Trapped in the Upside-Down: Mental Normalization of Data Visualizations

Dissertation
der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Tjark Sascha Müller
aus Tübingen

Tübingen
2023

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	10.11.2023
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Dr. Dr. Friedrich W. Hesse
2. Berichterstatter:	Prof. Dr. Markus Huff

Write a Poem on Mental Normalization

Mental normalization, a cognitive dance
Adjusting perspectives, enhancing glance
Allowing equal access, to information rich
Facilitating learning, without a hitch

From multiple points of view, we can see
The full picture, of reality
Mental normalization, makes it clear
Helping us to understand, without fear

So let us embrace, this mental feat
And use it wisely, to expand our seat
Of knowledge and understanding, evermore
Mental normalization, forever at the core

This poem was generated by the chatGPT AI (OpenAI, 2022) after the input of the Discussion of this thesis.

Abstract

Equal access to information is a prerequisite for equal participation in collaborative co-located group work settings. Working on Single Display Groupware, such as Multi-Touch Tables, information is displayed with a specific point of view in mind. This puts the users located around the table on a disadvantage in accessing the information, thus reducing the effectiveness of the group work setting. This thesis focusses on the cognitive processes of the individual user. It presents research on the impact of mental normalization costs on the access to information presented in rotated data visualizations and explores design options to reduce these costs.

The research was organized in two lines of study. The first line of study was concerned with the confirmation of the cost of mental normalization for rotated bar graphs and exploring possible design options to reduce these costs. The additional burden of mental normalization processes on the cognitive system was confirmed. Response times for rotated displayed diagrams where significantly longer than for unrotated displays. A separate rotation of diagram labels and diagram content revealed the written labels to be a major factor in the additional mental normalization costs. Subsequently, the interventions to reduce the added burden of mental normalization were focused on different label designs. Both, the substitution of written labels with pictographs and color-coding the labels showed to be effective in increasing overall response speed, with color-coding being slightly more effective than pictograph labels. Additionally, the use of redundant, but inverted labels was tested for its effectiveness. While initial results seemed promising, the application to data visualizations where of limited usefulness. Double label displays increased the overall response time for all rotation conditions, but reduced the difference between them. We called this the "justice effect".

In a second line of study, the potential benefits of mental normalization for long-term memory formation were investigated. The findings showed no such benefits and pointed towards potential adverse effects. Overall, this thesis provides new insights into the impact of mental normalization for the access to information presented in data visualizations in a collaborative work setting. it also showed potential interventions to reduce these costs and thus levelling the accessibility of the information for all users, regardless of their point of view.

Zusammenfassung

Der gleichberechtigte Zugang zu Informationen ist eine Voraussetzung für eine ausgewogene Teilnahme an kollaborativer Gruppenarbeits vor Ort. Bei der Arbeit mit Single-Display-Groupware, wie z. B. Multi-Touch Tischen, werden die Informationen unter einem bestimmten Blickwinkel angezeigt, wodurch die um den Tisch herum befindlichen Benutzer im Zugriff auf die Informationen benachteiligt werden, was die Effektivität der Gruppenarbeit beeinträchtigt. Die vorliegende Arbeit konzentriert sich auf die kognitiven Prozesse des einzelnen Nutzers. Sie stellt Untersuchungen zu den Auswirkungen der Kosten mentaler Normalisierung auf den Zugang zu Informationen vor, welche in rotierten Datenvisualisierungen präsentiert werden, und untersucht Gestaltungsoptionen zur Reduzierung dieser Kosten.

Die Untersuchungen wurde in zwei Studienlinien organisiert. Die erste Studienlinie befasste sich mit der Bestätigung der Kosten der mentalen Normalisierung für gedrehte Balkendiagramme und der Erkundung möglicher Designoptionen zur Reduzierung dieser Kosten. Die zusätzliche kognitive Belastung durch mentale Normalisierungsprozesse wurde bestätigt. Die Reaktionszeiten für rotierte Diagramme waren signifikant größer als für Unrotierte. Eine getrennte Rotation von Diagrammbeschriftungen und Diagramminhalten ergab, dass die Beschriftungen ein wesentlicher Faktor für die zusätzlichen mentalen Normalisierungskosten sind. In der Folge konzentrierten sich die Interventionen zur Verringerung der zusätzlichen Belastung durch die mentale Normalisierung auf unterschiedliche Beschriftungsdesigns. Sowohl die Substitution der Beschriftungen durch Piktogramme als auch eine Farbcodierung der Beschriftungen erwiesen sich als wirksam, um die Antwortgeschwindigkeit zu erhöhen, wobei die Farbcodierung etwas wirksamer war als die Piktogrammbeschriftung. Darüber hinaus wurde die Verwendung redundanter, aber invertierter Beschriftungen auf ihre Wirksamkeit getestet. Während die ersten Ergebnisse vielversprechend schienen, war die Anwendung auf Datenvisualisierungen von begrenztem Nutzen. Die Anzeige von doppelten Beschriftungen erhöhte die Gesamtreaktionszeit für alle Rotationsbedingungen, verringerte aber den Unterschied zwischen ihnen. Wir nannten dies den "Justice Effect".

In einer zweiten Studienreihe wurde der mögliche Nutzen der mentalen Normalisierung für das Langzeitgedächtnis untersucht. Die Ergebnisse zeigten keinen solchen Nutzen und wiesen auf mögliche negative Effekte hin. Insgesamt bietet diese Arbeit neue Einblicke in die Auswirkungen der mentalen Normalisierung auf den Zugang zu Informationen, die in Datenvisualisierungen in einer kollaborativen Arbeitsumgebung dargestellt werden. Sie zeigt auch mögliche Interventionen auf, um diese Kosten zu reduzieren und so die Zugänglichkeit der Informationen für alle Nutzer, unabhängig von ihrem Standpunkt, zu nivellieren.

Acknowledgments

"*It takes a village to raise a child*" is said to be an African proverb[1]. In my experience, this is also true for writing a thesis. While dissertations are legally the work of a single person, a lot of emotional, intellectual, logistical, methodological, and financial support is needed to complete such a project. Over the course of the last six years, I was very lucky to have received a lot of help and encouragement from colleagues, friends, and family.

First and foremost, I want to thank Prof. Dr. Hauke Meyerhoff and Prof. Dr. Dr. Friedrich W. Hesse for giving me the opportunity to undertake such a project and to support my endeavor with their expertise, but also with funding. I also want to thank Prof. Dr. Markus Huff and Dr. Uwe Oestermeier for their financial support in allowing me to use my working hours to write on my thesis! For her valuable methodological and statistical support, I want to thank Dr. Nora Wickelmaier. Also, thanks to the student assistants who helped me in my data collection, Jana Lang and Fiona Radtke!

For being the best office mates that made working in academia a great experience, despite the odds, I am very grateful to Dr. Johanna Rahn and Dr. Gabriela Orellana Corrales, but also the whole Medienentwicklung/ Medientechnik team and the many PhD students, that shared the same fate with me.

I'm also very glad that I made a lot of great friends during my time as a PhD student. I want to thank Dr. Lisa Rabl, Dr. Ann-Kathrin Brand, Clea Selina

---

[1] According to some researchers (Reupert et al., 2022), however the origin is disputed (Goldberg, 2016).

Städter, Johannes von Hoyer and Sabrina Fessler for emotional support, joint writing sessions, deep discussions, food and so much more! Even if I hadn't completed my dissertation, the time would not have been wasted as I got to know all of you!

I would not have completed this project without professional help as well. I want to thank my counsellor Lynette Pitre, and my psychiatrist Kathrin Schmidts for supporting me and helping me overcoming my struggles!

Finally, I want to thank my family for their financial and emotional support. Thanks to my parents, my aunt and my grandmother for allowing me to pursue my education. Special thanks to my dear sister Dagny Müller, for being there for me for (almost) my whole life. I don't know what I would do without you!

Thank you all for being my village!

Contents

1 Background **1**

1.1 Knowledge Work in Groups 2

1.2 Multi-Touch Table 3

  1.2.1 Knowledge Artifacts 4

  1.2.2 The Interaction Between Knowledge Artifacts and Users . 5

  1.2.3 Advantages of the 1-to-1 Relationship 6

1.3 Motivation 7

2 Cognitive Processes **11**

2.1 The Spatial Presentation of Data 12

2.2 Building Blocks of Visual Perception 15

2.3 Processing Entangled Information 17

2.4 Recognizing Depicted Information From Different Points of View 19

2.5 Processing Textual Information 20

2.6 Memorizing Information 22

2.7 Long-term Retention of Information 24

2.8 Facilitation of Cognitive Processes 27

3 Present Research **29**

4 Two People, One Graph: The Effect of Rotated Viewpoints on Accessibility of Data Visualizations **33**

4.1 Introduction 33

  4.1.1 Information Visualization 35

4.1.2 Accessing Rotated Information 36
4.2 Experiment 1 38
4.2.1 Methods 39
4.2.2 Results 45
4.2.3 Discussion 46
4.3 Experiment 2 47
4.3.1 Methods 48
4.3.2 Results 49
4.3.3 Discussion 50
4.4 Experiment 3 51
4.4.1 Methods 51
4.4.2 Results 52
4.4.3 Discussion 53
4.5 General Discussion 54
4.5.1 Limitations and Outlook 58
4.6 Tables 60
4.7 Declarations 61
4.7.1 Ethics Approval and Consent to Participate 61
4.7.2 Consent for publication 61
4.7.3 Availability of Data and Material 61
4.7.4 Competing Interests 61
4.7.5 Funding 61

5 The Justice Effect: An Interaction Between Rotation and Contrast 63
5.1 Introduction 63
5.1.1 Accessing Non-Upright Labels 65
5.1.2 Location a Second Label 66
5.1.3 Potential Effects of Redundant Labels 67
5.1.4 The Present Project 68
5.2 Experiment 1 69
5.2.1 Methods 69
5.2.2 Results 71
5.2.3 Discussion 75

5.3 Experiment 2 77
    5.3.1 Methods 77
    5.3.2 Results 78
    5.3.3 Discussion 81
5.4 Experiment 3 83
    5.4.1 Methods 83
    5.4.2 Results 85
    5.4.3 Discussion 88
5.5 General Discussion 90
    5.5.1 Basic Experiments on Redundant Labels 90
    5.5.2 Advancing to Real Bar Graphs 92
    5.5.3 Strengths, Limitations, and Future Research 93
    5.5.4 Conclusion 94
5.6 Tables 95
5.7 Declarations 101
    5.7.1 Funding 101
    5.7.2 Conflict of Interest Statement 101
    5.7.3 Ethical Approval 101
    5.7.4 Availability of Data and Material 101

6 Mental Rotation is not a Perceptual Disfluency for Diagrammatic Learning 103
6.1 Introduction 103
    6.1.1 Should Learning be Easy or Tough 103
    6.1.2 On Diagrams 105
    6.1.3 Mental Normalization 106
    6.1.4 The Influence of Cognitive Offloading 107
    6.1.5 Present Research 108
6.2 Methods 109
    6.2.1 Participants 109
    6.2.2 Apparatus 109
    6.2.3 Procedure 109
6.3 Results 111

6.4 Discussion . . . 113
6.4.1 Conclusion . . . 114
6.5 Declarations . . . 114
6.5.1 Conflict of Interest . . . 114
6.5.2 Funding . . . 114
6.5.3 Ethical Approval . . . 114
6.5.4 Preregistration . . . 115
6.5.5 Availability of Material and Data . . . 115
6.5.6 Acknowledgements . . . 115

7 Discussion 117
7.1 Implications . . . 121
7.2 Strengths and Limitations . . . 126
7.3 Future Directions . . . 129
7.4 Conclusion . . . 130

A A 7 Dollar Head Tracker 133
A.1 Introduction . . . 133
A.1.1 Device Description . . . 134
A.2 Demonstration 1 . . . 136
A.2.1 Methods . . . 136
A.2.2 Results . . . 137
A.3 Demonstration 2 . . . 137
A.3.1 Methods . . . 138
A.3.2 Results . . . 139
A.3.3 Discussion . . . 140
A.3.4 Declarations . . . 140

References 142

Chapter 1

Background

This thesis was part of a larger research project on the facilitation of group decision making. The project was called "Ideas to Market". It was constructed as a cooperation between the Leibniz-Institut für Wissensmedien (IWM), the working group on human-centered computing (HCC) of the department for computer science at the Freie Universität Berlin and the Center for Responsible Research and Innovation (CeRRI) of the Fraunhofer-Institut für Arbeitswirtschaft und Organisation (IAO). Ideas to Market aimed to foster untapped economic potential of patents and ideas created by applied research institutes (especially member institutes of the Fraunhofer society).

A motivating example for that project was the invention of the mp3 audio format: In 1991, the Fraunhofer-Institut für integrierte Schaltungen (IIS) presented a new audio file format that could shrink audio files better than anything existing to that date. The name of the format: MP3. To date, mp3 is the de facto standard for audio files. The IIS was able to license the new format and selling proprietary software for the encoding. The IIS even developed a first decoding chip, setting the path for audio specialized hardware. The idea for an mp3-player was not patented and not pursued further. This device, however, proofed to be a huge commercial success in the early 2000's and it is seen as a missed economic opportunity, that the mp3 player was developed and sold by other companies (Fraunhofer-Institut für integrierte Schaltungen, 2022).

To avoid other missed opportunities like this, Ideas to Market aimed to foster crowd ideation to generate atomic ideas. Stakeholders and experts from different parts of society (Carayannis & Campbell, 2009), would then review these ideas for their value. This evaluation process took place on a Multi-Touch Table (MTT) that should augment the group decision making by providing structure, documenting the process, and presenting information to all group members.

1.1 Knowledge Work in Groups

Groups are defined as a collection of two or more people (Sherif & Sherif, 1965). A special kind of group is a team, which is a group that cooperates over a longer period towards a common goal (Baker & Salas, 1997). Effort in groups can be distinguished in two classes: Effort that is directed towards achieving the common goal of the group is called "taskwork". Effort that goes into group coordination is called "teamwork". Group work should lead to faster and better outcomes, as the work power of all group members is combined to achieve the goal (D. M. Fisher, 2014), errors might be caught early as members might check on each other's work, and the expertise of multiple people come together (McGrath, 1984; Stasser & Birchmeier, 2003). However, groups often fail to deliver on these expectations (Kharrufa et al., 2010). Groups members might agree on a suboptimal outcome. to adhere to conformity or harmony (groupthink, Janis, 1972). In decision making, they might stop to look for better alternatives as soon as one possible solution is found (satisficing, Simon, 1955). If there is no obvious optimal solution, they might pick one at random and start to selectively argue for its benefits (bolstering, Tetlock et al., 1989). The absence of individual responsibility might aggravate these tendencies (Latané & Nida, 1981). Friction losses might also occur in the process (Stasser & Titus, 1985). Failure to share information with other group members, e.g., because of lack of opportunity to do so (production blocking), contributes to friction loss, as well as a heightened need for groupwork or conflict

(Bang & Frith, 2017). Lack of or insufficient access to information might skew the group decision making process towards a suboptimal outcome (Stasser & Titus, 1985). A faulty heuristic for the importance of some pieces of information also deters the group from the optimal outcome (Bause et al., 2018). Groups also need to establish a common ground to work together effectively (Clark & Schaefer, 1989).

While a group work process has the potential to foster higher quality outcomes, than single work, additional factors must be accounted for to achieve this. Technological aids were developed to promote the benefits of group work and to facilitate collaboration in group settings, (e.g., Bodemer et al., 2018; Järvelä et al., 2015; Phillips and Phillips, 1993). Multi-Touch Tables (MTTs) have been shown to be effective in supporting equality in and enthusiasm for participation in group work (Buisine et al., 2012).

1.2 Multi-Touch Table

In 1999, Mark Weiser published an article that would foreshadow the development of computer technology in the 21st century. He proposed many ideas for devices and networks, that are now an integral part of our everyday life, e.g., tablet computers, white boards, wireless home networks, etc. He also speculates about desktop sized horizontal interactive surfaces, that would replace ordinary desks. Although the technology already exists, this vision has not been fulfilled as of yet (Weiser, 1999). MTTs are a Single Display Groupware (SDG), that allows the simultaneous work of several users at the same screen. Today, Computer-Supported Collaborative Work (CSCW) takes mostly place in a distributed setting, with group members working on and being connected to each other through their own devices. This setup is also used in co-located settings like offices, despite the physical proximity of users. MTTs are rarely used, although they could enable new ways of interactions between the users, enrich their collaboration facilitate peer-learning (Stewart et al., 1999). To do so effectively, there are several affordances for co-located group ware that need to be taken into account, such as content orientation, spatial requirements of the collaborators, their position around the table, etc. (Shen et al., 2006; Westendorf et al., 2017).

Figure 1.1: A group working at the Multi-Touch Table. Some participants have to tilt their heads to ease their access to the presented information! ©IWM/ Sebastian Grotheloh

The MTT has some unique properties, that could be fostered for group work, but are absent in other tools that might be employed. To identify these idiosyncrasies, the MTT must be considered in comparison with other tools that serve the same purpose, i.e., information display and communication. In the context of this work, these tools are summarized under the term "knowledge artifacts". The following sections defines this term and the common properties of its' members.

1.2.1 Knowledge Artifacts

Popper (1978), following Frege (1918), distinguishes three worlds. World 1 is the physical world in which physical objects, such as rocks, plants, animals, mountains, etc. exist. World 2 is the world of consciousness, thoughts and feelings and

World 3 is, according to Popper, the world of products of human thought, such as literature, music, scientific knowledge, but also machines and tools. World 3 objects have the property that they can have a special connection to (physical) World 1. While there are some World 3 objects that are identical to World 1 objects, such as paintings or sculptures, there are other World 3 objects that are more difficult to locate. For example, a novel is simultaneously located in multiple World 1 objects. While each individual physical book of the novel is a separate World 1 object in its own right, together they all represent one and the same novel, the same story, which is a World 3 object. For this dissertation, the term *knowledge artifact* is defined as a World 1 object that has a relationship to a World 3 object.

1.2.2 The Interaction Between Knowledge Artifacts and Users

Most interactions with knowledge artifacts constitute a 1-to-1 relationship between user and artifact, i.e., knowledge artifacts are designed to interact with only one user at any given time; each book has exactly one reader and each computer has exactly one user. The information in the knowledge artifacts is presented in such a way that it is optimally accessible to that user.

The term *interaction* as used in this context needs further precision. An interaction consists of two channels, one for input, where the user manipulates the knowledge artifact, and one for output, where the knowledge artifact presents a piece of information. Indeed, many knowledge artifacts have a 1-to-N relationship on the output side: for example, in the case of advertising posters, presentations, and televisions. In these cases, the medium is designed to be consumed by multiple addressees simultaneously. The input channel, however, i.e., the part of the interaction in which the knowledge artifact is manipulated, nevertheless remains a 1-to-1 relation: a book can only be turned by one person at a time, a presentation is switched on by the presenter, the TV is only controlled by the person holding the remote control, the computer terminal can only be operated by the person at the keyboard, etc. In this sense, all these interactions remain a 1-to-1 relationship (There are also 1-to-0 relationships if no manipulation is possible).

There is a plethora of attempts to establish 1-to-2 or 1-to-N on devices designed for 1-to-1 interactions. For example, a computer game can be played in pairs on the same computer, so that the same artifact interacts with two different users. These solutions, however, remain an insufficient implementation of a 1-to-N interaction. In this example, multiple players can indeed interact at the same time on the same computer and keyboard, i.e., to play a game together. However, the interactions fall back on emulated 1-to-1 relationships. On the keyboard different key ranges might be defined that are used for the input of the respective players and/or the screen might be split to show the respective environment for each player.

1.2.3 Advantages of the 1-to-1 Relationship

The reason that the 1-to-1 relationship has prevailed in most interactions with knowledge artifacts probably lies in its solution to a fundamental problem in interaction design, namely the question of with whom the system interacts. Norman (1986) defines two hurdles in the interaction between artifact and user: the *gulf of execution* where the artifact responds to the user's input in a way that is consistent with the user's intention, and the *gulf of evaluation* where the artifact shows the user the result of the interaction in a way that the user understands the resulting changes that occurred in the artifact. Both hurdles are well-defined as long as there is exactly one user who interacts with and with whom the artifact interacts. The artifact can trivially match the input since there can be only one source of interaction. Similarly, the output does not need to be matched either, since there is only one user with a perspective on the artifact.

When more than one user interacts with the artifact at the same time, two questions arise that are difficult to resolve: From which user originated the input, and to which user does it need to communicate the change in the artifact? Both questions are non-trivial for an artifact because the usability of the artifact depends on how well it resolves them. There are several conceivable strategies that can now be used in the artifact design process as a solution to this problem:

1. The 1-to-1 relationship is established through segmentation. Input and output channels are segmented, and segments are attributed to individual users. For example, in the case of "local co-op games", the screen is split for players and buttons are defined that are assigned to individual players. In the case of MTT, interaction areas are defined for the individual users[1].

2. The knowledge artifact treats all users as one. In this case, it makes no difference to the knowledge artifact from whom the interaction originates. Each interaction is treated as if there were only one user.

3. The knowledge artifact identifies the different users. Once the knowledge artifact can assign from which user which input originates, the knowledge artifact can personalize the output accordingly. Ideally, the system also locates the user in space so that the orientation of the information can also be adjusted accordingly. Most MTTs do not have the inert ability to discern the different users. Occasionally, additional peripheral input is employed, such as fixed user positions (Dietz & Leigh, 2001), hand detection (Ramakers et al., 2012), or personalized tangibles (Kupke et al., 2019).

1.3 Motivation

This thesis project is concerned wit the display of information equally accessible to users around a MTT. Usually, information is presented *highly oriented*, as most forms of presentation are designed to be viewed from one specific angle or point of view. Some examples for this property of information displays might be paintings, that have a right way to be hung up at the wall[2], books (and texts in general) with an unambiguous optimal point of view and reading direction

---

[1] Like it is done in the MTT game "Protect the Exhibit!" by ijsfontein for the Leibniz research museums (IJsfontein, 2018)

[2] Sometimes, this is not obvious. The Museum of Modern Art (New York) displayed the piece "Le Bateau" by Henri Matisse upside-down for 47 days (Robertson, 1961).

(depending on the language, see also Chokron et al., 2009) or maps, having a compass rose for orientation and a main text orientation for legends and feature names[3]. This seems also be generally true for data visualizations, which often incorporate axes or data dimensions and labels, which indicate the orientation of the diagram.

Oriented information is generally highly beneficial for single user applications, providing quick orientation and reducing ambiguity. In a group setting however, these benefits might turn into losses. Each group member would be inclined to place themselves close to an optimal point of view. At the MTT, this would prompt all members to gather on one side of the table, forcing the group members to get closer than they would probably be comfortable with, due to the violation of social norms on personal space (Hall, 1963; Tse et al., 2004). This might also be one of the reasons, why single user displays are often abandoned in group work settings (Heath & Luff, 1992; Kruger et al., 2004). Furthermore, it might reduce the effectiveness of the group decision making process, as more coordination would be necessary, shifting the focus from the reaching the common goal to group coordination processes (D. M. Fisher, 2014). The group members could also decide to use all four sides of the MTT, despite working with highly oriented information. However, this would require additional mental effort from all group members (except the ones on the optimal side), as they are now forced to mentally rotate (also called *mental normalization* in this thesis, see also Chapter 2.4) the presented information (Shepard & Metzler, 1971). This additional mental effort would probably lead to disengagement with the task and thus deteriorate the outcome of the decision-making process (Stasser & Titus, 1985).

Therefore, presenting highly oriented information seems not to be a viable option for group work at the MTT. The next best idea might be to present the same information for all sides of the table. One could present the same (highly oriented) information multiple times on the table to make it accessible from all

---

[3] Early medieval maps of the Mediterranean Sea (so called portolan charts) did not indicate a clear orientation, as harbor names were often written perpendicular to the coastline. Consequently, the maps were probably rotated when used (Kretschmer, 1962).

sides. Unfortunately, the redundant display of information might introduce an importancy biases in the group decision making process (Bause et al., 2018).

An optimal solution, it seems, would be a "directionless" display of information that would provide equal access to all group members around the MTT, regardless of their point of view. Text is a highly oriented displaying method for information by its nature. More flexible forms of information presentation are graphics or data visualizations, as they rely less on a sequential display of information (e.g., writing), but rather use spatial relationships (Larkin and Simon, 1987, see also Chapter 2.1). This thesis investigates the effect of mental normalization processes on the access to information, presented as data visualizations. It explores different design options to mitigate adverse effects of mental normalization requirements and looks at potential benefits of mental normalization processes for long-term memory formation.

Chapter 2

Cognitive Processes

The affordances for collaborative work at the MTT are extensively discussed on a group level (e.g., Buisine et al., 2012; Mateescu et al., 2021; Rogers et al., 2009; Scott et al., 2004). Shen and colleagues (2006) did address design challenges for MTT interfaces, but concentrated on physical limitations and group coordination issues rather than individual perception. This thesis is concerned with design challenges at the MTT addressing not the group, but the individual user. The requirements and burden on the cognitive system, however, have scarcely been a subject of interest. Wigdor and Balakrishnan (2005) addressed the topic of mental rotation at the MTT, stating that the impact of a rotated display was present, but less severe than previously thought, thus partially contradicting earlier findings on rotated stimulus presentation by Koriat and Norman (1985) and Tinker (1956). However, they did not derive any design recommendations from their findings.

