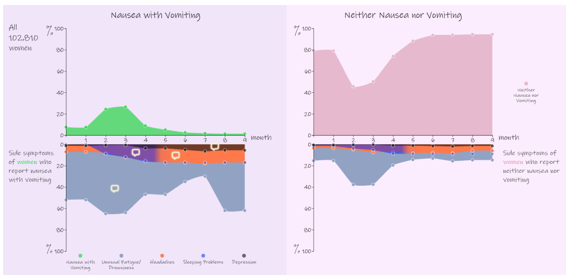
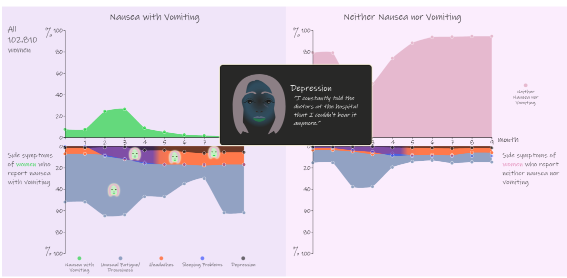
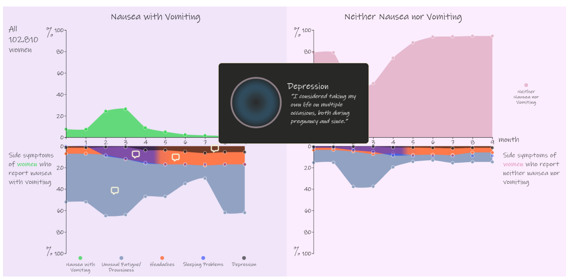
Individual story
General story
story piece
6
I felt sick all the time, sometimes I couldn't even drink.
I thought, I have to see my doctor !
They can feel sick all the time, sometimes cannot even drink.
Then it is highly recommended to see a doctor !
7
My doctor diagnosed me with "Hyperemesis Gravidarum",
a severe form of Nausea and Vomiting during Pregnancy.
A doctor can diagnose "Hyperemesis Gravidarum",
a severe form of Nausea and Vomiting during Pregnancy.
8
The doctor gave me a prescription for Anti-nausea Medication.
That helped a bit but my symptoms were still too severe.
Then doctors can prescribe Anti-nausea Medication.
But for some the symptoms are still too severe.
9
At just 6 weeks pregnant, I already had my first hospital visit.
I obtained fluids against dehydration, medication and treatment.
HG can feel quite the same as dying.
Affected women can experience multiple hospital visits
to obtain fluids against dehydration, medication and treatment.
HG can feel quite the same as dying.
Spiral of Escalation



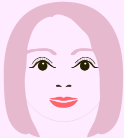
Individual story
General story
story piece
10
At the hospital, I was able to talk to experts who explained Hyperemesis Gravidarum (HG) to me.
Vomiting Center with GDF15 Receptors
Stomach
Placenta
Growth Factor (GDF15):
A Hormone released by the growing Placenta is transmitted to the brain via bloodstream where it triggers the receptors in the Vomiting Center.
[Fejzo 2019]
Researchers recently studied why some women are more affected by nausea and vomiting during pregnancy.
11
A new model explaining Hyperemesis Gravidarum
Recent Research found that, genetically, some women have a greater sensitivity to GDF15 which is linked to persistent Nausea and Vomiting.
[Fejzo 2024]
12
Data
The following representation uses data from
"The Norwegian Mother, Father and Child Cohort Study (MoBa)"
(*Norweg.* "Den norske mor, far og barn-undersøkelsen")
MoBa is an ongoing long-term prospective cohort study (Enrollment: 1999–2008).
In the following, we will visualize data of 102,810 pregnant Norwegian women.
13
Nausea & Vomiting
Sleeping Problems
Depression
Unusual Fatigue/ Drowsiness
Headache
How do women with HG feel ?
(we will use the colors in the following visualization)
14
Nausea with Vomiting
month
First, we look at women reported "Nausea with Vomiting". The x-axis refers to the month of pregnancy.
Explaining the Disease
Data-driven Insights


Individual story
General story
story piece
15
Nausea with Vomiting
Now, the y-axis shows the percentage of 102.810 mothers that participated in the MOBA study.
Nausea with Vomiting
Now, the y-axis shows the percentage of 102.810 mothers that participated in the MOBA study.
16
Nausea with Vomiting
102.810 women in total
This is an area chart over time (months). The green area shows the percentage of women suffered from "Nausea with Vomiting" during pregnancy.
Nausea with Vomiting
102.810 women in total
This is an area chart over time (months). The green area shows the percentage of women suffered from "Nausea with Vomiting" during pregnancy.
17
Nausea with Vomiting
102.810 women in total
Zoom in!
Red shows the amount of women "hospitalized" due to long-term Nausea and Vomiting (overall 3 %, in total 3084 women). However, researchers assume that many HG cases are not reported.
Nausea with Vomiting
102.810 women in total
Zoom in!
Red shows the amount of women "hospitalized" due to long-term Nausea and Vomiting (overall 3 %, in total 3084 women). However, researchers assume that many HG cases are not reported.
18
Nausea with Vomiting
102.810 women in total
As you can see, the percentage of women affected, decreases in the fourth month. But for women with HG it can last the entire pregnancy.
Nausea with Vomiting
102.810 women in total
As you can see, the percentage of women affected, decreases in the fourth month. But for women with HG it can last the entire pregnancy.
19
Nausea with Vomiting
All 102.810 women
Side symptoms of women who report nausea with vomiting
Unusual Fatigue/ Drowsiness
What is below the iceberg?
HG affects the whole body and mind. The most common symptom is unusual fatigue and drowsiness.
Note that in the chart below we only show the subgroup of women reported nausea with vomiting.
For example, if women reported nausea with vomiting in the first month, 52 percent of them also reported unusual fatigue and drowsiness in the first month.
Nausea with Vomiting
All 102.810 women
Side symptoms of women who report nausea with vomiting
Unusual Fatigue/ Drowsiness
What is below the iceberg?
HG affects the whole body and mind. The most common symptom is unusual fatigue and drowsiness.
Note that in the chart below we only show the subgroup of women reported nausea with vomiting.
For example, if women reported nausea with vomiting in the first month, 52 percent of them also reported unusual fatigue and drowsiness in the first month.
Data-driven Insights


Individual story
General story
story piece
20
20
21
21
21
Data-driven Insights
Nausea with Vomiting
Nausea and Vomiting over a longer period cause physical stress and can lead to life-threatening dehydration, which causes headaches.
Headache
Sleeping problems and depression are psychologically serious problems that require immediate medical and psychological help.
(Note that the data for sleeping problems from month 4 to 8+ is missing)
How do the symptoms look like in women who have not experienced HG?
Sleeping Problems
Depression

	Individual story	General story
story piece		
22		
23		
23		
24		
24		

Data-driven Insights

story piece	Individual story	General story	
25	 I will never forget how traumatic it was for me. But then my beautiful son was born. Through a very hard pregnancy my son never stopped growing and fighting. He was so strong throughout it all and gave me the strength to not give up.	It can be traumatic to experience such a hard pregnancy. But when the mothers finally hold their child in their hands, they are happy and proud.	Resolution
26	Don't hesitate to seek help, like Freja did ! Ask your doctor for advice, Antacids or Anti-nausea Medication that will alleviate your symptoms.	Don't hesitate to seek help if you have similar symptoms to these women ! Ask your doctor for Antacids or Anti-nausea Medication that will alleviate your symptoms.	Resolution
27	For more Information and Stories about HG experiences visit: www.hyperemesis.org	For more Information and Stories about HG experiences visit: www.hyperemesis.org	Sources & Credits
28	The Norwegian Mother, Father and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. We are grateful to all the participating families in Norway who take part in this on-going cohort study. MoBa	The Norwegian Mother, Father and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. We are grateful to all the participating families in Norway who take part in this on-going cohort study. MoBa	Sources & Credits
29	Data Visualization, Storytelling, Illustration, Animation, Design & Webdevelopment by Bea Budich (budich[at]ovgu.de) A supervised project by Universitetet i Bergen, Norway and Otto-von-Guericke-Universität Magdeburg, Germany.	Data Visualization, Storytelling, Illustration, Animation, Design & Webdevelopment by Bea Budich (budich[at]ovgu.de) A supervised project by Universitetet i Bergen, Norway and Otto-von-Guericke-Universität Magdeburg, Germany.	Sources & Credits

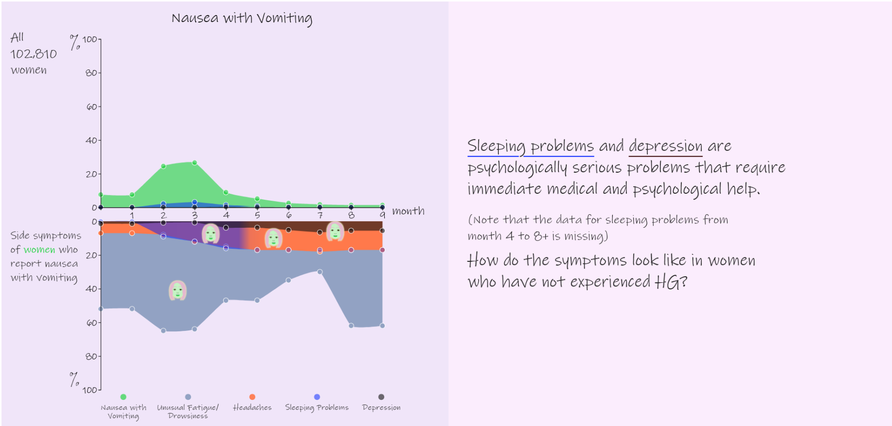


Figure 6.4: Individual story (storypiece 21): Using the Iceberg metaphor to show accompanying symptoms that affect the whole body and mind of affected women.

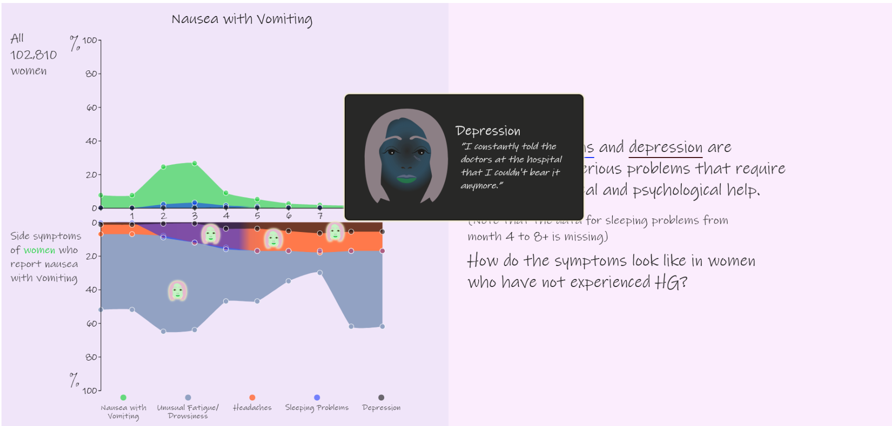


Figure 6.5: Individual story (storypiece 21): Interactive icons and tooltips with quotes from blogs to embed real stories in the area charts.

Storypiece 21 (Figure 6.4) shows the final visualization after incremental construction. The data visualizations, showing "Nausea with Vomiting", are consistently placed on the left, while the right is reserved for explanations of the visualizations or additional information about the symptoms and actions to take if they occur. Each symptom is semantically assigned to a color and annotated twice: in a legend below the chart and by tooltips. Each data point has an integrated tooltip that shows the percentage. Figure 6.5 represents the interactive view of the same story piece. Each area below the x-axis has an animated icon in the form of the protagonist's head with a glowing effect to indicate interactivity. When hovering over an icon, a quote from the blog entry illustrates how the protagonist feels. The illustration in the tooltips are designed based on the emotional impact of the symptom.

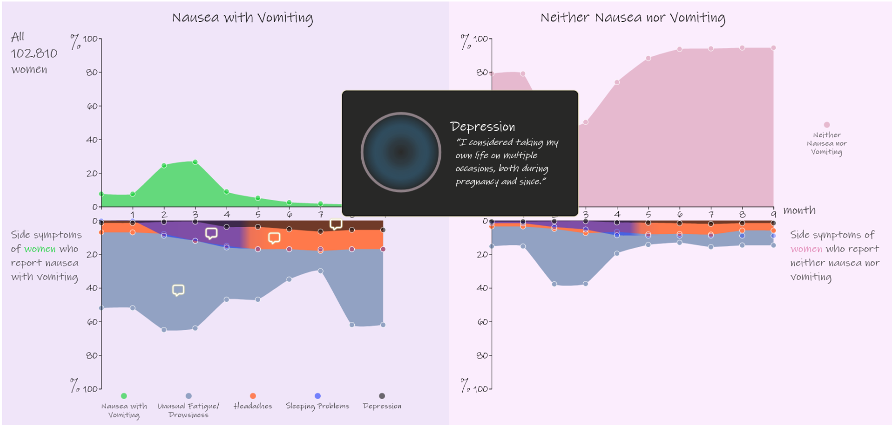


Figure 6.6: General story (storypiece 24): Icons and Quotes.

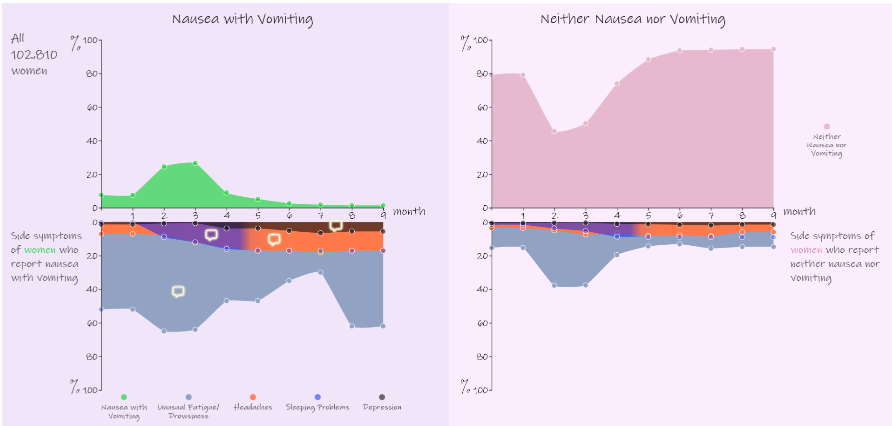


Figure 6.7: General story (storypiece 24): Comparative Visualization.

Figure 6.6 shows that the character icons have been replaced by word bubbles with the same glowing effect. As previously described, the disease is life-threatening due to metabolic imbalance and dehydration. This story also includes the psychological severity of the disease, citing a woman who considered taking her own life. It also indicates that the condition can have long-lasting effects on affected women, emphasizing the need for psychological therapy.

Storypiece 24 (Figure 6.7) uses a comparative visualization technique that juxtaposes two data subsets. On the left are women who experienced nausea and vomiting, along with other symptoms, in a given month. The right side shows women who did not experience nausea or vomiting in the same month. The lower chart shows that reported symptoms were less frequent in this group.

The next two chapters will describe the study design, including the measurement instruments (Chapter 7) and the experiment (Chapter 8).

7 INSTRUMENTS

This chapter describes the instruments and methods used in the study to measure emotional arousal and valence, and to assess user engagement, understanding, and memory.

7.1 ELECTRODERMAL ACTIVITY

This study uses electrodermal activity as the core instrument to measure emotional arousal continuously throughout each session.

Sensor Device: The experiment was conducted with the **EdaMove 4** sensor [41] developed by movisens GmbH, Karlsruhe, Germany [44]. This instrument records the Electrodermal Activity (EDA), i. e., skin conductance, as well as other parameters such as skin temperature, physical activity, position, and acceleration to capture motion. The sensor's wearable design is suitable for use during motion, e. g., while performing daily activities or exercising. Therefore, participants must not feel restricted in keeping their hand still during the experiment, as has been necessary in the past. The sensor uses the Exosomatic Measurement Method [29] and applies a Direct Current (DC) voltage of 0.5 V to the skin to measure skin conductance. The sensor receives a single channel signal from the current flow between the two silver–silver chloride (Ag/AgCl) electrodes. It is recommended that the electrodes be applied to the palms of the hands.

Software: Three software packages are needed to manage the sensor, preprocess, and analyze the data. The sensor was connected via USB to the sensor managing interface **SensorManger** [42]. The software was used to check the charging state, to start the sensor, to read out, and store the data. The raw data in unisens format is stored as *unisens.xml*. This file can be preprocessed and viewed with **UnisensViewer** [43] software, which provides line charts to give an initial idea of the data. To analyze the data the **DataAnalyzer** [40] from movisens can be integrated into the UnisensViewer. DataAnalyzer uses AI-based algorithms to decompose the raw EDA signal into tonic and phasic components. These derived components are called skin conductance level (SCL) and skin conductance responses (SCR), respectively. Artefact reduction on EDA data is performed by applying a built-in low-pass filter at <10 Hz in order to remove high-frequency noise. Overlapping skin conductance responses are identified by decomposing them into distinct components [23]. Additionally, it uses data from other integrated sensors that measure barometric pressure, temperature, and activity to help identify and isolate artifacts.

Resolution: The EDA resolution is 14 bits and the data analysis was done with the smallest possible time resolution that is 1 second.

Physical Unit: Microsiemens (µS) is the unit that indicates how well electricity is conducted through a liquid (sweat on the skin).

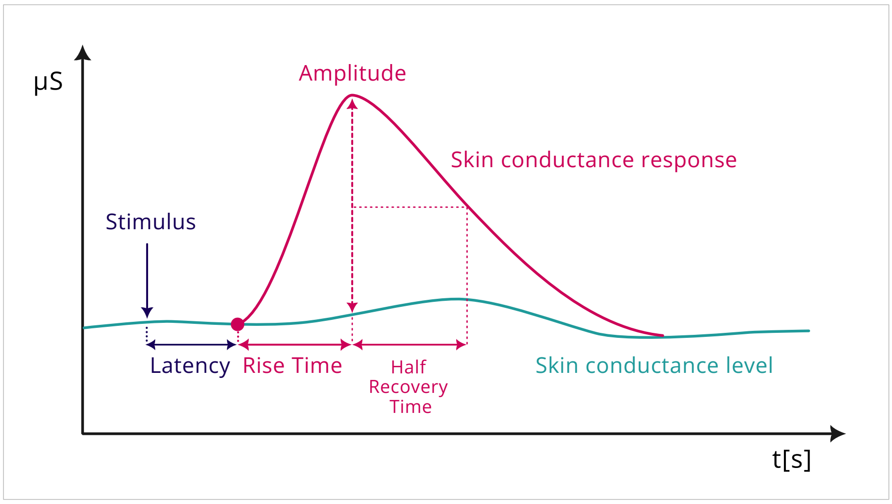


Figure 7.1: EDA signal and its variables [*created by the author*].

Variables: The following variables depicted in Figure 7.1 were considered in the data analysis:

- **EdaScrAmplitudesMean [µS]**: Skin conductance response is the fast increase in skin conductance and the added value onto the skin conductance level. This variable represents the height or amplitude of an emotional response and represents its intensity where higher values represent higher arousal. This variable is therefore the primary parameter in this experiment.
- **EdaScrCount** is the frequency of responses in a given time window (here 1 second).
- **EdaScrRiseTimesMean [s]** is the time in seconds that goes by from the onset of the response to reaching its maximum.
- **EdaSclMean [µS]**: Skin conductance level (SCL) represents the proportion of slowly changing skin conductivity that can be interpreted as the general mood of a person.
- **Time abs [hh:mm:ss]** and **Time rel [s]**: Absolute and relative time were used to synchronize the EDA data and eye tracking data, as well as for internal calculations.

Accuracy: Automated skin conductance response detection sensitivity: 92%. EdaMove slightly underestimates large peaks (>2 µS) by 5–10% due to 14-bit quantization and 32 Hz sampling.

7.2 eye tracking

The eye tracking sensor and software were used to capture and record eye gaze for each prototype view. The tool was primarily used to identify visual elements corresponding with EDA peaks and to label the EDA data according to story sections.

Sensor Device: The **Tobii Pro Spark** [107] is a remote eye-tracking device that was attached to the bottom of a monitor with a clip. The eye tracker captures eye gaze at a time resolution of 60 Hz, that is an alternating current (AC) power source completes 60 cycles per second. Three sensors automatically adjust to bright or dark pupil illumination. The eye tracker is suitable for people of any age, ethnicity, or who wear corrective lenses. Unlike older eye-tracking devices, which required participants to hold their heads still, modern devices like this one allow for natural head movements during the experiment while maintaining high performance. This is important in terms of engagement because users do not have to pay attention to behavioral limitations.

Software: The **Tobii Pro Lab** [106] software was used to calibrate the eye tracker for each participant and collect eye tracking data. The software can also be used to structure the experiment. I have created the following parts: (1) instructions for eye tracking calibration, (2) calibration, (3) an instructional slide indicating that the first story will start, and (4) an embedded link leading to the prototypes.All other instructional slides were integrated into the website, which incorporates both prototypes.

The software enables to store recordings, analyze them individually, and aggregate the data for quantitative analysis and visualization. Additionally, an export function provides access to the raw data. Exported data include variables such as gaze coordinates and absolute time stamps with microsecond or millisecond time stamp precision. The video recording was useful for identifying the time points at which participants scrolled forward or backward in the story. These timestamps were used to slice the EDA data into pieces of the story.

7.3 QUESTIONNAIRES

Three types of questionnaires were given to each participant in the order of the experimental setup (Chapter 8) and are attached in the appendix (Appendix A).

Questionnaire I: Understanding and Memory

Besides eliciting emotional engagement, the story has two other important goals: Understanding and memorability, which refer to recognizing and retaining the presented information. Questionnaire I (Section A.2) contains eleven multiple-choice questions about the story's main messages. Participants were asked about treatment options, symptoms, the biological cause of HG, and genetic factors, for example. There were also a four questions related to the data visualization, e. g., how to read the lower chart. One point was awarded for each question if all partial answers were selected correctly. If partial answers were missing, negative points were awarded according to the proportion. For example, if five answer options were correct and one was not selected, 0.2 points were deducted from the total score of 1, resulting in a score of 0.8.

Questionnaire II: Emotion Types

Questionnaire II (Section A.3) primarily aims to capture valence since EDA only records emotional arousal. To understand how emotional experiences depend on character usage, participants were asked to select emotion types for 14 out of 28 story pieces. To better compare the two prototypes and extract possible emotional patterns, the same page numbers from each story were chosen. These pages were carefully selected to represent all parts of the story (Section 6.3.5) and its most important story pieces. For example, pages that only provide textual information were excluded. This decision was made primarily to avoid overwhelming participants. The emotion types were arranged in a 5x4 matrix. Nine positive and one neutral emotion were represented on the top, and ten negative emotions were represented on the bottom. For each category, i. e., positive and negative emotions, the entire range of arousal levels from low to high was provided. To simplify the selection process, each emotion was also represented by an emoji that conveyed the emotion and its intensity. Since emotional experiences can be a mixture of different emotions, or since one story piece might evoke different emotions, participants were given the opportunity to select multiple emotion. They were instructed to mark the main emotion. For the analysis, I assigned a value of 2 to the main emotion or if only one was selected, and a value of 1 to all secondary emotions.

Questionnaire III: User Engagement

As discussed in the Related Work part (Section 5.3.3), physiological measurements can benefit from a multivariate methodology and be best interpreted with additional context. Besides valence, I also use self-reported methods to capture some aspects of user engagement. Questionnaire III (Section A.4) is adapted from Amini et al. (2018) and O'Brien & Toms (2009) [2, 79].

The following items were evaluated asking for the participant's agreement on 5-point Likert scales to each statement:

1. **Affective Involvement**:
   "I felt emotionally connected to the protagonist."
   "I think it is important to spread information about this condition."
   "I would recommend this story to mothers or their family members."

2. **Cognitive Involvement**:
   "It was difficult for me to process and understand the content."
   "The story deepend my understanding of the condition."

3. **Focused Attention**:
   "My thoughts drifted during the story."

4. **Novelty:**
   "The content of the story was totally new to me."

5. **Usability**:
   "I had trouble recognizing elements of the story."

Qualitative feedback

Most of the qualitative feedback was collected using questionnaires. The following questions were asked:

- *"Which elements had you trouble to recognize?"*
- *"Is there anything that was confusing about the story's content?"*
- *"Do you have any recommendation for improving the story?"*
- *"There is space here if you would like to share more thoughts!"*

After completing the questionnaires, many participants took the time to provide verbal feedback and answer additional questions. This feedback provided insights into why one story version was preferred and context for the physiological data. Some of this feedback could also be used to improve the data stories and the study design.

8 EXPERIMENT

Study Design

This study investigates whether a data story with an individual character evokes more emotional engagement and better performance in understanding and recall tasks than a data story with a general perspective. Additionally, this study examines whether emotional priming effects exist within subjects.

The study was primarily conducted as a ***Between-subjects study***. Participants were divided into two groups, A and B.

- ***Group A*** **first viewed the individual story with the individual character**, followed by the general story.
- ***Group B*** **viewed the general story first**, followed by the individual story.

The ***independent variable*** was defined by the presence or absence of an individual protagonist, and the ***dependent variables*** encompass emotional engagement, understanding, memory, and user engagement. Each participant was evaluated with only one story for understanding, memory, and user engagement. Since each participant viewed both prototypes, the EDA and eye tracking data were also analyzed from a ***Within-subjects study*** perspective.

8.1 PARTICIPANTS

Recruitment

All of the study participants were volunteers who were alternately assigned to Group A or Group B based on recruitment order. Besides a basic knowledge in English, I set no requirements for the study. Important terms were translated before the experiment began, and participants were asked to inquire about any unclear words. To include the perspective of mothers, I specifically sought out mothers and also had the opportunity to conduct the experiment with a former hyperemesis gravidarum (HG) patient. Overall, more women participated in this study. However, the subsample was perfectly balanced for the question of whether individual stories arouse more emotional engagement (six women and five men in each Group A and B).

Demographics

In total, 26 participants took part in this study.Fifteen participants were female, including five mothers, one former HG patient, and one participant was pregnant at that time. Eleven participants were male. Most of the participants (n = 15) were between 26 and 35 years old. Four were younger between 18 and 25 years old. The remaining subjects were in the middle-age range, including five participants were between 36 and 45 years old, and one between 46 and 55 years old.

Most of the participants had an university degree (n = 21), two received a degree from an university of applied science, three subjects had a high-school diploma, and one a secondary school diploma. In terms of occupation, I divide study participants into three groups: 1) Industry (n = 11), 2) Research (n = 11), 3) Students (n = 4). Six subjects were involved in both medical topics and data visualization in their profession or studies, and two were occupied with data visualization in their profession or studies.

In addition, I asked six women who have been pregnant about their symptoms during pregnancy. Three out of six reported nausea, and one reported nausea with vomiting. Three reported other symptoms: sleeping problems (n = 2), unusual fatigue/drowsiness (n = 1), and headaches (n = 1). The woman who was affected by HG stated that she had nausea with vomiting throughout her entire pregnancy and her doctor prescribed anti-nausea medication.

Samples

Data from all 26 study participants was integrated into the analysis of the emotion types and user engagement questionnaires. The memory and understanding scores were used for 25 subjects, excluding the data from the pilot study participant due to major changes in the questionnaire design.

As described in Section 5.3.2, unrealistically low values can occur during recording for various reasons. I describe how missing values were calculated and handled in Section 9.1. Three participants were excluded due to a high amount of missing EDA data for both stories: 43 %, 87 %, and 92 % respectively. Since all subjects were assigned to one group, three additional participants were recruited to balance the groups. In addition, three participants lost contact with the electrodes during the second story, resulting in 91 % to 100 % missing values.

To compare the general and individual story, I only used samples from the unbiased first story view. Additionally, the former HG patient was excluded for this comparison to reduce the bias that would be caused by the fact that only one participant had previous experience with the disease, leaving 12 participants who first read the individual story, and 10 subjects who first read the general story. The eye tracking analysis was conducted for story views of each participant: The analysis of the individual story and general story was conducted with data from 22 participants in each group.

8.2 EXPERIMENTAL PROCEDURE

The experiment took place in a climate-controlled laboratory. The same desk, monitor, and technical devices were used for each session. Participants were provided with drinks and refreshments. Participants needed 25 to 45 minutes to complete the experiment.

OVERVIEW: EXPERIMENTAL SETUP:

T0: Information about the experiment & consent

T1: Demographic questionnaire

T2: EDA preparation and application to the wrist

T3: Eye tracking calibration

T4: Story 1 (Group A: individual, Group B: general)

T5: Questionnaire I–III for Story 1

T6: Story 2 (Group A: general, Group B: individual)

T7: Questionnaire II for Story 2, liking

T8: Open discussion and informal feedback

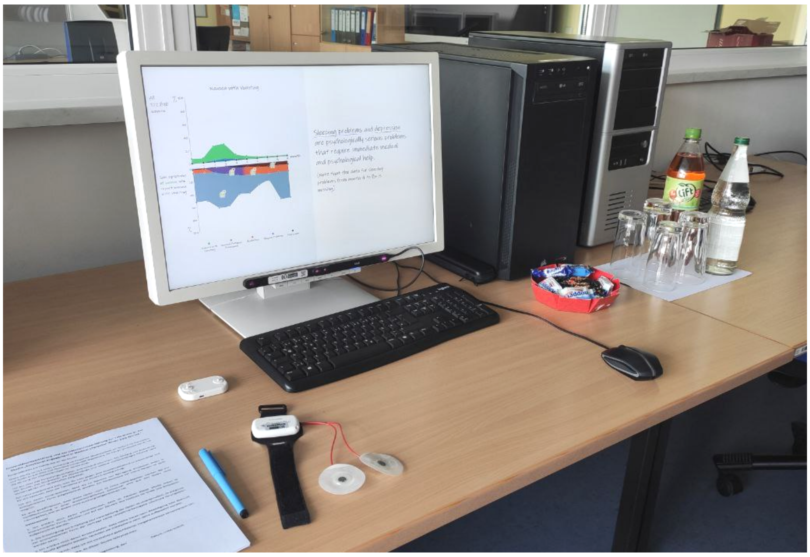


Figure 8.1: Experimental setup including EDA sensor, eye tracking device, questionnaires, and interactive slideshow prototypes in a lab of the Department of Simulation and Graphics.

PREPARATION:

T0: The participants were welcomed in the experimental space. Subjects were briefed on the nature, purpose, and methods of the study, as well as about the anonymization of the data, and how the data would be used. They were informed that participation is voluntary and that they may withdraw at any time. Before the experiment began, the participants were asked to give their consent. ⟶ T1: The participants were seated on comfortable chairs and asked to fill out the demographic questionnaire. They were informed of any important terms used in the stories. ⟶ T2: Then, they were equipped with the EdaMove 4 sensor on the wrist of their non-dominant hand, and electrodes were placed on their palm to record skin conductance. This way, their dominant hand is free to use the mouse and write between story views. Additionally, the probability of movement artifacts in the data is reduced. ⟶ T3: The eye tracking recording and calibration was then started via the Tobii Pro Lab software. Once the calibration result was approved, participants were informed that they were going to see the first story. They were given instructions on how to use the mouse to navigate and interact with the story. They were also encouraged to ask questions if anything was unclear during the session.

STORY VIEWING:

⟶ T4: Next, a Group A participant will view the individual story, while a Group B member will see the general story. There were no time restrictions for viewing the stories. ⟶ T5: After finishing the first story, instructional information was displayed on the monitor that a questionnaire is needed to be filled out. Questionnaire I was provided to assess understanding and retention of the content. Participants were also informed about how to complete Questionnaire II, which captured emotion types. Questionnaire III evaluated user engagement with this story. ⟶ T6: When participants were ready, they were asked to continue with the second story, i. e., the general story for Group A and the individual story for Group B. ⟶ T7: After completing the second story, participants were asked to select emotion types for this story in Questionnaire II. Finally, they were asked which story they liked most. ⟶ T8: In an open discussion, I asked about their reasons for preferring one story over the other and many participants reflected on the relevance of the story and provided feedback to the study design.

9 DATA ANALYSIS

9.1 DATA ANALYSIS WORKFLOW

This section will provide details on how the EDA and eye tracking data was further preprocessed and analyzed. In order to highlight differences between the story versions, sections of the signal curve were assigned to the individual and general story. This was done by information taken from eye tracking data. To illustrate the progression of emotional engagement throughout each story, the data was broken down into temporal sections, or story pieces. In this context, a story piece is defined as one slide of the web-based slideshow.

I derived four ***emotional engagement variables*** and ***time spent*** from the following variables: skin conductance response [µS] (mean value per second), skin conductance level [µS] (mean value per second), absolute time [hh:mm:ss], and relative time [s] (described in Section 7.1).

Emotional Engagement variables & Time spent:

- means of amplitude values *(per minute or story piece)*
- number of peaks *(per minute or story piece)*
- maximum peaks *(per story)*
- maximum sum of amplitudes *(in a story piece)*
- time spent *(per story or story piece)*

Skin conductance level was used as an indicator of missing values. Similar to Stuldreher et al. (2025) [104], I labeled skin conductance values below 0.3 microsiemens as missing. A detailed discussion of the influences, both internal and external, on the EDA signal is provided in Section 5.3. Although the recording is directly initiated from a computer device before applying the sensor to the hand, in two cases recording started approximately 1 min later. In some participants, recording started with missing values and stabilized in the first third of the first story. A practical explanation might be dry hands, which prevent the signal from being transmitted. I noticed missing values in the second story for some subjects. This could be due to loose electrodes. To prevent this in future studies, I recommend checking whether the electrode pad is firmly attached once in the middle of a longer experiment. If a participant has very sweaty hands, the pad may lose its contact to the skin. In this case, it is recommended that the pad be replaced. I would also plan for more time between applying the sensor and presenting the stimulus in a future study. Additionally, monitoring the signal could help resolve any issues during recording.

The following steps were taken to process, label, and summarize the EDA data with respect to the story view and story pieces.

I. Labeling

1. The synchronization point was extracted by exporting the eye tracking data to a raw data file. The first absolute time point (hh:mm:ss) marks the start of the eye tracking recording with the relative time point (00:00:00). The synchronization point was stored once for each session.

2. For each subject, all relative time points, where a transition between story pieces (i.e., a scrolling action) took place, were extracted from the eye tracking data. The number of the current story piece was entered into an array for each time point. The array of story pieces was organized in increasing order by time. *Note:* When a user moved back and forth, extra time points were extracted.

3. The following steps of data processing and labeling of the csv files are presented in pseudo code. I implemented these steps with python code using pandas and numpy library:

   a) Selecting and adding variables:

   ```
   input: table of each subject

   for each row (second):
       remove columns not needed for analysis
       add columns subject code, group
       add columns story no., story piece no.
   ```

   b) Time conversion:

   ```
   input: synchronization point, array = [start points in
       rel. time, story piece no.]

   for each relative time point i:
       [hour, min, sec] = synchronization point
           + relative time point
       clock = str(hour) + ":" + str(min) + ":" + str(sec)
       cache = [storypieces[i-1], clock]
       append absolute times and slide no. to final array
   ```

c) Label tables for each story:

```
input: table of each subject, array = [n absolute time
    points, n story piece no. for each story]

for each subject (0,23):
    for each row (second):
        find absolute time point:
            do label row with subject no., group
            do label row with story no.
            do label row with story piece no.
            fill remaining rows with the same labels
                until next time point is found
```

d) Find missing values where skin conductance level < 0.3 μS:

```
input: labeled table of each subject, threshold = 0.3

for each row (second):
    if scl < threshold:
        scl = 'nan'
        scr = 'nan'
```

e) Label skin conductance responses with stimulus by substracting rise time of the response and latency until onset:

```
input: labeled table, latency = 1.8 s

for each subject and for each story:
    for each row (second):
        find response:
            t = relative time - rise time - latency
            s = storypiece no. at t
            do store story piece no. in a new column
            do store t as absolute time in a new column
```

II. EDA Data Summaries

After preprocessing the EDA data, I created summaries to compare various features and groupings, such as the two participant groups (A and B), whether they saw a story for the first or second time (story 1 and story 2), and added a variable for gender. The data can also be analyzed at different levels of granularity, i. e., at story piece level.

At the **story level**, I calculated four variables for emotional engagement: Mean of skin conductance response amplitudes per minute (intensity), mean frequency of amplitudes per minute, maximum peak, and maximum sum of amplitudes in a story piece. The time calculation considered only seconds without missing values. At **story piece level**, I examined the distribution of the sum of amplitudes, frequency, and skin conductance level for the matrix visualization. For the final visualization, I only needed the sum of the amplitudes.

1. The first summary table aggregates the data for each subject on **story level**. To calculate the means, a separate time variable stores only seconds without missing values, called "eda time".

```
input: Labeled tables, demographic table

for each subject (0,23):
    for each story (0,2):
        for each story piece (0,28):
            for each row (second):
                store time spent
                if scl != 'NaN':
                    store eda time
                    store sum of amplitudes
                    store number of peaks
                    compare and store 3 maximum peaks
                    compare and store maximum sum
                    store skin conductance level
        calculate mean of amplitudes per min
        calculate mean frequency per min
        calculate missing values percentage
        store gender
        fill summary table
```

2. The second summary table summarizes the data for each subject on **story piece level**:

```
input: Labeled tables

for each subject (0,23):
    for each story (0,2):
        for each story piece (0,28):
            for each row (second):
                store amp_sum
                store scl sum
            calculate scl_mean
            if scl_mean < 0.3:
                amp_sum = 'NaN'
            fill summary table
```

III. Explorative Data Analysis & Statistics

I used R to gain an initial understanding of the data. I experimented with pirate plots/rainscloud plots, matrix visualization, area charts, line charts, scatter plots, and bar charts. For the final results, I refined some of these visualizations with consistent design choices. For example, I used the same color to represent data from individual stories only. Statistical analysis was also conducted in R Studio. First, I will present the workflow of the statistical analysis. Then, I will explain why I chose specific data visualization types to present the results.

Workflow for Statistical Analysis

1. First, I investigated whether a variable was characterized by **normal distribution**. The *Shapiro–Wilk test* evaluates whether a variable is not normally distributed if the p-value is below 0.05.

   ```
   Example: shapiro.test(story1 $ amp_per_min)
   ```

2. To test my hypotheses of **significant differences** between groups, e. g., between the individual and general story, I used two statistical tests, depending on the variable's distribution. For both tests, a confidence interval of 95% (alpha = 0.05) was chosen. Both tests compute a p-value, where $p = 0.05$ means that there is a 5% chance of a Type I error, i. e., falsely rejecting the null hypothesis. With $p < 0.05$, the null hypothesis is rejected.

   2.1. If a variable was not normally distributed, the *Mann-Whitney-U test*, also called *Wilcoxon rank-sum test*, was used. Mann-Whitney is a non-parametric test that compares ranks, not means, therefore extreme values, zero-inflation, and skewness have minimal impact.

   ```
   Example: wilcox.test(amp_per_min ~ story,
                data = story1, exact = FALSE)
   ```

   2.2. If a variable was normally distributed, the *Welch Two Sample t-test* was used.

   ```
   Example: t.test(max_peak ~ story, data = story1)
   ```

3. In the third step, I tested the **effect size**. Both Cohen's d and Cliff's delta are effect size measures used to quantify the magnitude of difference between two groups.

   3.1. The *Cliff's delta* indicates the effect size for non-normal distribution. This correlation coefficient returns a value in range $[-1, 1]$, where 0 = complete overlap, ± 1 = no overlap, $\mid \delta \mid < 0.147 \rightarrow$ negligible, $\mid \delta \mid < 0.33 \rightarrow$ small, $\mid \delta \mid < 0.474 \rightarrow$ medium, and otherwise large effect size.

   ```
   Example: effsize::cliff.delta(story1_ind $ amp_per_min,
                story1_gen $ amp_per_min)
   ```

   3.2. For normal distributed variables, it is common to quantify effect size with *Cohen's D* – a measure of the difference between the means of two groups standardized by dividing the difference by the standard deviation. Cohen's d is not bound to upper or lower boundaries, where the effect size is negligible for $d < 0.2$, small for $0.2 \leq d < 0.5$, medium for $0.5 \leq d < 0.8$, and large for $d > 0.8$.

   ```
   Example: cohen.d(max_peak ~ story, data = story1)
   ```

Data Visualization

I made extensive use of ***pirate plots*** also called ***raincloud plots*** due to their strength in combining multiple elements into a single, compact figure to provide a rich, transparent summary of group-wise continuous data distributions. First, raw data points (geometrical points) show individual observations and making it easy to identify clusters. Second, a density curve as violin plot is superimposed to reveal shape, skewness, and multimodality, which also allows to identify clusters or subgroups within a sample. Third, descriptive statistics are displayed as bar chart showing the mean superimposed with a rectangle to show the standard deviation.

For all visualizations presenting the results at the story piece level, I used a ***heat map*** visualization, where each column represents all observations (per participant or emotion type) within a story piece. Matrix visualizations provide a structured overview of denser occurrences in story pieces or parts.

I used ***parallel coordinates*** to explore changes and correlations between groups and ***bar charts*** to compare frequencies and values. I used ***scatterplots*** to analyze correlations and a ***line chart*** to visualize the *quadratic regression analysis*.

IV. Eye tracking & Story Element Analysis

For analyzing correlations between emotional engagement and story elements, I chose a selection criterion: For each participant, three maximum peaks were calculated in each story. The time points of the maximum peaks were converted into the relative time of the eye-tracking data and then labeled manually.

First, I assigned more detailed labels to each element to provide information about its meaning. For example, a data visualization was described detailed by elements, such as area, axis, annotation, icons, and quotes. Secondly, I categorized these detailed labels into visual element categories, such as text, data visualization, icons and quotes, chart animation, animated illustration, and illustration. Arousal can be caused by emotional or cognitive involvement. Therefore, I assigned each element depending on whether it was more story-related or descriptive also to the categories: *"emotional"* or *"cognitive."* If this was not obvious, I assigned it to the *"cognitive or emotional"* group.

Location Analysis Workflow

Input: Time stamps t of three maximum amplitudes for each participant in each story.

1. Convert time points of EDA peaks into relative time of the eye tracking data.

2. Find time point in eye tracking data and localize fixation point.

Output: Story element → Label → Categories I + II

9.2 QUESTIONNAIRE ANALYSIS

I. Emotion Types Analysis

The emotion types were acquired through a questionnaire for 14 story pieces of each story. The main emotions were weighted at 2, and all others were weighted at 1. I created a summary table of the emotion types, where I added up the weights for each story piece (row) and emotion type (column). To sort the twenty emotion types, I used a dual-criterion approach considering valence and arousal dimensions. For example, *joy* is highly positive and arousing and is therefore on the top, while *anxious* is very negative with high arousal and is placed at the bottom. *Fine* represents a more neutral state of low arousal, located in the middle.

As a result, positive emotions were sorted by positivity (decreasing) and intensity (decreasing), resulting in: *joy, awe, love, peaceful, connected, happy, excited, curious, calm, fine*.

The negative emotions were sorted by negativity (increasing) and intensity (increasing), resulting in: *bored, tired, disappointed, annoyed, sad, depressed, miserable, anxious, stressed, anger*.

II. Understanding, Memory & User Engagement

Quantitative data from these questionnaires was collected on paper and then entered into tables. I summarized the data with the means for each question and labeled the questions according to their category: *"Affective involvement," "Cognitive involvement," "Focused attention," "Usability,"* and *"Novelty."* Qualitative feedback was transferred to table documents according to the specific wording of written statements or information obtained from conversations.

In the next chapter, I will present and discuss the results and their interpretation in light of the research questions.

Part III

RESULTS & DISCUSSION

10 RESULTS

This study was driven by three research questions:

RQ1: *"Does a fictitious, individual character arouse greater emotional engagement in the story's viewers for a medical condition compared to a general story with no individual human protagonist?"* → H1–6

RQ2: *"Does experiencing a story in first-person perspective increase or decrease emotional engagement when the identical story is subsequently encountered in third-person perspective (and vice versa)?"* → H7

RQ3: *"Does a fictitious, individual character increases understanding and memory in the story's viewers for a medical condition compared to a general story?"* → H8

To address my research questions, I will validate the following eight ***alternative hypotheses***:

H1: Inclusion of a fictitious, individual character to embody the lived experience of a patient will elicit increased feelings of emotional engagement from the story audience.

H2: Women feel more engaged with the stories than men.

H3: Emotional engagement is not evenly distributed along the course of the story leading to more dense occurrences in certain story parts, such as the conflict presentation.

H4: Participants prefer the individual story over the general story.

H5: The individual story elicits more strong negative empathic emotions (sad, depressed, miserable, anxious) than the general story.

H6: The three highest EDA peaks for each participant and story correspond more often with character illustrations than with illustrations in the general story.

H7: Participants who read the story first in first-person and then in third-person will show higher emotional engagement on the second reading than participants who read it in the reverse order.

H8: The story that elicits more emotional engagement has better recall and understanding scores.

In the following, I will refer to ***null hypotheses***, which consist of there being no significant difference between two groups, until there is sufficient evidence to reject the null hypothesis. A null hypothesis will be rejected with a confidence interval of 95% (alpha value = 0.05), meaning that the alternative hypothesis is likely to be true.

10.1 EMOTIONAL ENGAGEMENT ON STORY LEVEL

In this section, I will present the results of the first hypothesis using the unbiased data of the first view. The null hypothesis states that there is no difference in emotional engagement between a story with and without an individual protagonist. In this exploratory study, I used four variables to measure emotional engagement. How I derived these from the data is described in Section 9.1. Additionally, I considered time spent as an interest indicator.

VARIABLE	*p-value*	*effect size*	Ind. Story *mean*		Gen. Story *mean*
Intensity [$\mu S/min$]	0.276	small	2.210	>	1.357
Frequency [n/min]	0.492	small	4.546	>	3.797
Max peak [μS]	**0.078**	**large**	1.054	>	0.727
Max sum [μS]	**0.060**	**large**	3.000	>	1.741
Time spent [s]	0.191	medium	519	>	441

Table 10.1: Results on emotional engagement & time spent by story version.

As summarized in Table 10.1, no significant difference was found in ***Intensity*** [$\mu S/min$] ($p = 0.276$, *Cliff's delta* = 0.28, 95 % CI [-0.23, 0.67] → *small effect size*). The Confidence Interval (CI) includes zero. Essentially no difference was found in ***Frequency*** of skin conductance responses per minute ($p = 0.492$, *Cohen's d* = 0.3, 95 % CI [-0.54, 1.14] → *small effect size*). The CI is very wide and centered near zero.

A large observed effect was found by the ***Maximum peak*** that narrowly missed $\alpha = 0.05$ ($p = 0.078$, *Cohen's d* = 0.81, 95 % CI [-0.07, 1.68] → *large effect size*). Here, the CI barely includes zero and suggests a real effect. The largest and most promising effect was found in the ***Maximum sum of amplitudes*** in a story piece ($p = 0.060$, *Cohen's d* = 0.86, 95 % CI [-0.03, 1.73] → *large effect size*). Again, the result barely missed statistical significance, but the lower bound of the CI is almost zero. Therefore, the maximum peak and maximum sum of amplitudes showed a large effect in favor of the individual story. Negative values in the CI indicate a too-small sample size. Although the null hypothesis cannot be rejected, there is strong evidence that a larger sample size would demonstrate that character-driven stories are more emotionally engaging. Notably, the means for all emotional engagement variables, which can be visually inspected in Figure 10.1, are higher for the individual story. Participants also spent more time on the character-driven story, at 519 seconds (8:39 *min*) versus 441 seconds (7:21 *min*) on the general story.