This chapter will provide an overview on the relevant literature for this field of research. The first section (Chapter 2.1) summarizes theoretical and empirical work on the characteristics of Data Visualizations and provides the reasons for the focus on this form of information presentation in the present thesis. The following sections focus on the cognitive processes of the individual user interacting with the MTT. The sections are arranged from lower to higher order visual processes, starting with the perception of color and contrast (Chapter 2.2), the most basic process in visual perception. The impact of suboptimal information presentation and the display of conflicting visual information as well as the cognitive processes to mitigate these are discussed in the chapter on the Eriksen Flanker task (Chapter 2.3) and Mental Normalization (Chapter 2.4). Chapter 2.5 is concerned with the processing of textual information, i.e., reading. This is followed by sections on the short- and long-term retention of information. The structure of the working memory model (Chapter 2.6) is presented, followed by the section on learning (Chapter 2.7). The last section is concerned with high-level cognitive strategies on Working Memory capacity management (cognitive offloading, Chapter 2.8) and their advantages and drawbacks.

2.1 The Spatial Presentation of Data

Data visualizations are preferable to other modes of data presentation. Their features can be tuned to accommodate cognitive processes, e.g., by using depictive features like icons or by employing other characteristics that are easily processed.

Compared with other cultural techniques like painting or writing, data visualization is a young form of communication. The earliest cave paintings known today are dating back more than 45.000 years (Aubert et al., 2019). The first written language, Sumerian, emerged around 5000 years ago (Gelb, n.d.). In comparison, the technique of data visualization is only dating back to the 18th century (Schnotz, 2001; Tufte, 1983). Data visualizations are defined as the visual presentation of (mathematical) data. (Cleveland & McGill, 1984; Wilkinson & Wills, 2011). They are used to show information, that is otherwise hard to grasp, such as information that includes large (or very small) time frames or large (or very small) physical dimensions (Moritz, 2019). Scholars have come up with various taxonomies to classify information presentations. Considering these provides a deeper understanding of the nature and variety of data visualizations.

Larkin and Simon (1987) distinguish between visual and sentential representation. All visual representations (diagram, picture, data visualization, etc.), have in common, that spatial positioning and relations are meaningful. This distinguishes them from sentential representations, that is information presented in sentences and text (Larkin & Simon, 1987). Although containing a certain amount of written information, e.g., in labels and legends, data visualization falls mainly in the category of visual representations, as the special relation of its features is crucial in its information display. Other researchers focus on the relationship between the actual object and its representation (Schnotz, 2001, 2002; Schnotz & Bannert, 2003). If they are linked by common structural characteristics, they are classified as depictive representations. If there is no recognizable link, like with abstract symbols, words or letters, the representation is called "descriptive". Data visualizations might fall in either category, depending on the choice of symbol. For example, an ISOTYPE-like bar graph (Haroz et al., 2015) would use pictograms as labels, which share common characteristics with their object and therefore be classified as "depictive". If these labels would be words, the same bar graph would fall in the category "descriptive".

A similar distinction is made between pictorial and semantic displays. Pictorial visual displays use images to communicate information, while semantic displays have the convention to use symbols (Carney & Levin, 2002; McCrudden & Rapp, 2017). Within the field of data visualization, Wilkinson and Willis (2011) categorize information displays by originality. They define graphs as defined categories of diagrams (such as choropleth maps, pie charts, bar charts, etc.) while they see graphics as a more general term for data visualizations, that might combine or exceed these categories. Finally, some researchers distinguish Information graphics from data visualizations, as they are defined as a combination of text, pictures and graphical means into a larger unit that aims to communicate information (Holsanova et al., 2009; Weber & Wenzel, 2013). Zwinger and Zeiller (2016) report three basic categories of information graphics: Principle representations, depicting causal relationships, cartographic infographics, that communicate space-related information and statistical charts, that report on quantitative information.

Using data visualization provides some benefits over other forms of data communication, like plain text or tables. The mere presence of an external representation may be enough to ease cognitive processing, e.g., by using data visualization (Scaife & Rogers, 1996). There is evidence, that people might get the "gist" of a diagram at first glance, but need to take time for deeper processing (Eitel et al., 2012). Generally, data visualizations are able to communicate patterns in data more efficiently and concise than other presentation formats (Cattaneo et al., 2007; Dambacher et al., 2016). The reasons for these benefits might be found in the pictorial nature of data visualizations. As pictures can contain much more information than words and more information can be kept in working memory. This is called the picture superiority effect (Cattaneo et al., 2007; Maisto & Queen, 1992; McBride & Dosher, 2002; Standing, 1973; Whitehouse et al., 2006). Also, pictures are more accurately remembered than words (Cattaneo et al., 2007; Standing, 1973) and response latencies are shorter for pictures than for words (Jenkins et al., 1967; Shor, 1971). Another explanation for the benefits of data visualizations might be drawn from the Feature Integration Theory, which describes basic visual features that the human visual system can process without strict capacity limitations (Treisman & Gelade, 1980). These basic features are therefore useful for the design of data visualizations (Nothelfer et al., 2017). Other researchers aimed to exploit the human capacity for facial recognition for the efficient visualization of multivariate data (Chernoff, 1973), but the success of this approach is debated (Lee et al., 2003; C. J. Morris et al., 1999).

The topic of data visualization design has gained attention of cognitive psychologists in recent years (Rensink, 2014, 2017; Rensink & Baldridge, 2010) and theories on human perception are informing data visualization design (e.g., Evergreen, 2017). There are different, sometimes conflicting standpoints on the best practice of data visualization design. The economist and member of the Vienna Circle Otto Neurath aimed to use data visualizations to educate the working class about economic data. He invented the ISOTYPE (International System of Typographic Picture Education) guidelines for data visualization design (Neurath, 1936; Neurath & Odgen, 1937). It promotes the use of icons and pictographs, reducing the need for labels and supports the depictive property of the data visualization. By doing so, it claims to make the data visualization more accessible. Recent research supports this claim, if ISOTYPE guidelines are applied properly (Haroz et al., 2015). Other scholars are critical of additional embellishments on data visualizations and promote a data-only design approach (Tufte, 1983). Indeed, for immediate information processing, embellishments seem to have an adverse effect (Bateman et al., 2010; Skau et al., 2015), while others found no impact (Kosara & MacKinlay, 2013). For long-term retention, embellishments seem to have a positive impact (Borgo & Abdul-Rahman, 2012; Borkin et al., 2013; Hullman et al., 2011). A big influence on the understandability of data visualizations seems to be the familiarity with the display, especially for lay people (Maltese et al., 2015). They found familiar visualization designs more attractive (Quispel et al., 2016) and were able to name and interpret them (Börner et al., 2016). Departing from familiar design patterns, such as color scales, might negatively impact the understandability of data visualizations (Christen et al., 2021). Training in basic statistics and data literacy might therefore improve the effectiveness of data visualizations (Aung et al., 2019).

If applied correctly, data visualizations are a powerful tool for information communication. The picture-like features can accommodate the natural visual processing capabilities and therefore be highly efficient. However, the audience is a key factor to be considered. For lay people, simple and familiar designs are to be preferred for data communication.

2.2 Building Blocks of Visual Perception

The process of visual perception relies on the most basic information, the visual system can detect in the environment. Before any higher order visual processes can be initiated, both, color and contrast need to be detected. This is true for any kind of stimulus, be it a written text, a landscape, a painting or, most relevant for this project, at a Multi-Touch Table. Color and contrast of the stimulus both have profound impact on the visual processings speed and accuracy.

The research on visual perception of contrast dates back to the early days of psychophysics (Fechner, 1859). Despite such a long tradition of research, it has not yet been possible to agree on a clear, universally accepted definition for the concept of "contrast" (Beghdadi et al., 2020). There are multiple reasons for the slow advancement in this area. First, there is the complex nature of the visual system. The process of visual perception spans over several levels of abstraction, beginning with the physical interaction between light an receptors in the eye, over preliminary information aggregation in the bipolar and ganglion cells to the information processing in the visual cortex (Grondin, 2016). Another reason is the intricate nature of "color". Scholars have proposed a multitude of color models (Fairchild, 2013). The one most commonly used in perception sciences is the HSB model (Grondin, 2016; Mausfeld, 2011). It was proposed as a color scheme for television (Valensi, 1939) and describes colors on the dimensions "hue" (the main property of color, e.g., yellow, blue, etc.), saturation (intensity of the color) and brightness (how light or dark the color is). However, Mausfeld (2011) also describes controversies around this subject that arose from different requirements, for example the HSB model can describe the perceptual properties of a colored surface but does not account for the true color. They therefore propose a two-factor model with illumination (light) and object/surface color, arguing that the color impression does not change by a different lighting in the room. This also highlights another problem concerning variations in experimental design. Some studies used surface reflection of light, while others use a computer screen setup (Beghdadi et al., 2020). Finally, the perceived contrast might be dependent on interactions between environmental and physiological factors (Hou et al., 2021). The combination of these factors leads to a multitude of different contrast measurement constructs; Beghdadi and colleagues (2020) list over thirty different formulas, customized to specific settings. The processing of color by the visual apparatus seems to be incredibly efficient (Treisman & Gelade, 1980). Color coding has been shown to be a very efficient cue in several visual tasks such as perceptual grouping (Palmer et al., 2003), guiding attention (Carter, 1982; Christ, 1975; Duncan & Humphreys, 1989) and visual search (Wolfe & Horowitz, 2004, 2017).

In the context of this project, the impact of contrast and color on the readability is the most relevant field of research. The first report of a systematic assessment of paper-typeface contrast were conducted by Babbage (1832), who studied the best way to print logarithmic tables. He found a favorability of high contrast (yellow or white page to black ink). Tinker and Paterson published an extensive article series called "Studies of Typographical Factors Influencing Speed of Reading", to assess various influences on readability. Experimenting with different fore- and background colors, they confirmed this by reaching the same conclusion (Tinker & Paterson, 1931). This also holds true for individuals with visual impairments (Sloan, 1969, 1977). Low contrast, on the other hand, seems to have a negative impact on readability (Howell & Kraft, 1959), but others found it negligible, except for very low contrast (van Nes & Jacobs, 1981) or unusual big or small font sizes (Legge et al., 1987). The polarity of contrast seems to have no impact on readability (Legge, Pelli, et al., 1985; Rubin & Legge, 1989), but conflicting evidence exists (Tinker & Paterson, 1931).

In summary, color and contrast play an important role in the visual processing. They influence the processing speed and accuracy of the individual user at the MTT. High contrast can help to speed it up. However, while black/white contrast conditions are researched in depth, color contrast is a very complex topic that can lead to counterintuitive results. Nevertheless, color and contrast need to be considered in designing effective data visualizations.

2.3 Processing Entangled Information

Traditionally, psychologists test the influence of a manipulation by presenting one stimulus at a time. While this clean experimental setup allows a precise manipulation of the stimulus, it might neglect interactions with and influences from the environment, thus limiting the generalizability of findings to real-world scenarios. Sometimes, relevant information needs to be discerned from irrelevant or contradicting information that accompanies it. Obviously, the information should be displayed unambiguously at the MTT, preventing additional strain on the limited resources of the users' working memory. However, it may be hard to identify potentially interfering properties of information presentations. The influence of additional stimuli on the processing of a main stimulus can be investigated in experiments with the *Eriksen Flanker paradigm*.

The Eriksen Flanker paradigm was first introduced by Barbara and Charles Eriksen (B. A. Eriksen & Eriksen, 1974). The now famous paradigm is used to investigate interferences occurring between stimuli. It consists of a target stimulus that should elicit a predefined response. This target stimulus is accompanied by additional, flanking stimuli that are either congruent (eliciting the same response), incongruent (eliciting a conflicting response) or neutral (eliciting no response). The time between stimulus presentation and response reaction is the indicator for the influence of the flanking stimuli. In their seminal work, Eriksen and Eriksen (1974) discovered, that the response times varied between these combinations. Response times were shortest for trials with congruent flankers, followed by trials with neutral flankers. Trials with incongruent flankers elicited the longest response times. This effect was further moderated by the distance between the target stimulus and the flanker stimuli. The closer the flanker was to the target, the higher was the influence of the flanker on the response time. The influence of flanking stimuli on the response on the main stimulus suggests a communally processing for all presented stimuli. This effect has been replicated many times with different sorts of stimuli, such as letters (B. A. Eriksen & Eriksen, 1974), arrows (Kopp et al., 1994), numbers (Lindgren et al., 1996), colors (Rafal et al., 1996), etc. It is not only used in the research on visual information processing and attention but also in a wide range of other areas such as clinical settings (Mullane et al., 2009), training (Aydmune et al., 2019) and second language impact (Salwei & de Diego-Lázaro, 2021). The original paradigm has also been expanded from visual perception to multimodal information processing (Ulrich et al., 2021).

The influence of flanking stimuli on the processing of a main stimulus was reliably replicated for a vast variety of different types of stimuli. For the design of the information display on the MTT, this influence must be considered. Whether the display of redundant or similar information acts as a congruent or incongruent flanker is hard to determine beforehand and needs to be subjected to experimental evaluation.

2.4 Recognizing Depicted Information From Different Points of View

An interesting question about visual processing is, how humans can recognize objects that have been rotated and now misalign with the initial position. With the movement, the retinal impression of the object is changed, and therefore could be something completely different, according to the pure sensory information. It is not trivial, that humans are able to perceive these two retinal impressions and still recognize the same object. At the MTT, rotation is an integral part of the user interface functionality and group communication (Kruger et al., 2004; Shen et al., 2006). Users are therefor constantly confronted with the task of mentally normalizing rotated displays to access the presented information.

In their seminal paper, Shepard and Metzler (1971) investigated how participants would judge the (in-) congruency of two presented shapes that were differently aligned. They found that the response time would increase proportional to the three-dimensional angular deviation of two congruent shapes. They therefore proposed some sort of mental rotation mechanism, that humans would use to mentally align the two shapes to judge their similarity. The effect of mental rotation on response time latencies was then replicated for other stimuli as well, such as two-dimensional random shapes (Cooper, 1975) and letters (Corballis & McLaren, 1984; Rüsseler et al., 2005). Indeed, whenever two rotated stimuli needed to be matched, response times would increase with the angle of deviation up to 180° and decrease again, until both displays would have the same rotation. Also, the response time increases with increasing complexity of the object to be rotated (Bethell-Fox & Shepard, 1988) while accuracy decreases (Meyerhoff et al., 2021).

The assumption of the mental rotation of a whole visual unit was soon to be contested. It was critically noted that other mental transformations would generate the same response time patterns, such as size comparison (Bundesen & Larsen, 1975) and contrast (O'Donell et al., 2010). Other authors proposed different mechanisms, such as storage of multiple points of view (Tarr & Bülthoff, 1998; Tarr & Pinker, 1989), piece-meal rotations of visual units (Just & Carpenter, 1985; Xu & Franconeri, 2015; Yuille & Steiger, 1982), the comparison to a mental 3D model of the object (Marr, 2010; Marr & Nishihara, 1978) or some sort of edge detection mechanism (Biederman, 1987). Each transformation proposal comes with their own evidence, and it seems, as if multiple mechanisms might be at work, depending on the task. A distinction between them needs the use of neuro-imaging procedures (Gauthier et al., 2002).

Apart from basic research, the cost of mental rotation has also been shown to have an impact in more applied settings. For example, Montello (2010) showed the real-life impact in errors an distress of misaligned you-are-here maps. Mental rotation is costly enough that humans tend to avoid it by offloading it on their body, e.g., by rotating their head (Risko et al., 2014). For the rotation of complex structures such as molecules, trained people are using additional information and properties to reconstruct the rotation (Stieff, 2007).

For the current work, there are three main take-aways: First, there is evidence for a general mental rotation cost, regardless of the underlying mechanism. Whenever humans are faced with a need to mentally rotate an object, it comes with additional strain on the visual processing system. Second, there might be a benefit from reduced complexity of the presented stimuli, which would reduce the additional rotation cost for the processing. Third, it might be interesting to gain some insight in the applied strategy for mental rotation of rotated bar graphs. A distinction between piece-meal and whole rotation might offer some insights for design recommendations.

2.5 Processing Textual Information

In the first section of this chapter, the advantages of the presentation of information as data visualizations instead of sequential text have been discussed (see Chapter 2.1). However, most data visualizations still rely on a certain proportion of their information to be communicated by text. They often contain some kind of labels, legends or annotations. As text is an integral part of data visualizations, the factors that influence the readability of a text are relevant for a informed design of information presentation at the MTT.

Reading is a highly complex cognitive task, and it involves several layers of processing. First, the visual representation (grapheme) must be recognized. Several physical factors can influence the "readability" of the grapheme, such as font (Bigelow, 2019; Paterson & Tinker, 1931a), color (Babbage, 1832; Legge et al., 1990), size (Howell & Kraft, 1959; Paterson & Tinker, 1929; Rudnicky & Kolers, 1984), contrast (Howell & Kraft, 1959; Ohnishi et al., 2020; van Nes & Jacobs, 1981), orientation (Byrne, 2002; Yu, 2010), lighting condition (Legge et al., 1990; Ohnishi et al., 2020) etc. Physiological characteristics of the reader, such as visual impairments might also have an impact (Legge, Rubin, et al., 1985)[1].

If the grapheme is physical and physiological recognizable, its visual information is passed through the visual cortex system, taking the ventral pathway - "recognition route" - (Goodale & Milner, 1992; Milner & Goodale, 1993) for identification (Pegado et al., 2014). Meaning is then extracted in the visual word form area (Dehaene & Cohen, 2011; McCandliss et al., 2003). Dyslexia (Richlan et al., 2011) and physiological impairments might have an impact on the grapheme-to-phoneme translation (Norton et al., 2007).

While the neurological processes are well researched, the body on the process of reading regarding the working memory model is rather sparse. originally, Baddeley (1979) summarized several results on acoustic and visual word processing tasks, concluding that the auditory system, specifically the "articulatory loop" (later a sub-system of the phonological loop) was probably not involved in reading, except for some special cases (see Chapter 2.6 for more information on the Working Memory model). However, there is evidence for a hand-over of read information to the phonological loop. Conrad (1964) found an acoustic confusion for letter sequences, based on their phonemes, rather than ther graphemes.

---

[1] There are two major lines of research on readability that cannot be completely referenced here. The earlier one is *Studies of typographical factors influencing speed of reading* by Tinker and Paterson (see also Sutherland, 1989), the second one is *Psychophysics of reading* by Legge and colleagues.

Zhang and Simon (1985) obtained similar results with homophone, but visually different Chinese characters. In 1986, Baddeley proposed an entry of non-speech information into the phonological loop (Baddeley, 1986), later highlighting the importance for second language learning (Baddeley et al., 1998). The next steps in the reading process would be higher order cognitive functions such as sentence comprehension and context creation. As these areas are outside the scope of this work, they will not be discussed!

The take-aways from reading science for this thesis are the following: First, the influence of contrast on the reading performance is important to keep in mind, when designing accessible information presentations for the MTT. The second major take-away is the unique handling of text in contrast to other visual information. At some point in the information encoded in the visual medium "text" is handed over to the phonological loop, meaning that it is treated like auditory information. This change of mode might indicate that text, as it is handled differently than other visual information, might be impacted differently by some of the presented influences than pictorial stimuli.

2.6 Memorizing Information

Once the user has successfully perceived the visual, the information needs to be processed for further utilization. This happens in a cognitive instance called the *working memory*. The concept of *working memory* (WM) evolved from the concept of short-term memory (STM). The concepts of long-term memory (LTM) and STM where solely focusing on information storage and did not include any mechanisms on how information would be received, transformed, and stored (Baddeley, 2012). Baddeley and Hitch (1974) proposed a three-component model, to include these processes. These components are the visuo-spatial sketchpad, the phonological loop, and the central executive (Baddeley & Hitch, 1974). Later, Baddeley added a fourth component to the model, the episodic buffer (Baddeley, 2000). Information enters the WM via visual or auditive channels on the visuo-spatial sketchpad or the phonological loop, respectively.

The *visuo-spatial sketchpad* is the part of the WM that is concerned with visual and spatial information processing (Baddeley et al., 2009). While it is mostly described as a single component (Buchsbaum, 2013), some researchers suggested a set of subsystems for it (Logie & Pearson, 1997). The general capacity of the visuo-spatial sketchpad is estimated of about three to four items (Luck & Vogel, 2013). The second component of the WM is the *phonological loop*, that contains auditory information for processing. It might contain two subsystems, the phonological store, a passive buffer, holding around two seconds of heard information and the articulatory control process, an inner voice to actively hold information in WM by rehearsing it (Baddeley & Hitch, 1974; Buchsbaum, 2013).

The third and most peculiar component of Baddeley's model of WM is the *central executive*. It is concerned with allocating resources on the other systems. It also decides, which information is passed on and stored for long-term retention in the LTM (Baddeley & Hitch, 1974). It is often criticized for being ill-defined while also taken the most important role in the WM model. Baddeley described it as a "homunculus" (Baddeley, 2012) that acts as a placeholder for future research to work it out (Baddeley, 2012). There was evidence mounting up, that information from the LTM is held in the WM without loading either the phonological loop or the visuo-spatial sketchpad, so Baddeley added another component to his working memory model, called the *episodic buffer* (Baddeley, 2000). This component is a place to hold information that is retrieved from the LTM and for processing multimodal integrated information from the visuo-spatial sketchpad and the phonological loop (Baddeley, 2000).

The overall capacity of the working memory is deemed to be seven plus minus two elements of information (Miller, 1956), but might be even more constraint to three to four items (Cowan, 2001; Marois & Ivanoff, 2005). To use this very limited capacity, humans can connect atomic elements to larger conglomerations called chunks (Miller, 1956). This process of chunking is shown in both, the phonological loop (Norris & Kalm, 2021; Norris et al., 2020) and the visuo-spatial sketchpad (Meyerhoff et al., 2021).

Baddeleys' is the most prominent working memory model, but it is not without criticism. Some researchers complained about missing information pathways of other sensory modes or proposed different models (Baddeley, 2012; Ward, 2001). Alternatives include the embedded processing model of working memory (Cowan, 1999) which is focused less on components and more on processes and cognitive architectures that comprise different conceptualizations of working memory, like ACT-R (Anderson & Lebiere, 1998) or EPIC (Kieras, 2016).

The main take-away for this project, is the general limited capacity of the working memory. As only a certain amount of information can be retained, the efficient use of this capacity, e.g., by chunking information is important. The proposed sub-routines, especially the different processing paths for visual and auditory information need also be taken into regard. As the subroutines have different storages, this could be utilized for a more efficient information processing. Once the information is successfully retained in the working-term memory, it can be handled in different ways, depending on the users' goals. It might be processed for long-term retention (learned, see Chapter 2.7), it can be processed deeper if it's not presented in an easy-to-digest format, e.g., containing conflicting information (see Chapter 2.3), looked at from an unfamiliar angle (see Chapter 2.4) or an artificial encoding like text (see Chapter 2.5). For easier processing, the information may also be stored in the body or the environment (see Chapter 2.8).

2.7 Long-term Retention of Information

If the user at the MTT intends to gain knowledge, the presented information should accommodate this goal. The best design for learning material for long-term retention is a long-standing debate in learning science. In the discourse, different theories emerged, aiming to describe and explain, how and why some study material or strategies work better than others. There are several models to guide the learning material design. In this section, five theories on learning material design that are suited to guide information presentation at the MTT are briefly summarized.

The *Cognitive Load Theory* (Sweller & Chandler, 1994) identifies three sources of load on the cognitive capacity, stemming from the learning material. As the capacity of the working memory is limited (Miller, 1956), it is crucial to act economical on it. *Intrinsic cognitive load* stems from the subject of the learning material itself. Complicated or complex subjects put a higher strain on the cognitive system than easy and simple material. This load can hardly be avoided. Bad material design, irrelevant information and repetitions put additional, *extraneous cognitive load* on the cognitive capacity of the learner. Careful material design should aim to keep it as low as possible, to free mental capacities for the last type of cognitive load. *Germane cognitive load* is defined as the effort, the learner makes to connect the new information to known concepts and material. This should be as high as possible. It can be guided through good instruction design.

Salomon (1984) proposed a different model that accounted for the surface difficulty of the learning material: The AIME (*Amount of Invested Mental Effort*) theory. If the learning material seems to be difficult (perceived demand characteristics), learners are inclined to invest more cognitive capacity in learning. The second factor influencing the amount of invested mental effort is the metacognition on the own capabilities of the learner (perceived self-efficacy). While being popular in different fields of learning science, some researchers are not convinced of its relevance as an independent theory (Schwab et al., 2018). Other researchers were unable to replicate the results of the original study (Beentjes, 1989).

Not every effort invested in studying is beneficial for learning (Bjork & Bjork, 2009). Therefore, it is important to identify *desirable difficulties*, that increase the load on the cognitive system, but also lead to better long-term retention. Some beneficial strategies were identified, such as using tests as a tool for learning (Kornell & Vaughn, 2016) or rephrasing the material in own words (Bertsch et al., 2007; Slamecka & Graf, 1978).

The *Dual Coding Theory* (Paivio, 1991) and its successor, the Cognitive Theory of Multimodal Learning (Mayer, 2005) take an integrative view on the design of learning material. Most other theories rely on only one type of learning material, e.g., text, video or graph, or are agnostic to it. Dual Coding Theory and Cognitive Theory of Multimodal Learning suggest a deeper processing of the learning material if it is presented in multiple modes. This would cause the brain to use different pathways to process the information, resulting in a more thorough learning. Good design for the integration for text and picture information is difficult (Eitel et al., 2012; Kombartzky et al., 2010; Scheiter et al., 2018).

Another concept on material design and its influence on cognition is *Cognitive fluency*. It describes how easy it is for our brains to process the presented information (Alter & Oppenheimer, 2009; Oppenheimer, 2008). The materials' ease of processing depends on various factors, e.g., familiarity, rhyme, and repetition (Oppenheimer, 2008; Reber et al., 2004). Cognitive fluency also impacts the metacognitve expectations (Song & Schwarz, 2008). Reducing the cognitive fluency by adding additional obstacles (*disfluencies*) to the material might trigger humans to invest more mental effort in general (Alter & Oppenheimer, 2009; Oppenheimer, 2008). Some studies found beneficial perceptual disfluencies for text retention like slight blurring of the text (Rosner et al., 2015), using an unusual font (Diemand-Yauman et al., 2011), or brief masking (Mulligan, 1996). However, the benefits of perceptual disfluencies remain debated (Magreehan et al., 2016; Rhodes & Castel, 2008, 2009; Xie et al., 2018; Yue, Bjork, & Bjork, 2013).

All theories have in common, that some mental effort needs to be invested for efficient learning. The concept of cognitive fluency is especially relevant for this project. Mental normalization (see Chapter 2.7) could act as a perceptual disfluency, prompting the user to invest more mental effort and result in better long-term retention of diagram information. It is also possible, that mental normalization acts as extraneous cognitive load (Cognitive Load Theory), occupying cognitive resources that otherwise could be invested in a deeper processing of the learning material.

2.8 Facilitation of Cognitive Processes

As the capacity of the working memory is limited (see Chapter 2.6), humans deal with this limitation by employing mitigation strategies. One of these is called *cognitive offloading*. Cognitive Offloading is defined as "the use of physical action to alter the information processing requirements of a task so as to reduce cognitive demand" (Risko & Gilbert, 2016). External representations save computation costs, memory capacity and might provide structure (Kirsh, 2010). Some examples for cognitive offloading in everyday life are finger counting (Alibali & DiRusso, 1999; Costa et al., 2011), note taking (Eskritt & Ma, 2014) or tilting the head to read rotated text (Risko et al., 2014). MTTs have also be shown to be suitable as an offloading device for WM intensive tasks (Brich et al., 2019, 2021).

For short-term tasks, cognitive offloading strategies boost performance (Beitzel & Staley, 2015; Grinschgl et al., 2021; Kirsh, 2010; Risko & Gilbert, 2016) but it might have a negative impact on long-term memorization. Sparrow and colleagues (2011) reported a loss in recall performance, when participants were made aware of internet search engines. However, these findings failed to be replicated (Hesselmann, 2020). Other researchers reported an adverse effect of cognitive offloading strategies for performance in follow-up trials in various areas, such as navigation (Fenech et al., 2010; Gardony et al., 2013, 2015), problem solving (van Nimwegen & van Oostendorp, 2009), skill acquisition (Casner et al., 2014; Ebbatson et al., 2010) and learning (Eskritt & Ma, 2014; Henkel, 2014; Kelly & Risko, 2019a, 2019b; Pyke & LeFevre, 2011). Some researchers proposed to use the interruption of cognitive offloading as a desirable difficulty (Beitzel & Staley, 2015; Morgan et al., 2009, 2013).

Cognitive Offloading is a viable strategy to free working memory capacities for other tasks. This can be beneficial for immediate task performance but might be adverse for long-term memory formation. The consequences of Cognitive Offloading can therefore support or hinder the user at the MTT in achieving their goal, depending on the nature of it.

Chapter 3

Present Research

The present research focusses on the cognitive processes involved in the work at the MTT. Users at the MTT can vary in the goal they try to achieve with it. On one hand, they might want to use the MTT for decision support or group work augmentation, both being short-term goals. On the other hand, the main goal of the group might be to learn about the presented information. Depending on the sizes of both the table and the group, some members will have to place themselves on the non-optimal sides of the MTT, eliciting the need to mentally normalize at least some information on it. This project investigates the short- and long-term impact of mental normalization on the cognitive processing of information in data visualizations

The first two chapters focus on the effect of mental normalization on short-term retention and its possible mitigation. Chapter 4 aims to provide insights on the short-term impact of mental normalization on the cognitive processing of information in data visualizations. Furthermore, a series of modifications on the data visualization are tested on their viability to reduce the mental normalization effect. The modifications include the use of pictographs as labels and redundant information coding with color. In Chapter 5, a different approach is tested. A flanker-paradigm inspired approach of a redundant display of rotated words is tested on its impact on response time alterations und mental normalization. This approach is subsequently applied to data visualizations as label manipulation.

The impact on long-term retention of mental normalization processes is investigated in Chapter 6. The possible use of mental normalization to facilitate long-term retention of information presented in data visualizations is tested. For this, a device to control for the influence of participants (involuntary) head tilt was developed and implemented. The specifics for this device can be found in Appendix A.

The chapters 4, 5, and 6 were designed to be stand-alone manuscripts. Therefore, the content shows some overlap with the general introduction and the general discussion of this thesis. Except for the parts with explicit statements, this thesis is completely my own work. Chapter 4 and 5 do both include the work of co-authors. Each chapter has a preceding declaration of co-authorships, stating the co-authors and their respective share of the work. All co-authors agreed on their stated share of work.

Declaration regarding § 5 Abs. 2 No. 8 of the PhD regulations of the Faculty of Science – Share in collaborative publications/ manuscripts

The following Chapter (Chapter 4) consists of a study that was designed together with Hauke S. Meyerhoff and Friedrich W. Hesse. The proportional contributions to this study are presented in the subsequent table.

Author	Author Position	Scientific Ideas %	Data generation %	Analysis & interpretation %	Paper writing %
Tjark S. Müller	First	70%	100%	85%	50%
Friedrich W. Hesse	Second	10%		5%	10%
Hauke S. Meyerhoff	Last	20%		10%	40%
Title of paper:	Two people, one graph: The effect of rotated viewpoints on accessibility of data visualizations				
Status in publication process:	published Müller, T., Hesse, F. W., & Meyerhoff, H. S. (2021). Two people, one graph: The effect of rotated viewpoints on accessibility of data visualizations. Cognitive Research: Principles and Implications, 6(1), 31. https://doi.org/10.1186/s41235-021-00297-y				

Author's contributions

All authors developed the study concept and contributed to the study design. TM coded the experiments and analyzed the data. TM and HSM drafted the manuscript. All authors provided revisions to the manuscript and approved the final version of the manuscript for submission.

Acknowledgements

We would like to thank Dr. Nora Wickelmaier for her counseling on statistics as well as Fiona Radtke and Jana Lang for their support in data collection and Dr. Gabriela Orellana Corrales for assistance in the problem demonstration.

Chapter 4

Two People, One Graph: The Effect of Rotated Viewpoints on Accessibility of Data Visualizations

4.1 Introduction

Imagine you are sitting opposite to someone who has spread out a newspaper across the table. One of the articles captures your interest, and despite the inverted orientation, you will be able to read the teaser of this article. Nevertheless, you might be slower and maybe less accurate than usual because the upside-down view on the article requires additional mental processes for recognizing and understanding the written text as well as the depictions (Hayward & Williams, 2000; Kolers, 1968). A similar situation arises when multiple users collaborate sharing a single technical device. The range of technical devices for such a scenario is large, ranging from smartphones on which multiple observers access the same information simultaneously to complex multi-touch tables which are explicitly designed to serve as a collaboration tool for co-located groups.

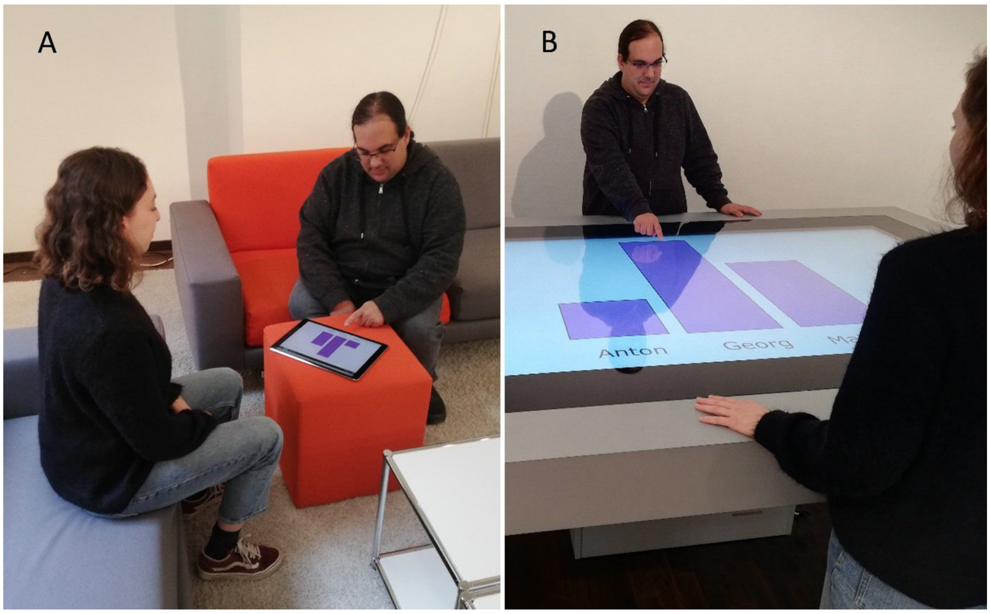

Figure 4.1: Illustration of the collaborative settings with shared displays. A: Two users study the same information on an iPad. B: Two users collaborate using a multi-touch table, an interactive tool for group collaborations. In both scenarios, the display is rotated for one of the observers.

With regard to the perception and recognition of objects, a substantial body of previous research has confirmed the existence of canonical viewpoints from which depicted information can be accessed the most efficiently in terms of errors and speed. Deviations from this canonical viewpoint (i.e., non-canonical views) typically come along with increasing access costs (Diwadkar & McNamara, 1997; Palmer et al., 1981; Tarr, 1995).

With regard to the collaborative view on the bar graphs depicted in Figure 4.1, the canonical view is the upright presentation of the graph. Even in rather simple scenarios with only two users such as those depicted in Figure 4.1, the viewpoint of one of the observers will deviate from the canonical upright view, as collaborators tend to position themselves on distinct sides of the display (Tang et al., 2006). In other words, (at least) one of the collaborators will have a non-canonical view, which is suboptimal for accessing the displayed information. In the present project, we studied how deviations from non-canonical views decrease the accessibility of the depicted information. Following an approach of use-inspired research, our aim was to investigate ways to reduce the detrimental effects of non-canonical viewpoints based on psychological theories of feature and information processing.

4.1.1 Information Visualization

With the tremendous increase of data and accessibility of data starting in the late 20th century, data presentation has become an important topic of research, as data visualizations can communicate patterns in data more clearly, efficiently, and concisely than other presentation formats (Cattaneo et al., 2007; Dambacher et al., 2016). In the present project, we chose to study bar graphs as one type of data visualization which is commonly used to summarize frequentist information and therefore reflects a plausible representation of data in collaborative group settings. The challenge of designing intuitive (Neurath, 1936) as well as effective visualizations (Cleveland & McGill, 1984) of data has been investigated across scientific disciplines (e.g., Montello, 2010; Stieff, 2007). The topic of optimizing data visualization has also gained attention from cognitive psychologists (Rensink, 2017; Rensink & Baldridge, 2010), and basic theories on human perception have influenced guidelines for the design of data visualizations (e.g., Evergreen, 2017). A good example of a psychological theory informing data-visualization designers is the feature integration theory (Treisman & Gelade, 1980), which describes a set of basic features that human observers can process without strict capacity limitations. These basic features therefore reflect useful features for graphs (Nothelfer et al., 2017). Another good example is the "picture superiority effect" (Cattaneo et al., 2007; Maisto & Queen, 1992; McBride & Dosher, 2002; Standing, 1973; Whitehouse et al., 2006), which describes the phenomenon that pictures are easier to learn (and retrieve) than words. In the visualization literature, this finding matches the ISOTYPE (International System of Typographic Picture Education) guidelines. Emerging from early work of Neurath (1936; see also Neurath and Odgen, 1937), this framework suggests using pictographs rather than written labels to represent units, concepts, and frequencies in visualizations. This framework has received empirical support from a study by Haroz, Kosara, and Franconeri (2015), who showed that pictographs indeed improve information processing relative to simple bar graphs with written labels. Whereas these approaches have made substantial progress in improving the accessibility of visualized information for individual observers, hardly any research has attempted to optimize visualizations for non-upright viewing conditions. This is not too surprising, as the pressing need for such research mostly arose with relatively new presentation technology, such as multi-user touch-tables. However, a substantial body of cognitive psychology research has investigated the question of how human observers access and compare visual information from different viewpoints, such as in the case of image or display rotations.

4.1.2 Accessing Rotated Information

Accessing visualized data and viewing and recognizing depictions, representations, and text from unusual angles are common tasks in everyday life. Aside from reading inverted news articles, this encompasses activities such as trying to orient oneself on a city map (Aretz & Wickens, 1992) or reading shop signs on a window from the inside of the shop. In some cases, failure to correctly identify two objects as the same or different from different viewpoints could have serious consequences, for instance, chemists map depictions of molecules (Stieff, 2007) in order to identify enantiomers, which could make the difference between a cure and a poison. Further it is also a key skill for a physician's success in laparoscopic surgery (J. Conrad et al., 2006). Different mental processes have been suggested for accomplishing the mapping operation between rotated views as well as for accessing information from non-canonical viewpoints (see Peissig and Tarr, 2007, for a review). In their seminal work, Shepard and Metzler (1971) asked participants whether two cube structures, which were presented from different viewpoints, were identical or mirrored images of each other. They observed a linear relationship between the time that is necessary for an accurate decision and the angular disparity between both cube structures. From these results, they inferred a mental rotation process (with a constant rotation speed) to align the different views for comparison. The capacity of this mental rotation process, however, remains debated, with some researchers arguing for a holistic rotation of the stimulus (Cooper & Podgorny, 1976), whereas other researchers have reported evidence for piecemeal rotations suggesting a limited capacity (Just and Carpenter, 1985; Yuille and Steiger, 1982; see also Xu and Franconeri, 2015). Whether mental rotation tasks are solved using holistic or piecemeal strategies depends on various factors, such as mental rotation ability (Khooshabeh et al., 2013), sex (Heil & Jansen-Osmann, 2008), and stimulus familiarity (Bethell-Fox & Shepard, 1988). Further, stimulus attributes such as the compressibility of the depicted information influence the amount of simultaneously rotated information (Meyerhoff et al., 2021).

A central limitation in attributing the linear increase in response latencies to a mental rotation process, however, is that such linear relationships also arise in more generalized theoretical conceptualizations of the mapping of information across different viewpoints (Jolicoeur, 1990; Tarr, 1995; Tarr & Pinker, 1989). Among others, these conceptualizations encompass an internalization of 3D-models (Marr, 2010; Marr & Nishihara, 1978), an edge-detection mechanism (Biederman, 1987), or an internalization of multiple viewpoints (Tarr & Bülthoff, 1998; Tarr & Pinker, 1989). In particular, there is neuroscientific evidence that distinct mental operations compensating for deviations in the viewpoint (indexed by distinct neural activity) can hardly be distinguished on a purely behavioral level, as they result in indistinguishable response patterns such as increases in response latency (Gauthier et al., 2002). The goal of our present research is not to disentangle the different mental processes contributing to the compensation of deviations in viewpoint. Derived from the work of Risko et al. (2014), we will therefore use the theoretically neutral term mental normalization as an umbrella for all mental operations contributing to solving this task. While the ubiquitous occurrence of linear decreases in performance makes it difficult to isolate particular mechanisms of mental normalization, it also emphasizes that – irrespective of the exact underlying mechanism – the costs that come along with deviating viewpoints likely affect a substantial number of practical tasks.

In the present experiments, we studied mental normalization for information summarized in bar graphs. Although we are not aware of any direct test of viewpoint costs on accessing information in bar graphs, there are good reasons to propose that these costs follow the same pattern as all previously described tasks. Most relevant here are observations demonstrating that individual elements of bar graphs themselves are subject to costs of mental normalization when their presentation deviates from upright viewing conditions. For instance, such detrimental effects have been consistently reported for shape information (Cooper, 1975), cube figures (Shepard & Metzler, 1971), letters (Corballis & McLaren, 1984; Rüsseler et al., 2005), words (Koriat & Norman, 1985), sentences (Risko et al., 2014), and pictures (Tarr & Pinker, 1989). As most of these effects were observed in prolonged response latencies, it appears likely that extracting information from bar graphs is also prolonged when the presentation deviates from the upright view.

In our experiments, we followed a use-inspired rationale. First, we aimed to establish the costs of mental normalization for accessing information depicted in bar graphs. Second, we strove to understand whether such detrimental effects arise from the rotation of the graph itself or merely due to the rotation of written labels. Third, and finally, we aimed at reducing the costs of mental normalization for information depicted in bar graphs by transferring effects from knowledge-driven basic psychological research into our use-inspired scenario. We arranged our experiment to closely resemble the situation of a collaborative setting as depicted in Figure 4.1. For instance, the participants solved the task on a horizontal touch display which has the same functionality as a smartphone or a multi-touch table. Nevertheless, we tested participants individually in order to allow us to isolate the effect of display rotations with full experimental control.

4.2 Experiment 1

With the first experiment, we aimed to demonstrate that rotated views of a bar graph interfere with the extraction of the depicted information (i.e., the costs of mental normalization). We focused mainly on response latency as a proxy for task performance with prolonged response latencies signaling the proposed interference. We asked our participants to answer comparative questions by extracting the corresponding information from the bar graph. We hypothesized that a rotated presentation of the bar graph would elicit longer response latencies.

A rather obvious candidate for the expected detrimental effect of graph rotation on response latencies are the written labels which identify the individual bars. To investigate the impact of such labels further, we introduced two additional manipulations. First, we compared words versus letters as labels. We expected that the more simplistic shapes of the letters could be identified faster across rotations, thus reducing the detrimental effects of display rotations. Second, we compared two variants of presenting the labels: co-rotation in which the labels rotate with the bar graph (similar to a physical rotation of a sheet of paper) versus reformatted labels which maintain their upright orientation relative to the observer (see Figure 4.2 A for an illustration). If the detrimental effect of display rotations on response latencies emerges only from difficulties in recognizing the rotated labels, the conditions with reformatted labels should not differ from the unrotated view.

4.2.1 Methods

Power Considerations

The most relevant manipulations in our experiments were the label rotation scheme (reformatted vs. co-rotated labels) as well as the label type (e.g., different labels such as text and letters in Exp. 1). Introspectively, both manipulations had an influence on response onsets suggesting substantial effect sizes ($\eta_p^2 > .25$). However, there was no directly related prior work which could have served for a more precise estimate for effect sizes. The closest related study investigated response latencies for rotated textual stimuli (Risko et al., 2014) and observed large effect sizes ($\eta_p^2 = .81$ in their Experiment 1; $\eta_p^2 = .68$ in their Experiment 2). However, this study did not involve representations of data such as ours so that these estimates are probably too large. Consequently, as our study is the first of its kind, we intended to power it appropriately for lower effect sizes ($\eta_p^2 = .10$; assuming correlations among repeated measures of $r = .50$). A corresponding power analysis, $(1 - \beta) > .95$ at $\alpha = .05$) suggested a minimum sample size of 32 participants (G*Power, Faul et al., 2007). In order to compensate for potential data exclusions, we slightly overpowered this sample size resulting

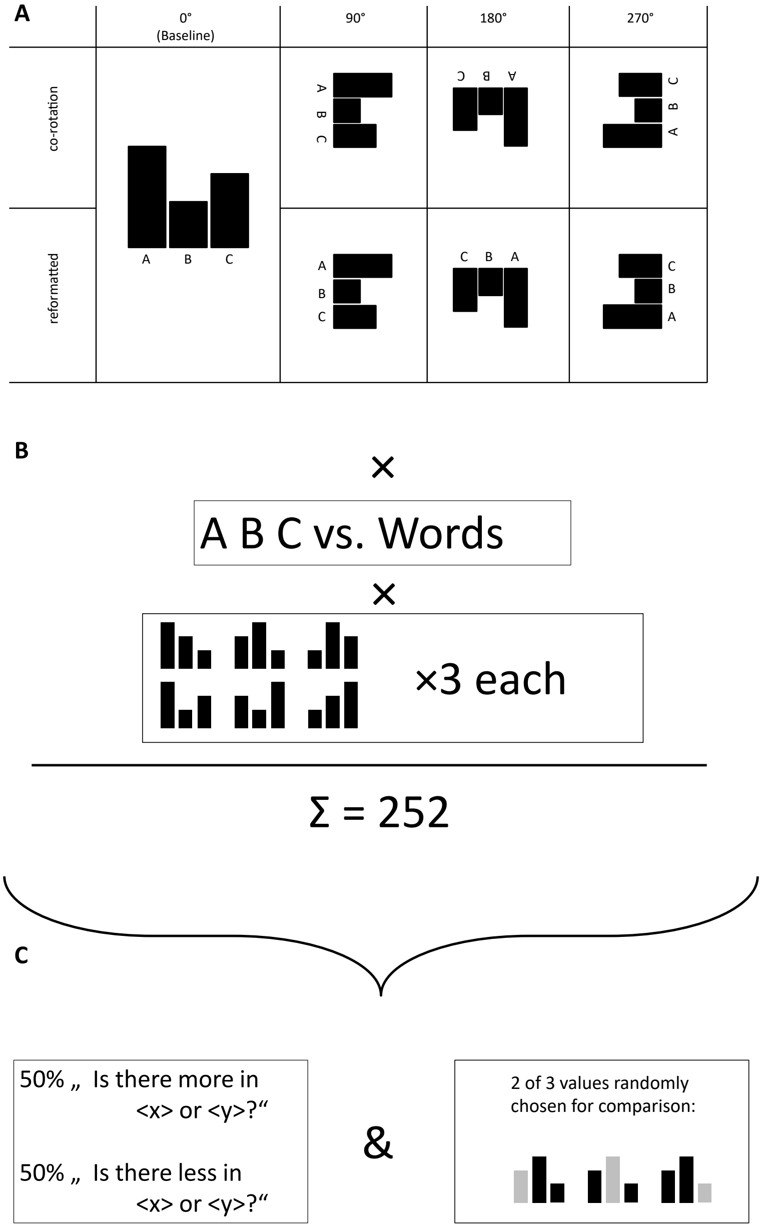

Figure 4.2: Illustration of the design of Experiment 1. A: The baseline condition is an unrotated bar graph labeled either with words or letters. Relative to this baseline, we investigated three different angular rotations (90°, 180°, and 270°) and two different categories of labels (co-rotating vs. reformatted). B: We manipulated the labels of the bar graph (words vs. letters). There were six different arrangements of the bars (each repeated three times), resulting in a total of 252 trials. C: In half of the trials (selected randomly), the participants had to indicate the smaller bar in the comparison. In the other half the participants indicated the larger bar. The two bars involved into the comparison were selected randomly. Each participant received a unique set of stimuli.

in 33 participants in Experiment 1, and 35 participants in Experiment 2. In Experiment 3, we accidentally overrecruited the sample resulting in 41 participants.

Participants

Thirty-three students (22 female, 18-35 years) from the University of Tübingen, recruited via an online platform for volunteer participants for experiments, took part in Experiment 1. They received a compensation of 5€ for 40 minutes of their time. The experimental procedure was ethically approved by the institutional review board of the Leibniz-Institut für Wissensmedien, Tübingen, and all participants provided informed consent prior to testing.

Apparatus

The experiment was conducted on a horizontal $23''$ touch sensitive monitor (Dell Panel Monitor S2340Tt) controlled by a HP Elitebook 8530p. The experimental scripts were coded in Python using the PsychoPy libraries (Version 1.85.1; Peirce, 2007, 2008). The unrestricted viewing distance was approximately 60 cm.

Materials and Procedure

Our participants solved a series of 252 simple questions for which the answers were depicted in a bar graph (see Figure 4.3). Each trial started with a question informing the participant of the names of two values that she/he should subsequently compare (e.g., "Is there more in A or B?" or "Is there less in A or B?"). This question was presented in font size 40 in the middle of the screen. The participants preceded with the trial by putting their index finger down onto the start position, which was horizontally in the middle of the screen, approx. 6.7 cm from the bottom, and 4.5 cm in diameter. Next, a bar graph (15.2 x 15.2 cm), consisting of three bars, appeared in the center of the screen. Two of the bars corresponded to the values that the participants were asked to compare. The third bar corresponded to the values of a related category but was irrelevant for answering the question. It was added so that the participants would have to identify and select the relevant bars first (see Figure 4.2 C).