Hypothesis 1

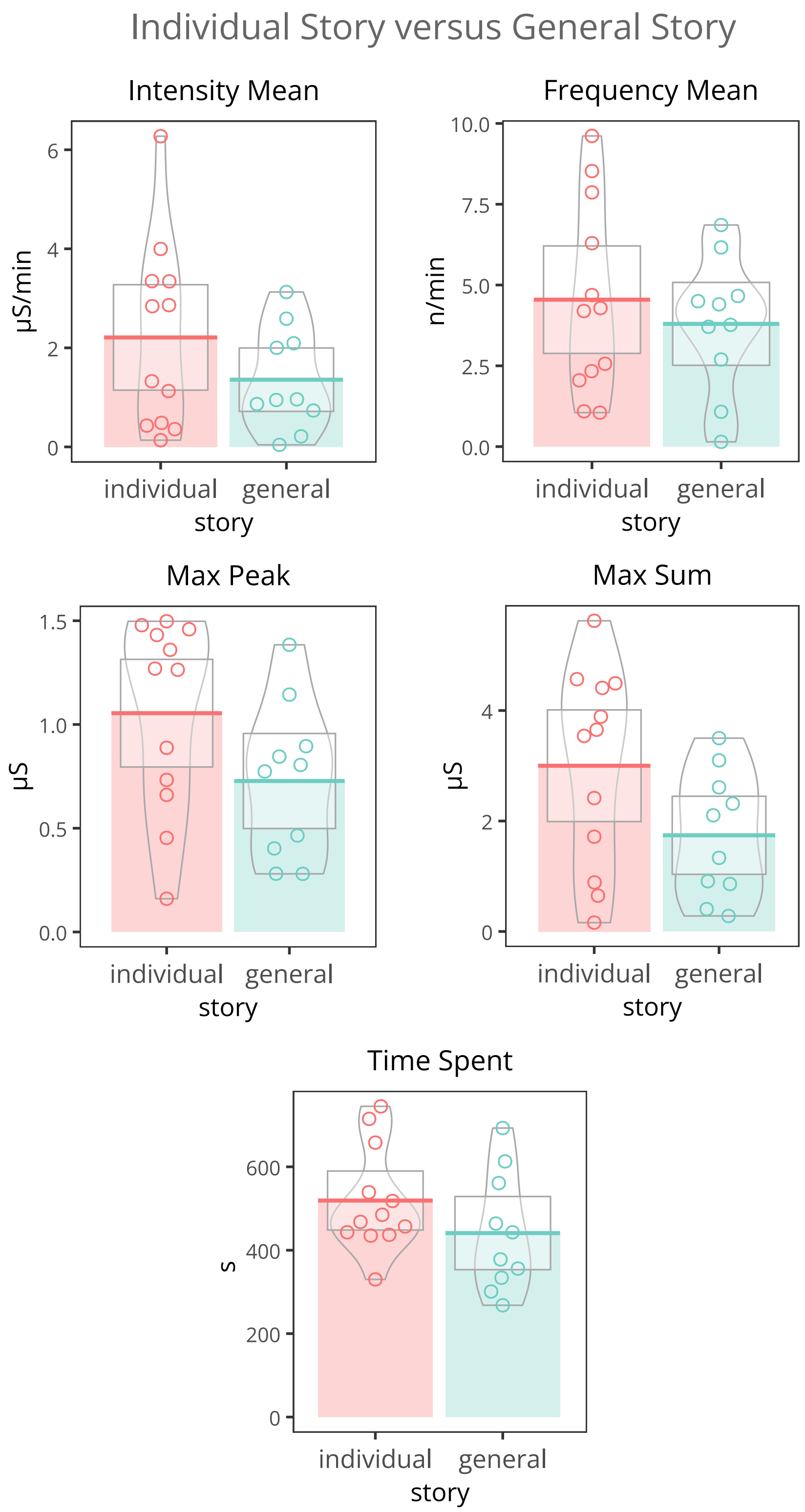


Figure 10.1: Results on emotional engagement & time spent by story version.

10.2 GENDER DIFFERENCES IN EMOTIONAL ENGAGEMENT

Since, the pregnancy topic might be more emotional or relevant for female participants, it is worth investigating the emotional effect in gender differences. Figure 10.2 uses Pirate Plots to visualize each sample, means, and distribution.

VARIABLE	*p-value*	*effect size*	Females *mean*		Males *mean*
Intensity [$\mu S/min$]	**0.0005*****	**large**	2.798	>	1.150
Frequency [n/min]	**0.007****	**large**	5.410	>	3.475
Max peak [μS]	**7.7e-05******	**large**	1.217	>	0.717
Max sum [μS]	**0.001001****	**large**	3.036	>	1.651
Time spent [s]	0.970	negligible	418	>	416

Table 10.2: Emotional engagement in females and males.

Hypothesis 2

Indeed, female participants showed significantly more emotional engagement than male participants. As shown in Table 10.2, all four emotional engagement parameters had highly significant to very highly significant p-values and *large effect sizes*:

Intensity [$\mu S/min$] (p = 0.0005***, *Cliff's delta* = 0.63, 95% CI [0.27, 0.83]), ***Frequency*** [n/min] (p = 0.007**, *Cohen's d* = 0.9, 95% CI [0.25, 1.54]), ***Max peak*** [μS] (p = 0.00007****, *Cliff's delta* = 0.72, 95% CI [0.42, 0.87]), and ***Maximum sum of amplitudes*** (story piece) [μS] (p = 0.001001**, *Cohen's d* = 1.1, 95% CI [0.47, 1.79]).

The most significant variables were maximum peak and skin conductance response intensity, where the sample size of N = 41, with 21 samples from females and 20 samples from males, was sufficient to show a large positive effect. Therefore, the second hypothesis can almost certainly be rejected in favor of the alternative hypothesis that women were more emotionally engaged by the stories.

Interestingly, female and male participants spent nearly the same amount of time on the stories on average (422 seconds for females versus 416 seconds for males). Additionally, females felt slightly more engaged with the character-guided story. The means of all four emotional engagement variables, as well as the time spent, were higher for this story. I found a *small effect size* in the frequency and maximum sum of amplitudes in a story piece in this group. Although there was very little difference between the story versions for males, three out of four engagement variables were slightly higher for the neutral story. Time spent was also slightly higher for this story.

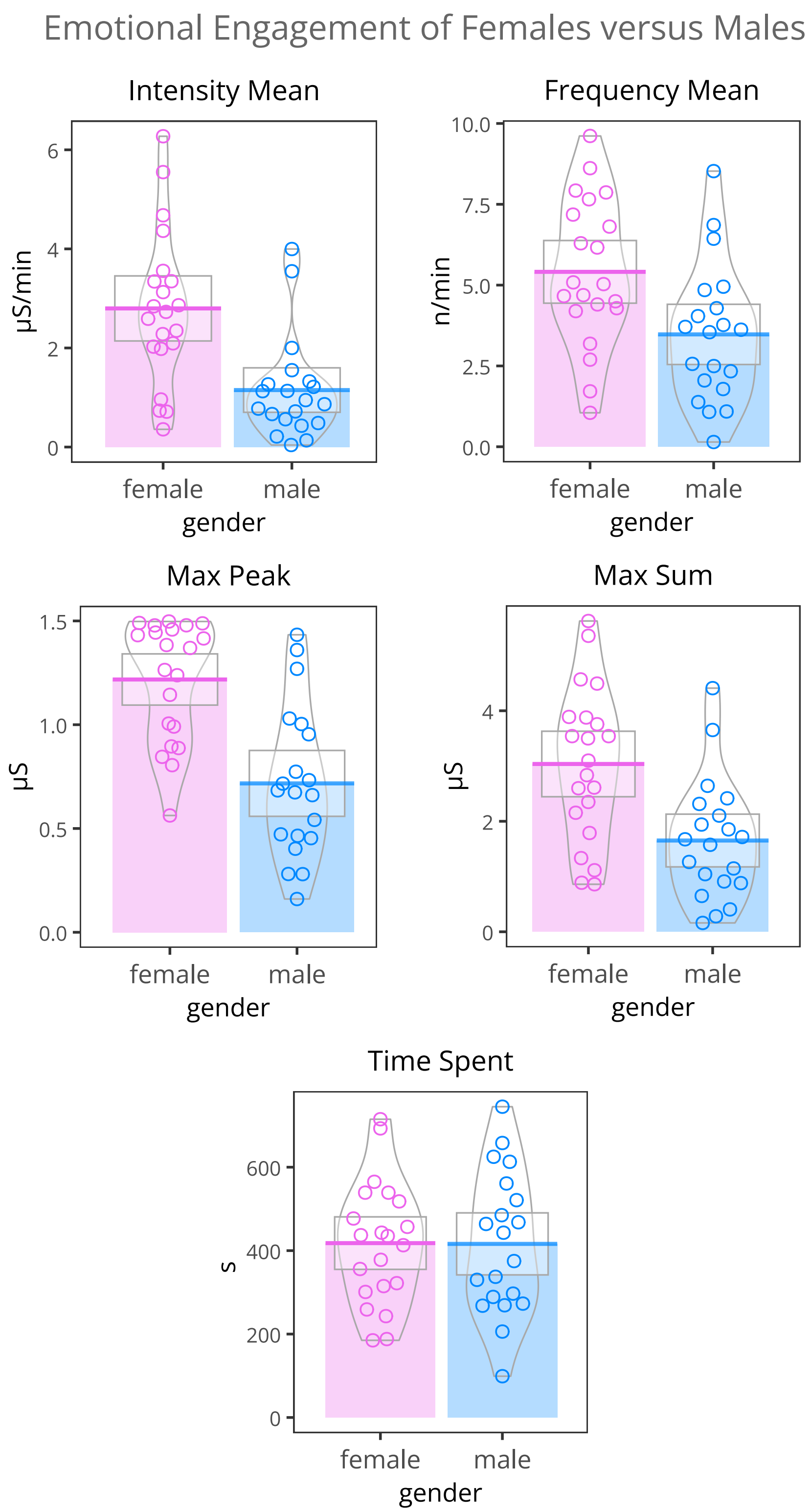


Figure 10.2: Emotional engagement & time spent in females and males.

10.3 EMOTIONAL ENGAGEMENT ON STORY PIECE LEVEL

Next, I will present more detailed results at the story piece level and demonstrate how emotional engagement evolved throughout the story. Data from participants with missing values throughout a story were removed. I chose heat maps to display the results. In Figure 10.3 and Figure 10.4, each row represents a participant and each column represents a story piece. A gradient depicts the intensity of emotional arousal based on the sum of EDA amplitudes. If a story piece contained no amplitudes, the square is filled with white. If a story piece was had a mean skin conductance level below 0.3, the square is filled with light gray (missing value). Participants with similar results were placed closer to each other.

By only considering story pieces without missing values, 60.66 % of the character-driven story pieces elicited emotional arousal. The maximum in this story was found in storypiece 23. When participants viewed the general story, 47.70 % of the story pieces showed amplitude activity, with the maximum occurring in storypiece 21. These results align with Hypothesis 1, which states that the story with an individual character is more emotionally engaging than a general story. Both distributions seem to be denser in the center of the story, while emotional arousal seem to decrease more in the general story towards the end.

A secondary analysis of the **maximum peaks** in the individual story showed that these peaks often occurred during the story part *"Data-driven insights (Data Vis)"* ($n = 11$). Here, four peaks occurred during the data visualization introduction. Four maximum peaks were localized in the "Explaining the Disease" part, while all other story parts evoked at least one maximum peak. In the general story, the most maximum peaks also occurred during *"Data-driven insights (Data Vis)"* ($n = 8$), with five peaks occurring during the data visualization introduction. Same as in the personal story, four maximum peaks were found in the *"Explaining the Disease"* part. The introduction was more arousing with the individual story counting four peaks. The remaining story parts had one maximum peak each.

Considering the question of which story pieces were the most arousing based on the **maximum sum of EDA amplitudes**, two stand out. Storypiece 10 aroused seven participants the most during the explanation and illustration of the disease's mechanisms. Furthermore, storypiece 19 was more arousing for the individual story ($n = 6$) than for the general story ($n = 2$). In the introduction, one participant was most emotionally moved when the protagonist said she couldn't drink and had to see her doctor. Three maximum sums occurred during storypiece 13 (data visualization intro), storypiece 17 (zoom animation), and the juxtaposing data visualization showing data for women with nausea and vomiting compared to women without.

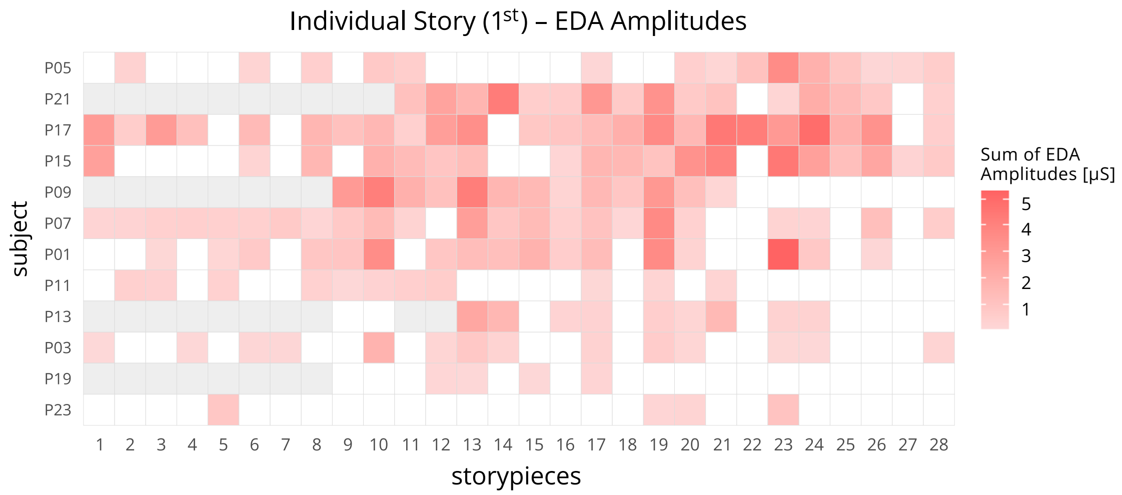


Figure 10.3: Emotional arousal throughout the individual story.

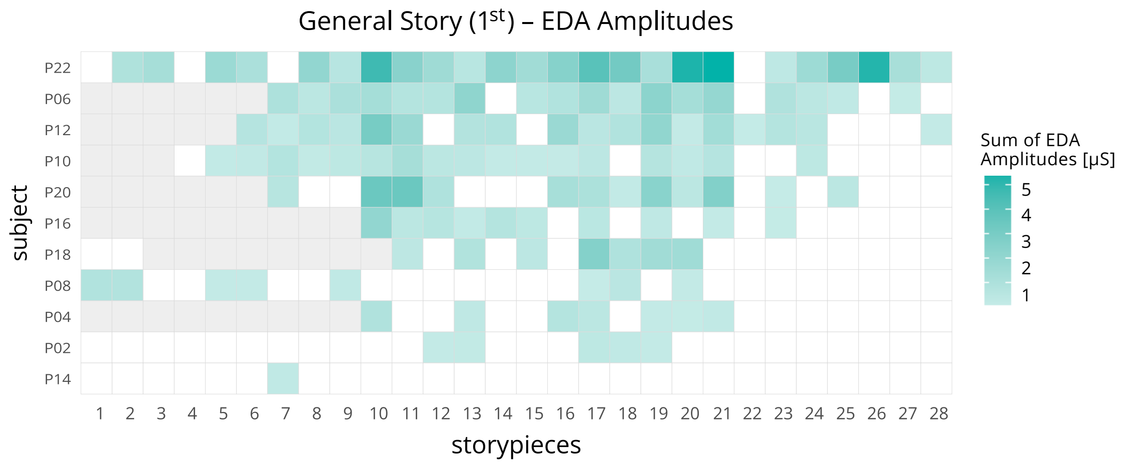


Figure 10.4: Emotional arousal throughout the general story.

Introduction	Conflict	Explain	Data-driven Insights	Resolution	Credits
1–3	4–9	10–11	12–24	25–26	27–28

Table 10.3: Story parts and pieces.

To synthesize the data from all participants and analyze which story pieces evoked the strongest emotional responses relative to the time spent on them (excluding seconds with missing values), I aggregated the data for each story piece. Figure 10.5 shows change over time. Uncertainty arises when there are many missing values. Of the 28 story pieces, the personalized story had a higher amplitude sum in 16 of them. Both stories had seven pieces that reached an amplitude sum of over 3 $\mu S/min$. While all seven story pieces with higher arousal were localized in the first third of the general story, i. e., introduction, conflict, and explaining the disease story parts; story pieces with higher arousal were more broadly distributed over the course of the

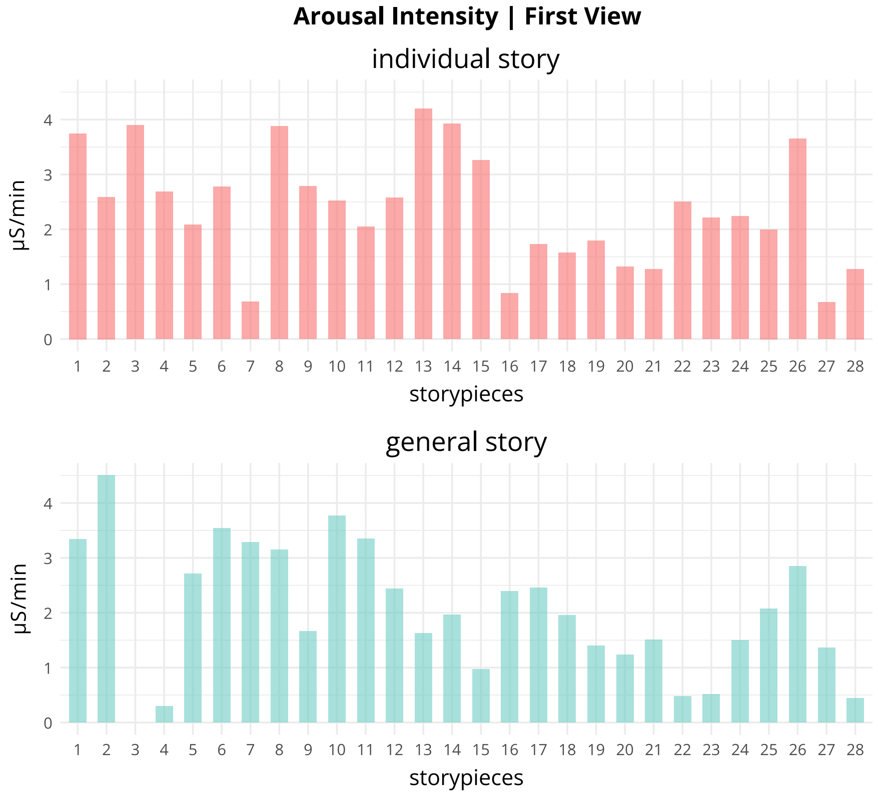


Figure 10.5: Average arousal intensity over story progress.

individual story. Notably, the introduction to the data visualization (story piece 13) was highly arousing when shown with an illustration of the character but less so when presented unpersonalized. The call-to-action (story piece 26) also evoked stronger emotional responses in the context of an individual character. In the general story, only a former HG patient P22 showed an emotional response.

Hypothesis 3

This analysis at a more granular level shows that emotional engagement is not evenly distributed throughout a story, leading to more dense occurrences in certain story pieces and parts. In the individual story, arousal intensity was denser in the (1) data visualization intro (3.91 $\mu S/min$), (2) introduction (3.56 $\mu S/min$), (3) resolution (2.72 $\mu S/min$), and (4) conflict (2.53 $\mu S/min$). In the general story the most dense story parts were: (1) Explaining the disease (3.86 $\mu S/min$), (2) introduction (3.33 $\mu S/min$), (3) conflict (2.70 $\mu S/min$), and (4) resolution (2.31 $\mu S/min$). In contrast, the least arousing part in both stories was the information section at the end (1.13 $\mu S/min$ and 0.89 $\mu S/min$, respectively). In summary, the parts of the story that evoked the most emotional arousal were the data visualization intro, explaining the disease, call-to-action, introduction, and the conflict situation. These results are at an exploratory stage. Since there are many missing values at the beginning of both stories, the dataset must be balanced to draw a conclusion about statistical significance.

Figure 10.6: Average time spent in a story piece.

Figure 10.6 shows the distribution of time spent in story pieces. Although participants spent overall more time in the individual story, the distribution are very similar between the story versions. Participants spent less time in the introduction and conflict part. Moderate time was spent in story piece 10 where the disease's biological cause was conveyed. Another peak was found in story piece 17 where an interactive zoom animation was applied to the data visualization.

Participants spent the most time on story piece 19, which explained the unconventional data visualization concept, especially how to interpret the area chart below the x-axis. This story piece also contained the first quote and the most text overall. Additionally, participants spent more time on story pieces 20, 21, 23, and 24, which provided more complex data visualizations.

The different distributions of Figure 10.5 and Figure 10.6 show that there was no positive correlation between time spent and emotional arousal. In the next section, I will present the results regarding preference for one story version over the other.

10.4 LIKING & EMOTIONAL AROUSAL

In this section, I will present the results on the fifth hypothesis and correlate the preference for a story with the corresponding emotional arousal. Additional insight into why participants preferred a story was gathered through qualitative feedback in the form of free-response questions and informal interviews.

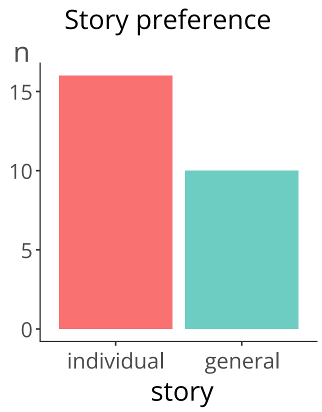


Figure 10.7: "Which story did you like most?"

Hypothesis 5

Sixteen participants preferred the story with an individual character, while ten preferred the general story (Figure 10.7). No significant difference was found in liking a more personal or neutral design of the story (p = 0.1, *Cliff's delta* = 0.23, 95 % CI [-0.05, 0.47] → *small effect size*). However, this result indicates a tendency of a preference for stories with individual characters.

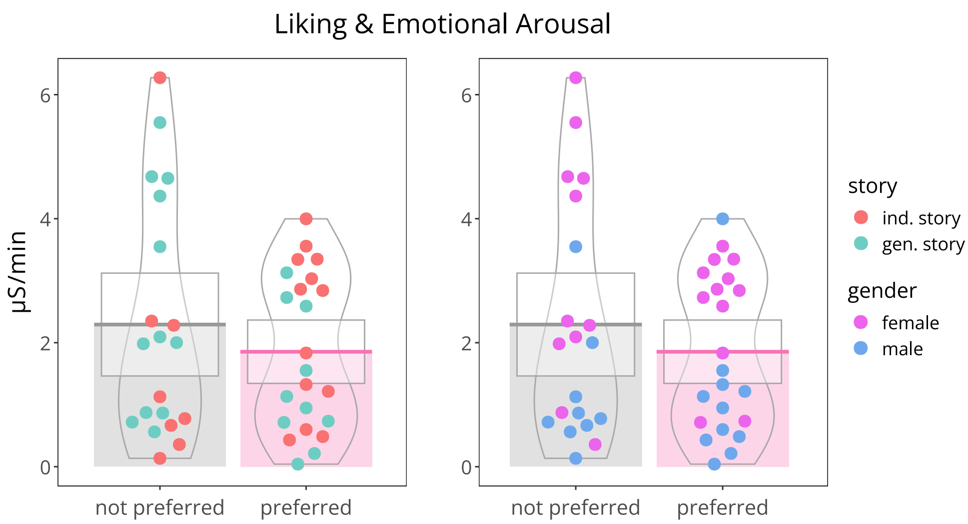


Figure 10.8: Liking and emotional arousal by story version and gender.

Figure 10.8 relates average emotional arousal to liking. Only participants with sufficient EDA data are represented here. Filled circles show the story version (left) and gender (right), providing more context. Most of the preferences for the individual story are located in an upper cluster above 2.5 $\mu S/min$, while the lower cluster is more mixed.

Regarding gender, there are two clusters for the preferred story. The first cluster is formed by female participants and has a central value of around 3 $\mu S/min$ for the most liked story. The lower cluster displays most of the male participants, with a central value of slightly <1 $\mu S/min$. The mean roughly separates the clusters.

Notably, the less preferred story elicited more emotional arousal overall. In five cases the arousal mean was highest for the not preferred story involving only female participants where four were assigned to Group A. Four out of these five samples belong to the general story, while the highest value occurred while a female read the individual story first. On the other hand, most of the data points form a cluster in the lower area around 0.5 $\mu S/min$.

Participants gave the following reasons for liking the personal story:

Qualitative Feedback

- P13: *"[The individual story] (Freja) is much more personal and creates a direct connection and more empathy.
I like general statements about treatments and prevalence, but the personal story is more engaging."*
- P15: *"[The individual story] was a different story to what I know and I liked the illustrations more."*
- P18: *"Fascinating how a person can turn the understanding just by creating empathy. Good example."*
- P25: *"The personal story makes it more relateable and more interesting to follow."*

Most participants that rated for the ***individual story*** said that they did because of two main reasons: (1) the story is different, more personal, and engaging; (2) the illustration of the character were appealing and emotional. Although some participants did prefer the general story, they mentioned also qualities of the individual story. For example, participant 23 liked the *details in the character illustration that expressed the character's emotional change*.