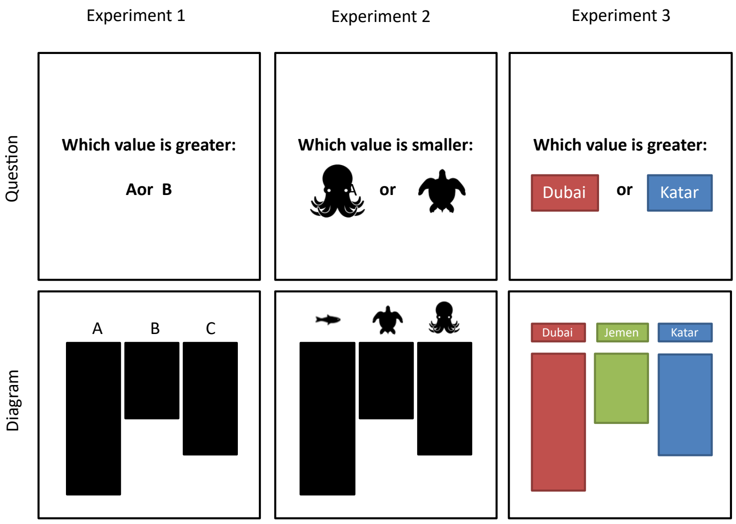

Figure 4.3: Schematic depiction of stimuli for the for experiments of study 1. Experiment 1 used letter and word labels. Experiment 2 employed word and pictograph labels. In Experiment 3, the coloring of the bars was manipulated. All bars had either the same or different colors.

All bars clearly differed in height, as all bars had a constant height and constant height differences, with a bar height pool of 3.1 cm, 7.1 cm, and 10.9 cm. To rule out potential influences from the relative bar positions, we presented all possible permutations of the bars in each condition to each participant. The bars that we asked participants to compare were chosen randomly. In one half of the trials, the participants were asked to choose the larger of the two values. In the remaining half of the trials, the participants had to choose the smaller of the two values (see Figure 4.2 C). We randomly generated a new set of trials for each participant.

We manipulated the presentation of the bar graph. First, we manipulated the orientation of the bar graph itself (0°, 90°, 180°, and 270°, clockwise with 0° being the standard upright bar graph view). Second, we manipulated whether the labels of the bars rotated with the bars (co-rotate) or whether they remained upright to the participant (reformatted). Third, we manipulated whether the labels of the bars consisted of words or letters. We sampled the word labels from a list of 20 topics containing three labels each (e.g., topic "male first names" with the labels "Anton", "Malte", and "Georg"). To control for word length, we included only words with two syllables consisting of a total of five letters. In the condition with letters as labels, the letters were randomly sampled from the following 17 letters: A, B, C, F, G, H, I, K L, O, R, S, T, U, W, X, Z. The remaining letters were excluded as their appearance is too similar to other letters when being rotated.

The participants responded by moving their finger to one of two response boxes below the bar graph, each of which corresponded to one of the values in question. The response boxes were positioned at 9.5 cm from the bottom and 20.2 cm from the left and right side of the screen and measured 2.4 cm in diameter. As the dependent variable, we captured the latencies of the initiation of a response by the participants (i.e., the time difference between onset of the bar graph and the start of the movement of the index finger towards the response box). For an illustration of the trial setup, see Figure 4.4.

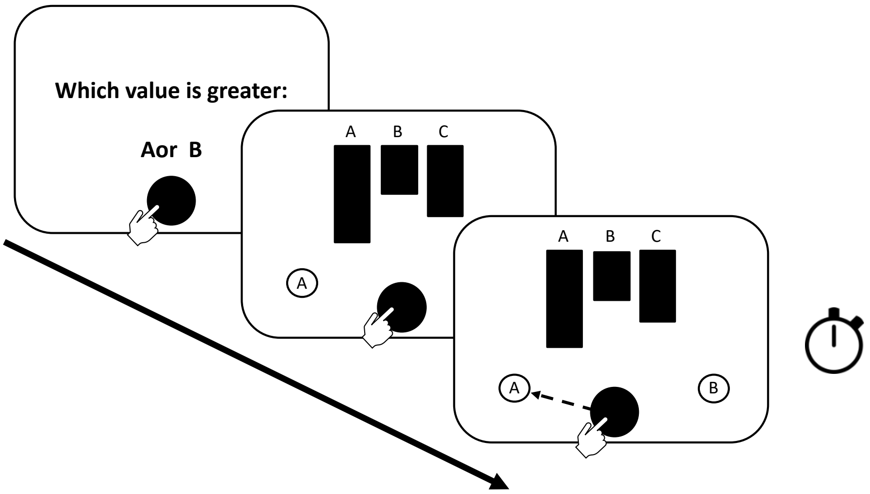

Figure 4.4: Experiment 1 flow diagram. For each trial, participants were first presented with a question. After putting their finger on the button, a diagram for answering the question appeared. Participants gave their response by sliding the button onto one of two options. The time to start the movement of the correct trials was taken as the dependent variable.

Design

All factors were manipulated in a within-subject design. Please note that in the 0° rotation condition both the reformatted and the co-rotated text looked identical. Therefore, this condition serves as a baseline. The response latencies of this baseline are subtracted from each other condition in order to isolate the effect of the rotation. The remaining combinations follow a 3 (angular rotation: 90°, 180°, 270°) x 2 (label rotation scheme: reformatted vs. co-rotated text) x 2 (label type: words vs. letters) within-subject design. Each factor combination was repeated 18 times (including three repetitions of each permutation of the bar heights), cumulating to 252 trials per participant (see Figure 4.2 A & B).

Analysis Plan

Our research questions focused on the relative difference between conditions with rotations of the bar graphs and the baseline condition without such rotations. In order to isolate these effects of the rotation, we subtracted the mean of the 0° rotation condition (individually for each participant). Our analyses of the data focused on two questions. First, we analyzed how label rotation scheme and label type affected performance across the remaining angular rotations. For this analysis, we conducted a repeated measures ANOVA on the differences in response latencies between the conditions with rotation and the baseline without rotation. Second, we analyzed whether the conditions with rotations differ from the baseline without rotation (i.e,. whether the difference scores differ from zero). We tested this with a series of t-tests. As analyses with multiple tests are prone to alpha error cumulation, we used a Bonferroni-corrected alpha level of $p = .00416$ for this series of 12 t-tests. Please note that this is a rather conservative correction which comes along with the risk of incorrectly classifying meaningful results as insignificant. The challenge for the current project was that insignificant deviations from the baseline might signal the successful prevention of prolonged response latencies; we therefore attempted to prevent that such a conclusion would only be based on insignificant tests. Beyond pure significance, we therefore also considered effect sizes as an indicator for the relevance of a particular result.

In accordance with Cohen (1969, p. 25), we considered effects with a size of $d_z < 0.2$ as negligible[1]. Therefore, for evaluating the reduction of costs associated with rotations, we only interpreted insignificant results with $d_z < 0.2$ as a successful reduction of rotation costs, whereas the results which were insignificant (after the correction) with $d_z > 0.2$ cannot be interpreted either way.

4.2.2 Results

Overall, accuracy was very high (M = 95.42%, SD = 3.1%). Accuracy levels for all experimental conditions ranged from 94.8% to 95.9%. These discrepancies were deemed negligible. For the analysis of response latencies, all trials with incorrect responses were removed from the data set. Further, we excluded trials for which we registered the motion onset of the index finger less than 250 ms after stimulus onset, as they likely reflect anticipations, involuntary movements, and measurement errors rather than regular task processing (6.46%) as well as trials of which the response latency deviated more than 3 SD from the personal mean (i.e., outliers, 1.32%).

Following the exclusions, we calculated the difference score in response latencies between the experimental conditions with rotations and the baseline without rotation (raw response latency data for all three experiments is available in Table 4.1). Finally, we aggregated this difference score for all conditions separately for each participant.

In order to examine differences between the conditions with rotations, we conducted a repeated measures ANOVA with angular rotation (90°, 180°, 270°), label rotation scheme (co-rotating, reformatted), and label type (word, letter) as the independent variables, as well as the difference score in response latency as the dependent variable. This analysis revealed a main effect of the label rotation scheme, $F(1, 32) = 16.19$, $p < .001$, $\eta_p^2 = .34$, indicating larger deviations of the response latencies from those of the baseline when the label orientation rotated with the bar graph rather than when it remained upright relative to

---

[1] Bootstrapped 95% confidence intervals (10,000 repetitions) for all effect sizes are available in Table 4.1.

the observer (see Figure 4.5). Further, we observed a main effect of the label type, $F(1, 32) = 49.21$, $p < .001$, $\eta_p^2 = .61$, indicating that words revealed larger deviations from the baseline than letters. There was no main effect of the three different angular rotations of the bar graph, $F(2, 64) = 0.11$, $p = .89$ and neither the two-way interactions, all Fs$(2, 64) < 1$, all ps $> .376$ and $F(1, 32) = 1.723$, $p = .198$, nor the three-way interaction, $F(2, 64) = 0.137$, $p = .872$, reached significance.

Furthermore, we compared the difference scores of all rotation conditions to zero with a series of t-tests (Bonferroni-corrected $p = .00416$). All conditions in which words served as labels showed significant differences to the baseline (all ts$(32) > 3.64$, all ps $< .001$, all d_zs > 0.634). For the conditions in which letters served as labels, none of the t-tests showed significant difference (after the corrections) to the baseline (all ts$(32) < 2.32$, all ps $> .02$). The effect sizes for letter label conditions varied: Reformatted letter labels showed negligible effects (all d_zs < 0.122), while co-rotated letter labels showed small effect sizes (all d_zs > 0.21 and < 0.41).

4.2.3 Discussion

The results of this experiment highlight that the written label provided a central challenge in accessing information from rotated bar graphs. Labels consisting of written words (rather than letters) that co-rotated with the bar graph (rather than being reformatted) revealed the largest mental normalization costs. The attribution of normalization costs to written labels is further supported by the results in the conditions with letters as labels. In particular, we did not observe substantial costs when the letters remained upright to the observers, signaling that it was not the rotation of the bars that induced the mental normalization costs. These results are particularly challenging, as written labels that co-rotate with the bar graph reflect the most authentic instance of mental normalization "in the wild", as reformatted labels cannot be realized for more than one viewer, and letters are hardly ever appropriate descriptions of depicted data. In the following experiments, we therefore implemented modifications of the labels and

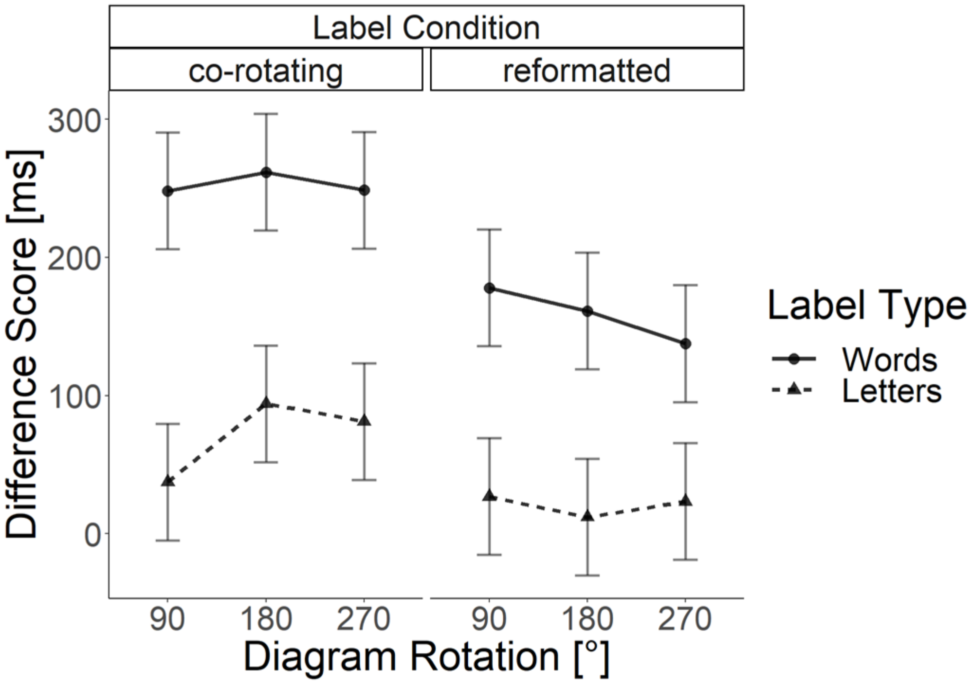

Figure 4.5: Results of Experiment 1. Mean differences of the response onset relative to the baseline without rotation. Error bars are based on Fisher's Least Significant Difference (R. Fisher, 1935).

the presentation of the bar graph that were intended to improve the accessibility of the depicted information during rotated views.

4.3 Experiment 2

As Experiment 1 revealed that word labels seem to be the Achilles' heel of mental normalization processes for information depicted in bar graphs, we explored the potential of pictorial labels for overcoming such normalization costs. We opted to try pictographs, as pictures seem to have a general superiority for information transportation. They are more accurately remembered than words (Cattaneo et al., 2007; Standing, 1973). Furthermore, response latencies are shorter for pictures than for words (Jenkins et al., 1967; Shor, 1971). In the context of information visualization, ISOTYPE-esque pictographs have been demonstrated to promote the understanding of bar graphs (Haroz et al., 2015). These studies suggest that pictures may act as good substitutes for word labels in data visualizations. In Experiment 2, we therefore explored whether the benefit of pictographs as labels for bar graphs also applies to mental normalizations. Although the costs of mental normalization for image stimuli have been demonstrated previously (Quaiser-Pohl, 2003), it remains possible that these costs are less pronounced for very simplistic pictographs (for an example of the pictograph label condition, see Figure 4.3). Also, it is unclear how potential mental normalization costs of pictograph labels relate to those of written labels. As in Experiment 1, we manipulated whether the label rotated with the bar graph or whether it maintained its orientation relative to the observer. If normalization costs arise from the rotation of the labels, the conditions with reformatted labels should be at the level of the unrotated baseline (i.e., a difference score around 0 ms). Finally, we again manipulated the angular disparity between the presented bar graph and its upright orientation. Although this manipulation had no effect in Experiment 1, we decided to maintain this manipulation for exploratory purposes as well as consistency across experiments.

4.3.1 Methods

Participants

Thirty-five new students (28 female) from the University of Tübingen (18–33 years), recruited via an online platform for volunteer participants in studies took part in this experiment. They received 5€ for approximately 40 minutes of their time.

Stimuli and Procedure

All stimuli and procedure were identical to the first experiment with the following exceptions: As label types, we compared written words with pictographs as labels (see Figure 4.6). We used the pictographs reported in the study by Haroz et al. (2015). The written labels were chosen to match the meanings in the pictographs.

Design

The experiment followed a 3 (angular rotation: 90°, 180°, 270°) x 2 (label rotation scheme: reformatted vs. co-rotated text) x 2 (label type: words vs. pictographs) within-subject design. As in the first experiment, we subtracted the baseline without any rotation from all other conditions in order to isolate the effect of display rotation.

4.3.2 Results

The overall subject accuracy was 94.37% (SD = 4.6%). Accuracy levels for the different factors fell between 93.65% and 95.16%. Trials with a motion onset of the index finger below 250 ms were excluded (8.7%) as were outliers (1.4%). The analysis plan was identical to Experiment 1.

In order to examine differences between the conditions with rotations, we conducted a repeated measures ANOVA with angular rotation (90°, 180°, 270°), label rotation scheme (co-rotating, reformatted), and label type (words, pictographs) as the independent variables, as well as difference scores in response latency as the dependent variable. Again, the rotation of the bar graph did not create a significant increase in the difference score, $F(2, 68) = 0.94$, $p = .39$. However, the label type showed a significant difference, $F(1, 34) = 22.42$, $p < .001$, $\eta_p^2 = .40$, indicating larger difference scores in response latencies for text than pictographs. Further, there was an effect of the label rotation, $F(1, 34) = 12.71$, $p = .001$, $\eta_p^2 = .272$, indicating larger difference scores with co-rotated than reformatted labels. No interaction reached significance with Fs$(2, 68) < 2.48$, ps $> .091$ and $F(1, 34) = 1.03$, $p = .317$. The results are depicted in Figure 4.6.

Furthermore, we tested whether the difference score in response latencies deviated from zero with a series of Bonferroni-corrected t-tests ($p = .00416$). In the word condition, the difference score for the word labels that maintained their upright orientation to the observer in the 90° and 270° diagram rotation condition did not reach significance with $t(34) = 1.825$, $p = .076$, $d_z = 0.31$ and $t(34) = 2.73$, $p = .01$, $d_z = 0.46$, respectively, whereas the remaining four conditions reached significance, all ts$(34) > 3.12$, ps $< .0037$, all d_zs > 0.527.

In the conditions with pictorial labels, none of the tests reached significance, all ts$(34) < 2.03$, all ps $> .05$. The stably displayed pictograph labels showed negligible effect sizes with all d_zs < 0.2, while the co-rotated pictograph labels showed small effect sizes ($0.28 <$ all d_zs < 0.34).

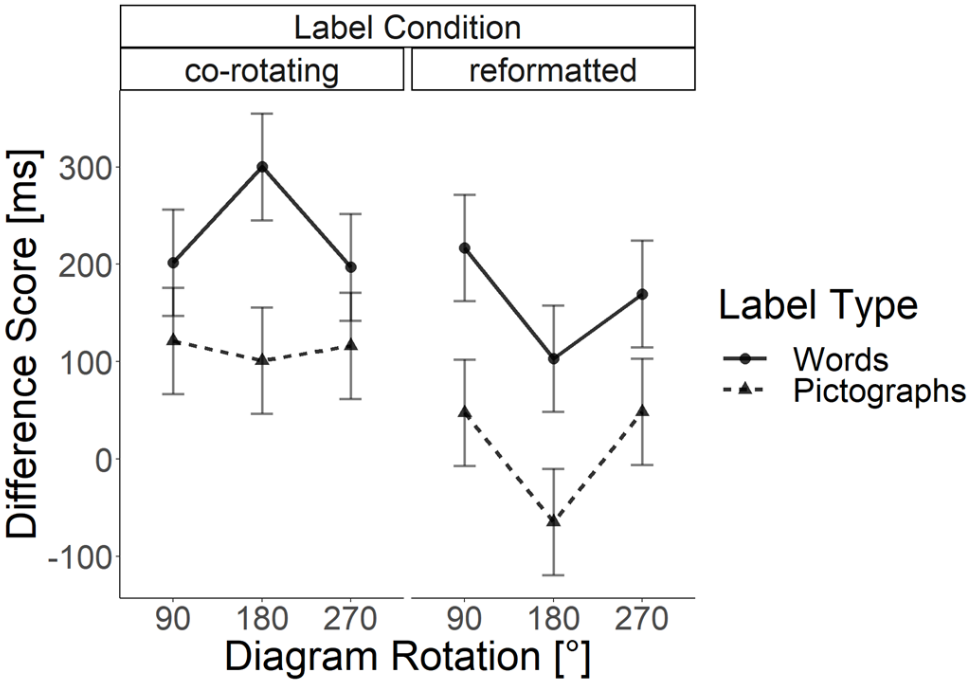

Figure 4.6: Results of Experiment 2: Mean differences in response onset relative to the baseline without rotation. Error bars indicate Fisher's Least Significant Difference.

4.3.3 Discussion

While the rotation of the diagram itself did not cause significant increases in response time relative to the unrotated condition, rotating the label did. Consistent with Experiment 1, there was a significant difference between reformatted and co-rotated labels, signaling that identifying the labels from rotated views was the weak spot in accessing depicted information from rotated bar graphs. With regard to the label type, there was a significant difference between word labels and the corresponding pictograph labels. Pictograph labels deviated less from the unrotated baseline. Furthermore, the comparison of each condition with the baseline (i.e., the difference score) showed a similar pattern of Cohen's d values as in Experiment 1. In particular for the pictographic labels, only the conditions with stably oriented labels to the observer revealed negligible effect sizes, whereas there remained substantial effects sizes for the conditions in which the pictographs rotated with the bars of the graph. As reformatted labels cannot be realized for more than one viewer at a time, using pictographs as labels is therefore not sufficient to (fully) compensate for the effects of mental normalization in bar graphs.

4.4 Experiment 3

In this experiment, we tested the potential of colored text labels for reducing the detrimental effects of graph and label rotation. As color is considered to be processed efficiently (i.e., in parallel for the entire display) (Treisman, 1986), coloring labels might not impose an additional load on working memory but may support a faster mental normalization instead. An example for the color intervention stimuli is given in Figure 4.3, Experiment 3.

4.4.1 Methods

Participants

Forty-one students (33 female) from the University of Tübingen, aged 18-27 and recruited via an online platform for voluntary student participants, took part in this experiment and were paid an allowance of 8€ for 40 minutes of their time.

Stimuli and Procedure

Stimuli and procedure were identical to Experiment 1 with the following exceptions: The label type factor consisted of a "multicolor" and a "monocolor" condition. In the multicolor condition, each bar was filled with a different color and the associated word labels were highlighted in the same color, in the question as well as in the diagram (see Figure 4.6). The colors were selected randomly in order to avoid systematic influences from semantic associations between the selected color and the corresponding label (S. Lin et al., 2013). In the monocolor condition, all bars were filled with the same color and their labels where not highlighted.

With regard to the monocolor condition, we expected to replicate Experiment 1: Diagrams with co-rotating labels should show longer response latencies than diagrams with reformatted labels (and all conditions should differ from the baseline). Further, we expected the multicolor condition to (at least partially) reduce the costs of mental normalization. Furthermore, we explored how this expected reduction of mental normalization costs relates to the baseline without rotations.

4.4.2 Results

The overall subject accuracy was 94.9% (SD = 3.1%). Accuracy levels of the factors were between 93.43% and 95.49%. Trials with a motion onset of the index finger below 250 ms were excluded (9.5%) as were outliers (1.5%).

Analogous to the previous experiments, we conducted a repeated measures ANOVA with three factors: angular rotation (90°, 180°, 270°), label rotation scheme (co-rotating, reformatted), and coloring type (monocolor, multicolor). The difference score in response latency was again used as the dependent variable. In contrast to Experiments 1 and 2, the label-rotation did not reach significance in this experiment, $F(1, 40) = 2.91$, $p = .096$. However, the coloring type yielded significant differences, $F(1, 40) = 57.05$, $p < .001$, $\eta_p^2 = .588$, indicating a reduced deviation from the baseline for multicolor than monocolor labels. Finally, the difference score in response latencies increased with angular deviation in this experiment, $F(2, 80) = 10.11$, $p < .001$, $\eta_p^2 = .202$. None of the two- or three- way interactions reached significance, Fs$(2, 80) < 2.69$, all ps $> .07$ and $F(1, 40) = 1.09$, $p = .302$. The results are depicted in Figure 4.7.

To further explore these results, we tested the difference score in response latencies against zero with a series of Bonferroni-corrected t-tests ($p = .00416$). The difference scores for response latencies reached significance in all monocolor conditions, all ts$(40) > 4.925$, all ps $< .0001$, all d_zs > 0.76. However, the difference score did not reached significance in any of the multicolor conditions, all ts$(40) < 1.5$, all ps $> .029$. All effect sizes but one were negligible with all d_zs < 0.2. Only the multicolor condition with 270° diagram rotations and co-rotated word labels showed a small effect size with $d_z = 0.237$.

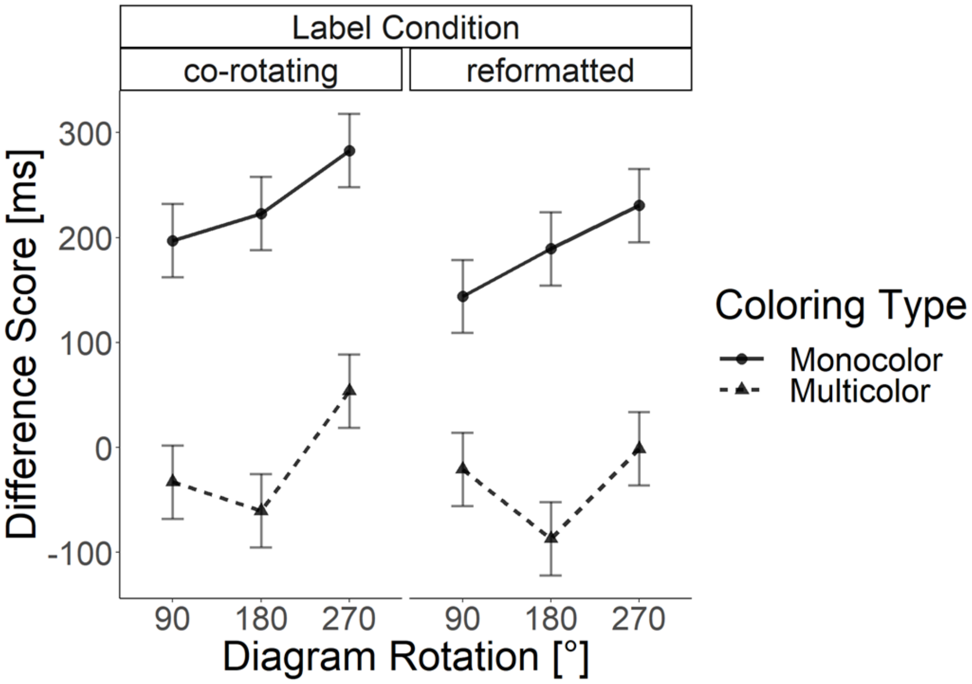

Figure 4.7: Results of Experiment 3: Mean differences of the response onset relative to the baseline without rotation. Error bars are based on Fisher's Least Significant Difference.

4.4.3 Discussion

Contrary to the previous experiments, the rotation of the label had no significant effect on response latencies in this experiment. Numerically, it seems as if there was a small effect on the label rotation scheme in the monocolor condition but clearly absent in the multicolor condition. However, as there was no significant label type x label rotation scheme interaction, such a conclusion cannot be drawn directly from the data of this experiment. What also deviates from the preceding experiments, is that there was an effect of the angular rotation of the graph which we did not observe in the previous experiments. A (rather speculative) explanation for this is that the coloring reduced the contrast between the labels and the background what might have reduced the readability of the labels. In return, this might have induced additional normalization costs which might increase with the angle of rotation. Nevertheless, what is consistent with the preceding experiments (numerically as well as statistically) in this experiment is that in the multicolor conditions, the difference score in response latencies hardly increased with rotations of both graphs and labels (in terms of negligible effect sizes). Multicolored bars and labels therefore appear to be a promising approach in reducing the costs of mental normalizations in bar graphs.

4.5 General Discussion

In this study, we investigated how non-upright viewing conditions impact the extraction and processing of information from bar graphs. We followed a use-inspired research approach by converting a practical problem (i.e., that only one user could have upright viewing conditions on shared graphs in collaborative settings) into a laboratory research paradigm.

A first major result of our study is that bar graphs in non-upright viewing conditions indeed pose an additional burden on cognitively processing the depicted information. In all three experiments, the standard condition (i.e., word labels) yielded significant normalization costs (i.e., the difference in response latency relative to the baseline). With regard to our practical collaboration setting outlined in the introduction (see Figure 4.1), this finding confirms that in most cases a collaborator with non-upright viewing conditions would have to invest additional cognitive resources in order to access the same information as the other collaborator with upright viewing conditions.

A second major result of our study is identifying the rotation of the labels as the "Achilles' heel" in terms of normalization costs for compensating for non-upright viewing conditions. This conclusion stems from the observation that normalization costs were remarkably reduced (in most cases to negligible effect sizes) when the labels maintained their upright orientation relative to the observer.

Whereas these conditions with reformatted labels are helpful in identifying the origin of the largest proportion of the normalization costs, it appears unlikely that this finding could be part of an intervention aiming to reduce the normalization costs. The reason for this is because maintaining the upright orientation of the labels can be realized for multiple collaborators only by introducing additional, redundant labels with different orientations so that each observer has his/her own label with an upright orientation. Such an approach, however, would be challenging technically as well as psychologically. From the technical perspective, it would be necessary to track the number as well as the location of collaborators with non-upright viewing conditions and add the necessary labels at their corresponding orientation. While this technical challenge can be met (e.g., Dietz and Leigh, 2001; Ramakers et al., 2012 for multi-touch tables), the psychological challenge is more difficult. Psychologically, the inflation of labels at different orientations would provide an additional burden on the perceptual and cognitive processing of the depicted information. Given the known capacity limitations for attentional processing (e.g., Meyerhoff et al., 2017), it seems unlikely that human observers can deal with the inflation of labels. Consequently, all collaborators would have to ignore the additional labels while focusing only on those labels presented upright from their current point of view in order to reduce the information load. However, an extensive body of research has shown that irrelevant stimuli in the visual field cannot always be ignored even when processing them is harmful for performance in the task at hand (B. A. Eriksen & Eriksen, 1974; C. W. Eriksen & Schultz, 1979; Kopp et al., 1994).

As an alternative approach to reducing the costs of mental normalization in bar graphs, we therefore investigated two candidate modifications of the labels and/or bars. In Experiment 2, we investigated pictographs as labels for which we observed a substantial reduction in normalization costs. However, in particular when the pictographs rotated with the bar graph, the effect sizes were – at least for the critical "co-rotation" condition - not in the negligible range (although not significant due to the Bonferroni-corrections). Please note that we changed only the label to its corresponding pictograph, not the bars themselves, as has also been done in previous work (Haroz et al., 2015). It therefore remains possible that changing the bars themselves into pictographs would further diminish normalization costs to the negligible range. However, it appears more likely that pictographs would persist in showing small normalization costs even under these circumstances, as pictures clearly have canonical viewpoints facilitating their identification (Palmer et al., 1981). As picturizing labels is restricted to contents that can be turned into intuitive pictographs, we instead investigated the coloring of bars and labels in the remaining third experiment. In line with the intensive research demonstrating the efficiency of color processing in terms of visual search (Wolfe & Horowitz, 2004, 2017), perceptual grouping (Palmer et al., 2003), as well as attentional efficiency (Carter, 1982; Christ, 1975; Duncan & Humphreys, 1989), we also observed that this coloring manipulation reduced mental normalization costs to the negligible range of effect sizes. In sum, both pictographs and coloring reduce normalization costs, but coloring appears to be slightly more efficient.

Although our experiments were not designed to disentangle different processes involved in the processing of rotated displays, there are two findings in our results which we would like to elaborate. First, in contrast to the many previous studies exploring normalization costs for rotated stimuli (for an overview see Khooshabeh et al., 2013), the angular deviation from upright viewing conditions played only a minor role in our results. Although the rotated display was generally harmful in all three experiments, normalization costs only increased with angular deviation in Experiment 3. The most plausible reason for this reduced relevance of angular rotation is that we studied orthogonal or opposite viewpoints which are known to be less prone to normalization costs than angular deviations between these orientations (Hintzman et al., 1981; Montello, 1991). In the present project, however, we focused on the orthogonal and opposite viewpoints because these seem to be the most common perspectives in the collaborative group settings with multi-touch tables (e.g., Bause et al., 2018; Higgins et al., 2011, 2012; Rogers et al., 2009), which inspired our project.

Second, our results revealed less pronounced normalization costs when the labels remained in their upright orientation to the observer. This observation is consistent with a piecemeal mental normalization strategy rather than a holistic mental normalization strategy. The reason for this is because a holistic mental normalization would result in the opposite effect. Such a holistic normalization would turn the labels away from an upright orientation, thus increasing response latencies. Overall, this finding is consistent with the literature on classical mental rotation which has observed holistic strategies mostly for shape information (Cooper, 1975; Cooper & Podgorny, 1976; Cooper & Shepard, 1973) but not for stimuli that required the integration of multiple visual feature (Hochberg & Gellman, 1977; Xu & Franconeri, 2015).

Given the complexity of a graph (relative to simple shapes), it appears likely that multiple processes contribute to the overall performance as well as the normalization costs. One candidate mechanism that might speed up the processing of the depicted information is feature-based visual search (e.g., Treisman, 1986; Wolfe and Horowitz, 2017). In particular, our final experiment showed that color-coding reduced normalization costs toward the negligible range. Combining a selection of relevant information based on the color with a piecemeal rotation strategy might allow participants to mentally normalize only relevant information thus reducing the overall costs. Future research is necessary to disentangle such combined processes of mental normalization. Such a line of research would probably have to combine behavioral and psychophysiological methods in order to isolate the individual components of rather complex mental normalizations (Gauthier et al., 2002). Another promising venue would be to investigate individual differences, as many mental normalization tasks are known to be subject to individual differences arising from spatial abilities (Collins & Kimura, 1997; Khooshabeh et al., 2013; Peters et al., 2006; Tapley & Bryden, 1977; Tarampi et al., 2016), general intelligence (Varriale et al., 2018), or expertise with the stimulus materials (Stieff, 2007). In the present project, we neglected these differences (we drew the sample from our rather homogenous group of students), as we focused on the basic effects that most likely arise across all participants (although they might be differently pronounced). For the same reason, we neglected the differences for sex (Burnett, 1986; Jones & Anuza, 1982; Semrud-Clikeman et al., 2012) and age (Hertzog & Rypma, 1991), which have been reported for many tasks.

4.5.1 Limitations and Outlook

To the best of our knowledge our project is the first to address the impact of mental normalization as it would appear in collaborate settings dealing with the extraction of depicted information from graphs. Given the novelty of this research approach, there of course are limitations which should be further explored in future research. A first limitation addresses the coloring intervention. Although coloring clearly reduced normalization costs, it seems likely that color coding works only for a limited number of depicted elements. With an increasing number of elements in a graph, the colors will, by nature, become more similar in the color space, making it more difficult to distinguish the colors efficiently (Cahill & Carter, 1976). Thus, there will certainly be a threshold after which color coding of graph elements will not be helpful anymore. Furthermore, coloring the bars uniquely gets more challenging when the bar graph depicts more than one data dimension. As such bar graphs typically uses color or shading information to group bars, adding (further) color cues to link the labels with the bars might be confusing. Thus, future research should attempt to replicate the usefulness of the color intervention with more complex data structures. Relatedly, it would be useful to extend the general research approach to other types of data graphs in order to ensure that the normalization costs which we observed in the present experiments as well as the success of the interventions generalize beyond bar graphs.

A further interesting venue for prospective research would be to extend our methodology to generally more complex problems. In the present study, we intended to isolate normalization costs and effects of interventions, and therefore used rather simplistic tasks. As our tasks could be solved at a single glance (we intended to get high accuracy in order to study response latencies), it would be worth to study whether the observed effects would also emerge in experiments studying accuracy as dependent variable, or whether they would scale up in displays which contain more complex data structures or combinations of multiple graphs (see Moritz et al., 2020).

Finally, our project has established normalization costs as well as interventions to reduce these costs on the level of an individual observer. As our research question has been inspired by a collaborative setting in which two (or more) observers access the same depicted data from different viewpoints, an urging follow-up research question would be how the observed normalization costs as well as the interventions alter subsequent collaborative processes such as communication (Lyons, 2009) or joint problem solving (Bause et al., 2018).

4.6 Tables

Absolute latencies for response onsets by condition				*test statistics for differences to baseline*		
intervention	diagram rotation	label condition	mean (sd)	t (df)	p	d_z [95% CI]
Experiment 1						
letter	baseline		1574 (283)			
letter	90	co-rotate	1693 (344)	1.210 (32)	.235	0.21 [-0.12, 0.63]
letter	180	co-rotate	1750 (420)	2.322 (32)	.027	0.40 [0.08, 0.78]
letter	270	co-rotate	1737 (380)	2.152 (32)	.039	0.37 [0.06, 0.71]
letter	90	reformatted	1683 (408)	0.702 (32)	.487	0.12 [-0.25, 0.45]
letter	180	reformatted	1668 (417)	0.360 (32)	.721	0.06 [-0.31, 0.40]
letter	270	reformatted	1679 (389)	0.700 (32)	.489	0.12 [-0.24, 0.43]
word	baseline		1736 (371)			
word	90	co-rotate	1904 (408)	6.883 (32)	<.001	1.20 [0.85, 1.75]
word	180	co-rotate	1917 (433)	5.334 (32)	<.001	0.93 [0.63, 1.36]
word	270	co-rotate	1904 (414)	6.535 (32)	<.001	1.14 [0.76, 1.67]
word	90	reformatted	1834 (390)	4.705 (32)	<.001	0.82 [0.54, 1.19]
word	180	reformatted	1817 (390)	4.533 (32)	<.001	0.79 [0.46, 1.23]
word	270	reformatted	1793 (358)	3.643 (32)	.001	0.63 [0.32, 1.03]
Experiment 2						
pictograph	baseline		1496 (425)			
pictograph	90	co-rotate	1703 (581)	1.943 (34)	.060	0.33 [-0.02, 1.00]
pictograph	180	co-rotate	1682 (539)	2.029 (34)	.050	0.34 [0.00, 0.90]
pictograph	270	co-rotate	1698 (570)	1.705 (34)	.097	0.29 [-0.05, 0.89]
pictograph	90	reformatted	1629 (458)	1.169 (34)	.250	0.20 [-0.12, 0.66]
pictograph	180	reformatted	1517 (444)	-1.124 (34)	.269	-0.19 [-0.43, 0.16]
pictograph	270	reformatted	1630 (517)	1.036 (34)	.308	0.18 [-0.15, 0.57]
word	baseline		1636 (465)			
word	90	co-rotate	1783 (507)	4.189 (34)	<.001	0.71 [0.31, 1.38]
word	180	co-rotate	1882 (592)	6.007 (34)	<.001	1.02 [0.64, 1.63]
word	270	co-rotate	1778 (559)	3.120 (34)	.004	0.53 [0.10, 1.58]
word	90	reformatted	1798 (577)	4.433 (34)	<.001	0.75 [0.50, 1.13]
word	180	reformatted	1684 (484)	1.826 (34)	.077	0.31 [-0.04, 1.11]
word	270	reformatted	1751 (555)	2.726 (34)	.010	0.46 [0.07, 1.32]
Experiment 3						
multicolor	baseline		1306 (356)			
multicolor	90	co-rotate	1385 (448)	-0.928 (40)	.359	0.14 [-0.48, 0.16]
multicolor	180	co-rotate	1357 (462)	-1.607 (40)	.116	-0.25 [-0.57, 0.05]
multicolor	270	co-rotate	1472 (470)	1.520 (40)	.136	0.24 [-0.07, 0.60]
multicolor	90	reformatted	1397 (469)	-0.549 (40)	.586	0.09 [-0.40, 0.22]
multicolor	180	reformatted	1331 (399)	-2.252 (40)	.030	-0.35 [-0.70,-0.05]
multicolor	270	reformatted	1416 (450)	-0.035 (40)	.972	0.01 [-0.32, 0.31]
monocolor	baseline		1526 (469)			
monocolor	90	co-rotate	1615 (526)	4.955 (40)	<.001	0.77 [0.45, 1.19]
monocolor	180	co-rotate	1641 (487)	6.924 (40)	<.001	1.08 [0.81, 1.47]
monocolor	270	co-rotate	1701 (551)	7.065 (40)	<.001	1.10 [0.79, 1.56]
monocolor	90	reformatted	1562 (439)	4.925 (40)	<.001	0.77 [0.45, 1.18]
monocolor	180	reformatted	1607 (499)	5.403 (40)	<.001	0.84 [0.55, 1.24]
monocolor	270	reformatted	1648 (466)	6.915 (40)	<.001	1.08 [0.82, 1.45]

alpha-level bonferroni-corrected to .00416

Table 4.1: Statistical details for the experiments in Chapter 4.

4.7 Declarations

4.7.1 Ethics Approval and Consent to Participate

The experimental procedure was approved by the institutional review board of the Leibniz-Institut für Wissensmedien (LEK 2017/049). All participants signed informed consent prior to testing.

4.7.2 Consent for publication

All depicted persons in Figure 4.1 agreed with the publication of the image.

4.7.3 Availability of Data and Material

Study materials, data, and analyses are available at the Open Science Framework (https://doi.org/10.17605/OSF.IO/YMC35).

4.7.4 Competing Interests

The authors declared they have no conflicts of interest with respect to their authorship and the publication of this article.

4.7.5 Funding

Funding by the German Federal Ministry of Education and Research (BMBF) 2016-2019 under grant number 01IO1616.

Declaration regarding § 5 Abs. 2 No. 8 of the PhD regulations of the Faculty of Science – Share in collaborative publications/ manuscripts

The following Chapter (Chapter 5) consists of a study that was designed together with Hauke S. Meyerhoff and Friedrich W. Hesse. The proportional contributions to this study are presented in the subsequent table.

Author	Author Position	Scientific Ideas %	Data generation %	Analysis & interpretation %	Paper writing %
Tjark S. Müller	First	70%	100%	75%	60%
Friedrich W. Hesse	Second	10%		5%	10%
Hauke S. Meyerhoff	Last	20%		20%	30%
Title of paper:	The Justice Effect: An Interaction Between Rotation and Contrast				
Status in publication process:	submitted				

Author's contributions

MÜLLER, TJARK wrote most of this paper, designed and conducted the experiments, and conducted the data analyses
MEYERHOFF, HAUKE S. wrote parts of the paper and provided valuable feedback and supervision on the experiment design, data analysis, and writing.
HESSE, FRIEDRICH W. received the grant for these experiments and provided valuable feedback and supervision.

Acknowledgements

We want to thank Dr. Nora Wickelmaier for statistical counseling.

Chapter 5

The Justice Effect: An Interaction Between Rotation and Contrast

5.1 Introduction

Imagine you are one of the two collaborators depicted in Figure 5.1 A. Your task is to discuss the implications of the data depicted on the multi-touch Table between you and your collaborator. Which of the two collaborators would you like to be? Based on the layout of the depicted bar graph, you would probably prefer being the collaborator at the right edge of the image. In this case, the bar graph would be upright to you, making it easy to identify the depicted information. For your collaborator, however, the task is much more difficult, as all labels are upside down from his point of view. These suboptimal viewing conditions will slow down your collaborator and most likely also affect your collaboration. As previous research has traced back the detrimental effect of non-upright viewing conditions in such collaborative settings to the inverted labels (Müller et al., 2021), a rather straightforward attempt to solve this problem would be to include redundant upright and inverted labels on the graph (see Figure 5.1 B). By doing so, both collaborators in our example would be able to access the label in an upright manner. In this report, we investigate how such additional labels affect the time that is needed to extract the meaning of the label for both viewing conditions.

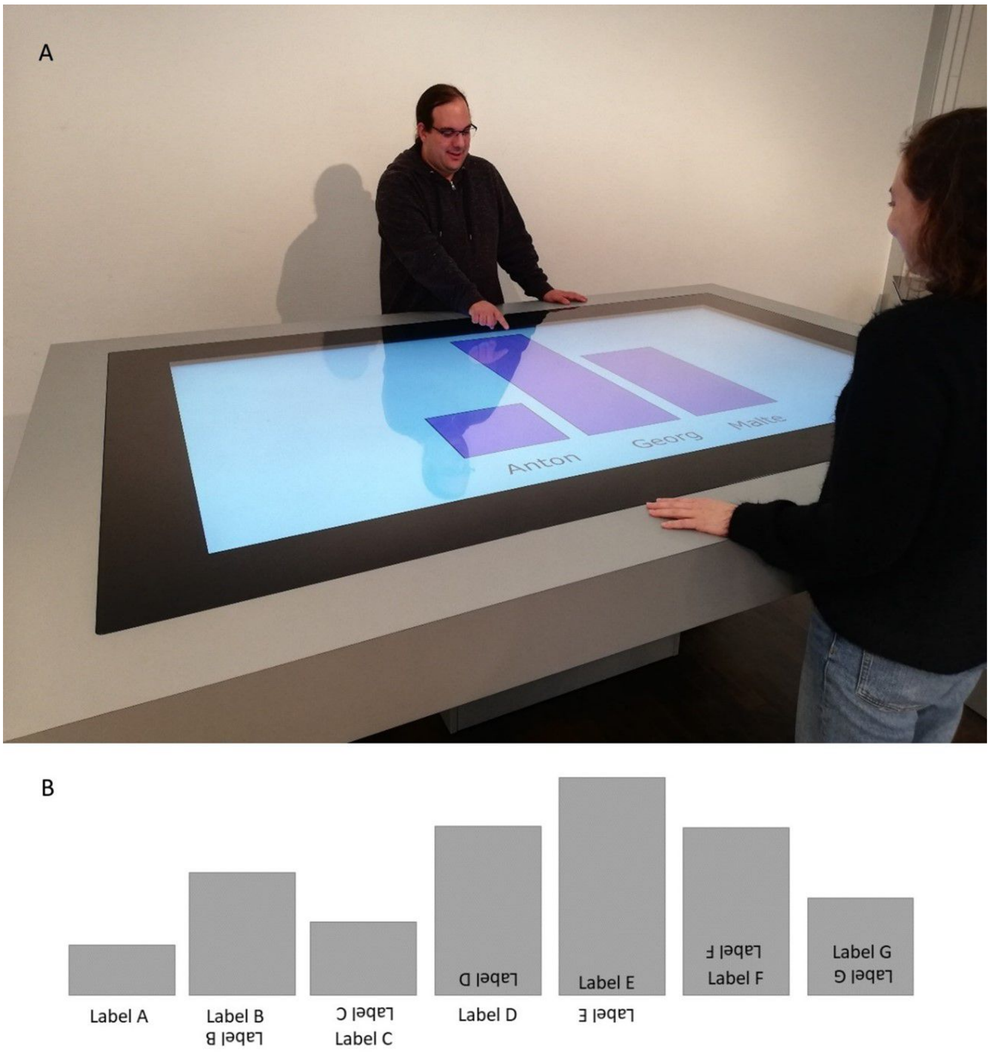

Figure 5.1: A: Illustration of a collaborative setting in which one of the collaborators perceives the depicted bar graph from an upright viewpoint, whereas the other collaborator sees the bar graph upside down. B: Illustration of the different variants of labels which might improve the accessibility of the depicted information from upright as well as non-upright viewing conditions.

5.1.1 Accessing Non-Upright Labels

Numerous reports have repeatedly demonstrated that accessing information with deviations from non-upright viewing conditions (Diwadkar & McNamara, 1997; Palmer et al., 1981; Tarr, 1995) as well as deviations in the viewpoint between two views (Bundesen & Larsen, 1975; Jolicoeur & Besner, 1987; Shepard & Metzler, 1971) is more effortful and/or error-prone than accessing information without such deviations. Regarding to parts of the graphical depictions of data, such detrimental effects of deviations in the viewpoint have been reported for 2D shapes (Cooper, 1975), letters (Cooper & Shepard, 1973), words (Koriat & Norman, 1985; Risko et al., 2014), or maps (Montello, 2010).

In their seminal work, Shepard and Metzler (1971) observed a linear relationship between the response latency in identifying objects across rotations and the angular deviation between the two views. From these results, they inferred a mental rotation process which would be a cognitive equivalent to a physical rotation. However, similar patterns in response latencies have also been observed in broader conceptualizations of object recognition between views (Besner, 1983; Bundesen & Larsen, 1975; Ellis et al., 1989; Jolicoeur, 1987; Jolicoeur & Besner, 1987). For instance, Biederman (1987) proposed an edge detection mechanism, Marr suggested a 3D model storage (Marr, 2010; Marr & Nishihara, 1978), or Tarr proposed the simultaneous storage of multiple points of view (Lawson et al., 1994; Tarr & Bülthoff, 1998; Tarr & Pinker, 1989). As it is almost impossible to distinguish these different processes with purely behavioral data (Gauthier et al., 2002), we use the theoretically agnostic term mental normalization as an umbrella for all processes contributing to overcoming deviations in viewpoints. One striking observation in mental normalization is that only parts of an object or a graphical representation might be normalized at a time (i.e. piecemeal vs. holistic processing, Just and Carpenter, 1985; Xu and Franconeri, 2015; Yuille and Steiger, 1982), although grouping mechanisms can improve performance (Meyerhoff et al., 2021).

Regarding more applied stimuli such as information depicted in data visualizations, different parts of the stimulus might also be more susceptible to normalization costs than others. For instance, Müller et al. (2021) recently studied such a scenario for bar graphs. They asked participants to solve simple tasks based on the information depicted in a standard bar graph. They observed increasing response latencies when the orientation of the bar graph deviated from an upright viewpoint. Most importantly, this research has identified the labels of the bar graphs to be the Achilles' heel of accessing information in bar graphs from non-upright viewing conditions. When Müller et al. rotated the bar graph but reformatted the labels so that they maintained the upright orientation to the observers, the mental normalization costs were rather negligible. The latter finding suggests that the largest proportion of the mental normalization costs in the study of Müller et al. (2021) emerged from the need to access the inverted labels. In the present project, we aimed at reducing these mental normalization costs by introducing a redundant label rotated by 180°. This additional label allowed the collaborator on the top edge of the multi-touch table in Figure 5.1 A to read this label upright, thus circumventing this critical proportion of normalization costs.

5.1.2 Location a Second Label

As illustrated in Figure 5.1 B, there are several options for the implementation of a redundant, inverted label into a bar graph. A first consideration addresses the question of which of the two labels should appear in the upright versus inverted orientation. On the one hand, it seems reasonable to plot the label closest to the corresponding collaborator in the upright orientation (Labels C, D, & F in Figure 5.1 B). In this case, the lower label would be upright, whereas the upper label would be inverted. On the other hand, it also might be reasonable to place the upright label above the inverted label (Labels B, E, & G in Figure 5.1 B), as subjects from western cultures tend to start reading the topmost text (Chokron et al., 2009).

A second considerations addresses the location of both labels which has an immediate effect on the contrast between the labels and the background. First, both labels could be placed below their corresponding bar (Labels B & C in Figure 5.1 B). In this case, both labels would have a high contrast. Second, one label could be placed below the bar, whereas the other could be placed within the bar (Labels D & E in Figure 5.1 B). In this case, the label below the bar would have a high contrast, whereas the other would have a lower contrast. Third and finally, both labels could be placed within the bar (Labels F & G in Figure 5.1 B). In this case, both labels would have a low contrast. This is relevant as the contrast between label and background could affect the accessibility of the label. In the article series "Studies of Typographical Factors Influencing Speed of Reading", Tinker and Paterson extensively studied various combinations of ink color and paper color. They observed the fastest reading speed in the condition with the highest contrast (i.e., black ink on white paper; Tinker and Paterson, 1931). This confirmed earlier findings of Babbage (1832) and also matched later reports investigating patients with visual impairments (Sloan, 1969, 1977). Further research points to a negative influence of lower contrast, as it may increase reading time (Howell & Kraft, 1959); however, subsequent research pointed out that this detrimental impact might be negligible unless for very small or very large letters (Legge et al., 1987) or only for very low contrast (van Nes & Jacobs, 1981). The polarity of the contrast, however, seems not to have a systematic influence on the readability of textual stimuli (Legge, Pelli, et al., 1985; Rubin & Legge, 1989), but there exists also conflicting evidence (Paterson & Tinker, 1931b).

Given that the contrast between text and a background might influence reading performance under at least some circumstances, we considered it relevant to test the impact of contrast with regard to our redundant labels. Beyond affecting readability, the contrast of the labels might also influence how the presence of the redundant label affects the processing of the upright label or how both labels might interfere with each other.