On the other hand, participants explained their preference for the general story as follows:

- P02: *"Facts are at the forefront."*
- P19: *"[The general story] was visually more appealing and better to understand."*
- P20: *"The order of the stories probably influenced which one I liked better. In [the general story], I had the feeling that the quotes came from different women, which I liked more, as well as the more neutral impression. But if I had seen the character/ persona first, I would probably have missed her in Story 2."*
- P26: *"The [individual] story was something different, something you hadn't seen before. All in all, however, I felt less stressed during the [general] story and therefore more comfortable."*
- P23 preferred the individual story aesthetically. However, she mentioned two main reasons for preferring the general story: (1) *While in the individual story, she focused a lot on the character, in the general story, she could focus more on herself and what this pregnancy disease could mean for her. She mentioned the topic of pregnancy risk.* (2) *Reading reports from several women in the data visualization part made the story more interesting, and these quotes touched her more deeply.*

In summary, some participants preferred the ***general story*** because it was more neutral and less emotional. Therefore, they found it easier to read the stories and focus on the content. While the individual story only provided one quote from the protagonist for each symptom, several participants stated it more interesting and enriching to find quotes from different women for each symptom.

Some participants mentioned different levels of emotional engagement with the stories. I found that, for some participants, their self-reports matched the emotional arousal data; for other participants, the self-reports did not match their data.

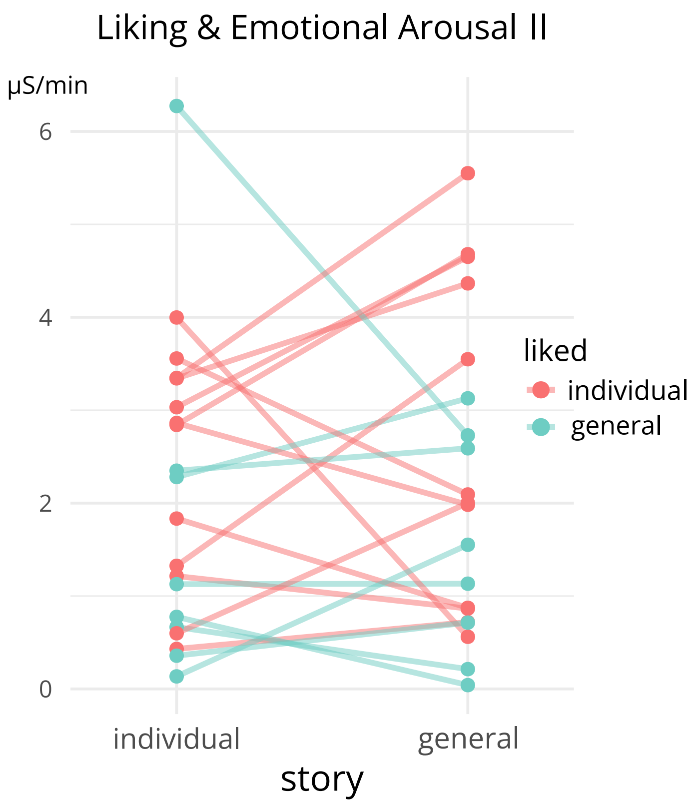


Figure 10.9: "Link between liking and emotional arousal?"

To quantify whether a calmer state is preferred over emotional arousal, I visualized the emotional arousal means for both stories on the y-axes of a parallel coordinates plot (Figure 10.9) where the color indicates the preferred story. Each line represents a participant.

If a line increases and is colored teal, it means the general story is preferred and evokes more emotion. Five participants experienced this, while three were less emotional with the general story but still preferred it. Conversely, of the twelve participants who preferred the individual story, five also felt more emotionally engaged with it, while seven felt less engaged. The results were mixed, however, with a larger sample this question could be investigated in more detail.

The next section will present the results of the analysis of the self-reported emotions experienced while viewing the stories, as requested by the Emotion Types Questionnaire in Section A.3.

10.5 EMOTION TYPES ANALYSIS

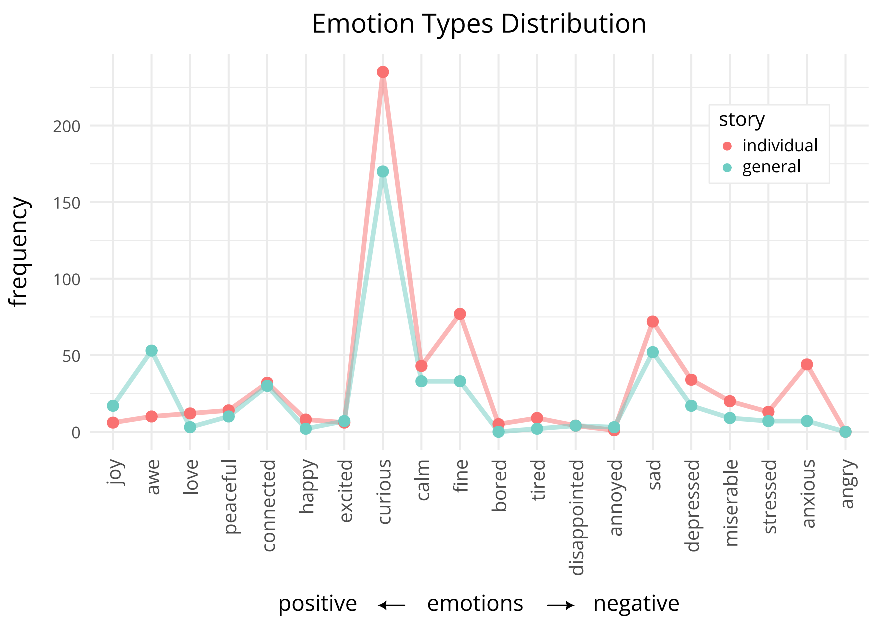


Figure 10.10: Overview of the distribution of emotion types in the first story.

Emotion	*p-value*	*Cliff's delta* [95% CI]	Ind. story | Gen. story *mean*		
awe	0.0001***	0.81 [0.28, 0.96], large	0.71	<	3.78
anxious	0.0002***	0.78 [0.42, 0.92], large	3.14	>	0.50
joy	0.02*	0.44 [0.03, 0.72], medium	0.42	<	1.21

Table 10.4: Significant differences in emotion types (first story).

The overall distribution of the twenty emotion types is displayed in Figure 10.10 for the first story that participants saw. The emotions are sorted by valence (from positive to negative) and by arousal (from least arousing in the centre to most arousing on the edges). All emotions were selected except *angry*. Participants used the option of selecting more than one emotion differently for the stories. Group A selected 64.5 % additional emotions, while Group B reported on the general story 49.0 % additional emotions. *Curious*, *fine*, and *sad* being the most frequently chosen emotions for the individual story, with the most frequently reported emotions regarding the neutral story were *curious*, *awestruck*, and *sad*. The positive emotions *awestruck* and *joy* showed significant p-values with $p = 0.0001^{***}$ (*large effect*) and $p = 0.02^{*}$ (*medium effect sizes*), respectively; with higher means in the general story. On the other hand, *anxiety* ($p = 0.0002^{***}$, *large effect size*) was more often experienced in the personal story. *Curiosity* only missed statistical significance with ($p = 0.05$, *Cliff's delta* = 0.28, 95 % CI [-0.23, 0.67] → *medium effect size*) in favor of the personal story.

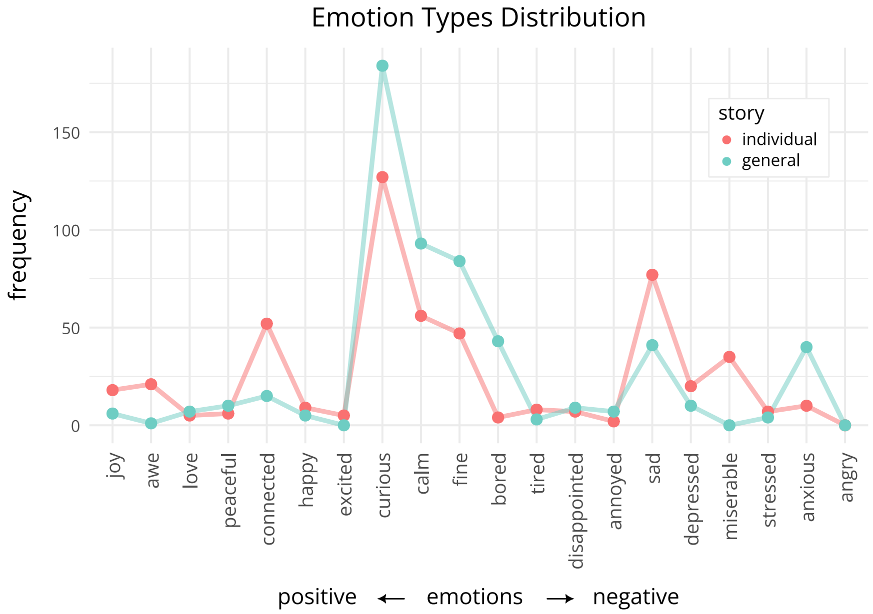


Figure 10.11: Overview of the distribution of emotion types (second story).

Figure 10.11 illustrates that larger differences in certain emotions between the participant groups occurred in the second story. In Table 10.5, nine out of twenty emotions showed significant p-values, with eight having *large effect sizes* and one having a *medium effect size*. Notably, participants reported feeling more *connected* with the personal story after reading the general story. There was also a significant increase in reports of *boredom* among participants who saw the neutral story after the individual story. A similarity across the views was that the *'trio of melancholy'* (*sad, depressed, and miserable*) was still more often selected for the personal story, regardless of viewing sequence.

Emotion	*p-value*	*Cliff's delta* [95% CI]	Ind. story | Gen. story *mean*		
awe	0.0002***	0.73 [0.38, 0.90], large	1.50	>	0.07
miserable	0.0005***	0.64 [0.30, 0.83], large	2.50	>	0.00
bored	0.0008***	0.70 [0.12, 0.92], large	0.28	<	3.07
connected	0.0008***	0.72 [0.34, 0.90], large	3.71	>	1.07
fine	0.012*	0.55 [0.08, 0.81], large	3.35	<	6.00
anxious	0.0183*	0.50 [0.08, 0.77], large	0.71	<	2.85
calm	0.020*	0.51 [0.08, 0.78], large	4.00	<	6.64
curious	0.027*	0.49 [0.04, 0.77], large	9.07	<	13.14
sad	0.047*	0.43 [-0.0007, 0.73], med.	5.50	>	2.92

Table 10.5: Significant differences in emotion types (second story).

Valence

No significant difference was found in *positive emotions* (p = 0.180, *negligible effect size*). However, given that the story is designed to evoke empathy in the audience for women suffering from the disease, I assumed that the personal nature of the ***serious story*** in particular would make the experience more relatable and therefore elicit emotions that correspond with empathy.

Empathy

First, I examined emotions connected with empathy, as listed in Table 10.6. These emotions are consistently and strongly associated with empathy in psychological research (affective empathy/empathic concern). Overall, the individual story scored higher for the empathetic emotion *connected*, with 84 points versus 35 points (Figure 10.15). The results of the user engagement questionnaire support this finding: *"I felt emotionally connected to the protagonist"* resulted in a mean score of 3.57, compared to 3.34 for *"feeling connected to women with the condition."* However, connectedness is not the only emotion associated with empathy. Feeling sad when someone else is sad is called *'sympathetic sadness.'* This is direct emotional contagion, or mirroring, of another's sadness. Feeling *anxious* or *concerned* can also be a sign of empathy by reliving someone's distress. *Concern* can also be connected to *love*. For example, it is common to feel *happy* or *proud* when someone else experiences success or happiness. For instance, *excitement* about pregnancy was more often reflected in relation to the character's experience. Sometimes, participants expressed *disappointment* when the character did not enjoy the pregnancy as much as she had hoped, which could be interpreted as empathy as well. A more extreme form would be *'empathetic anger.'* This would, however, require a high level of personal involvement, such as when someone has been wronged or treated unfairly.

Empathic emotions

Based on the emotions listed in Table 10.6, I evaluated the question whether there is a difference between both stories in empathic emotions, including positive and negative emotions. I obtained a significant p-value of 0.016*. However, the effect size was small (*Cliff's delta* = 0.16, 95% CI [0.02, 0.30]), indicating that the individual story showed a tendency to elicit more empathy than the general story with little practical impact ($mean(ind.\ story) = 3.761 > mean(gen.\ story) = 2.158$).

Hypothesis 5

Negative empathic emotions

Hypothesis 5 states that the personal story would elicit more strong negative emotions that correspond with empathy, i. e., *sad, depressed, miserable*, and *anxious*. The mean for these emotions was 3.035 for the individual story and 1.517 for the general story which already shows a tendency. I conducted a significance test to verify this hypothesis and found a highly significant difference between the two story versions (p = 0.0009***, *Cliff's delta* = 0.27, 95 % CI [0.06, 0.46] $\rightarrow$ *small-to-moderate effect size*. With a sample size N = 112, the sample is not overpowered to the point of detecting meaningless microscopic effects, therefore there is very strong evidence that the two stories

EMOTION	TYPE OF EMPATHY	CONTEXT
sad	Sympathetic sadness (core of affective empathy)	*feeling sad because someone else feels sad*
depressed	Compassion	*Lowered mood when seeing another's depression*
miserable	Compassion	*Feeling another's suffering*
connected	Shared emotion	*feeling "with" someone*
love	Empathic concern, Affection	*Caring, triggered by someone's vulnerability*
joy / happy	Positive empathy	*Sharing joy or happiness of another person*
anxious	Emotional contagion	*resonantly feeling another person's anxiety, personal distress*
stressed	Emotional contagion	*vicariously feeling tension, personal distress*

Table 10.6: Emotion types associated with empathy.

truly differ in negative empathic emotions. However, the magnitude of that difference is small-to-moderate, and could be quite small in practical terms if the true effect size is close to the lower bound of CI. Therefore, the null hypothesis can be rejected under the consideration of a modest practical impact, in favor of the alternative hypothesis that the character-driven story elicits more strong negative emotions. This finding further supports Hypothesis 1.

In the following, I will show the results at the story piece level. The progression of emotions throughout the stories are visualized by heat maps. The emotion types are sorted in the same order as before. Here, positive emotions are displayed on the upper half, negative emotions in the lower half, and the arousal level increases from the center to the top and to the bottom. I used the absolute values and a *log* scale to make finer differences in frequencies more visible – since some emotion types were selected much more often. High values are represented by warm colors ranging from yellow to orange, moderate values are represented by green, and lower values are displayed by blue to violet. Dark blue indicates that a particular emotion was not selected by a single participant. First, I will consider the data from the first story participants saw.

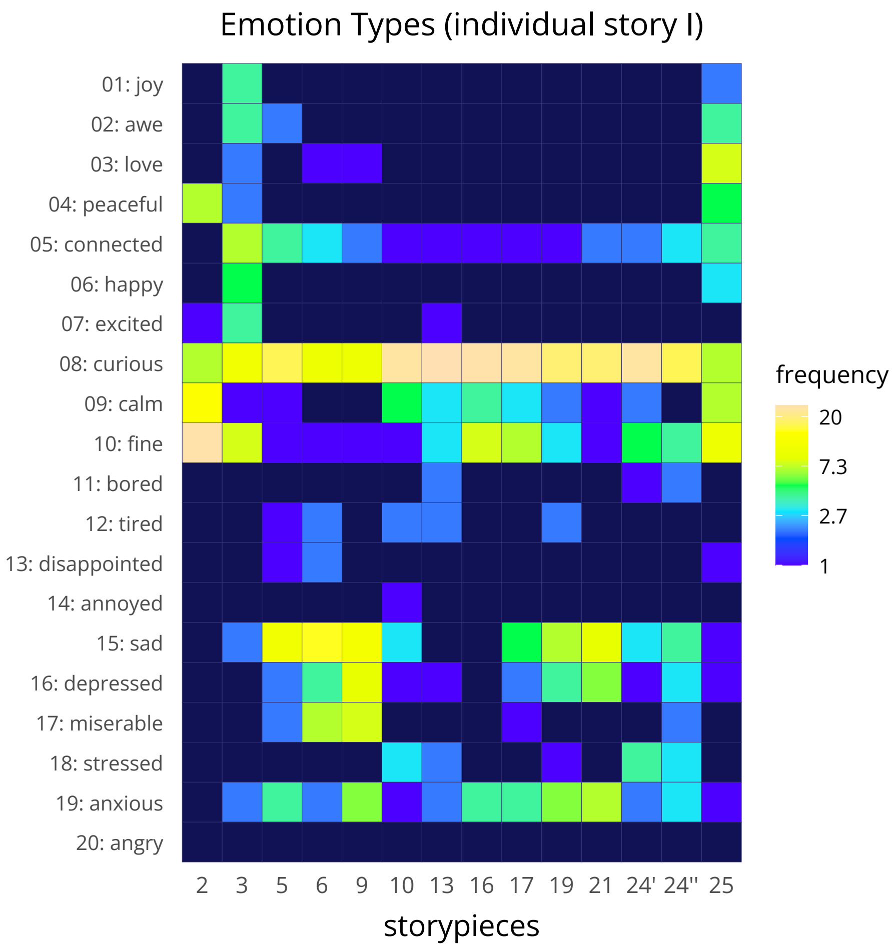


Figure 10.12: Frequency of emotion types over the course of the individual story.

Similarities

The heat maps Figure 10.12 and Figure 10.13 show that positive emotions were experienced more often in the introduction and resolution, especially in storypiece 25, which concludes with a happy ending. In the resolution of both stories, the most frequently selected emotions were *awestruck*, *love*, *peaceful*, and *connected*. *Curious* was selected most frequently throughout both stories, peaking around the explanation of the disease. Negative emotions tend to be more prevalent in the *'spiral of escalation'* (storypieces 5, 6, 9) and in the *data visualization section* (storypieces 17, 19, 21, 24). Notably, storypiece 24 elicited more negative emotions when the quote appeared. In this story piece, positive emotions also increased, with more subjects selecting *connected* in response to the interactive element.

Differences

A key difference between the stories was the different distributions of positive emotions. The individual story showed only strong positive emotions in the introduction and resolution, and a higher amount of negative emotions in the *spiral of escalation* and *data visualization part*, such as *sadness*, *depressed*, *miserable*, and *anxious*, than the responses

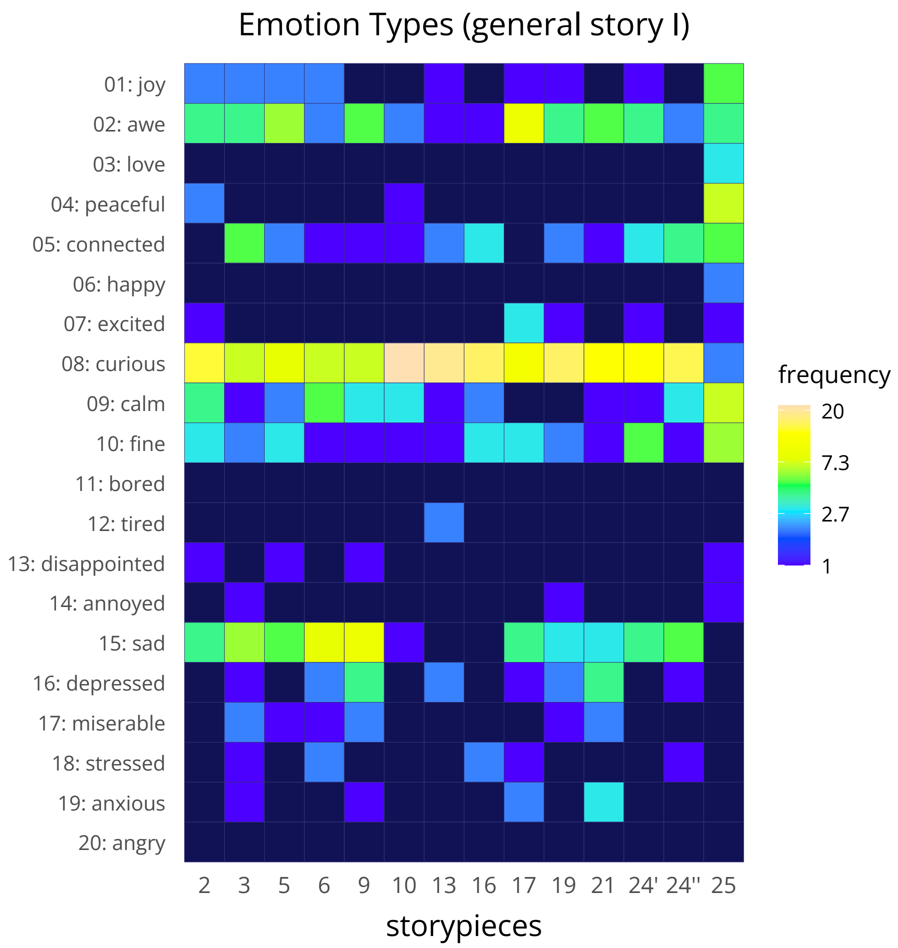


Figure 10.13: Frequency of emotion types over the course of the general story.

to the general story. Many participants felt strong positive emotions during storypieces 2-3, demonstrating that the character introduction was highly positive arousing. In contrast, the introduction did not have a comparable effect on participants who read the general story, but showed stronger positive emotions throughout the story, e. g., *awestruck* and *joy*. In the personal story, *connected* revealed a decreasing gradient after the introduction that increased again towards the end. The neutral story did show a more irregular pattern. While *curiosity* dropped suddenly in the last story piece of the general story, *curiosity* about the individual character's story ending was still moderate.

Visual symmetry & Rhythm

When comparing both heatmaps, the individual story shows a symmetry pattern: Positive emotions at the beginning and the end, strong negative emotions in the first third, and last third, and calmer emotions in the middle. Another interpretation could be that there is an emotional engagement rhythm: Highly positive – highly negative – calmer – highly negative – highly positive. Whereas the general story shows a more horizontal pattern, except for the resolution.

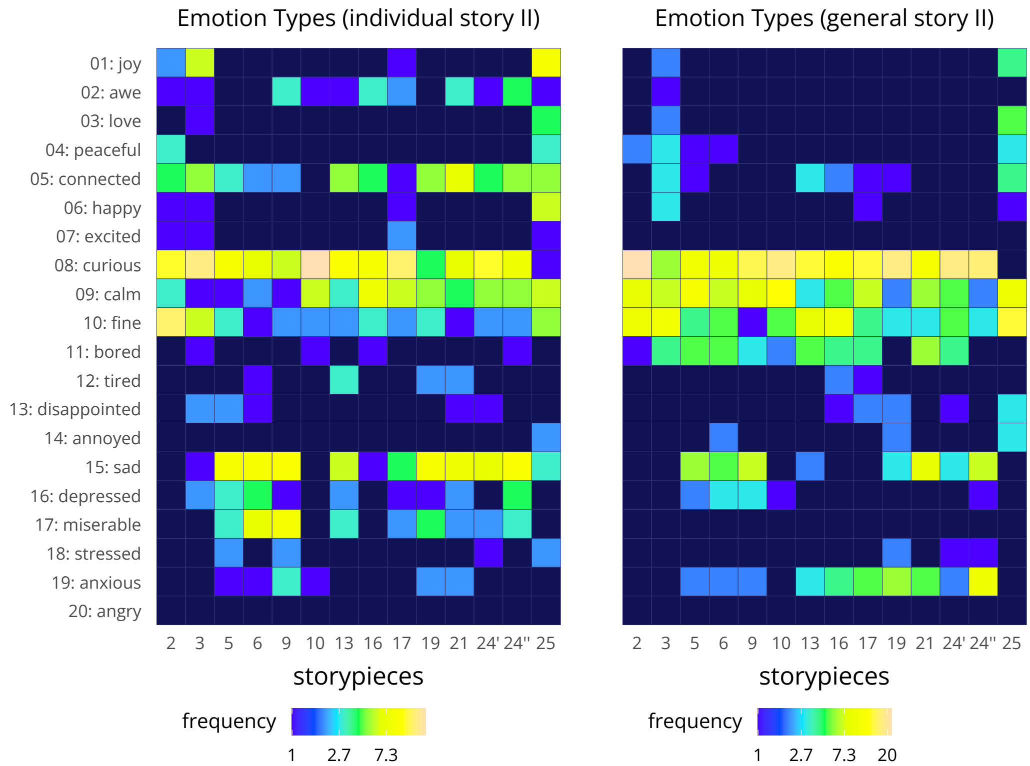


Figure 10.14: Frequency of emotion types throughout story 2.