5.1.3 Potential Effects of Redundant Labels

Although the redundant labels are designed so that each collaborator needs to read only the label which is upright from his/her viewing position, it seems likely that both labels also influence each other when considering research on the effect of flanker stimuli. This consideration emerges from the seminal work of Eriksen and Eriksen (B. A. Eriksen & Eriksen, 1974), who set up the first version of what today is called the *Eriksen Flanker paradigm*. In this paradigm, participants respond to a letter stimulus while ignoring flanking letters which could either elicit neutral, congruent, or incongruent responses with the response to the target letter. Effects of such flanking stimuli have been observed with various kinds of stimuli including arrows (Kopp et al., 1994), colored tiles (Rafal et al., 1996), and numbers (Lindgren et al., 1996).

In the original study of Eriksen and Eriksen (1974), the presence of all flanking stimuli interacted with responding to the target stimuli when target and flanking stimuli were spatially close, but incongruent flankers were more harmful for performance than neutral flankers. Congruent flankers produced even better performance and were capable of accelerating response times in comparison to neutral flankers. The redundant labels in our study are similar to congruent flankers, as they refer to the same information (despite that the rotation reduces their surface similarity). As the redundant label reflects a rather applied variant of the flanker task (we are not aware of a similar preceding study), we decided to compare the effect of all variants of redundant labels with a standard condition of just one label (upright as well as inverted). To highlight the structural differences between our redundant labels and the classical flankers, we call them *pseudo-flankers*.

5.1.4 The Present Project

In the present research project, we followed a use-inspired basic research approach. To do so, we translated the applied question of the effects of redundant labels in graphical representations of data such as bar graphs into a basic research paradigm that allows for full experimental control. We investigated how fast observers can extract the meaning out of a stimulus that mimics a unique or a redundant label of our applied scenario. Inspired by the examples depicted in Figure 5.1 B, we investigated word stimuli that appeared either in isolation or were accompanied by an inverted version of the same word either at the same or different contrasts. Considering the practical scenario of the two collaborators at the multi-touch table, our first research question was whether the additional label would improve the performance of the collaborator seeing the display upside-down. In our experimental study, such a beneficial effect would appear, as faster response times for word stimuli with pseudo-flankers than single inverted words. We expected such an improvement in identification performance given that the second label allowed for circumventing mental normalization costs. As a second research question, we intended to study how the addition of a second label would affect the identification performance of the collaborator who sees the depicted information in an upright manner. In our study, this effect can be investigated by comparing response times for single upright labels with the response times of redundant labels. The prediction here was less clear. It seems unlikely that the redundant label improves the identification performance; however, both the absence of an effect (if participants can successfully process only the upright word) as well as a detrimental effect of the redundant label (i.e., additional noise in the display) on response times seemed equally likely.

5.2 Experiment 1

In the first experiment, we investigated whether the presentation of redundant labels would alter response times relative to rotated or unrotated labels. In our design, we included the three factors (pseudo-flanker, contrast, rotation) that we considered to possibly influence response times. For our analysis, we investigated the factors using ANOVAs as well as the individual factor combination using a multiple regression model.

5.2.1 Methods

Participants

The final sample consisted of 158 students (102 females, 55 males, 1 diverse; 18 – 35 years, M = 23.9 years, SD = 3.3 years) of the University of Tübingen. The experiment was conducted online. The participants were recruited from a mailing list for voluntary study participants. The experimental procedure was ethically approved by the institutional review board of the Leibniz-Institut für Wissensmedien, Tübingen. All participants provided informed consent prior to testing and confirmed that their data could be used for scientific purposes at the end of the study. Data from participants who did not complete the experiment were excluded from the analyses. Additionally, we excluded the data of 17 participants, as their average response times deviated more than 1.5 times the interquartile range from the median of all participants. In return for their participation, the participants had the opportunity to enter a lottery of 8 x 20€ vouchers for a big online marketplace at the end of the experiment.

The sample size resulted from the following considerations/constraints. A power-analysis with G*Power 3.1 (Faul et al., 2007, 2009) revealed a suggested sample size of 91 participants for ANOVAs with within-factors having a small effect size ($f = 0.1$) and standard power of 0.8 for 8 measurements. We chose to power our study for such a rather small effect size, as we are not aware of any study which would provide sufficient evidence for a larger effect size. In order to achieve the corresponding rather large number of participants, we recruited our participants for an online experiment using an open mailing list at the University of Tübingen. Due to an unexpected resonance, this procedure resulted in an overshot of the intended sample size.

Apparatus

The participants used their personal computers to complete the experiment online. The responses were entered with the corresponding keyboard. The experiment was coded in lab.js (Henninger et al., 2019) which runs on all common browsers and delivers reliable response times (Bridges et al., 2020).

Stimuli and Procedure

The task of the participants was to answer simple categorization questions. At the beginning of each trial, a new question asking for the membership to a particular category was presented centrally on the screen for 2,000 ms (e.g., "Is this an animal?", or "Is this a country?", font size 32 pt). Immediately after this question, the target word (font size 32 pt) for which the participants had to answer the question appeared onscreen and remained onscreen until a response was recorded. The spatial location of the target word randomly varied on an invisible circle

(300 pixel in diameter) around the center of the screen. The participants were instructed to press the "y"-key of their keyboard for confirming or the "n"-key for rejecting that the target word was part of the questioned category. Thereafter, the participants received feedback regarding their accuracy. If the response was correct, a screen filling green frame appeared for 100 ms. However, if the response was incorrect, the German word "Falsch!" (engl. wrong) with a surrounding red frame appeared for 3000 ms. The increased presentation time of the feedback following incorrect answers was intended to encourage accurate task performance. As the dependent variable, we measured response times (i.e., the latency between onset of the target word and the registration of a response; see Figure 5.2 for a depiction of the experimental flow).

We manipulated the appearance of the target word along three dimensions (see Figure 5.3). The target word could appear in high contrast (web color value #000000) or low contrast (web color value #DDDDDD). Further, the target word could appear upright or upside down (i.e., rotated by 180°). Finally, the target word could be accompanied by a matching but rotated flanker (i.e., the same word rotated by 180° presented below the target word). As this flanker was both more complex and less controlled than in usual flanker task experiments, we named this factor "pseudo-flanker" to provide a more concise description. Taken together, the experiment followed a 2 (contrast; high vs. low) x 2 (orientation; upright, upside-down) x 2 (pseudo-flanker; present vs absent) within-subject design. Each factor combination was repeated 20 times, resulting in a total of 160 experimental trials. Prior to the test trials, each participant completed three practice trials which were accompanied by a rehearsal of the central elements of the instructions (task, response keys, etc.).

5.2.2 Results

As intended, response accuracy was high (M = 94.01%, SD = 3.93%). For the subsequent analysis of the response times, we only included the trials which were answered correctly. Further, we excluded trials for which the response time deviated more than ±3 SDs for the corresponding participant and condition (1.1% of all trials).

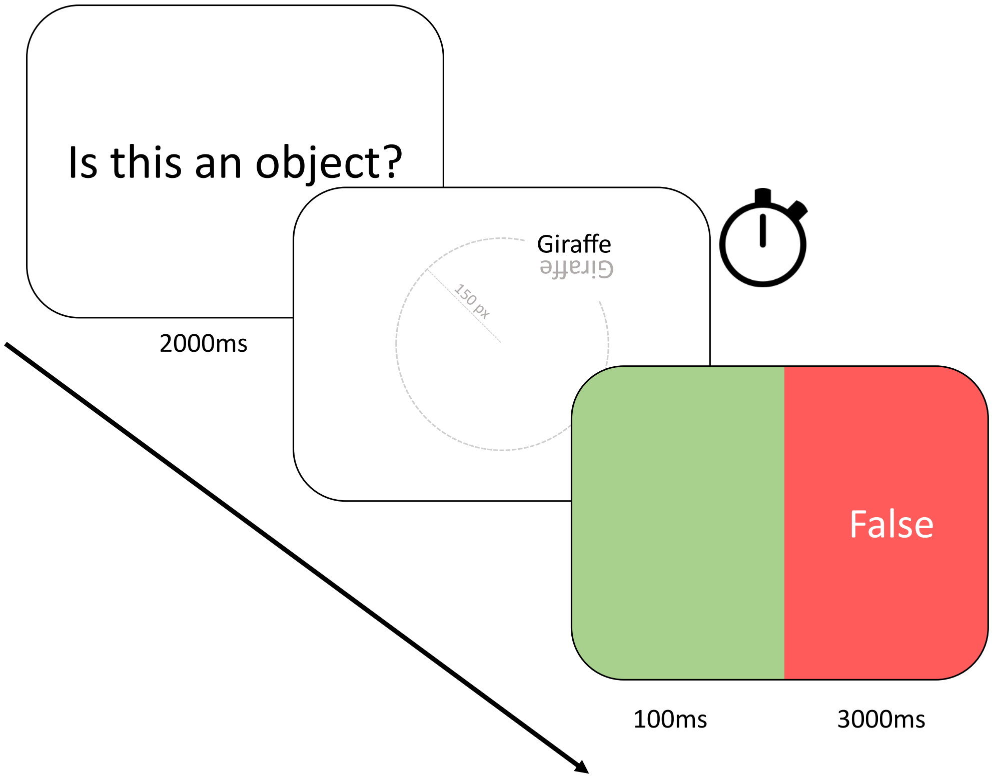

Figure 5.2: Illustration of the trial sequence in Experiment 1. The participants received a categorization question followed by a target word for which they completed the categorization. The target word appeared randomly on an invisible circle around the center of the screen. Following the response, the participants received accuracy feedback.

We conducted a repeated measures ANOVA with the three factors (pseudo-flanker, rotation, and contrast) as independent variables and with response time as dependent variable. The ANOVA revealed a three-way interaction between stimulus rotation, contrast, and the presence of pseudo-flankers, $F(1, 157) = 11.04$, $p = .001$, $\eta_p^2 = .07$. Additionally, we observed a two-way interaction between contrast and the presence of pseudo-flankers, $F(1, 157) = 43.06$, $p < .001$, $\eta_p^2 = .22$, as well as a two-way interaction between rotation and the presence of pseudo-flankers, $F(1, 157) = 234.68$, $p < .001$, $\eta_p^2 = .60$. The remaining two-way interaction between contrast and rotation was not significant, $F(1, 157) = 0.21$, $p = .649$. There were, however, significant main effects of the factors rotation, $F(1, 157) = 413.33$, $p < .001$, $\eta_p^2 = .72$, contrast, $F(1, 157) = 114.53$, $p < .001$,

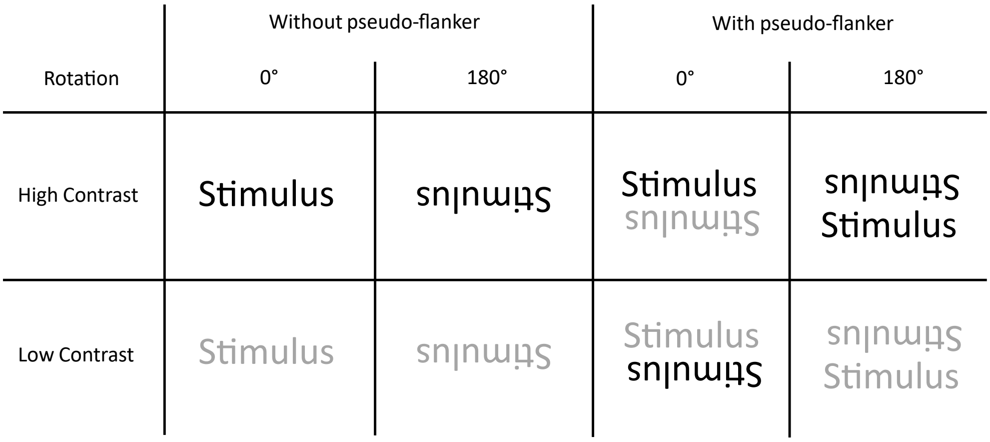

Figure 5.3: Illustration of the target word presentation in Experiment 1. The appearance of the target words varied with regard to contrast (high, low), orientation (upright, upside-down), and whether it was accompanied by a matching rotated pseudo-flanker (with, without).

$\eta_p^2 = .42$, and the presence of pseudo-flankers, $F(1, 157) = 140.02$, $p < .001$, $\eta_p^2 = .47$ (see Table 5.1 for an overview).

A visual inspection of the depicted means in Figure 5.4 suggests that the three-way interaction is driven by the rotation influencing the two-way interaction between contrast and the presence of pseudo-flankers. To confirm this impression, we ran two reduced ANOVAs, one for the data of unrotated trials and one for the rotated trials. In both ANOVAs, the two-way interaction between contrast and the presence of pseudo-flankers was significant, but it appears that the rotation increased the impact of the contrast manipulation in trials without pseudo-flankers, $F(1, 157) = 36.84$, $p < .001$, $\eta_p^2 = .19$, whereas it reduced the impact of the contrast manipulation in trials with pseudo-flankers, $F(1, 157) = 7.53$, $p = 0.007$, $\eta_p^2 = .05$.

With regard to the interpretation of the data, however, the most relevant finding is the strong interaction between the factors rotation and presence of pseudo-flankers. When pseudo-flankers are absent, stimuli in both the low and high contrast conditions are susceptible to the effect of rotations (low contrast: $t(157) = -17.34$, $p < .001$, $d = -1.37$; high contrast: $t(157) = -17.23$, $p < .001$, $d = -1.30$). When pseudo-flankers are present, however, the influence of rotation appears to be absent or starkly reduced (low contrast: $t(157) = -1.03$, $p = .30$; high contrast: $t(157) = 4.59$, $p < .001$, $d = -0.24$).

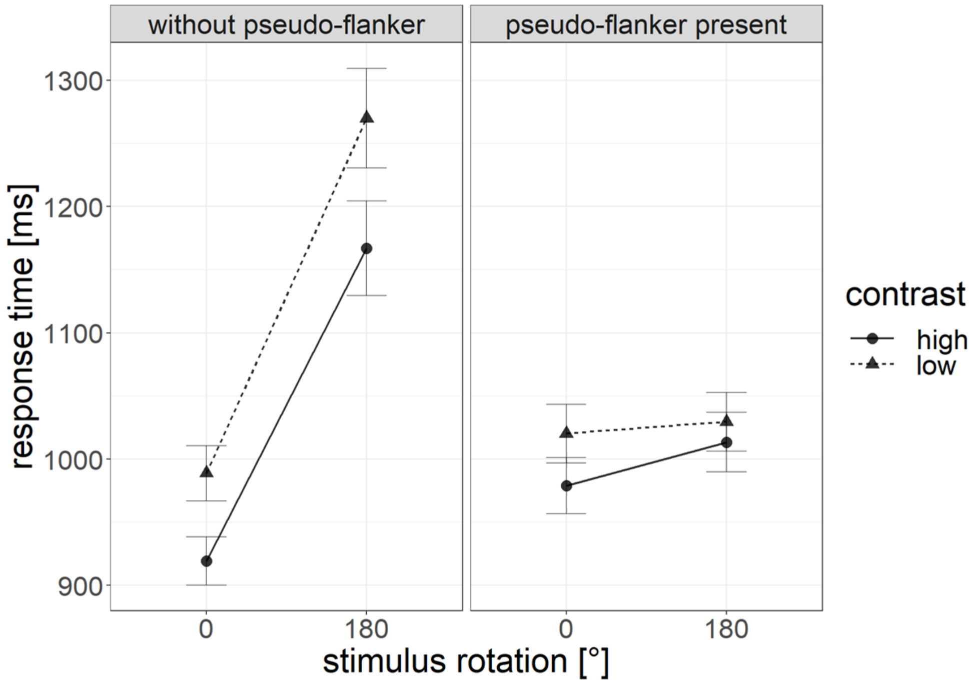

Figure 5.4: Results of Experiment 1 presented as factorial design with the three independent factors: contrast (high vs. low), orientation (upright, upside-down), and pseudo-flanker (present vs absent). Error bars indicate 95% confidence intervals.

Regarding the practical implications of our findings, it is of relevance how the response times for the individual factor combinations relate to those conditions that reflect the standard views in our collaborative scenario. For the collaborator with the upright view, this would be the single upright label (i.e., no pseudo-flanker, 0° rotation, high contrast), whereas it would be the single inverted label (i.e., no pseudo-flanker, 180° rotation, high contrast) for the collaborator facing the bar graph upside down. To explore this question, we have replotted the factor combinations in the order of the elicited response latencies (see Figure 5.5). As visible in this plot, the single upright label elicited the fastest responses, whereas the condition mimicking the inverted label elicited almost the slowest responses (with only the low contrast variant being even slower). To probe this pattern statistically, we conducted two mixed linear regressions. The regressions included individual intercept and slope parameters for each participant. Both regressions were conducted with response time as the dependent variable, and the type of stimulus as the independent variable. The only difference between both regressions was the definition of the reference category.

The first linear regression used the condition mimicking the inverted label as the reference category. This regression confirmed that response times in all conditions differed from the response time in the condition without a pseudo-flanker, 180° rotation, and high contrast, all $|z|$s > 9.881[1], all ps $< .001$. In other words, all relevant conditions elicited faster response latencies than the condition with mimicking the inverted label. The second linear regression used the condition mimicking the upright label as a reference category. This regression confirmed that all other conditions elicited longer response latencies than the condition without pseudo-flankers, 0° rotation, and high contrast, zs > 7.017, all ps $< .001$ (see Table 5.2 for full statistical details on both regressions).

5.2.3 Discussion

The main analysis showed large effects for all main factors (contrast, rotation, and presence of a distractor). The most relevant effect, however, appears to be the two-way interaction between the rotation and presence of a pseudo-flanker, hinting toward a beneficial effect of a second, rotated label. This pattern of results is in accordance with our hypothesis that a second label would reduce response time (as it probably reduces the mental normalization costs for inverted labels). The interaction between the contrast manipulation and the presence of pseudo-flankers is also of relevance. The presence of pseudo-flankers seemed to compensate for the lower readability in the low contrast condition. This potentially could be the result of the presence of more high contrast displays on trials with pseudo-flankers. For instance, it might indicate a preference for reading high contrast labels even if this implies the necessity of a mental normalization. Due to the design of the experiment, the interpretation of this effect is hard and might be contaminated

---

[1] Due to high degrees of freedom (> 151) and problems in interpretations of degrees of freedom in linear mixed models (Baayen et al., 2008), we decided to report z-values instead of t-values.

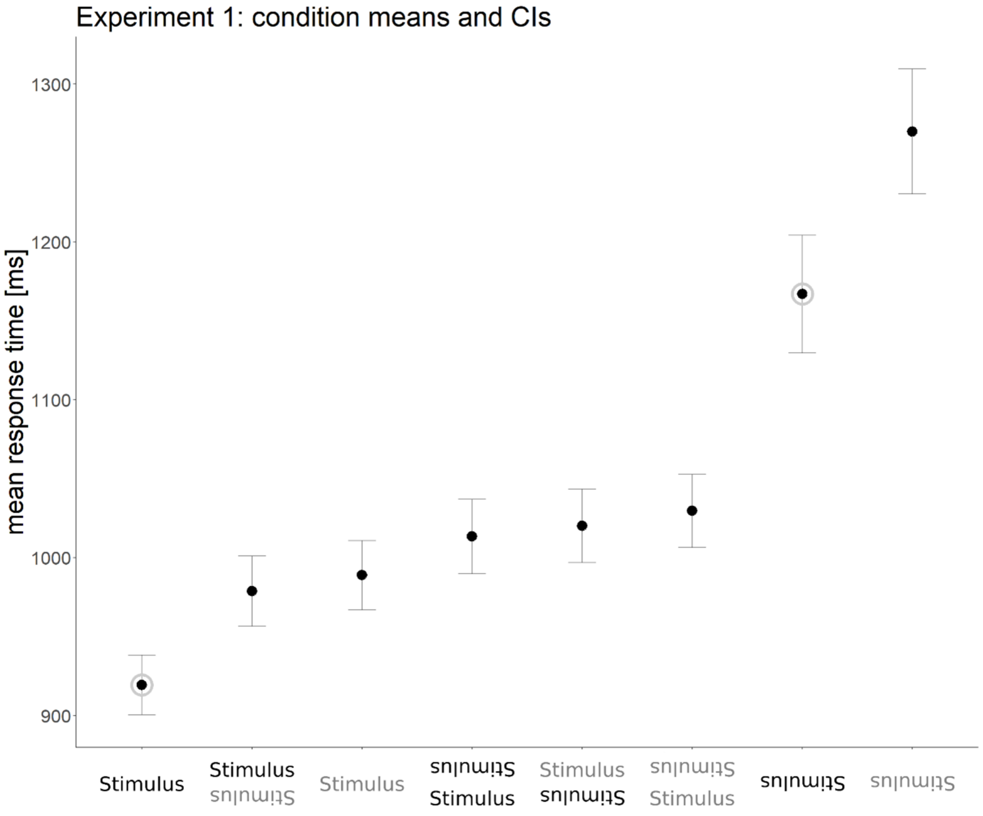

Figure 5.5: Mean response time by condition of Experiment 1. Gray cycles indicate the baselines for both linear regressions. The error bars indicate 95% confidence intervals.

(see Experiment 2 for further considerations). The interaction between contrast and rotation was inversely pronounced with and without pseudo-flankers, which likely also caused the significant three-way interaction. For the interpretation of our results, however, this three-way interaction seems not to be of relevance so that we will not discuss it any further.

Regarding the use-inspired research question whether it would be sensible to include a rotated second label in bar graphs to improve the accessibility of the depicted data from multiple viewpoints, our results show that this decision requires situation-dependent considerations. For observers with upright viewing conditions, the additional label is harmful in terms of processing time (i.e., in the unrotated label with high-contrast stimulus condition, the pseudo-flankers acted as distractors). For an observer with rotated viewing conditions, however, the additional label is beneficial (i.e., when compared with the 180°-rotated, single-display high contrast condition, the pseudo-flankers acted as facilitators). This observation suggests that in the case of two viewers, the additional label could establish a compromise equalizing the processing time for the label for both viewers (this is what we refer to as the justice effect). Please note, however, that Experiment 1 did not include all possible combinations of high- and low-contrast labels due to the factorial design. We therefore explored all possible combinations in Experiment 2.

5.3 Experiment 2

The results of Experiment 1 showed that an additional label improves performance for the viewer with the rotated viewpoint but decreases performance for the viewer with upright viewing conditions. It therefore might be sensible to implement a second, inverted label when two viewers attend to the same bar graph from opposing viewpoints. With regard to the exact design of such a redundant label, however, Experiment 1 left some of the possible combinations of high and low-contrast labels untested. Most importantly, we did not explore the option that a low contrast pseudo-flanker could appear upside down above an upright high-contrast target word. This seems to be of interest, however, because if the upper label would be presented within the bar, it would appear at a lower contrast, and in this scenario the observers might be inclined to read the label closer to their own viewing location. We therefore tested all potential combinations of high and low contrast labels in Experiment 2.

5.3.1 Methods

Participants

The final sample consisted of 155 students (110 females, 44 males, 1 diverse; 18 – 34 years, M = 23.2 years, SD = 4.0 years) of the University of Tübingen. We excluded the data of an additional 10 participants, as their average response time deviated more than 1.5 times the interquartile range from the median of all participants. As Experiment 1, this experiment was conducted online. Following the protocol of Experiment 1, we again used the open mailing list at the University of Tübingen to recruit our participants. The participants had the opportunity to enter a lottery of 8 x 20€ vouchers for a big online marketplace in return for their participation. This procedure again resulted in an over-recruitment of participants.

Apparatus, Stimuli, and Procedure

All apparatus, stimuli, and procedure were identical to those of Experiment 1 with the following exceptions. We implemented ten different variants of the target word(s) mimicking the labels; in contrast to Experiment 1, they were not arranged along experimental factors but chosen for practical considerations.

The conditions with isolated upright (0°) or rotated (180°) target words were the same as in Experiment 1. In the conditions in which the target words were accompanied by pseudo-flankers, we realized combinations of high/low, low/high, and high/high contrast. For these contrast combinations, we also realized both possible orientations (i.e., the upright word above or below). However, we dropped the low/low contrast condition from this experiment, as it was unlikely to be reasonably applied to real-world scenarios and therefore was deemed uninformative (for all tested combinations, see Figure 5.6).

5.3.2 Results

The overall response accuracy was high (M = 94.42%, SD = 3.01%). For the subsequent analysis of the response latencies, we only included the trials which were answered correctly. From the remaining data, we excluded the trials for which the response latency deviated more than ±3 SDs for the corresponding participant and condition (1.0% of all trials).

For the analysis of the response times, we followed the approach for the practical implications of Experiment 1. For this approach, it is relevant how the individual factor combinations relate to those conditions that reflect the standard views in our collaborative scenario. As in Experiment 1, we conducted

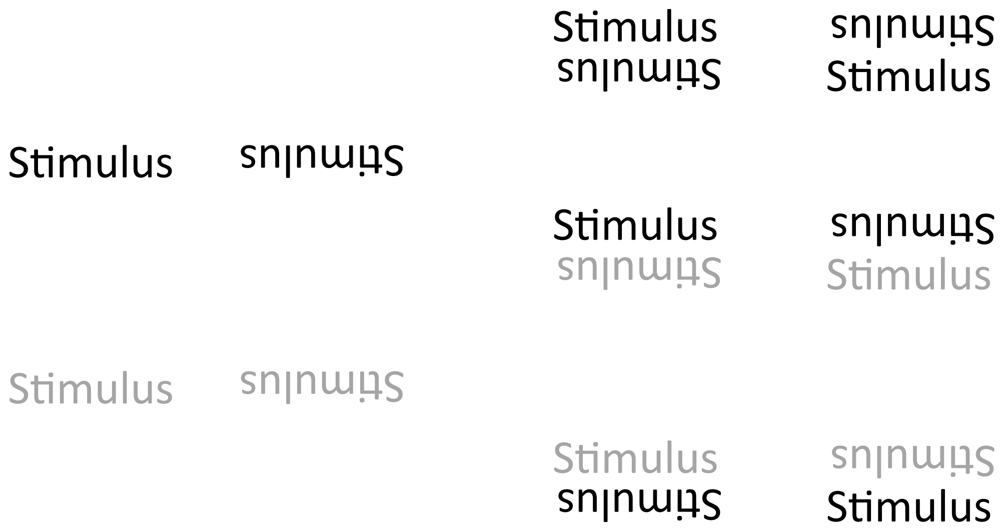

Figure 5.6: Illustration of the target word presentation in Experiment 2. The appearance of the target words varied regarding contrast (high, low), orientation (upright, upside-down), and whether it was accompanied by a matching rotated pseudo-flanker (with, without). In conditions, with pseudo-flanker, we also manipulated whether the upper or lower word was upside down or inverted. Note that his experiment does not follow a factorial design but merely extends the findings of Experiment 1 to additional conditions which are of practical relevance.

two mixed linear regressions with individual intercept and slope parameters for each participant. Again, the only difference between both regressions was the definition of the reference category (i.e., single upright vs. single rotated label; see Figure 5.7).

The first linear regression used the condition mimicking the inverted label as the reference category. This regression confirmed that response times in all but one condition differed from the response latency in the condition without a pseudo-flanker, 180° rotation, and high contrast, all $|z|$s > 5.71, all ps $< .001$. The only exception was the condition without a pseudo-flanker, 180°-rotated, and low-contrast, $z = 1.78$, $p = .077$ (see Table 5.2 for full statistical details). Replicating Experiment 1, this shows that all relevant conditions (i.e., those with a pseudo-flanker) elicited faster response times than the condition mimicking the inverted label (see Table 5.2 for full statistical details).

The second linear regression used the condition mimicking the upright label as the reference category. This regression confirmed that all but one condition elicited longer response times than the condition without a pseudo-flanker, 0° rotation, and high contrast, all zs > 4.70, all ps $< .001$. The only exception was the condition in which an inverted low-contrast pseudo-flanker appeared below the upright high-contrast target word, $z = 0.75$, $p = .454$ (see Table 5.3 for full statistical details).

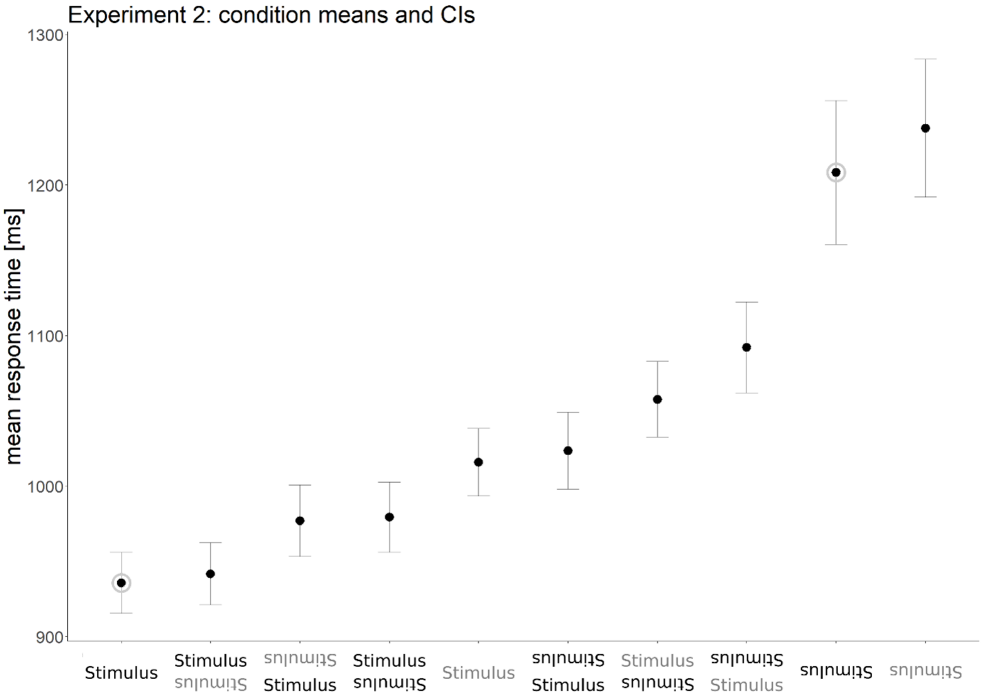

Figure 5.7: Mean response times by condition of Experiment 2. Gray cycles indicate the baselines for both linear regressions. The error bars indicate 95% confidence intervals.

The Justice Effect

One striking observation from Experiment 1 was that the additional label could potentially reduce the performance differences between accessing the bar graph in an upright versus. inverted manner. We labelled this the justice effect as the reduction in performance differences emerged from faster responses for inverted viewing conditions that were accompanied by slower responses for upright viewing conditions. To further explore this justice effect, we conducted an additional analysis of the relevant conditions of this experiment. In particular, our experiment involved four pairs of matching label conditions, allowing a direct investigation of the justice effect[2]. These pairs of matching conditions are those that constitute rotations of each other (inward rotated with one label in high contrast and the other in low contrast, outward rotated with one label in high and low contrast, the single label in high contrast, and the single label in low contrast; see the labels of Figure 5.8). For these pairs of conditions, we analyzed the difference scores in response times, as they simulate upright versus. inverted viewing conditions.

We conducted a repeated measures ANOVA on the difference scores in response time between the matching conditions. The ANOVA was significant, $F(3, 462) = 38.57$, $p < .001$, $\eta_p^2 = .20$ (see Table 5.4 for detailed results). Post-hoc pairwise comparisons showed significant differences between both double display differences and both single display differences, all ts > 5.65, all ps $< .001$, a significant difference between the two single display differences, $t(154) = 3.24$, $p = .001$, $d = 0.20$, but no significant difference between both conditions with redundant labels, $t(154) = 0.07$, $p = .942$, $d = 0.01$ (see full results in Table 5.5).

5.3.3 Discussion

There are two key insights that emerge from this experiment. First, we were able to replicate the principal findings of Experiment 1, as the overall pattern of the results suggests that the additional label increased performance relative to inverted viewing conditions but was harmful for performance relative to upright viewing conditions. This suggests that the design of labels for bar graphs in collaborative scenarios needs to adhere to situational considerations, as a second label could have beneficial as well as detrimental consequences.

Second, this experiment provides evidence that the pseudo-flankers were the least harmful when the upright display was presented in high contrast. Regarding practical applications, this finding suggests that in cases in which a second label is

---

[2] The two conditions with both labels in high contrast would be self-similar in 180° rotation, so we could not compute any differences between them. Therefore, they were excluded from the subsequent analysis.

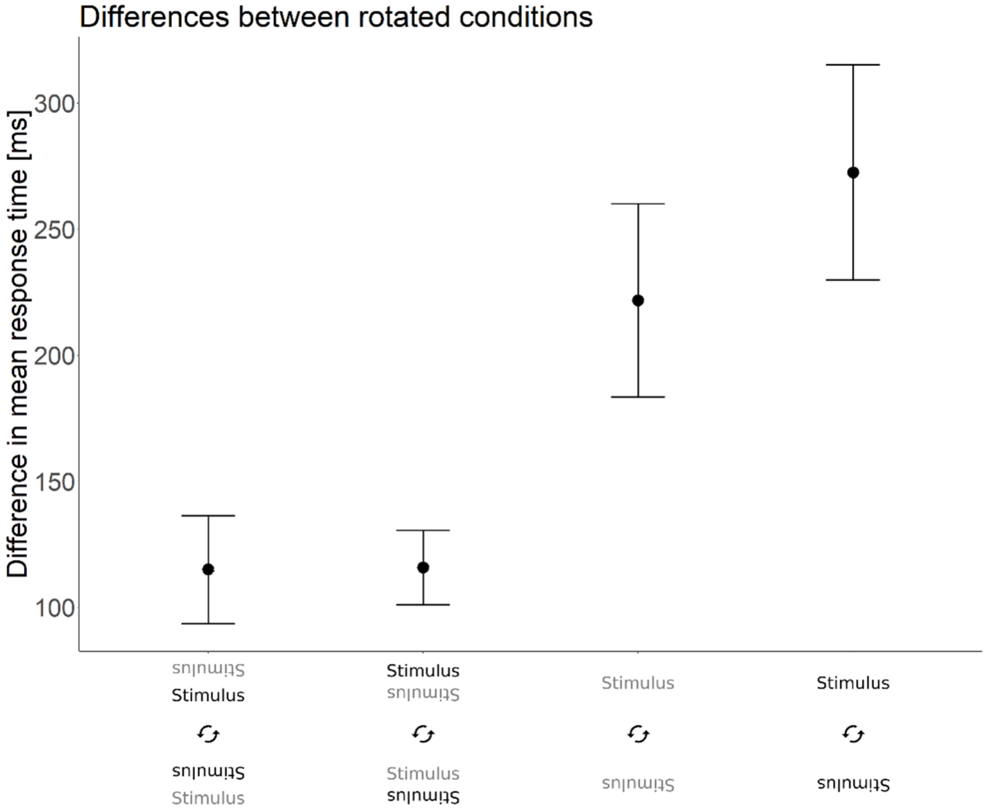

Figure 5.8: Mean difference scores for response times between the rotationally symmetrical conditions. Error bars indicate 95%-confidence intervals.

considered to be of benefit, presenting the unrotated one in high contrast would be a sensible design choice. Remarkably, we did not observed evidence for any differences between "inward" and "outward" redundant labels. This suggests that there is no strong preference for accessing the upper or the lower part of a redundant label.

The additional analyzes of the difference scores in the response times of those conditions that would be rotational symmetric further highlighted the advantages of using redundant labels. The differences between the simulated upright and inverted views were significantly smaller than for single display stimuli. With regard to the collaborative setting that inspired the present research project, this finding suggests that the redundant label would reduce the performance differences when two participants approach the depicted graph with upright versus. inverted viewing conditions.

5.4 Experiment 3

In the first two experiments, we used words as stimuli which mimicked the corresponding label conditions. Consistently across both experiments, we observed that a redundant label at an inverted orientation could potentially equalize the processing time of two people who assess the depicted information from opposing viewpoints such as illustrated in Figure 5.1 A. In this final experiment, we aimed at testing whether (and to what extend) these findings transfer to knowledge work with real bar graphs. We therefore implemented redundant labels similar to those of Experiments 1 and 2 into bar graphs that the participants accessed in an upright or inverted manner. Based on the depicted information, the participants answered simple frequency questions to make sure that they actually had to process the labels as well as the magnitude of the corresponding bar. Based on the promising effects in Experiments 1 and 2, we expected to observe a similar pattern within this transfer experiment; that is, we expected that the additional label would be beneficial relative to a traditional inverted view on the bar graph but likely harmful relative to a traditional upright view on the same graph.

5.4.1 Methods

Participants

The final sample consisted of 139 students (102 females, 36 males, 1 diverse; 18 – 34 years, M = 23.2 years, SD = 3.6 years) of the University of Tübingen. The recruitment of the participants was identical to Experiments 1 and 2 and again resulted in an over-recruitment of participants. We excluded the data of an additional 14 participants, as their average response time deviated more than 1.5 times the interquartile range from the median of all participants. Participants gave their informed consent at the beginning of the experiment. Incomplete data (including participants who did not confirm that their data could be analyzed for scientific purposes after the experiment were discarded). In return for their participation, the participants who completed the experiment had the opportunity to enter a lottery for 8 x 20€ vouchers for a big online marketplace.

Apparatus

As both preceding experiments, this experiment was conducted as an online experiment coded with lab.js (Henninger et al., 2019).

Stimuli and Procedure

In each trial, the participants attended to a bar graph (584 x 438 pixels) consisting of three bars of different magnitude (101 px, 202 px, 302 px). Based on the depicted information, the participants answered simple comparison questions that involved two of the bars (e.g., "which value is bigger: leopard or hummingbird?"; "Which value is smaller: Antenna or Obelisk?"). At the beginning of each trial, the question was presented for 2500 ms. Then, the bar graph appeared and remained onscreen until a response was recorded. The participants indicated their responses with the y- and n-keys (y-keys on German keyboards are at the location of the z-key on American keyboards). Feedback was provided identically to Experiments 1 and 2 (see Figure 5.9 for an illustration of the trials). Within each graph, the labels of the bars emerged from a common theme (e.g., "country" or "animal"). Each label (13 px; 17 px with descenders) consisted of three syllables / seven letters (except for one out of the 58 different labels which accidentally consisted of eight letters).

Following Experiments 1 and 2, we manipulated the design of the labels. Because we investigated real bar graphs in this experiment, we rotated the bar graph by 180° in one half of the trials (rather than mimicking different rotations with variations of isolated labels). The labels could be placed either inside the bars (low contrast) or outside of the bars (high contrast). Beyond the contrast, we also manipulated whether the labels were rotated inwards or outwards. (see Figure 5.10 for an overview displaying all conditions).

Each permutation of the three bars was displayed three times, with each possible comparison appearing once. The direction of the comparison was deter-

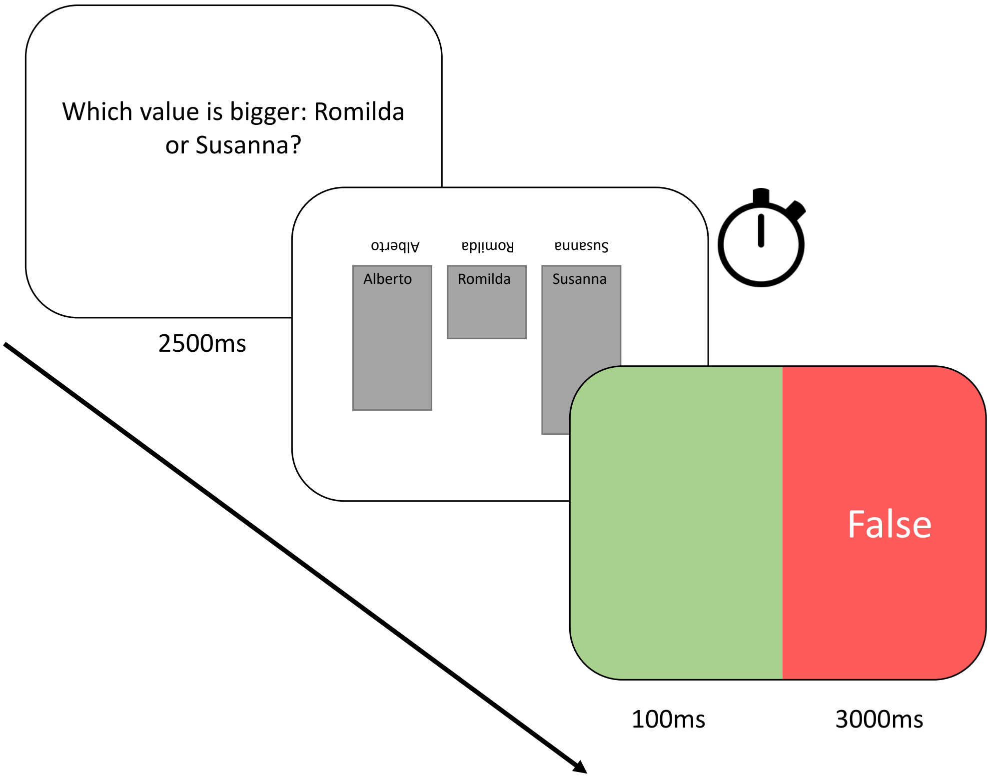

Figure 5.9: Illustration of the trial structure of Experiment 3. First a question appears onscreen, which the participants answer based upon a subsequently presented bar graph. We manipulated the design of the labels as well as the rotation of the bar graph.

mined randomly (bigger vs. smaller value). This resulted in 18 trials per label condition (i.e., 180 trials in total).

5.4.2 Results

Overall, task accuracy was high (M = 88.22%, SD = 7.6%). For subsequent analyses, only correct trials were included. Additionally, we excluded outliers for which response latencies deviated more than 3 SD from the individual mean of the corresponding condition (0.7% of all trials). Identically to the first two experiment, we conducted two linear mixed models to analyze the response latencies that differed only regarding the reference condition (see Figure 5.11 for an overview of the results).

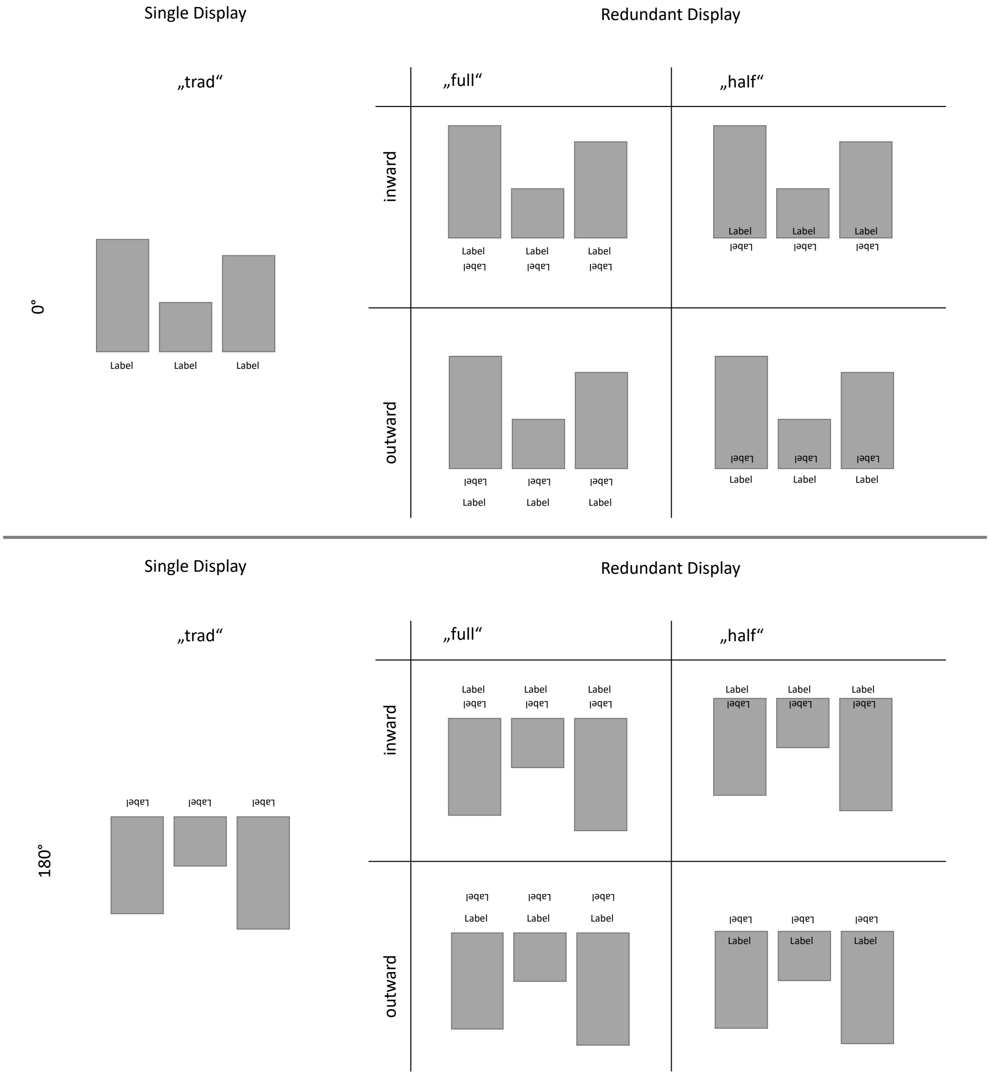

Figure 5.10: Illustration of the stimuli used in Experiment 1. Redundant label stimuli were presented upright (0°) and upside down (180°).

In the first model, the (traditional) single upright condition was used as the reference condition. All other conditions different significantly from this reference; that is, they revealed longer response times than the single upright label (all zs > 6.67, all ps $< .001$, see Table 5.6 for an overview).

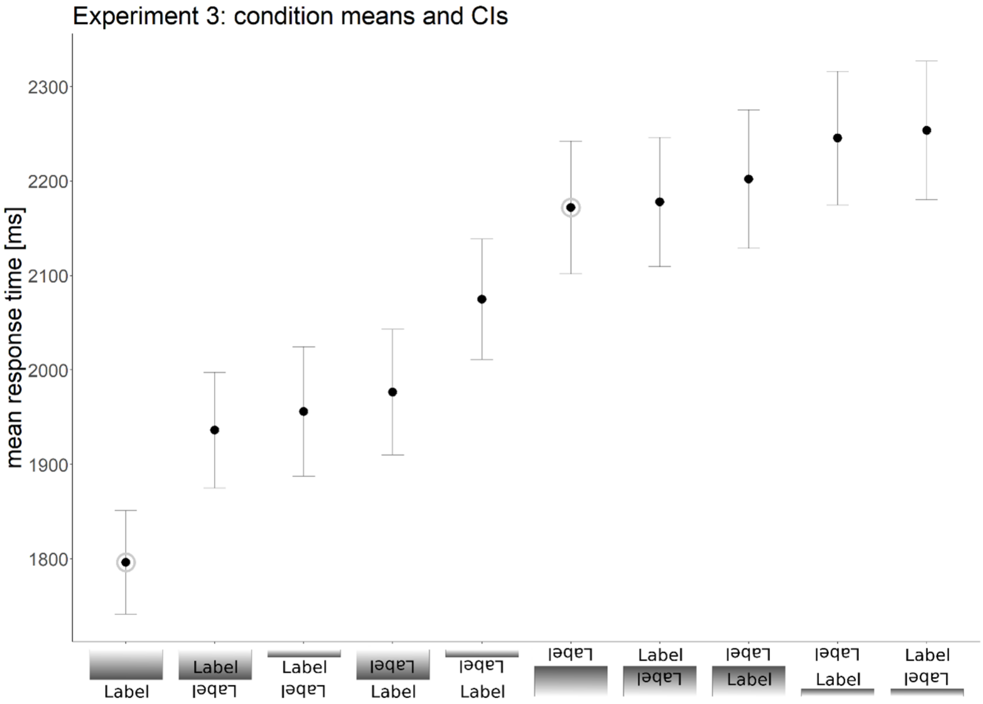

Figure 5.11: Experiment 3. Mean response times separated by condition. The circles refer to the reference conditions for the two linear regressions. The error bars refer to 95% confidence intervals.

In the second model, the single, 180°-rotated condition was set as the reference. This model showed a more complex result pattern. Four conditions elicited significantly smaller response times than the reference condition: The single upright label ($z = 13.89$, $p < .001$), the unrotated both-labels-inside-bar condition ($z = 8.60$, $p < .001$), the unrotated one-label-inside-one-label-outside condition ($z = 8.88$, $p < .001$ and $z = -8.05$, $p < .001$) as well as the unrotated stimulus condition with both labels outside the bars ($z = 3.86$, $p < .001$). The 180°-rotated both-labels-inside-bar condition elicited significant larger response latencies than the reference condition ($z = 3.41$, $p < .001$) as did the 180°-rotated both-labels-outside-bar condition ($z = 3.41$, $p < .001$).

Both the 180° rotated half-inside, half-outside stimulus conditions did not show significant differences, neither if the labels were rotated inwards ($z = 0.21$, $p = .832$) nor if they were rotated outwards ($z = 1.32$, $p = .189$, see also Table 5.6).

Justice Effect

As in Experiment 2, we computed the difference scores for response times between matching conditions (i.e., the conditions mimicking upright vs. inverted viewing conditions on otherwise identical labels). This analysis again aimed at investigating whether some conditions can reduce the difference in access costs to the presented information (i.e., which label condition would elicit the most similar performance for participants with upright vs. inverted viewing conditions).

As in Experiment 2, we conducted a repeated measures ANOVA for which the five difference scores of the matching conditions served as the dependent variable (see Figure 5.12). This ANOVA indicated significantly different difference scores across the matching conditions, $F(4, 552) = 13.95$, $p < .001$, $\eta_p^2 = .09$ (see Table 5.7). Most importantly, Bonferroni-corrected post-hoc tests showed that all but one condition differed significantly from the single upright label reference condition, ts < 5.32, ps $< .001$. Only the inward rotated labels placed outside the bar did not show such a significant difference, $t(138) = 2.43$, $p = .017$, $d = .26$. Further, the outward rotated labels which are located outside the bars showed significant differences not only to the reference condition but also to the inward rotated labels placed outside the bar, $t(138) = 4.06$, $p < .001$. This pair of matching conditions also showed the smallest difference score in response latencies. The remaining difference scores were not significant, ts < 2.60, ps $> .01$. (see Table 5.8 for a complete list of post-hoc pairwise t-test results).