Story 2 A main difference in the second viewing of the story (Figure 10.14) arises in the emotion *connected*. Participants who read the neutral story first often felt connected to the character-driven story throughout, except for the storypiece 10 where the disease was explained and the character was not shown by an illustration. Controversially, the other positive emotions, *awestrucked* and *joy*, were now less frequently selected by Group B. Another main difference was that *bored* was now frequently selected when the story was repeated in a neutral manner but rarely when it was told more personally. When comparing the emotion selection within Group B, the horizontal pattern was mostly maintained. A difference became obvious in the *'Trio of Melancholy'*. Here, the pattern was similar to the group that first saw the individual story, peaking in storypieces 5-9 and storypieces 19-24. In this group of emotions, the conflict situation of the protagonist had the most effect. Group A maintained certain selection patterns. For example, they selected nearly the entire range of strong positive emotions in storypiece 3 and 25, but almost none in the middle of the story. However, the group selected continuously feeling *curious*, *calm*, *fine*, and *bored*. *Anxious* and *sad* were selected more often in the data visualization part which aligns with some participants stating that different women's quotes were engaging and moving.

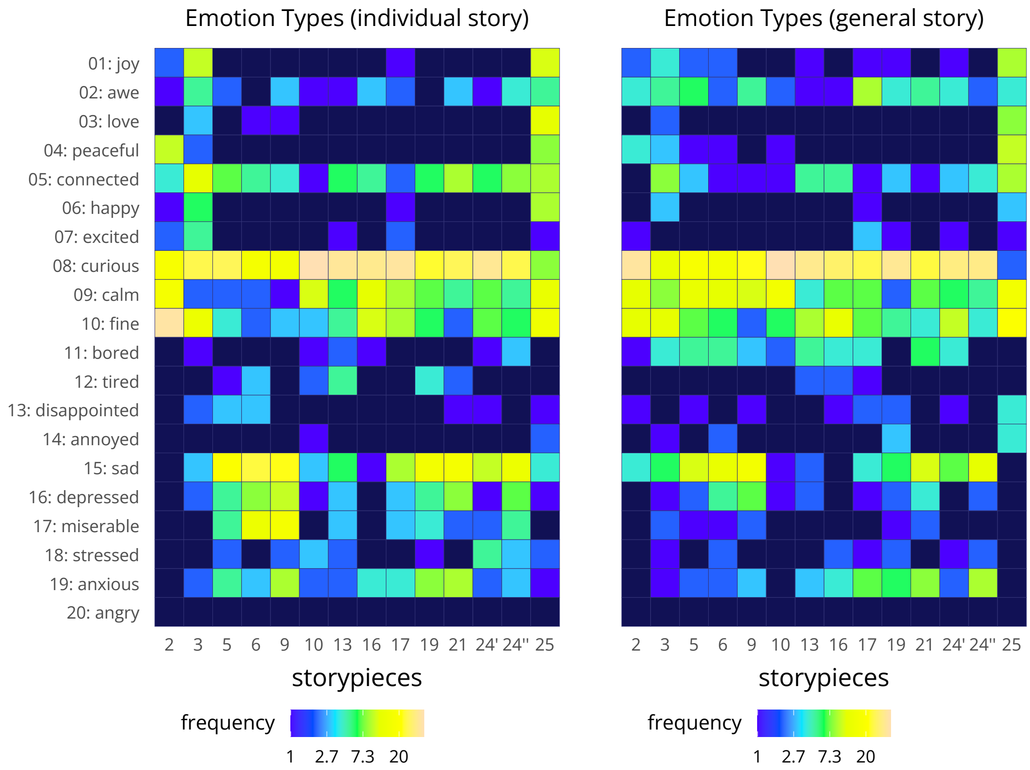


Figure 10.15: Frequency of emotion types throughout both views.

As seen in Figure 10.15, the distributions of emotion types related to story pieces converge when all the data from both views are analyzed together. This is similar to the effect seen in the EDA data, though the character-driven story elicits slightly more emotional engagement overall. This is evident in the frequency and variety of emotion types, as well as in the selection of fewer emotion types with lower arousal. Instead, more intense negative emotions were chosen for certain story parts, such as the conflict situation.

In summary, while some emotions, such as curiosity, remained consistent throughout the story, most did not. This emphasizes that the stories were well-designed to evoke different emotions and produce emotional tension. The character-driven story in particular evoked more emotional engagement and empathy.

10.6 LOCATION ANALYSIS

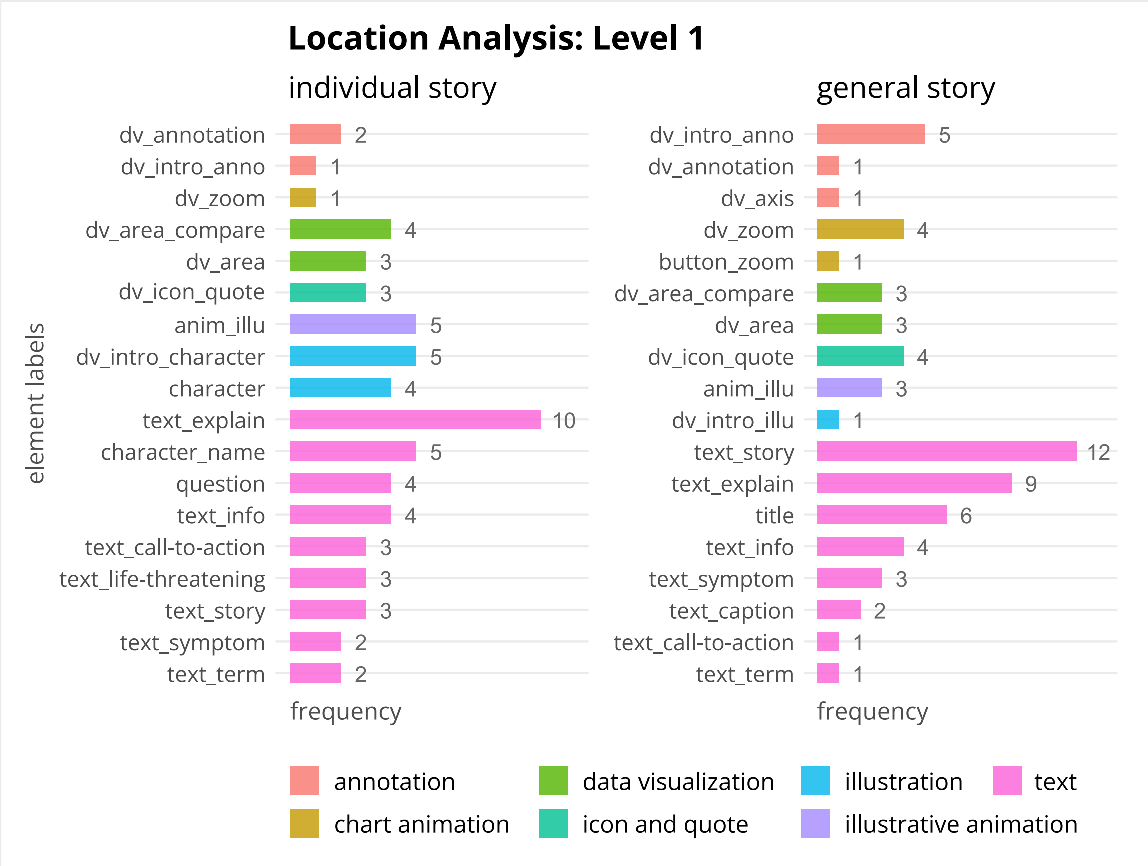


Figure 10.16: Labeling and categorizing of visual story elements of six maximum peaks for each participants (where available).

The eye tracking data provided additional context for interpreting EDA peaks. For the location analysis, I labeled all elements associated with each participant's three highest peaks in each story (Figure 10.16). Since EDA responses can be caused by emotional stimuli or cognitive tasks, I created more specific labels at the first level to categorize them later into emotional or cognitive responses at the second level as shown in Figure 10.17. Each element was considered based on its association with higher cognitive processing, such as understanding explainable text or interpreting data visualization annotations. Other elements were assigned to the cognitive category due to their lack of emotional content, such as informational text. Conversely, elements with weaker associations to cognitive effort and stronger associations to emotional aspects of the story, such as character representation, story-based text, icons, and quotes, were assigned to the emotional category. For some elements, the association was more ambiguous. For instance, understanding data visualizations or comparing two juxtaposed visualizations requires cognitive effort. On the other hand,

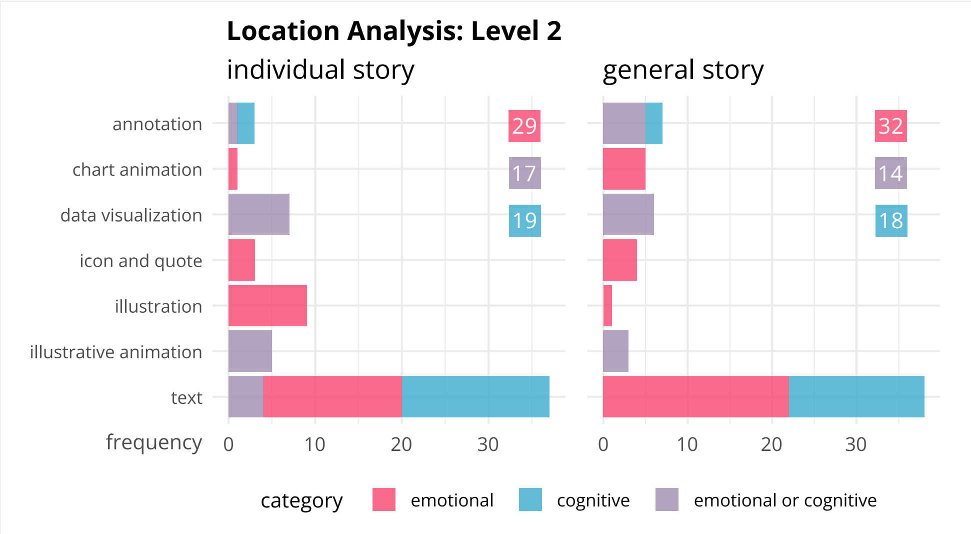


Figure 10.17: Categorizing of elements with emotional or cognitive involvement association.

recognizing meaningful insights from data related to human experience or the effect of visual design can also have an emotional impact. These labels must be considered in light of the psychological interactions of emotional and cognitive engagement and the uncertainty surrounding the actual source of the peak cause. Peaks do not necessarily occur due to an element that was looked at; they can also occur due to thoughts about elements seen before, not task-related thoughts, or mind wandering.

text elements

Eye tracking analysis revealed that most of these peaks were related to emotional stimuli and that participants spent a significant amount of time reading. Text elements were also the main source of the maximum peaks: 55.38 % in the character-driven story and 59.37 % in the more neutral story. Most story-based text responses occurred in the general story, whereas question- and call-to-action-based responses occurred more frequently in the personal story. There may be a tendency to perceive a call to action as more urgent when a character is involved. Notably, the character's name accounted for 7.69 % of the responses in the character-driven story, underscoring the importance of naming a character. I identified three instances of strong emotional responses to the message that the disease threatens the protagonist's life. However, I found no peaks corresponding to this message in the general story.

Character name

Hypothesis 6

High peaks of emotional responses were significantly more aligned with character illustrations (p = 0.0015**, *Cliff's delta* = 0.16, 95 % CI [0.06, 0.26] $\rightarrow$ *small effect size*) than with illustrations in the general story. Therefore, the null hypothesis can be rejected in favor of the alternative hypothesis that illustrations corresponding to personal stories influence emotional engagement. However, the effect size was small, likely due to the small sample size. The practical effect could be higher with an analysis of additional peaks. This result corresponds with some participants expressing appreciation for the character illustrations. On the other hand, only one person mentioned preferring the visual style of the neutral story.

Data Visualization & Animation

Responses to the data visualizations, including the animated chart, were very similar between the two stories. There was slightly more response to the general story (13.81 % versus 12.63 %). Both stories contained an animated data visualization and an animated illustration. These visuals were associated with six responses (9.23 %) in the personal story and seven responses (10.93 %) in the general story. Considering that only two animations were used, animation positively affected the emotional experience. Some study participants explicitly stated that they were surprised by the zoom-in effect. Many of them repeatedly pressed the zoom button to see it again. This underscores the fact that integrating playfulness into data visualization can evoke emotional responses.

Quotes from women suffering from the disease corresponded similarly with emotional responses in both stories. Overall, 5.38 % of all responses were associated with the interactive "icon and quote" component of the data visualization.

At this point, the eye tracking analysis is in its preliminary stage, considering only 10.35 % of all EDA amplitudes. The large number of peaks, totaling 1,246, requires an automated method to detect and label visual elements. A concept idea is described in Chapter 11. Further analysis of the additional peaks would provide more insight into the distribution of elements associated with emotional responses.

10.7 PRIMING EFFECT ON EMOTIONAL ENGAGEMENT

This section examines the between-subject dynamics in the two participant groups and whether the sequence of the stories mattered. To recap, Group A read the individual story first, while Group B read the general story first. As described in Section 4.4, emotional priming suggests that an emotional experience depends on a previous stimulus; in this case, the first story. Figure 10.18 provides an impression that Group A was more emotionally involved. Participants introduced to the topic by an individual character had higher means in all emotional engagement variables. Significant differences were found in the maximum peak (p = 0.023*, *Cliff's delta* = 0.42, 95 % CI [0.05, 0.68] → *medium effect size*) and maximum sum of amplitudes in a story piece (p = 0.0107*, *Cohen's d* = 0.83, 95 % CI [0.18, 1.47] → *large effect size*). Both CI's do not contain zeros, however, the max sum has higher practical relevance. The sample size of N = 41 (A = 23, B = 18) was sufficient to show an effect by maximum EDA peak features.

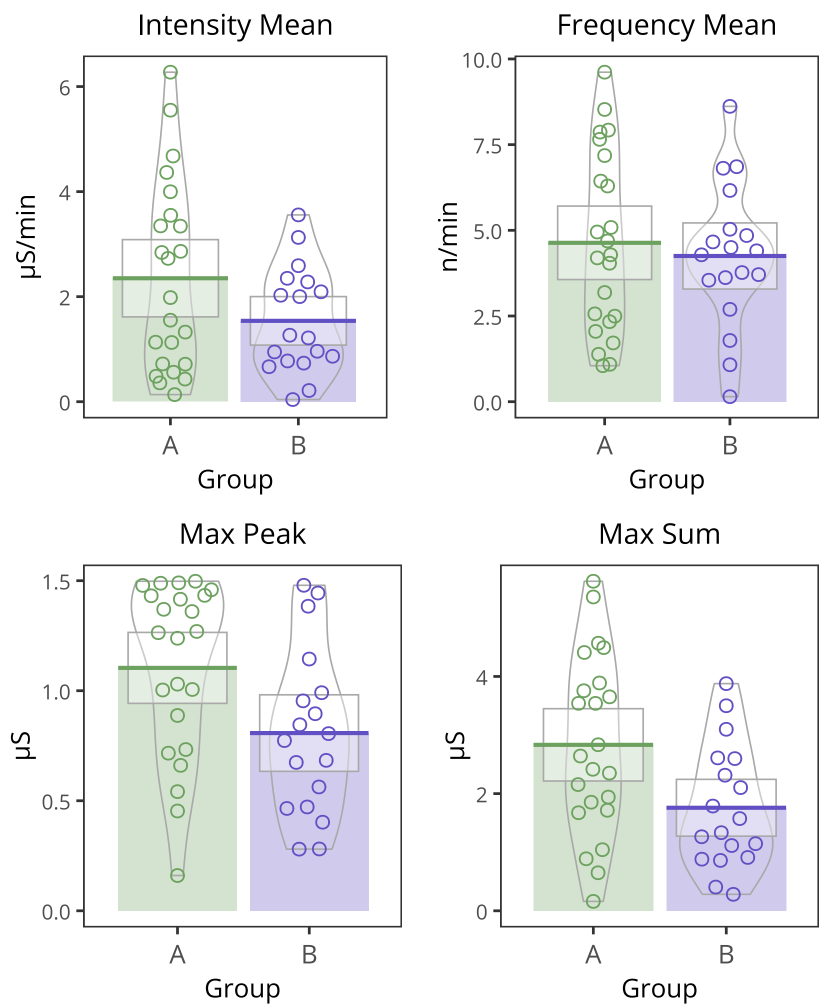


Figure 10.18: Emotional engagement in participant groups.

These findings indicate that Group A was more emotionally primed resulting in overall higher emotional engagement. On average, Group A spent 7:33 minutes on the stories, while Group B spent 6:10 minutes (p = 0.091, *Cohen's d* = 0.56, 95 % CI [-0.07, 1.19] $\rightarrow$ *medium effect size*). Time spent was significantly higher for the first story in both groups, which is reasonable given that the content of the second story was mostly the same. Group A spent an average of 8:39 minutes on the first story and 6:22 minutes on the second story. Group B, on the other hand, started with an average time of 07:21 min and 04:39 min on the second story. This suggests that participants in Group B were more likely to lost interest in the story during the second session. The parallel coordinates plots in Figure 10.19 visualize the dynamics between story views. Each line represents a participant. The left axis of each plot depicts the value of average emotional arousal in the first story viewed by a participant, and the right axis depicts the value of the second story viewed. The derivative of each line shows the change in emotional engagement. When the derivative is positive, the line is colored red, showing an increase in emotional arousal. Blue indicates a negative derivative, or a decrease in arousal.

	Group A				Group B			
VARIABLE	$\nearrow$		$\searrow$	*tendency*	$\nearrow$		$\searrow$	*tendency*
Intensity	8	$>$	3	increasing	4	$<$	5	decreasing
Frequency	7	$>$	4	increasing	3	$<$	6	decreasing
Max peak	7	$>$	4	increasing	2	$<$	7	decreasing
Max sum	6	$>$	5	increasing	5	$>$	4	increasing

Table 10.7: Change in emotional engagement between first and second story.

Hypothesis 7

In table Table 10.7, the results of both participant groups are summarized. Group A showed a tendency of increased emotional engagement from the first story to the second story, while Group B showed rather a decreasing trend in emotional engagement. Hypothesis 7 states that an emotionally primed person will show increased emotional engagement with a subsequent story, while a less emotionally involved person will show decreased emotional engagement. Given the higher emotional engagement with the individual story analyzed in Section 10.1, Group A was more emotionally primed. The results showed a tendency of increased emotional engagement with the second story. Group B, on the other hand, which was more neutrally primed, showed a tendency of decrease in emotional engagement.

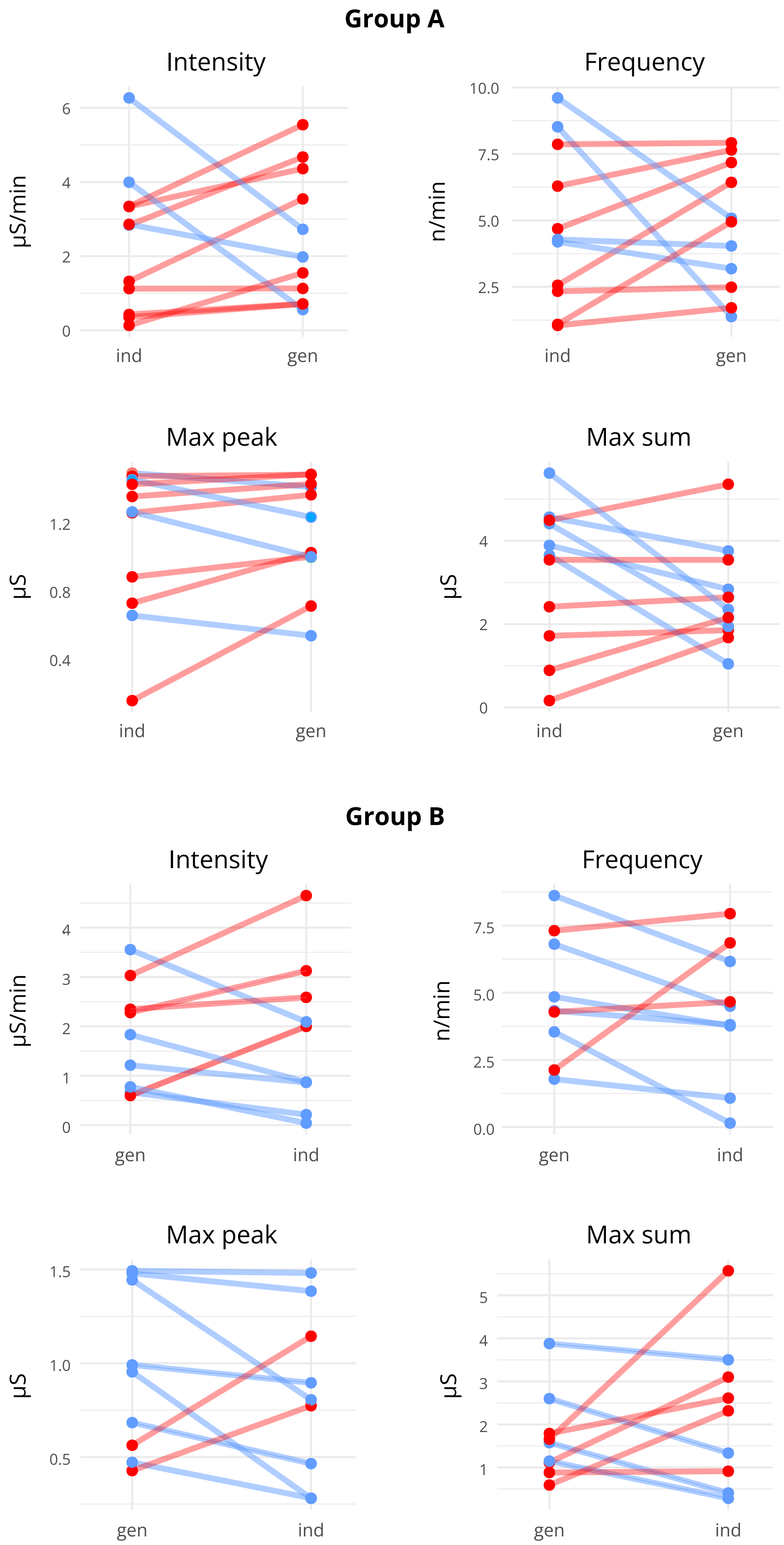


Figure 10.19: Emotional priming: Increase (red lines) and decrease (blue lines) in emotional engagement between stories.

10.8 UNDERSTANDING & MEMORY

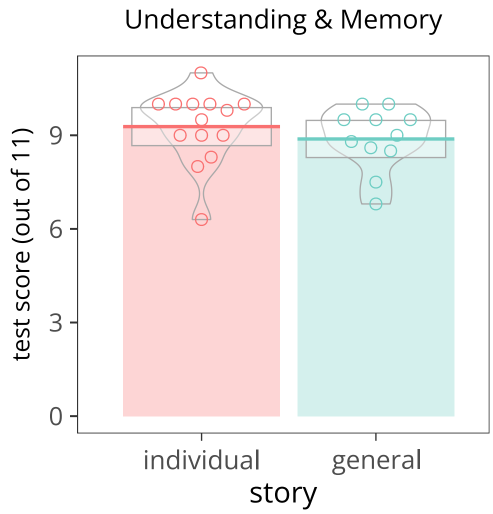


Figure 10.20: Scores of the understanding and short-term memory test for both stories. The test contained 11 questions a 1 point.

The understanding and short-term memory questionnaire (Section A.2) contained eleven questions. Overall, the scores presented in Figure 10.20 were positive in both stories, supporting the idea that the stories successfully conveyed its intended messages. The mean score was slightly higher for the individual story (μ = 9.28) than for the general story (μ = 8.88). One person in Group A achieved a perfect score of 11 points after reading the personal version. The scores for this story had a slightly larger range of 4.7 points and contained also the overall minimum score of 6.3 points. The other story version resulted in a range of 3.2, with a minimum of 6.8 and a maximum of 10. This suggests that a more neutral presentation might result in a more consistent performance across individuals.

Hypothesis 8

There was no significant difference found in recall performance between the stories (p = 0.254, *Cliff's delta* = 0.27, 95 % CI [-0.19, 0.64] → *small effect size*). However, I found a significant difference in the responses to the question of whether women with HG often have to visit the hospital (p = 0.0064**, Cliff's delta = 0.45, 95 % CI [0.09, 0.71] → *medium effect size*). All participants in Group A responded correctly, suggesting that the protagonist's experience was more memorable than simply stating the fact. Another interesting observation was that Group A performed better on four out of five content-related questions, while Group B performed better on all four questions related to data visualization (Q4, Q9–11).

In Figure 10.21, the test score of each participant with sufficient EDA data is correlated with the amplitude mean per minute. The distribution is certainly nonlinear, with an R-squared value of 0.049 and

Figure 10.21: Scores of the understanding and short-term memory test correlated with emotional response and superimposed with a quadratic regression model.

a p-value of 0.32. A quadratic regression model fits the data slightly better because its R-squared value of 0.237 means the model explains more of the data's variability. The p-value of 0.076 was also closer to 0.05. Moderate emotional responses were associated with lower test scores, while higher test scores showed greater variation in amplitude height per minute, ranging from higher to lower emotional arousal. Lower emotional intensity appears to be beneficial for understanding and memory tasks. About half of the data points are clustered in the lower right corner of the chart. This corresponds to the self-reports of some participants who stated that a calmer state of mind helped them concentrate better when processing information.

At this point, it is difficult to claim any correlation between EDA data and recall. With more data, one could investigate whether the correlation between emotional engagement and recall test score is funnel-shaped along a quadratic model — meaning that some participants perform better with high emotional engagement, while others perform better with low emotional engagement.

In conclusion, the individual story produced slightly better results overall and on content-related questions, while the general story produced better results on data-related questions. Higher recall scores tend to be associated with lower emotional involvement, but a person can be highly emotionally involved and have high test scores.

10.9 USER ENGAGEMENT & QUALITATIVE FEEDBACK

At the end of the results presentation, I will compare the participants' reports on user engagement to investigate differences in story experience in terms of self-reported *Affective Involvement*, *Cognitive Involvement*, *Focused Attention*, *Usability*, and *Novelty*.

In Figure 10.22, the results of each question from the User Engagement Questionnaire Section A.4 are presented. The top bars depict the mean for the individual story (ind) and the lower bars depict the mean for the general story (gen). Notably, all questions were rated higher with respect to the individual story.

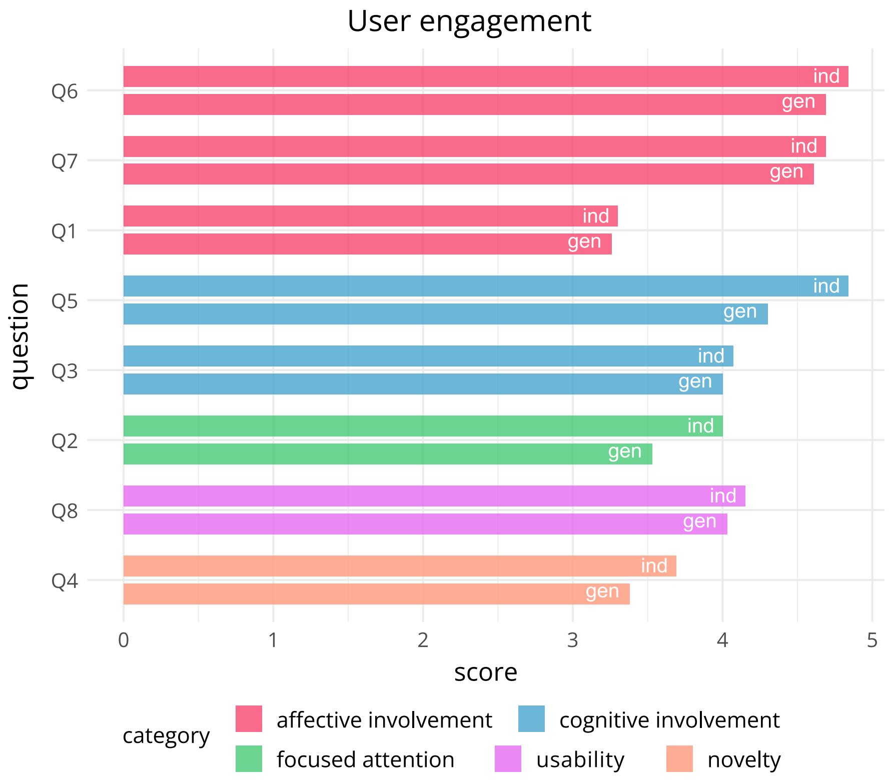


Figure 10.22: User engagement reports.

Novelty & User experience

Most participants stated that the content was completely new to them (Q4). Some study participants explained they were aware of the disease because women in their social networks, such as friends or family members were affected by it or in one case the participant herself. Overall, participants had few issues recognizing elements of the interface in both stories (Q8). However, I received written and informal feedback on how to improve the data visualization. For example, the slightly transparent colors of the area chart confused some subjects, and some participants wished for an additional cue indicating that the data points were interactive.

Focus attention & Cognitive involvement

On average, participants said that their thoughts drifted less during the individual story (Q2). Furthermore, participants reported that the content in both stories was equally easy to process and understand (Q3). However, participants stated that the character-driven story deepened their understanding more than the neutral story (Q5). Additionally, some mentioned that they would like to learn more about the disease.

There was no significant difference in self-reported affective involvement ($p = 0.521$), the effect size was negligible, and the average score was only slightly higher for the character-based story ($\mu = 4.282$ versus $\mu = 4.192$). The majority of participants expressed surprise at the impact of the disease in informal feedback and considered it important to spread information about it (Q6). Consequently, many would recommend the stories to their family members (Q7). Feeling connected to women with the disease or the protagonist was reported as moderate for both stories (Q1). The results for the five dimensions are summarized below Figure 10.23.

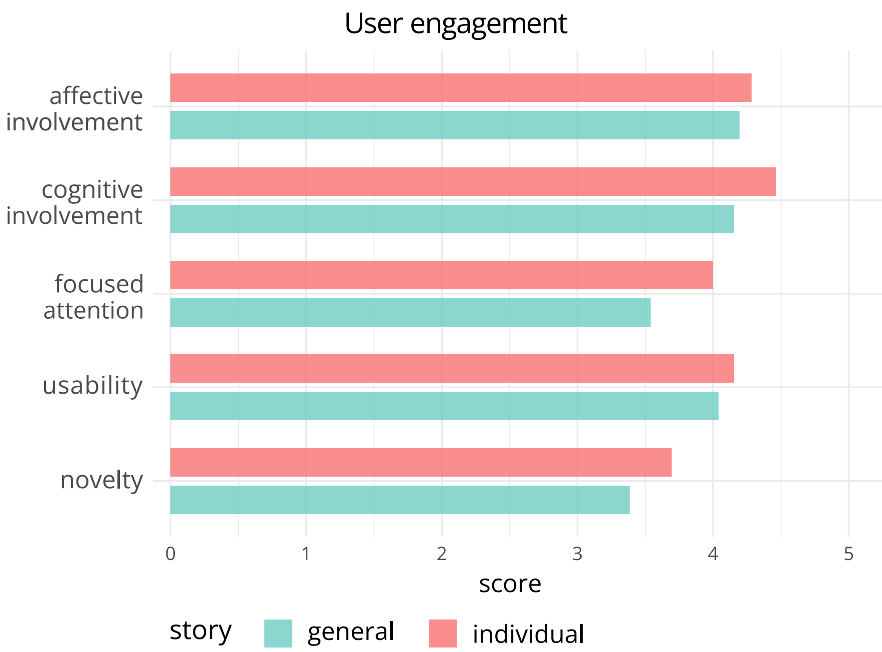


Figure 10.23: User engagement summary.

11

DISCUSSION

One of the most important goals of Narrative Medical Visualization is to evoke emotional engagement with a medical topic or disease in order to increase interest, attention, recall, and behavioral changes. Narrative characters play a crucial role in this context. However, previous studies investigating engagement had several limitations, including reliance on self-reports and high-level analysis of overall engagement. This study proposes a mixed-methods approach that includes physiological, continuous measurements to address the challenge of measuring emotional engagement in medical data storytelling.

11.1 KEY FINDINGS & INTERPRETATION

In this section, I will restate the research questions and summarize what was found related to each hypothesis. I also will interpret the findings and relate them to previous work in the field of Narrative Medical Visualization.

11.1.1 *Emotional Engagement*

This subsection discusses the results on the first research question.

RQ1: *"Does a fictitious, individual character arouse greater emotional engagement in the story's viewers for a medical condition compared to a general story with no individual human protagonist?"*

EDA analysis on story level

Electrodermal indices of average emotional arousal (i. e., average amplitude and frequency of skin conductance responses per minute) showed no meaningful differences between the stories (*small effect sizes*, p-values > 0.27). In contrast, measures capturing peak emotional arousal revealed *large effects* favoring the character-driven story: Maximum peak amplitude ($p = 0.078$, *Cohen's* $d = 0.81$, 95 % CI [-0.07, 1.68]) and maximum sum of amplitudes on the most arousing story piece ($p = 0.060$, *Cohen's* $d = 0.86$, 95 % CI [-0.03, 1.73]). Although these effects did not reach the conventional $\alpha = 0.05$ threshold, the large effect sizes and confidence intervals that only barely include zero indicate that the absence of statistical significance is very likely due to the limited sample size (N = 22). Here, the achieved power is ≈ 45 %. The results provide promising evidence that the manipulation in character usage selectively enhanced peaks in emotional engagement. This pilot study showed that the maximum peak features are most promising in identifying differences between story versions. The sample size, however, have to be larger to detect large effects with higher certainty. To detect these large effect sizes, $d = 0.84$ (average of both large effect sizes) with 90 % power ($\alpha = 0.05$, two-tailed) in a future study, the sample size should be $N \approx 62$ (31 per group). Alternatively, with 80 % power, the sample size had to include around 50 participants (25 per group). A replication or extension with 25–31 participants per group would probably turn the current $p = 0.06$ into $p < 0.01$–0.05.

A comparison with the results of a previous study [75] which also compared a patient story with a base condition story found no difference in emotional response using self-reports, the results of this pilot study revealed that there might be a significant difference with a larger sample size.

There was a notable trend that almost all measures in this study reached higher results for the character-driven story, i. e., on the four emotional engagement variables, number of story pieces with higher emotional arousal, time spent, liking, number of emotion types, curiosity, score of strong negative and empathetic emotions, recall, and user engagement.

I compared participants' self-reports of feeling more emotionally engaged with one story or another with the corresponding EDA data. In some cases, the self-reports matched the data; in others, they did not. Additionally, the results of the user engagement questionnaire, which measured self-reported affective involvement, were very similar for both story versions ($p = 0.521$), showing negligible effect size and indicating that a powered follow-up would not reveal significant differences. Given that physiological measurements are more objective and reliable than self-reports, these findings highlight that sole reliance on participants' high-level interpretation of emotional engagement may overlook certain differences. These results underscore the need for more objective measures in Narrative Medical Visualization.

Time spent & Interest

Participants spent on average more time in the personal story. This difference cannot be explained by additional content to introduce the protagonist as Mittenentzwei et al. (2023) [75] assumed because the length of the investigated stories was almost the same. One exception was storypiece 2 where the character was briefly introduced by its name and an illustration, while in the general story the disease was introduced with its prevalence and an infographic with slightly more information to process. A small difference in time spent was found here. Although the content was the same in all other story pieces, participants still spent more time in each story piece of the personal story. Higher time spent was previously interpreted as higher interest in a story. First-person narratives typically produce longer reading times than third-person versions of the same story when length and complexity are controlled [45], which aligns with my findings.

Dixon & Bortolussi (2011) found that passages that were of high interest to a reader resulted in longer reading duration [26] and Nell (1988) found a strong positive correlation between reading time and emotional engagement [76]. Therefore, time spent is a well-validated, reliable, and objective behavioral measure of interest in a narrative. Based on these findings, spending more time could indicate greater interest in and emotional engagement with the character-driven story. High fixation duration has also been associated with interest [51]. The correlation between reading time, emotional engagement, and fixation duration could be investigated in more depth by analyzing the eye tracking data alongside emotional arousal peaks in EDA data. However, some participants mentioned that it was easier to focus on the content of the more neutral story because it was less emotional. Therefore, longer duration could also imply that more time was needed to process the content when the story was more emotional. Mittenentzwei et al. (2023) assumed that including a protagonist in a story would make it objectively longer. These two story prototypes proved that story length can be objectively the same when information and character experience are skillfully combined.

Gender differences

To the best of my knowledge, no study investigated gender differences in emotional engagement in the context of data storytelling. While the level of interest in terms of time spent was similar, I found significant differences in the emotional engagement parameters derived from EDA data between female and male participants, with a large effect size. The second null hypothesis must be rejected in favor of the alternative hypothesis that women felt more emotionally engaged with the stories. A previous study found higher tonic and phasic EDA parameters in females both under stress and relaxation, but no significant difference between females and males [8]. One possible explanation for the results of my study would be that pregnancy and the disease are more relevant topics for women, as they are the ones who could experience them. This could explain why they elicit more empathetic feelings. To investigate this objective, it would be useful to compare these results with those of a medical topic that is not gender-specific.

According to all four parameters, females also felt slightly more emotionally engaged with the individual story. By contrast, males reported higher arousal in two out of four parameters. To prevent gender bias in the evaluation of emotional engagement, it is crucial to balance the participant group in terms of gender. In this study, six women and six men read the individual story first, while five women and five men read the general story first. Therefore, any gender-related influences on hypothesis 1 can be excluded.

EDA analysis on story piece level

The analysis at the story piece level revealed that 60.66 % of individual story pieces evoked emotional responses, which is a 12.96 % difference to the more neutral story (47.70 %). In the character-driven story, higher arousing story pieces were distributed more consistently throughout than in the general story. After the first third, there was a tendency for engagement to decrease in the latter. Denser occurrences of emotional arousal were found in the data visualization intro, explaining the disease, resolution (call-to-action), introduction, and the conflict situation. Given that tension would be felt as emotional arousal intensity, Freytag's classical model, which assumes an increase in tension to the climax followed by a decrease, cannot be applied here. The main purpose of this investigation was to explore differences between general and character-based stories at a more granular level. However, these results are at an exploratory stage. Since there are many missing values at the beginning of both stories, the dataset needs to be more balanced in order to make a claim about statistical significance and practical effects.

Liking

16 out of 26 study participants preferred the individual story over the general story. This shows a tendency for preferring the personal story (61.5 % vs. 38.5 %), but the difference did not reach statistical significance ($p = 0.1$, *small effect size*). Qualitative feedback provided further insight into why a particular story design was liked. Participants who rated the individual story provided the following reasons for doing so: (1) the story was more personal and engaging, and (2) the character illustrations were appealing and emotional. Other participants preferred the general story because it was (1) more neutral and less emotional, and (2) easier to read and focus on the content. Several subjects explicitly stated that the quote design in the general story was more interesting due to the different perspectives of several women. The analysis revealed that people's preferences can differ: some prefer information to be presented in an emotional context, while others prefer a more neutral approach.

Interestingly, the less preferred story elicited slightly more emotional arousal overall. Of the five cases located above all data points, the participants were female. Four of them did not prefer the general story and also four belonged to Group A. I will discuss the possible influence of emotional priming on this objective in the next section. Another possible explanation, which aligns with some participants' statements, is that higher emotional arousal can be associated with stress, which some individuals may find uncomfortable. Female participants, in particular, seemed to prefer a moderate level of emotional arousal (in this study, around 3 *µS/min*). In comparison, there were no data points around 3 *µS/min* in the less preferred story. For males, less clear patterns emerged regarding this correlation. Assuming a difference of around 0.5 *µS/min* between the left and right clusters, males may prefer higher emotional arousal. However, there might be a tendency that overly high emotional arousal is not preferred. Since I could only use data from participants with sufficient EDA data for both stories, the sample size decreased to 20. It would be interesting to investigate this question with more data in a future study.

Emotion types

The stories evoked a wide range of emotions reported by a selection-based questionnaire for half of the story pieces covering all story parts. *Curiosity* was the most frequently reported emotion and remained consistent throughout. This emphasizes that the stories were well-designed to generate interest. The variety of emotions, with a concentration in certain parts of the stories, indicates that the stories produced emotional tension. In particular, the character-driven story evoked more curiosity and empathy. A previous study found no difference in curiosity between a patient story and a story without a protagonist [75]. While this study only used questions that were related to the whole story, I used a more concrete approach where participants do not have to do high-level interpretations. I also did not directly ask whether they felt curious which might have lowered the social desirability bias. Using the selection-based method with 20 emotion types for a presented slide, curiosity was close to a significant difference in favor of the patient story ($p = 0.058$, *Cliff's delta* = 0.42, 95 % CI [-0.03, 0.73]) with a *medium effect size*. Mirroring the protagonist's negative emotions could imply higher empathy. I found significant differences in empathy-related emotion types between the stories; in particular in negative empathic emotions ($p = 0.0009^{***}$, *Cliff's delta* = 0.27, 95 % CI [0.06, 0.46] $\rightarrow$ *small-to-moderate effect size*). As a result, the personal story evoked more empathy. The results also suggest that participants experienced deeper emotional engagement with the character-driven story, as they displayed a greater variety of emotions and more mixed emotions. Previous studies evaluated emotion types for the whole story and did not provide such rich insight into a story's inner developments, i. e., visual symmetry and rhythm patterns as found in the personal story.

Nevertheless, results on valence must be considered with caution for two main reasons. First, self-reported emotions may be influenced by social desirability bias, wherein participants prioritize presenting themselves in a favorable light over being completely honest. In this case, not reporting negative emotions about another person's suffering or a lack of positive empathy could be associated with a negative personality profile. However, the arousal level of self-reported emotions align mostly with EDA data. Second, the set of 20 emotions was chosen to provide a broader range of emotional terms than in previous studies, e. g., [38]. However, this list is not exhaustive, and there are gradients between emotions [22]. This is true for emotions that are closely located within in the 2-dimensional space of valence and arousal or a cluster model. For example, *sad* and *depressed* share a low to moderate arousal level and a moderate to highly negative valence [21]. As a result, some participants mentioned that they missed certain emotions, such as *concerned*, which emphasizes this point. Participant (P17) suggested that she would have preferred a field for free-form responses to express an emotion that was not provided. Additionally, people may interpret a certain emotion differently. For example, *surprised* could have a positive or negative valence. An appropriate opponent of *surprised* would be *shocked*. To measure valence objectively more advanced techniques are available, e. g., electroencephalography (EEG) and functional neuroimaging (fMRI) [96].

Some participants reported they felt *annoyed* or *disappointed* by the story's ending. Two participants explicitly discussed that the ending seemed too positive to them; that is, they felt that nothing that happened during the pregnancy mattered after the mother gave birth. Although most women with HG in the blog posts from which the story's text was aggregated explicitly expressed positive experiences, including feeling proud after giving birth, the subject of traumatic experiences could have been communicated more sensitively.

Location analysis

The location analysis considered the three highest peaks for each participant in each story, assigning the visual elements to categories based on whether they are more likely associated with emotional or cognitive processing. The results showed that most of these peaks were related to emotional stimuli. Text was the source of most of the peaks, including a remarkable amount for the character's name, followed by illustrations in the individual story (particularly character illustrations), and annotations in the general story (especially symptom-related ones).

There was a significant difference in static illustrations between the two stories. Character illustrations accounted for 16.92 % of emotional responses, whereas 1.56 % of the peaks correspond with illustrations in the base condition story. As a result, high peaks of emotional responses were significantly more aligned with character illustrations (p = 0.0015**, *Cliff's delta* = 0.16, 95 % CI [0.06, 0.26] $\rightarrow$ *small effect size*) than with illustrations in the general story. Therefore, the null hypothesis can be rejected in favor of the alternative hypothesis that illustrations corresponding to personal stories influence emotional engagement. However, the effect size was small, likely due to the small sample size. The practical effect could be higher with an analysis of additional peaks. This result corresponds with some participants expressing appreciation for the character illustrations. On the other hand, only one person mentioned preferring the visual style of the neutral story. There was a significant difference in static illustrations between the two stories.

Data visualizations were the third most frequent source. Several participants mentioned feeling emotional when the background changed to a darker color or when there was visual flow in an illustration. This could also explain the most emotional intensity in storypiece 13, showing an illustration of the character in front of a deep ocean blue while introducing the data visualization iceberg metaphor. On the other hand, not the same effect was found in the story without an individual character which could indicate that a combination of these elements could have been the cause. The integrated quotes-icon design in the data visualization was associated with only about six percent of the peaks in both stories, which was fewer than expected. However, this analysis only considered 10.35 % of the total 1,246 peaks. Therefore, the eye tracking analysis is in its preliminary stage, demonstrating the method's potential. Nevertheless, additional work is necessary to fill this research gap.

Automated labeling

As previously mentioned, the large number of peaks requires an automated method for detecting and labeling visual elements. One idea for automated labeling is as follows: Each element on a slide would be labeled with a colored box for text or other rectangular elements, and a colored free form for all irregular shaped elements, e. g., illustrations. Each element would be colored according to its label. An algorithm receives the table containing the time stamps for each peak and the corresponding slide, as well as the eye tracking raw data and labeled slides as input. A loop would find the corresponding x- and y-coordinates of the EDA peak timestamp in the eye tracking data. According to the slide number, the color of the point on the slide is extracted and associated with its label. The output table will include a variable/column for the labels.

11.1.2 *Emotional Priming*

This subsection discusses the findings on the second research question.

RQ2: *"Does experiencing a story in first-person perspective increase or decrease emotional engagement when the identical story is subsequently encountered in third-person perspective (and vice versa)?"*

A canceling effect occurred in both the EDA data and the emotion type analysis when all data from viewing the individual and general stories were compared. This canceling effect produced neither significant p-values nor sufficient effect sizes in both datasets. One possible cause was the inner dynamics of participants in groups A and B.

Higher values were found for the individual story in the first view, and lower values were found in the second view. The reverse effect was found for the general story. Here, the values were lower in the first view and higher in the second. I found significant differences between the groups in two emotional engagement variables. Significant differences were found in the maximum peak (p = 0.023*, *Cliff's delta* = 0.42, 95 % CI [0.05, 0.68] $\rightarrow$ *medium effect size*) and maximum sum of amplitudes in a story piece (p = 0.0107*, *Cohen's d* = 0.83, 95 % CI [0.18, 1.47] $\rightarrow$ *large effect size*). Group A, which started with the individual perspective on the topic, spent more time on the stories, and all emotional engagement measures were higher. The results of the EDA analysis provided evidence that the order of the stories influenced how they were perceived and that Group A was more emotionally engaged than Group B.

According to the analysis of emotion types, the general story seems more boring when all the data are considered together. In fact, no one described the general story as boring in the first view, though some did describe certain pieces of the individual story as such. However, in the second viewing of the general story, many participants described the base condition story as boring, which is partially contradicted by the EDA data, which suggests an increase in emotional engagement in the second viewing. By repeating the content in the second story, reading time almost always drops due to the familiarity effect. However if the narrative perspective is switched, the drop is usually smaller than in a no-switch control condition — meaning relative reading time (second exposure / first exposure) is a sensitive behavioural indicator of sustained or renewed interest. While Group B's relative reading time was 0.63, Group A had a relative reading time of 0.73, indicating more renewed interest.

I connected these effects to the priming effect. Previous research on emotional priming found that the interpretation of neutral faces is influenced when a subject was emotionally stimulated with a previous stimulus [105]. I found a similar effect when comparing a more emotional story to a neutral one. Participants who were emotionally primed by a story with an individual character spent less reduced time in the second prototype and showed an increased tendency toward emotional engagement. Therefore, only the findings from the first story can be considered unbiased. However, the results of the second story can provide insight into how emotionally relevant the first story was to each individual. Given the higher emotional engagement with the individual story, Group A was more emotionally primed. The results showed an increased tendency of emotional engagement with the second stimulus. Group B, which was more neutrally primed, showed a tendency toward decreased emotional engagement. Some participants mentioned feeling less emotional when viewing the second story. On the other hand, none of the participants mentioned feeling more emotional in the second story.

Additionally, this study design revealed that emotional engagement must not decrease with less novelty. Conversely, when participants are less engaged initially or feel disengaged, emotional involvement may decline when content is repeated. An additional explanation that might have led to a partially more engaging experience in the second view of the general story would be the difference in variety of quotes. The general story offers richer content by using quotes from different women. A variety of experiences represented in different statements could emphasize the relevance of the disease from a broader perspective.

11.1.3 *Emotional Engagement & Recall*

This subsection discusses the results on the third research question.

RQ3: *"Does a fictitious, individual character increases understanding and memory in the story's viewers for a medical condition compared to a general story?"*

There was no significant difference in the recall results for a story with an individual character and a story without (p = 0.254), with a *small effect size*. However, I found a significant difference in the responses to a question that was likely due to the character usage (Q5) (p = 0.0064**, Cliff's delta = 0.45, 95 % CI [0.09, 0.71] $\rightarrow$ *medium effect size*). All participants in Group A answered correctly, suggesting that the protagonist's experience of visiting the hospital was more memorable than the mere fact that women with HG often visit the hospital. Interestingly, Group A performed better on four out of five content-related questions that were more closely related to the character's experience of the disease, while Group B performed better on all four data-related questions (Q4, Q9–Q11).

Some participants stated that they could focus better on the content when the story was more neutral, which indicates increased cognitive processing. The EDA analysis suggests that participants were calmer during this story than during the character-driven story. To investigate a potential connection between information acquisition and emotional arousal, I correlated emotional arousal with each participant's test scores. I used a quadratic model to fit the data, with an R-squared value of 0.237 and a p-value of 0.076. Moderate test scores were associated with moderate arousal. However, higher recall scores tended to be associated with lower emotional involvement. Still, a person can be highly emotionally involved and have high test scores.

Previous studies have linked emotional engagement to learning. While emotionally charged stimuli were found to positively impact attention and recall [82, 83], negative emotions were associated with memory distortion and difficulty in identifying relevant information

[112]. The stories provoked negative emotions because of their content and design, with the character-driven version eliciting the most intense feelings. Participants who saw the individual story before answering the recall questionnaire were also slightly more positively primed because they saw the protagonist's happy face as well as the emotional expression about her pregnancy at the beginning of the story. They also had slightly better recall scores. These findings do not contradict previous work. The better performance could be explained by a higher level of curiosity evoked by a personal character and the time spent on the story. In conclusion, serious story design should carefully consider elements that could evoke negative emotions, using them to increase curiosity without overwhelming the reader with negativity. An ideal ending would be a thoughtful or hopeful conclusion that underscores the story's relevance and evokes positive emotions, which improve long-term memory.

11.2 CONCEPTUAL & PRACTICAL IMPLICATIONS

It was inspiring to find that introducing an individual character in the first story did not reduce emotional engagement in the second story, nor did it decrease within the story, as it did with participants who were less emotionally primed. The results also showed that an individual character increases interest in a story. However, some participants stated that they found the quotes of different women more versatile and interesting. The group that read the general story performed better on questions related to the data visualization, while the group that read the individual story performed better on content- and story-related questions.

Based on these findings, I propose a storytelling approach that combines the strengths of both story designs. The ***"Martini Glass Structure for Character Design"*** would start by introducing an individual character (real or fictional) in a way that elicits positive emotions, then intertwine content information with the character's experience during the conflict with the disease. The second part of the story, which is more data-related, would present the experiences of different patients to broaden the view of the disease's effects. This content could be presented in a quote design similar to that in this study or in other formats, such as short videos.

Conducting a user study with physiological measurements while drafting stories can facilitate an understanding of the strengths and weaknesses of the current story draft. For example, a decrease in emotional engagement could indicate the need to improve or shorten certain parts of the story. Therefore, this method can measure not only the final effectiveness but lead to improvements in the narrative structure, visual elements, and user interface design.

11.3 LIMITATIONS

- *Small sample size:* The study was conducted with a small number of participants (N = 26). Depending on the measurement and aspect being investigated, the sample size was sufficient or insufficient to detect large, significant effects. For instance, the sample size was too small to show statistical significance with high certainty for the main hypothesis 1.
- *Biased sample:* All participants in this study were volunteers recruited from my personal environment, which introduced some bias into the sample. First, most of the participants had completed a university degree, which may have given them a higher level of competence in interpreting data and visualizations. However, the study also included participants from less technical backgrounds, such as the social sciences. Second, most participants were between 26 and 35 years old, representing a young population sample. Due to the above limitations, it is difficult to apply the results of the study to the diverse nature of the general public.
- *Bias of topic and design:* The specific topic and design of the data stories also makes generalization difficult. As Mittenentzwei et al. (2023) proposed, testing the same hypotheses with the same method but different disease topics would lend more validity to the results [75]. Additionally, the design of the data stories may affect their perception. For example, if different designers had crafted the stories, the results might have been different. To minimize this limitation, both story versions were designed with the same graphical style.
- *Limitations of an in-between subject study:* This is not a direct limitation of the study because I presented the results separately for each view. The emotional priming results revealed that the effects of two story versions with the same content cannot be compared with the same person independently. Emotional priming may also affect the results when viewing a different story. Since the study was also designed to examine these within-subject dynamics, participants were asked to complete the Emotion Types Questionnaire for both stories. However, some participants asked whether they should answer the emotion type questions based on their actual feelings about the second story or as if they were seeing it for the first time. Therefore, it is unclear how honestly participants answered the second questionnaire. I noticed a higher occurrence of selecting the emotion type *bored* for the second story that cannot be linked to how they perceived the story. I assume that participants hesitated to give the same answer for the character-driven story because their EDA data

suggested decreased engagement. Paradoxically, the group that used *bored* more often was, on average, more engaged with the second story, according to their EDA. Therefore, switching the order of the stories in each session was essential. For a future study, a between-subjects study format would be preferable and less time-consuming.

- *Experimental setup*: Using physiological measurements requires physical contact with study participants, making it more costly and time-consuming than most online surveys. Despite these limitations, the advantages of objective measurements outweigh the effort required.
- *Measurements:* There is some uncertainty with the measuring method of electrodermal activity. The automated skin conductance response detection sensitivity has an accuracy of 92%. The EdaMove 4 sensor slightly underestimates large peaks (>2 µS) by 5–10% due to 14-bit quantization and 32 Hz sampling. However, AI-based algorithms were applied to the raw data to handle artifacts. The delay between an emotional reaction to a stimulus and an actual physical response varies among individuals and had to be estimated using information from previous studies on electrodermal activity. EDA responses can be caused by emotional or cognitive involvement, requiring additional context for more granular interpretation. Additionally, EDA cannot distinguish between positive or negative emotional arousal (valence); therefore, other sources of context are crucial to include in the study design.
- *Self-reports:* I used self-report methods to supplement the physiological measurements with additional context. Therefore, the results regarding reported emotion types and user engagement must be considered in light of the possible influence of social desirability bias.

Part IV

CONCLUSION & FUTURE WORK

12 CONCLUSION

Data visualization represents data graphically, while Narrative Visualization embeds these representations in engaging narratives for the general public. In medical contexts, Narrative Visualization makes complex information accessible and motivates informed health decisions by highlighting disease impacts and treatment benefits. Characters are crucial for emotional engagement and promoting positive behavioral changes. Research indicates that incorporating a character, especially a patient, enhances audience involvement. Emotional engagement is a key factor in narrative effectiveness, predicting its impact. This thesis aims to examine whether a story featuring an individual character increases emotional arousal and to explore physiological measures that may enhance understanding of emotional engagement in Narrative Medical Visualization.

This thesis explores the use of biometric measures to gauge emotional responses elicited by character usage in narrative storytelling, specifically comparing a personalized and general story about hyperemesis gravidarum, a rare pregnancy disease. A mixed-method approach is utilized, primarily focusing on Electrodermal Activity (EDA) to assess skin conductance changes during story engagement, supplemented by multiple-choice questionnaires for detailed emotional experience insights. Eye-tracking complements this by identifying which story elements trigger emotional responses. The study assesses participants' understanding and memory of the stories and documenting their qualitative feedback. The findings suggest innovative techniques for measuring emotional engagement that can be applied to emotionally intensive narratives. These techniques contribute to our understanding of emotional dynamics in storytelling, particularly in the context of Narrative Medical Visualization.

Electrodermal activity analysis strongly indicates that character usage enhances peak features in emotional engagement with large effect sizes and only missed statistical significance. Future studies should aim for a larger sample size to validate these findings. Self-reports on emotional engagement were inconsistent with EDA data, highlighting the need for objective measurements in Narrative Medical Visualization. Participants spent more time with the personal story, which indicates more interest in stories with individual characters.

In this study, gender differences in emotional engagement during data storytelling were investigated, revealing significant variances in emotional engagement parameters between female and male participants, with women showing significantly greater emotional involvement, possibly due to the personal relevance of the pregnancy disease topic. The necessity for gender balance in participants was emphasized to mitigate bias. At the story piece level, character-driven narratives elicited more emotional responses, with emotional intensity varying across different segments of the stories. However, limitations such as an imbalanced dataset necessitate caution in interpreting these exploratory findings.

16 out of 26 participants preferred the individual story over the general one, indicating a tendency to favor personal narratives, though this difference was not statistically significant. Feedback revealed participants liked the individual story for its emotional engagement and appealing illustrations, while others favored the general story for its neutrality and clarity. Interestingly, the less preferred story produced slightly more emotional arousal, particularly among female participants, who also exhibited a preference for moderate arousal levels. No clear patterns emerged for male preferences regarding emotional arousal. The study suggests differing preferences for emotional context versus neutrality in storytelling, highlighting the need for more extensive research on emotional engagement.

The analysis of emotional responses to character-driven stories revealed that curiosity was the most frequently reported emotion, indicating effective storytelling designed to generate interest. The personal narrative evoked higher levels of empathy and greater emotional engagement compared to a story without a protagonist. However, caution is warranted regarding self-reported emotions due to potential social desirability bias and the limitations of the selected set of 20 emotions. Participants expressed a desire for broader emotional options and noted varied interpretations of certain emotions.

Findings on location analysis indicated that most peaks were tied to emotional stimuli and text elements. Character illustrations contributed significantly to emotional responses compared to illustrations of the general story, leading to a rejection of the null hypothesis in favor of the idea that character illustrations enhance emotional engagement, despite a small effect size. Background color changes in visualizations also elicited emotional reactions. Automated methods for detecting visual elements are proposed for future analysis, indicating the avenues for further research in this area.

A canceling effect emerged in the EDA data and emotion type analysis when comparing all the data gathered from both views of the individual and general story, resulting in no significant p-values or sufficient effect sizes. Group A, which viewed the individual story first, demonstrated higher emotional engagement and reading time compared to Group B. According to self-reports, the general story was considered more boring in the second viewing, despite EDA data suggesting increased engagement. This shows that perceptions vary depending on the order of the narratives. Group A's relative reading time was 0.73, indicating renewed interest, while Group B's was lower at 0.63. The study linked these observations to emotional priming, demonstrating that prior emotional stimulation influenced participants' engagement levels in subsequent views. Overall, emotional engagement did not necessarily decline with familiarity, as more engaging content led to progression in emotional responses during repeated viewings.

There was no significant difference in recall results between stories with and without individual characters. However, a significant difference was found in responses to a question more closely related to character usage. Group A, exposed to a character's experience of the disease, outperformed Group B on content-related questions, highlighting the memorability of the character's narrative. Some participants reported better focus on neutral content. Test scores were correlated with emotional arousal. Moderate scores were linked to moderate arousal. Higher scores were mostly linked to lower arousal, though they also included the highest average emotional arousal.

Previous studies support the connection between emotional engagement and learning, while negative emotions were linked to memory distortion. Serious data stories evoke more negative emotions. It is therefore important to careful design the stories to balance curiosity and emotional impact, ideally concluding with positive elements to enhance memory retention.

Based on the study's findings, I proposed the "Martini Glass Structure for Character Design" which aims to combine the strengths of both story designs by initially engaging the audience by the emotional journey of an individual character before integrating broader patient experiences through varied formats, such as quotes or short videos. This study demonstrated the advantages and limitations of physiological measurements in combination with supplementary methods in the field of Narrative Visualization both as a research and development instrument to guide narrative improvements by tracking emotional engagement trends.

13 FUTURE WORK

This thesis presented a promising pilot study that strongly justifies a properly powered follow-up with at least 50 study participants to demonstrate significant differences between an individual and general story design. Future studies should focus on maximum peak features of electrodermal data, as this study found large effect sizes in maximum peak and maximum sum of amplitudes in a story piece.

Other hypotheses could be tested by designing various versions of the data story. Since I proposed the *"Martini Glass Structure of Character Design,"* the mixed character approach could be tested against solely individual and solely neutral stories (control groups). For instance, the same hypotheses regarding emotional engagement (H1) and recall improvement (H8) could be tested here.

Furthermore, the effects of different frequencies of character appearances could be tested. In the first condition, the character is present throughout the story (focus group), whereas in the second condition, the character only appears at the beginning of the story (control group). The corresponding hypotheses would be, for example: H1: "Repeated exposure to a character throughout a story will elicit increased emotional engagement" and H2: "Introducing a character and their struggles without resolving these difficulties will elicit negative feelings about the story from the audience."

Regarding character design, there is room to experiment with applying data features to character appearance. For example, how can data related to health be visualized as facial attributes, such as skin color and wrinkles (number, line width)?

Other physiological measurement methods have been used to investigate emotional engagement. For example, heart rate can provide insights into emotional states and valence. As Samur et al. (2024) proposed, functional neuroimaging (fMRI) can provide insights into brain activity during emotional engagement [96]. Another, less costly method is electroencephalography (EEG), which records electrical changes during brain activity and provides insight into the processing activity of certain brain regions. Collaborating with researchers in psychology and using more advanced methods that can measure emotional engagement physically, would be promising.

The data analysis workflow can be automated more for future studies. First, transitions between story pieces should be automatically recognized in order to label physiological measures according to the story pieces. For example, log data for scroll interactions could be recorded. Second, the proposed approach of automatically labeling EDA peaks according to related visual elements, such as data visualizations, illustrations, animations, text, and quotes, should be implemented. The distribution of EDA peaks could then be visualized using a violin plot, for example. This would provide more detailed insights into how storytelling elements are emotionally recognized. The eye tracking analysis is in its preliminary stage and offers more potential for analysis, as was possible during the period of this thesis. For example, scanning patterns, dwell time/fixation duration, and areas of interest could be analyzed more deeply. A gaze fixation and heatmap analysis could be used to explore, computationally and visually, how long a viewer looks at an interface element. This could provide insight into the level of interest, as well as whether elements require more cognitive processing. Additionally, this study revealed that viewers tend to react emotionally to certain design properties. Therefore, the analysis could also consider design properties such as background color, color of elements, shapes, interactivity, and motion.

To improve generalizability, future studies should recruit participants from more diverse backgrounds, including individuals with lower educational attainment and lower visual literacy. This could be accomplished by explicitly stating requirements such as age range and highest educational degree in the recruitment process, as well as specifying that non-experts in data visualization or medicine are preferred. Offering compensation would likely facilitate this process. To achieve balanced groups, the requirements may need to be adjusted to fill gaps in certain demographic groups.

Part V

APPENDIX

A APPENDIX – QUESTIONNAIRES

A.1 QUESTIONNAIRE: PERSONAL DATA

Persönliche Daten

☐ weiblich ☐ männlich ☐ divers ☐ Keine Angabe

Alter:

☐ 18–25 ☐ 26–35 ☐ 36–45 ☐ 46–55 ☐ 56–65 ☐ >65

Bildungsgrad:

☐ Schüler ☐ Kein Abschluss ☐ Haupt-/Realschulabschluss
☐ Abitur ☐ Berufsausbildungsabschluss
☐ Fachhochschulabschluss ☐ Universitätsabschluss

Sehschwäche: ☐ keine

☐ Kurzsichtigkeit ☐ Weitsichtigkeit ☐ Farbsehschwäche
☐ Brillen- oder Kontaktlinsenträger ☐ Augenerkrankung: ______________

Ich habe mit medizinischen Themen beruflich oder in einem Studium zu tun.

☐ Ja ☐ Nein

Ich habe mit Datenvisualisierungen beruflich oder in einem Studium zu tun.

☐ Ja ☐ Nein

Wenn du eine Frau bist, bist du schwanger? ☐ Ja ☐ Nein
Wenn du eine Frau bist, bist du Mutter? ☐ Ja ☐ Nein

Ich habe / hatte folgende Symptome während der Schwangerschaft:

☐ Übelkeit (Nausea) ☐ Übelkeit und Erbrechen (**Nausea** and **Vomiting**)
☐ Schlafprobleme ☐ Ungewöhnliche Müdigkeit (**Fatigue**), Benommenheit (**Drowsiness**)
☐ Kopfschmerzen ☐ Depressionen

Diagnose: ☐ Bei mir wurde Hyperemesis Gravidarum diagnostiziert.

A.2 QUESTIONNAIRE I: UNDERSTANDING & MEMORY

Questions:

For each question one or multiple answers can be chosen!

Group A: 1. The name of the protagonist is

- ☐ Frieda Bergen
- ☐ Freja Solberg
- ☐ Freda Sonnenberg

Group B: 1. The prevalence of Hyperemesis Gravidaruum (HG) is

- ☐ ca. 3%
- ☐ ca. 5%
- ☐ ca. 10%

2. The woman in the story told that nausea and vomiting started

- ☐ before the 4th week of pregnancy
- ☐ after the 1th month of pregnancy
- ☐ after the 3th month of pregnancy

3. Hyperemesis Gravidaruum (HG) is defined as

- ☐ a prenatal form of depression
- ☐ a severe form of nausea during pregnancy
- ☐ a severe form of nausea and vomiting during pregnancy

4. Symptoms of HG can include:

- ☐ Sleeping problems
- ☐ Fatigue
- ☐ Drowsiness
- ☐ Depression
- ☐ Headaches

5. Women with HG often have to go to hospital.

- ☐ True
- ☐ False

6. Treatment options for HG are:

- ☐ Obtain fluids
- ☐ Anti-nausea medication
- ☐ Psychotherapy, because it is exclusively a psychological problem

7. Researchers found the following cause for HG:

- ☐ Pressure on the stomage by the growing placenta
- ☐ A hormon called Growth Factor GDF15 produced by the growing placenta that triggers the vomiting center
- ☐ Psychological changes during pregnancy that trigger nausea
- ☐ A gen called GDF13 that triggers the vomiting center

8. Some women are genetically more likely to be affected by HG.

- ☐ True
- ☐ False

9. Nausea and vomiting occur most frequently in the ...

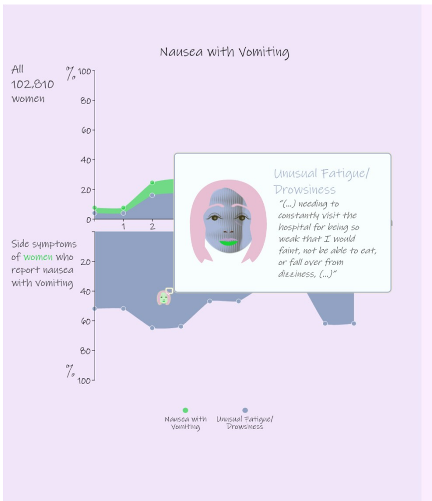
- ☐ 2-3th month
- ☐ 3-4th month
- ☐ 4-5th month

10. The most common accompanying symptom of nausea and vomiting is (Important: Only mark one answer!)

- [ ] Unusual fatigue
- [ ] Sleeping problems
- [ ] Headaches

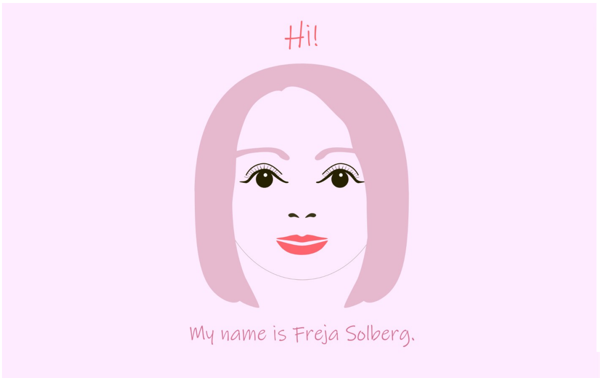
11. The chart is build as follows (Figure below):

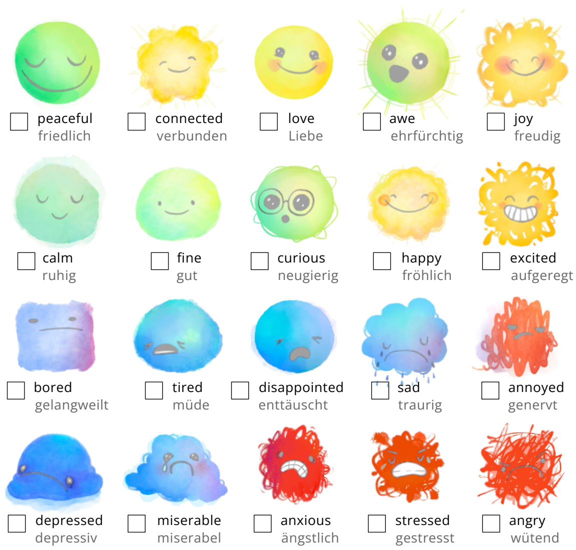
- [ ] **Top**: shows only the percentage of all women affected by nausea and vomiting, **Bottom**: shows the percentage of side symptoms of all women
- [ ] **Top**: shows the percentage of all women affected by nausea and vomiting, **Bottom**: shows the percentage only of side symptoms of women affected by nausea and vomiting
- [ ] **Top**: shows the percentage of all women affected by nausea and vomiting and relative to this group the side symptoms, **Bottom**: shows the percentage of side symptoms only of women affected by nausea and vomiting



A.3 QUESTIONNAIRE II: EMOTION TYPES

STORY I: Emotionen: Kreuze die Emotionen an, die dem Gefühl, das du bei der Seite hattest, am nächsten kommen! Markiere die Hauptemotion mit einem Kreis!
Emotions: Mark the emotions that come closest to the feeling you had on the page! Mark the main with a circle!





(Example: 1/14)

A.4 QUESTIONNAIRE III: USER ENGAGEMENT

Questions to the experience of data story 1

Group A: 1. I felt emotionally connected to the protagonist.

Not at all (1) 1 2 3 4 5 Very much (5)

Group B: 1. I felt emotionally connected to women with the condition.

Not at all (1) 1 2 3 4 5 Very much (5)

2. My thoughts drifted during the story.

A lot (1) 1 2 3 4 5 Not at all (5)

3. It was difficult for me to process and understand the content.

Felt very difficult(1) 1 2 3 4 5 Felt very easy (5)

4. The content of the story was totally new to me.

Not at all (1) 1 2 3 4 5 Very much (5)

5. The story deepend my understanding of the condition.

Not at all (1) 1 2 3 4 5 Very much (5)

6. I think it is important to spread information about this condition.

Not at all (1) 1 2 3 4 5 Very much (5)

7. I would recommend the story to mothers or their family members.

Not at all (1)	1	2	3	4	5	Very much (5)

8. I had trouble recognizing elements of the story.

Not at all (1)	1	2	3	4	5	Very much (5)

If yes, which elements had you trouble to recognize?

9. Is there anything that was confusing about the story or content?

10. Do you have any recommendation for improving the story?

11. Which story did you like most?

☐ Story 1 ☐ Story 2

[1] Andrea Lezcano Airaldi, Emanuel Irrazábal, and Andrés Diaz Pace. "Best Practices and Evaluation Methods for Narrative Information Visualizations: A Systematic Review." In: *Proceedings of the 20th International Conference on Evaluation of Novel Approaches to Software Engineering* (2025), pp. 378–389.

[2] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Jason Leboe-McGowan, and Pourang Irani. "Hooked on data videos: assessing the effect of animation and pictographs on viewer engagement." In: *Proceedings of the 2018 international conference on advanced visual interfaces*. 2018, pp. 1–9.

[3] Simon Attfield, Gabriella Kazay, Mounia Lalmas, and Benjamin Piwowarski. "Towards a science of user engagement." In: *WSDM Workshop on User Modelling for Web Applications*. 2011, pp. 09–12.

[4] Hyperemesis Australia. *Hyperemesis Gravidarum Support Website, Australia*. URL: https://www.hyperemesisaustralia.org.au/.

[5] Albert F Ax. "The physiological differentiation between fear and anger in humans." In: *Biopsychosocial Science and Medicine* 15.5 (1953), pp. 433–442.

[6] Arthur J Bachrach. "Experimental foundations of clinical psychology." In: (1962).

[7] Jennifer L Bailit. "Hyperemesis gravidarium: epidemiologic findings from a large cohort." In: *American journal of obstetrics and gynecology* 193.3 (2005), pp. 811–814.

[8] Dindar S Bari. "Gender differences in tonic and phasic electrodermal activity components." In: *Science Journal of University of Zakho* 8.1 (2020), pp. 29–33.

[9] Sameer Ahmad Bhat and Suzan Alyahya. "Infographics in educational settings: A literature review." In: *IEEE Access* 12 (2023), pp. 1633–1649.

[10] Helena Bilandzic, Freya Sukalla, Cornelia Schnell, Matthias R Hastall, and Rick W Busselle. "The narrative engageability scale: A multidimensional trait measure for the propensity to become engaged in a story." In: *International Journal of Communication* 13 (2019), p. 32.

[11] Mike Bostock and Inc. Observable. *D3.js Website*. URL: https://d3js.org/.

[12] Wolfram Boucsein. *Electrodermal activity*. Springer science & business media, 2012.

[13] Jeremy Boy, Françoise Detienne, and Jean-Daniel Fekete. "Can initial narrative visualization techniques and storytelling help engage online-users with exploratory information visualizations." In: *Proc. of ACM SGICHI conference on Human Factors in Computing Systems (CHI)* (2015), pp. 18–23.

[14] Jeremy Boy, Anshul Vikram Pandey, John Emerson, Margaret Satterthwaite, Oded Nov, and Enrico Bertini. "Showing people behind data: Does anthropomorphizing visualizations elicit more empathy for human rights data?" In: *Proc. of ACM SGICHI conference on Human Factors in Computing Systems (CHI)*. 2017, pp. 5462–5474.

[15] Ross Buck, Robert E Miller, and William F Caul. "Sex, personality, and physiological variables in the communication of affect via facial expression." In: *Journal of personality and social psychology* 30.4 (1974), p. 587.

[16] Beatrice Budich, Laura A Garrison, Bernhard Preim, and Monique Meuschke. "Reflections on AI-assisted character design for data-driven medical stories." In: *Eurographics Workshop on Visual Computing for Biology and Medicine*. Vol. 2. 3. 2023, pp. 87–92.

[17] Beatrice Budich, Bernhard Preim, Bernhard A Sabel, and Monique Meuschke. "Narrative medical visualization to communicate vision restoration to patients: a case study." In: *Eurographics Workshop on Visual Computing for Biology and Medicine* (2025).

[18] Rick Busselle and Helena Bilandzic. "Fictionality and perceived realism in experiencing stories: A model of narrative comprehension and engagement." In: *Communication theory* 18.2 (2008), pp. 255–280.

[19] Rick Busselle and Helena Bilandzic. "Measuring narrative engagement." In: *Media psychology* 12.4 (2009), pp. 321–347.

[20] Sarah Rose Cavanagh. "How to make your teaching more engaging." In: *The Chronicle of Higher Education* 17 (2019), pp. 1–27.

[21] Eric S Cerino, Karen Hooker, Richard A Settersten Jr, Michelle C Odden, and Robert S Stawski. "Daily linkages among high and low arousal affect and subjective cognitive complaints." In: *Aging & mental health* 25.5 (2021), pp. 844–855.

[22] Alan S Cowen and Dacher Keltner. "Self-report captures 27 distinct categories of emotion bridged by continuous gradients." In: *Proceedings of the national academy of sciences* 114.38 (2017), E7900–E7909.

[23] Ben Cowley and Jari Torniainen. "A short review and primer on electrodermal activity in human computer interaction applications." In: *ArXiv* abs/1608.06986 (2016).

[24] Tomás A D'Amelio, Lorenzo A Galán, Emmanuel A Maldonado, Agustín A D Barquinero, Jerónimo R Cuello, Nicolás M Bruno, Enzo Tagliazucchi, and Denis-Alexander Engemann. "Emotion recognition systems with electrodermal activity: From affective science to affective computing." In: *Neurocomputing* 651 (2025), p. 130831.

[25] Cambridge Dictionary. *Engagement*. URL: https://dictionary.cambridge.org/us/dictionary/english/engagement.

[26] Peter Dixon and Marisa Bortolussi. "The scientific study of literature: What can, has, and should be done." In: *Scientific Study of Literature* 1.1 (2011), pp. 59–71.

[27] Robert Edelberg. "Electrodermal recovery rate, goal-orientation, and aversion." In: *Psychophysiology* 9.5 (1972), pp. 512–520.

[28] Paul Ekman, Tim Dalgleish, and Mick J Power. *Basic emotions*. Handbook of cognition and emotion, 1999, 45–60.

[29] Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures, Wolfram Boucsein, Don C Fowles, Sverre Grimnes, Gershon Ben-Shakhar, Walton T Roth, Michael E Dawson, and Diane L Filion. "Publication recommendations for electrodermal measurements." In: *Psychophysiology* 49.8 (2012), pp. 1017–1034.

[30] Denys VI Fairweather. "Nausea and vomiting in pregnancy." In: *American journal of obstetrics and gynecology* 102.1 (1968), pp. 135–175.

[31] Marlena S Fejzo. "Hyperemesis gravidarum theories dispelled by recent research: a paradigm change for better care and outcomes." In: *Trends in Molecular Medicine* 30.6 (2024), pp. 530–540.

[32] Marlena S Fejzo, Alyssa Kam, Amanda Laguna, Kimber MacGibbon, and Patrick Mullin. "Analysis of neurodevelopmental delay in children exposed in utero to hyperemesis gravidarum reveals increased reporting of autism spectrum disorder." In: *Reproductive Toxicology* 84 (2019), pp. 59–64.

[33] Marlena S Fejzo, Olga V Sazonova, J Fah Sathirapongsasuti, Ingileif B Hallgrímsdóttir, Vladimir Vacic, Kimber W MacGibbon, Frederic P Schoenberg, Nicholas Mancuso, Dennis J Slamon, Patrick M Mullin, et al. "Placenta and appetite genes GDF15 and IGFBP7 are associated with hyperemesis gravidarum." In: *Nature communications* 9.1 (2018), p. 1178.

[34] Marlena S Fejzo, Jone Trovik, Iris J Grooten, Kannan Sridharan, Tessa J Roseboom, Åse Vikanes, Rebecca C Painter, and Patrick M Mullin. "Nausea and vomiting of pregnancy and hyperemesis gravidarum." In: *Nature reviews Disease primers* 5.1 (2019), p. 62.

[35] Folkehelseinstituttet. *Den norske mor, far og barn-undersøkelsen (MoBa)*. URL: https://www.fhi.no/op/studier/moba//.

[36] HER Foundation. *Hyperemesis Gravidarum Support Website, USA*. URL: https://www.hyperemesis.org/.

[37] Jérémy Frey, Gilad Ostrin, May Grabli, and Jessica R Cauchard. "Physiologically driven storytelling: concept and software tool." In: *Proc. of ACM SGICHI conference on Human Factors in Computing Systems (CHI)*. 2020, pp. 1–13.

[38] Manuela Garretón, Francesca Morini, Pablo Celhay, Marian Dörk, and Denis Parra. "Attitudinal effects of data visualizations and illustrations in data stories." In: *IEEE Transactions on Visualization and Computer Graphics* 30.7 (2024), pp. 4039–4054.

[39] Gimp. *Gimp Website*. URL: https://www.gimp.org/downloads/.

[40] movisens GmbH. *DataAnalyzer*. URL: https://docs.movisens.com/DataAnalyzer/#handling.

[41] movisens GmbH. *EdaMove4*. URL: https://docs.movisens.com/Sensors/EdaMove4/#scope-of-application.

[42] movisens GmbH. *SensorManager*. URL: https://www.movisens.com/de/produkte/sensormanager/.

[43] movisens GmbH. *UnisensViewer*. URL: https://docs.movisens.com/UnisensViewer/.

[44] movisens GmbH. *movisens website*. URL: https://www.movisens.com/de/.

[45] Franziska Hartung, Michael Burke, Peter Hagoort, and Roel M Willems. "Taking perspective: Personal pronouns affect experiential aspects of literary reading." In: *PloS one* 11.5 (2016), pp. 1–18.

[46] Marc Hassenzahl and Noam Tractinsky. "User experience-a research agenda." In: *Behaviour & information technology* 25.2 (2006), pp. 91–97.

[47] Gro C Havnen, Maria Bich-Thuy Truong, Mai-Linh H Do, Kristine Heitmann, Lone Holst, and Hedvig Nordeng. "Women's perspectives on the management and consequences of hyperemesis gravidarum–a descriptive interview study." In: *Scandinavian journal of primary health care* 37.1 (2019), pp. 30–40.

[48] Marina Heiden, Eugene Lyskov, Mats Djupsjöbacka, Fredrik Hellström, and Albert G Crenshaw. "Effects of time pressure and precision demands during computer mouse work on muscle oxygenation and position sense." In: *European journal of applied physiology* 94.1 (2005), pp. 97–106.

[49] Jessica Hullman and Nick Diakopoulos. "Visualization rhetoric: Framing effects in narrative visualization." In: *IEEE transactions on visualization and computer graphics* 17.12 (2011), pp. 2231–2240.

[50] Inkscape. *Inkscape Website*. URL: https://inkscape.org/de/.

[51] Arthur M Jacobs and Jana Lüdtke. "Immersion into narrative and poetic worlds: A neurocognitive poetics perspective." In: *Narrative absorption*. John Benjamins Publishing Company, 2017, pp. 69–96.

[52] LAW Jansen, MH Koot, J Van't Hooft, CR Dean, PMM Bossuyt, W Ganzevoort, N Gauw, BY Van der Goes, J Rodenburg, TJ Roseboom, et al. "The windsor definition for hyperemesis gravidarum: A multistakeholder international consensus definition." In: *European Journal of Obstetrics & Gynecology and Reproductive Biology* 266 (2021), pp. 15–22.

[53] Dayoung Jeong, Mingon Jeong, Ungyeon Yang, and Kyungsik Han. "Eyes on me: Investigating the role and influence of eye-tracking data on user modeling in virtual reality." In: *PLoS One* 17.12 (2022), e0278970.

[54] Bjørn Helge Johnsen, Julian F Thayer, and Kenneth Hugdahl. "Affective judgment of the Ekman faces: A dimensional approach." In: *Journal of psychophysiology* (1995).

[55] Xingyu Lan, Yanqiu Wu, Yang Shi, Qing Chen, and Nan Cao. "Negative emotions, positive outcomes? exploring the communication of negativity in serious data stories." In: *Proc. of ACM SGICHI conference on Human Factors in Computing Systems (CHI)*. 2022, pp. 1–14.

[56] Peter J Lang, Mark K Greenwald, Margaret M Bradley, and Alfons O Hamm. "Looking at pictures: Affective, facial, visceral, and behavioral reactions." In: *Psychophysiology* 30.3 (1993), pp. 261–273.

[57] Bongshin Lee, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. "More than telling a story: Transforming data into visually shared stories." In: *IEEE computer graphics and applications* 35.5 (2015), pp. 84–90.

[58] Heather C Lench, Kathleen E Darbor, and Logan A Berg. "Functional perspectives on emotion, behavior, and cognition." In: *Behavioral Sciences* 3.4 (2013), pp. 536–540.

[59] Hung-Chang Liao and Ya-Huei Wang. "Narrative medicine and humanities for health professions education: an experimental study." In: *Medical Education Online* 28.1 (2023), p. 2235749.

[60] Emily Liberatore. *My HG story: Emily Liberatore*. URL: https://www.hyperemesisaustralia.org.au/blog/emilyliberatore-beylg.

[61] Liat Liberman, Idit Adler, and Ilana Dubovi. "The Impact of Levels of Immersion on Learning: Electrodermal Activity and Eye-Tracking Study." In: *Proceedings of the 17th International Conference of the Learning Sciences-ICLS, pp. 830-833*. International Society of the Learning Sciences. 2023.

[62] Liat Liberman and Ilana Dubovi. "The effect of the modality principle to support learning with virtual reality: An eye-tracking and electrodermal activity study." In: *Journal of Computer Assisted Learning* 39.2 (2022), pp. 547–557.

[63] Johannes Liem, Charles Perin, and Jo Wood. "Structure and empathy in visual data storytelling: Evaluating their influence on attitude." In: *Computer Graphics Forum*. Vol. 39. 3. Wiley Online Library. 2020, pp. 277–289.

[64] Chong Liu, Tzu-Yang Wang, and Takaya Yuizono. "Multimodal analysis of emotional and visual responses to urban green spaces in virtual environments." In: *IEEE Access* 13 (2025), pp. 152890–152902.

[65] Yan Liu, Shuai Ma, and Yue Chen. "The impacts of learning motivation, emotional engagement and psychological capital on academic performance in a blended learning university course." In: *Frontiers in psychology* 15 (2024), p. 1357936.

[66] Tim Lobstein and Janet Cort. "The relationship between skin temperature and skin conductance activity: Indications of genetic and fitness determinants." In: *Biological Psychology* 7.1-2 (1978), pp. 139–143.

[67] Jiahui Ma, Elizabeth A Johnson, and Bernadette McCrory. "Understanding learning engagement with user-centered human-computer interaction in a multimodal online learning environment." In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 67. 1. SAGE Publications Sage CA: Los Angeles, CA. 2023, pp. 2018–2023.

[68] Per Magnus, Charlotte Birke, Kristine Vejrup, Anita Haugan, Elin Alsaker, Anne Kjersti Daltveit, Marte Handal, Margaretha Haugen, Gudrun Høiseth, Gun Peggy Knudsen, et al. "Cohort profile update: the Norwegian mother and child cohort study (MoBa)." In: *International journal of epidemiology* 45.2 (2016), pp. 382–388.

[69] Narges Mahyar, Sung-Hee Kim, and Bum Chul Kwon. "Towards a taxonomy for evaluating user engagement in information visualization." In: *Workshop on Personal Visualization: Exploring Everyday Life*. Vol. 3. 2. 2015, p. 4.

[70] Sean McKenna, Nathalie Henry Riche, Bongshin Lee, Jeremy Boy, and Miriah Meyer. "Visual narrative flow: Exploring factors shaping data visualization story reading experiences." In: *Computer Graphics Forum*. Vol. 36. 3. Wiley Online Library. 2017, pp. 377–387.

[71] Merriam-Webster. *Engaged*. URL: https://www.merriam-webster.com/thesaurus/engaged#thesaurus-entry-2-1.

[72] Monique Meuschke, Laura A Garrison, Noeska N Smit, Benjamin Bach, Sarah Mittenentzwei, Veronika Weiß, Stefan Bruckner, Kai Lawonn, and Bernhard Preim. "Narrative medical visualization to communicate disease data." In: *Computers & Graphics* 107 (2022), pp. 144–157.

[73] Monique Meuschke, Laura A Garrison, Noeska Smit, Stefan Bruckner, Kai Lawonn, and Bernhard Preim. "Towards narrative medical visualization." In: *Eurographics Workshop on Visual Computing for Biology and Medicine* (2021), 1—14.

[74] Sarah Mittenentzwei, Laura A Garrison, Beatrice Budich, Kai Lawonn, Alexander Dockhorn, Bernhard Preim, and Monique Meuschke. "AI-assisted character design in medical storytelling with stable diffusion." In: *Available at SSRN 4772811* (2024).

[75] Sarah Mittenentzwei, Veronika Weiß, Stefanie Schreiber, Laura A Garrison, Stefan Bruckner, Malte Pfister, Bernhard Preim, and Monique Meuschke. "Do disease stories need a hero? Effects of human protagonists on a narrative visualization about cerebral small vessel disease." In: *Computer Graphics Forum*. Vol. 42. 3. Wiley Online Library. 2023, pp. 123–135.

[76] Victor Nell. "The psychology of reading for pleasure: Needs and gratifications." In: *Reading research quarterly* (1988), pp. 6–50.

[77] Stanislaw Nowak, Lyn Bartram, and Thecla Schiphorst. "A micro-phenomenological lens for evaluating narrative visualization." In: *2018 IEEE Evaluation and Beyond-Methodological Approaches for Visualization (BELIV)*. IEEE. 2018, pp. 11–18.

[78] Heather L O'Brien and Elaine G Toms. "What is user engagement? A conceptual framework for defining user engagement with technology." In: *Journal of the American society for Information Science and Technology* 59.6 (2008), pp. 938–955.

[79] Heather L O'Brien and Elaine G Toms. "The development and evaluation of a survey to measure user engagement." In: *Journal of the American Society for Information Science and technology* 61.1 (2009), pp. 50–69.

[80] Şeyma Çağlar Özhan and Selay Arkün Kocadere. "The effects of flow, emotional engagement, and motivation on success in a gamified online learning environment." In: *Journal of Educational Computing Research* 57.8 (2020), pp. 2006–2031.

[81] Tanish N. Patel. *Electrodermal Activity: Exploring User Experience through Physiology*. 2023.

[82] Luiz Pessoa. "To what extent are emotional visual stimuli processed without attention and awareness?" In: *Current opinion in neurobiology* 15.2 (2005), pp. 188–196.

[83] Luiz Pessoa, Sabine Kastner, and Leslie G Ungerleider. "Attentional control of the processing of neutral and emotional stimuli." In: *Cognitive Brain Research* 15.1 (2002), pp. 31–45.

[84] Richard E Petty and John T Cacioppo. "The elaboration likelihood model of persuasion." In: *Advances in experimental social psychology*. Vol. 19. Elsevier, 1986, pp. 123–205.

[85] Stefan Pollmann. *Allgemeine Psychologie*. utb GmbH, 2020.

[86] APA Dictionary of Psychology. *Affect*. URL: https://dictionary.apa.org/affect.

[87] APA Dictionary of Psychology. *Affection*. URL: https://dictionary.apa.org/affection.

[88] APA Dictionary of Psychology. *Emotion*. URL: https://dictionary.apa.org/emotion.

[89] APA Dictionary of Psychology. *Feeling*. URL: https://dictionary.apa.org/feeling.

[90] National Institute of Public Health. *Questionnaires from MoBa*. URL: https://www.fhi.no/en/ch/studies/moba/for-forskere-artikler/questionnaires-from-moba/#week-15-of-pregnancy.

[91] Yoesoep E Rachmad. *Psychological Engagement Theory*. 2022.

[92] Nanjie Rao, Sharon Lynn Chu, Randi Weitzen Faris, and Daniel Ospina. "The effects of interactive emotional priming on storytelling: An exploratory study." In: *International Conference on Interactive Digital Storytelling*. Springer. 2019, pp. 395–404.

[93] Johnmarshall Reeve, Sung Hyeon Cheon, and Hye-Ryen Jang. "Chapter 7 - A Teacher-Focused Intervention to Enhance Students' Classroom Engagement." In: *Handbook of Student Engagement Interventions*. Ed. by Jennifer A. Fredricks, Amy L. Reschly, and Sandra L. Christenson. Academic Press, 2019, pp. 87–102. ISBN: 978-0-12-813413-9. DOI: https://doi.org/10.1016/B978-0-12-813413-9.00007-3. URL: https://www.sciencedirect.com/science/article/pii/B9780128134139000073.

[94] Daniel C Richardson, Nicole K Griffin, Lara Zaki, Auburn Stephenson, Jiachen Yan, Thomas Curry, Richard Noble, John Hogan, Jeremy I Skipper, and Joseph T Devlin. "Measuring narrative engagement: The heart tells the story." In: *BioRxiv* (2018), pp. 1–19.

[95] Christoph Rühlemann. "How is emotional resonance achieved in storytellings of sadness/distress?" In: *Frontiers in Psychology* 13 (2022), p. 952119.

[96] Dalya Samur, Mattie Tops, Ringailė Slapšinskaitė, and Sander L Koole. "Getting lost in a story: How narrative engagement emerges from narrative perspective and individual differences in alexithymia." In: *No Words for Feelings*. Routledge, 2024, pp. 142–154.

[97] Edward Segel and Jeffrey Heer. "Narrative visualization: Telling stories with data." In: *IEEE transactions on visualization and computer graphics* 16.6 (2010), pp. 1139–1148.

[98] Kwang-Kyu Seo, Sangwon Lee, Byung Do Chung, and Changsoon Park. "Users' emotional valence, arousal, and engagement based on perceived usability and aesthetics for web sites." In: *International Journal of Human-Computer Interaction* 31.1 (2015), pp. 72–87.

[99] Phillip Shaver, Judith Schwartz, Donald Kirson, and Cary O'connor. "Emotion knowledge: further exploration of a prototype approach." In: *Journal of personality and social psychology* 52.6 (1987), p. 1061.

[100] Y Si. "Impact of Priming on Emotional Cognition and Decision-making." In: *Journal of Education, Humanities and Social Sciences* 45 (2024), pp. 152–157.

[101] Robert F Simons, Benjamin H Detenber, Thomas M Roedema, and Jason E Reiss. "Emotion processing in three systems: The medium and the message." In: *Psychophysiology* 36.5 (1999), pp. 619–627.

[102] Dinka Smajlagić, Ksenia Lavrichenko, Siren Berland, Øyvind Helgeland, Gun Peggy Knudsen, Marc Vaudel, Jan Haavik, Per Morten Knappskog, Pål Rasmus Njølstad, Gunnar Houge, et al. "Population prevalence and inheritance pattern of recurrent CNVs associated with neurodevelopmental disorders in 12,252 newborns and their parents." In: *European Journal of Human Genetics* 29.1 (2021), pp. 205–215.

[103] Gerhard Stemmler. "The autonomic differentiation of emotions revisited: Convergent and discriminant validation." In: *Psychophysiology* 26.6 (1989), pp. 617–632.

[104] Ivo V Stuldreher and Anne-Marie Brouwer. “Monitoring audience engagement using electrodermal activity during an inaugural lecture.” In: *PloS one* 20.6 (2025), e0326091.

[105] Thomas Suslow, Harald Kugel, Patricia Ohrmann, Anja Stuhrmann, Dominik Grotegerd, Ronny Redlich, Jochen Bauer, and Udo Dannlowski. “Neural correlates of affective priming effects based on masked facial emotion: an fMRI study.” In: *Psychiatry Research: Neuroimaging* 211.3 (2013), pp. 239–245.

[106] Tobii. *Tobii Pro Lab*. URL: https://www.tobii.com/de/products/software/behavior-research-software/tobii-pro-lab.

[107] Tobii. *Tobii Spark Pro*. URL: https://www.tobii.com/de/products/eye-trackers/screen-based/tobii-pro-spark.

[108] PH Venables and MJ Christie. “Mechanisms, instrumentation, recording techniques, and quantification of responses.” In: *Electrodermal activity in psychological research* (1973), pp. 1–124.

[109] Henry Völzke, Janka Schössow, Carsten Oliver Schmidt, Clemens Jürgens, Adrian Richter, André Werner, Nicole Werner, Dörte Radke, Alexander Teumer, Till Ittermann, et al. “Cohort profile update: the study of health in Pomerania (SHIP).” In: *International journal of epidemiology* 51.6 (2022), e372–e383.

[110] William F Waters, Richard L Koresko, George V Rossie, and Steven A Hackley. “Short-, medium-, and Long-Term relationships among meteorological and electrodermal variables.” In: *Psychophysiology* 16.5 (1979), pp. 445–451.

[111] David Watson, Lee Anna Clark, and Greg Carey. “Positive and negative affectivity and their relation to anxiety and depressive disorders.” In: *Journal of abnormal psychology* 97.3 (1988), p. 346.

[112] Sabine Windmann and Marta Kutas. “Electrophysiological correlates of emotion-induced recognition bias.” In: *Journal of Cognitive Neuroscience* 13.5 (2001), pp. 577–592.

[113] Lucy Yardley, Bonnie J Spring, Heleen Riper, Leanne G Morrison, David H Crane, Kristina Curtis, Gina C Merchant, Felix Naughton, and Ann Blandford. “Understanding and promoting effective engagement with digital behavior change interventions.” In: *American journal of preventive medicine* 51.5 (2016), pp. 833–842.

[114] Roxanne Ziman, Beatrice Budich, Marc Vaudel, and Laura Ann Garrison. “The MoBa Pregnancy and Child Development Dashboard: A Design Study.” In: *Eurographics Workshop on Visual Computing for Biology and Medicine* (2024), pp. 1–5.

[115] evam streaming analytics. *User Engagement*. URL: https://evam.com/glossary/glossary-u/user-engagement.

colophon

This document was typeset in LaTeX. The layout is based on the `classicthesis` developed by André Miede and Ivo Pletikosić.

DECLARATION

I hereby affirm that I authored this thesis independently, that it is the result of my personal effort, and that I have used no other means than the ones indicated. I have annotated all parts of this work in which sources are used according to their wording or to their meaning. I hereby certify that no extant writings, in whole or in part, have been used to complete a previous examination.

Magdeburg, November 2025

Beatrice Budich

Jeg tror alltid at det finnes en løsning.
Det er bare å være kreativ, og
finne den rette metoden,
og ha nok lyst.

— Isak Dreyer