5.4.3 Discussion

We set up this experiment in order to transfer the results from text stimuli mimicking labels (i.e., words without any associated graph) to scenarios which involve actual bar graphs. The participants had to compare bars that were

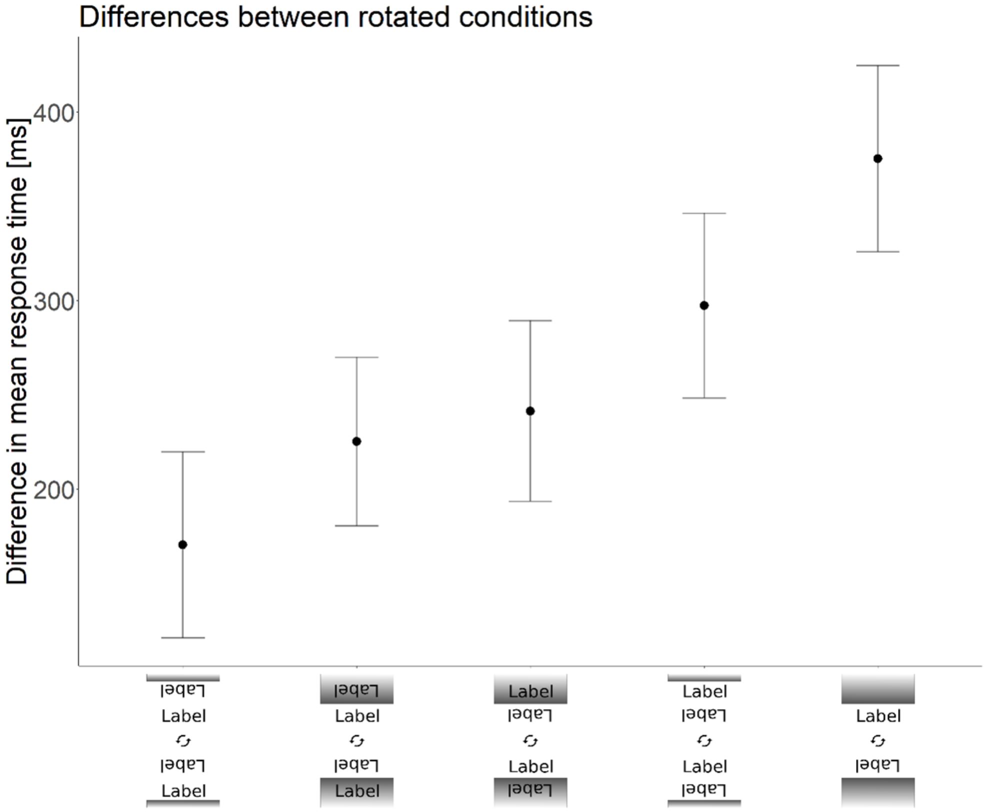

Figure 5.12: Difference scores in response times between the rotationally symmetrical conditions. The error bars refer to 95% confidence intervals.

indicated by traditional labels or the promising variants of the double labels introduced in Experiments 1 and 2. The results were mixed. On the one hand, the conditions with the redundant labels again reduced the differences in response latencies between the upright and inverted viewing conditions. On the other hand, however, this more equalized performance arose to the level of a singular but inverted label. Transferred to a collaborative scenario this finding is challenging, as the results imply that the redundant label has a detrimental influence on the performance of the collaborator with upright viewing conditions, whereas there appears to be no equivalent beneficial impact on the performance of the collaborator with the inverted viewing conditions. There are numerous reasons that could be responsible for this pattern of results. We will return to this within the General Discussion.

5.5 General Discussion

Collaborative work on a depiction of data such as the setting in Figure 5.1 A implies that at least one of the collaborators accesses the graphical information from non-upright viewing conditions. Previous research has demonstrated that such non-upright viewing conditions come along with detrimental performance in accessing the depicted information and further has identified the rotated written labels as a key factor for this effect (Müller et al., 2021). In the present study, we followed up on this finding. We investigated whether adding a second, redundant label at an inverted orientation would be able to reduce this discrepancy in performance between the two collaborators. We hypothesized that such an additional label would limit the deviation from upright viewing conditions to a maximum of 90° (instead of 180°) and therefore should reduce the maximal mental normalization costs. In order to investigate this question, we divided our research project into two parts. In the first part (Experiments 1 and 2), we approached the impact of this intervention from basic psychology research. In the second part (Experiment 3), we applied the findings from the first part to an investigation of bar graphs.

5.5.1 Basic Experiments on Redundant Labels

In the first two experiments, we used a basic research approach to investigate the potential of a redundant label isolated from the more complicated interplay with graphical information (Nothelfer et al., 2017). In other words, we studied how fast our participants were able to categorize target words that mimicked different variants of labels. As outlined in the introduction, our experiments were inspired by a set of findings from related basic research (e.g. B. A. Eriksen and Eriksen, 1974; Shepard and Metzler, 1971). First, a redundant label appears to be promising from the viewpoint of mental normalization processes. In the case of a non-upright viewing condition, such a redundant (inverted) label would remove the necessity to perform mental normalization. Further, research on flanking stimuli suggested that a redundant label might generally speed up processing, as both labels imply the same response (B. A. Eriksen & Eriksen, 1974; C. W.

Eriksen & Schultz, 1979) (although the inversion of one of the labels might diminish this effect). Beyond the presence of a redundant label, we also varied whether the orientation of the labels faced inwards or outwards as well as the contrast between the labels and the background. These two manipulations were implemented for the practical purpose of informing the subsequent application of the redundant label in a bar graph (Experiment 3).

In general, Experiments 1 and 2 showed remarkably matching results. Most importantly, they showed that a redundant label most likely cannot be introduced without costs for the participant with the upright viewing conditions. This is because all conditions that mimicked a set of redundant labels resulted in longer response latencies than the condition with a singular upright target word that mimicked a traditional bar graph label. In scenarios in which only one person works with the depicted information (with an upright viewpoint), introducing a redundant label therefore seems not only to be pointless but also detrimental with regard to performance. However, relative to an inverted target word mimicking a singular label from an upside-down viewpoint, all conditions mimicking a redundant label revealed shorter response times. Thus, in a collaborative setting with two or more people working with the same bar graph, a redundant label potentially might help those collaborators who access the information upside down (M. R. Morris et al., 2006). As the conditions mimicking a redundant label accelerate one of two potential collaborators while slowing down the other, we refer to this effect as the justice effect (i.e., both collaborators become more equal in their level of performance).

With regard to the manipulation of the orientation of the redundant target words (i.e., inward vs. outward) and the contrast between the target words and the background, the results of the basic experiments did not reveal a clear pattern of results. This lack of consistent evidence suggests that at least on the basic level in which participants are required to classify a target word, the impact of the orientation (Koriat & Norman, 1985; Risko et al., 2014; Shepard & Metzler, 1971) as well as perceptual contrast (C.-C. Lin, 2003) appear to be of minor relevance at best, but no clear recommendations can be derived from these findings.

5.5.2 Advancing to Real Bar Graphs

A central motivation of this project was to investigate the impact of the redundant label on identifying depicted information in bar graphs from an inverted viewpoint. With regard to this purpose, the results from the basic Experiments 1 and 2 were encouraging, as they indicated that the redundant label might balance the performance between two collaborators with opposing views (i.e., the justice effect). In the final Experiment 3, we therefore transferred our experimental paradigm to real bar graphs. Instead of identifying target words mimicking labels, the participants had to compare two bars of a bar graph in which we implemented different variants of the singular and redundant labels.

The main finding of this experiment was that the redundant label again balanced the performance from both viewpoints; however, this balanced performance was at the absolute level of a singular inverted label. In other words, instead of increasing performance from the inverted viewpoint at the costs of the performance from the upright viewpoint, there appear to be only costs from the upright viewpoint without any benefits. This finding shows that the observed justice effect from the target word categorization task in Experiments 1 and 2 does not directly transfer to a bar comparison task that involves real bar graph tasks.

There are some technical differences between the first two and the third experiment that could potentially account for the different result patterns with regard to absolute response times. While participants in the first two experiments were asked to classify a target word, participants in the third experiment were asked to compare bar heights and indicate the larger or smaller one. The differences in the tasks might trigger different ways to compensate for the necessary mental normalization. The higher amount of cognitive load from the visualized information that needed to be extracted could have also contributed to the differences in the experiments' results. This also showed up in the lower accuracy of the answers with only about 88% of the trials being answered correctly. Additionally, in the first two experiments, we also controlled for familiarity of stimulus location by placing the stimuli randomly on the screen. With real bar graphs, however, this was not possible. Thus, familiarity with uptight but not with inverted labels may have contributed to the prolonged response times with redundant label conditions.

Besides these differences emerging from the tasks, there are at least two further possibilities how the lack of direct transfer can be explained. First, it of course is possible that the additional label overloads the amount of information presented in the bar graph (Meyerhoff et al., 2021) which might result in a detrimental extraction of the depicted information. In this case, the redundant label would not be able to offer a solution for inverted viewpoints on graphical information such as in the collaborative scenario depicted in Figure 5.1 A. Second, the redundant label rather clearly deviates from the standard layout of bar graphs, which likely everybody in the western world has seen multiple thousands of across her/his lifespan. Therefore, routines that have likely been developed across the previous encounters with bar graphs (Börner et al., 2016; Maltese et al., 2015) might not work with the variant with the redundant label. In this case, the acquisition of new routines would be necessary to bring back performance to the optimal level. However, this would likely require substantial training (Börner et al., 2016) and might be worth it only for those individuals who are involved in collaborative work on graphs on a regular basis.

5.5.3 Strengths, Limitations, and Future Research

In this project, we aimed at investigating whether a redundant label in a bar graph could potentially improve the information processing performance of a person accessing the information upside down (e.g., as part of an in-person collaboration). One central strength in our approach is that we built it upon basic psychology findings and then attempted to transfer these findings to more applied stimuli. Doing so, we observed that the redundant label works on the level of word classification but that it is yet unclear how to transfer this beneficial effect to real bar graphs. The central limitation of our study refers to the generalizability of the findings, which is a common limitation for almost every attempt to transfer a finding from basic psychology research to an applied problem. We only tested one instance of transferring the redundant label to a bar graph; and of course, there might be other implementations which could potentially have more success.

A similar critique might apply for the sample, which consisted of students who likely have a lot of experience with depicted information in general. It thus remains possible that the effect of the redundant label could be different across varying backgrounds of the participants (e.g., the redundant labels might have more success with rather inexperienced participants). In any case, our study is in accordance with other projects highlighting that a transfer from an effect in basic research to an applied problem cannot be taken for granted (LaFortune & Macuga, 2018; Lleras et al., 2017; Sobkow et al., 2019). For future research (including our own), one might consider two relevant lines of research. The first line should attempt to investigate whether practice in reading with bar graphs with redundant labels can unfold the effectiveness of the redundant label that we have observed in the basic word classification task. The second line of research should take a step back and explore other possibilities to improve the accessibility of depicted information from inverted viewpoints that might work without further practice.

5.5.4 Conclusion

In this project, we investigated whether a redundant label is beneficial for processing the depicted information in a bar graph from an inverted viewpoint as it could arise from collaborative scenarios. On the level of the identification of a target word in basic psychology experiments, the redundant label was indeed beneficial for performance from the inverted viewpoint; however, it came along with costs from the upright viewpoint (i.e., the justice effect). Applied to a more ecologically valid information extraction task with bar graphs, however, the redundant label was mostly harmful from the upright view without increasing performance from the inverted view. Therefore, redundant labels might only be useful in rather specific settings (e.g. Chvátil, 2015). Further research therefore needs to explore whether practice might establish the beneficial effect of redundant labels in bar graphs or whether other interventions might improve the accessibility of information from inverted viewing conditions.

5.6 Tables

ANOVA Experiment 1

	DF_n	DF_d	**F**	η_p^2	**p**	
contrast	1	157	114.53	.42	<.001	*
rotation	1	157	413.33	.72	<.001	*
distractor	1	157	140.02	.47	<.001	*
contrast*rotation	1	157	0.21	.00	.649	
contrast*distractor	1	157	43.06	.22	<.001	*
rotation*distractor	1	157	234.68	.60	<.001	*
contrast*rotation*distractor	1	157	11.04	.07	.001	*

Table 5.1: Results of the ANOVA on the data of Experiment 1.

Linear Mixed Models Experiment 1

	Estimate	SE	DF	t	p	
single display high contrast unrotated reference						
(Intercept)	919.78	10.26	171.20	89.62	<.001	*
single_180_low	350.41	16.28	162.48	21.52	<.001	*
single_180_high	247.01	14.82	165.06	16.67	<.001	*
single_0_low	69.23	8.46	1050.01	8.19	<.001	*
double_both_high	94.64	8.62	829.99	10.98	<.001	*
double_upper_high	58.99	8.41	1290.26	7.02	<.001	*
double_lower_high	101.35	8.69	431.36	11.67	<.001	*
double_none_high	109.67	8.66	566.87	12.67	<.001	*
single display high contrast upside-down reference						
(Intercept)	1166.47	18.36	156.22	63.53	<.001	*
single_180_low	103.83	8.35	2736.16	12.43	<.001	*
single_0_high	-246.70	14.23	163.58	-17.33	<.001	*
single_0_low	-177.44	14.67	151.06	-12.10	<.001	*
double_both_high	-152.08	14.49	155.00	-10.50	<.001	*
double_upper_high	-187.57	13.84	164.68	-13.55	<.001	*
double_lower_high	-145.32	12.61	163.64	-11.52	<.001	*
double_none_high	-137.09	13.87	154.27	-9.88	<.001	*

Table 5.2: Overview of the results of both linear mixed model regressions performed on the data of Experiment 1.

Linear Mixed Models Experiment 2

	Estimate	**SE**	**DF**	**t**	**p**	
single display high contrast unrotated reference						
(Intercept)	935.90	10.83	166.21	86.40	<.001	*
single_0_low	80.92	9.25	505.54	8.74	<.001	*
inner_both_high	44.10	8.90	966.38	4.96	<.001	*
inner_lower_high	122.14	9.71	326.38	12.58	<.001	*
inner_upper_high	6.52	8.70	1681.71	0.75	.454	
outer_both_high	88.41	9.55	239.57	9.26	<.001	*
outer_lower_high	41.40	8.80	1248.18	4.70	<.001	*
outer_upper_high	155.79	11.97	170.86	13.01	<.001	*
single_180_high	272.54	21.97	155.46	12.40	<.001	*
single_180_low	301.98	21.53	155.01	14.03	<.001	*
single display high contrast upside-down reference						
(Intercept)	1208.44	24.37	153.73	49.59	<.001	*
single_180_low	29.44	16.54	155.14	1.78	.077	
inner_both_high	-228.45	22.15	154.70	-10.31	<.001	*
inner_lower_high	-150.40	20.25	155.08	-7.43	<.001	*
inner_upper_high	-266.03	22.11	155.04	-12.03	<.001	*
outer_both_high	-184.13	21.51	154.09	-8.56	<.001	*
outer_lower_high	-231.14	22.11	154.12	-10.45	<.001	*
outer_upper_high	-116.75	20.43	153.96	-5.71	<.001	*
single_0_high	-272.54	21.98	155.17	-12.40	<.001	*
single_0_low	-191.62	22.27	154.56	-8.60	<.001	*

Table 5.3: Results of the two linear mixed model regressions performed on the data of Experiment 2.

ANOVA on rotationally symmetrical conditions in Experiment 2

	DF_n	DF_d	**F**	η_p^2	**p**	
(Intercept)	1	154	240.05	.61	<.001	*
condition	3	462	38.57	.20	<.001	*

Table 5.4: Results of the ANOVA on the "rotation twin" condition differences on the data of Experiment 2.

Post-hoc t-tests on rotationally symmetrical conditions in Experiment 2

	$Mean_{diff}$	SD_{diff}	**SE**	**95% CI**	**t**	**DF**	**d**	**p¹**	
Difference to single display with high contrast									
single_low	50.74	194.15	15.59	[19.94, 81.55]	3.25	154	0.20	.001	*
inner_upper_high	156.65	257.50	20.68	[115.79, 197.51]	7.57	154	0.77	<.001	*
outer_lower_high	157.47	260.33	20.91	[116.16, 198.78]	7.53	154	0.74	<.001	*
Difference to single display with low contrast									
inner_upper_high	105.91	233.28	18.74	[68.89, 142.92]	5.65	154	0.57	<.001	*
outer_lower_high	106.72	231.75	18.61	[69.95, 143.50]	5.73	154	0.54	<.001	*
Difference to inward rotated double display									
outer_lower_high	0.82	139.76	11.23	[-21.36, 23.00]	0.07	154	0.01	.942	

[1] alpha level was bonferroni-corrected

Table 5.5: Results of the post-hoc tests on the "rotation twin" condition differences of Experiment 2.

Linear Mixed Models Experiment 3

	Estimate	**SE**	**DF**	**t**	**p**	
single display high contrast unrotated reference						
(Intercept)	1797.32	28.90	144.88	62.20	<.001	*
half-inner_0	140.57	21.08	2770.14	6.67	<.001	*
half-inner_180	378.57	25.92	223.44	14.61	<.001	*
half-outer_0	177.58	22.48	417.64	7.90	<.001	*
half-outer_180	404.14	27.41	202.85	14.74	<.001	*
full-inner_0	161.87	22.57	251.84	7.17	<.001	*
full-inner_180	455.46	26.98	209.53	16.88	<.001	*
full-outer_0	277.70	21.97	336.45	12.64	<.001	*
full-outer_180	452.30	26.55	203.35	17.03	<.001	*
trad_180	374.14	26.72	219.45	14.00	<.001	*
single display high contrast upside-down reference						
(Intercept)	2171.41	36.42	140.48	59.62	<.001	*
half-inner_0	-233.45	26.28	170.95	-8.88	<.001	*
half-inner_180	4.50	21.27	6179.49	0.21	.832	
half-outer_0	-196.52	24.41	196.01	-8.05	<.001	*
half-outer_180	30.13	22.88	282.29	1.32	.189	
full-inner_0	-212.19	24.67	175.25	-8.60	<.001	*
full-inner_180	81.37	23.89	201.99	3.41	.001	*
full-outer_0	-96.41	25.01	186.32	-3.86	<.001	*
full-outer_180	78.32	22.96	233.00	3.41	.001	*
trad_0	-374.09	26.94	175.65	-13.89	<.001	*

Table 5.6: Results of the two linear mixed model regressions performed on the data of Experiment 3.

ANOVA on rotationally symmetrical conditions in Experiment 3

	DF_n	DF_d	**F**	η_p^2	**p**	
(Intercept)	1	138	270.11	.66	<.001	*
condition	4	552	13.95	.09	<.001	*

Table 5.7: Results of the ANOVA performed on the ”rotation twin” condition differences of Experiment 3.

Post-hoc t-tests on rotationally symmetrical conditions in Experiment 3

	$Mean_{diff}$	SD_{diff}	**SE**	**95% CI**	**t**	**DF**	**d**	**p**[1]	
Difference to single display outside bar									
full-inner	77.98	379.10	32.15	[14.40, 141.56]	2.43	138	0.26	.017	
half-inner	133.90	296.66	25.16	[84.14, 183.65]	5.32	138	0.46	<.001	*
half-outer	150.12	317.86	26.96	[96.81, 203.43]	5.57	138	0.53	<.001	*
full-outer	204.86	352.90	29.93	[145.67, 264.04]	6.84	138	0.69	<.001	*
Difference to inward rotated completely outside bar									
half-inner	55.92	365.41	30.99	[-5.37, 117.20]	1.80	138	0.19	.073	
half-outer	72.14	327.68	27.79	[17.19, 127.10]	2.60	138	0.26	.010	
full-outer	126.88	368.42	31.25	[65.09, 188.67]	4.06	138	0.43	<.001	*
Difference to inward rotated half outside bar									
half-outer	16.23	350.51	29.73	[-42.56, 75.01]	0.55	138	0.06	.586	
full-outer	70.96	382.00	32.40	[6.89, 135.03]	2.19	138	0.24	.030	
Difference to outward rotated half outside bar									
full-outer	54.73	324.32	27.51	[0.34, 109.13]	1.99	138	0.19	.049	

[1] alpha level was Bonferroni-corrected

Table 5.8: Results of the post-hoc tests performed on the ”rotation twin” condition differences of Experiment 3.

5.7 Declarations

5.7.1 Funding

Funding was by the German Federal Ministry of Education and Research (BMBF) 2016-2019 under grant number 01IO1616.

5.7.2 Conflict of Interest Statement

The authors confirm that they have no conflict of interest to declare.

5.7.3 Ethical Approval

Ethical approval was granted from the IWM ethic board under LEK 2019/026

5.7.4 Availability of Data and Material

Study materials, data, and analyses are available at the Open Science Framework (https://osf.io/2b4cs/?view_only=f8973154c6984553b2a631e95757df50).

Chapter 6

Mental Rotation is not a Perceptual Disfluency for Diagrammatic Learning

6.1 Introduction

6.1.1 Should Learning be Easy or Tough

Remember being a student approaching the end of a term. While preparing for all sorts of tests, a friend informs you about an upcoming exam that you had completely forgotten. Now you need to squeeze the additional learning load into your already packed revision schedule. How can you achieve this? Most education researchers seem to agree, that learning needs to be "hard". They disagree, however, on which aspect should be effortful.

The *Cognitive Load Theory*, proposed that easy and accessible learning material frees cognitive capacities, that can then be used to encode the new information more thoroughly (Sweller & Chandler, 1994). Load on the working memory during learning has three possible sources. First, it can be generated by the material itself, by the complexity of the information to be learned. This "intrinsic cognitive load" can hardly be avoided. The second source of cognitive load stems from the design of the learning material. This "extraneous cognitive load" is put on the working memory by badly presented material e.g., with additional, irrelevant information and repetition. Careful design of the learning material can reduce this kind of cognitive load to a minimum. Finally, there is "germane cognitive load", the effort the student puts into connecting the new information to similar known subjects ("schemas"). This should be high and can be guided by good instruction design. As the capacity of the working memory is limited (Miller, 1956), it is crucial to reduce intrinsic and extraneous cognitive load to free capacities for germane cognitive load.

The *AIME theory* (Amount of Invested Mental Effort) by Salomon (1984) argued that television is perceived as a medium that is easy to consume. In contrast, text presentation seems to be a more demanding material. The amount of invested mental effort is influenced by two factors: the perceived demand characteristics (PDC) and the Perceived self-efficacy (PSE). TV imagery has a low PDC, and therefore, as it is perceived as an easy medium, little mental effort invested, thus barring deep processing of the presented information. In contrast, the effort needed to read and understand a text is perceived as far more demanding and therefore ensure a deeper processing. While the AIME-theory is often cited and applied in various contexts, some researchers doubt its relevance as an independent theory (Schwab et al., 2018). It also failed to replicate in a Dutch sample (Beentjes, 1989), weakening its plausibility or generalizability. Nonetheless, the notion of perceived characteristics influencing the effectiveness of learning material remains popular (Schwab et al., 2018).

Bjork and Bjork (2009) noticed that some efforts students make to achieve sustainable knowledge do not work as well as suspected, therefore one needs to identify fruitful strategies of making learning hard, but in the right way (called *desirable difficulties*). In this line of research, some useful strategies for sustainable learning have emerged, such as using tests as a tool for learning (Kornell & Vaughn, 2016) or reproduce the learned information in ones own words (Bertsch et al., 2007; Slamecka & Graf, 1978).

The *dual coding theory* (Paivio, 1991) states, that learning is more effective if information is presented both as text and as an image. This would lead to the activation of two information processing pathways (textual/phonological and visual) which would ensure deeper processing. In considering multiple pathways, the dual coding theory opens a broader view on what can be learned. Most of the presented research traditions are either focussed on text information (Cognitive Load Theory, Desirable Difficulty) or are agnostic to the media that presents the information (Cognitive Offloading). While Dual Coding Theory is extending this view, only AIME centers around the medium that presents the information. Multimodal learning has gained some track in recent years, with an extension of the *Cognitive Theory for multimodal learning* (Mayer, 2005) and extensive research on text picture integration (Eitel et al., 2012; Scheiter et al., 2018).

Another approach to make learning harder is derived from the already mentioned concept of *cognitive fluency* (Alter & Oppenheimer, 2009; Oppenheimer, 2008). To nudge humans towards investing more mental resources in learning, one could manipulate the perceived difficulty of the material. This concept is called "disfluency" (Alter & Oppenheimer, 2009; Oppenheimer, 2008). There is some evidence that e.g., text retention might be improved by making it harder to read by using an unusual font (Diemand-Yauman et al., 2011), blurring (Rosner et al., 2015) or briefly masking it (Mulligan, 1996). These perceptual disfluency effects should make the material seem and feel harder to learn than it actually is (Geller et al., 2020). However, the existence of this influence remains debated (Magreehan et al., 2016; Rhodes & Castel, 2008, 2009; Xie et al., 2018; Yue, Castel, & Bjork, 2013).

6.1.2 On Diagrams

Diagrams are visual depictions of quantitative information (Cleveland & McGill, 1984) and are widely used as they can be superior in communicating this information (Dambacher et al., 2016). This advantage of a graphical depiction of data over other forms of data presentation (such as tables) might be attributed to several aspects of both the medium and the cognitive processing of it. There is a generally described "picture superiority effect" for memory (Cattaneo et al., 2007; Standing, 1973; Whitehouse et al., 2006). The higher memorability of pictures might be due to more distinctive markers of the image than flowing text. There are also some aspects of visual information that can be processed almost effortlessly, such as color and in some cases shape (Treisman & Gelade, 1980). Additionally, in contrast to e.g., tables, data visualizations can transport additional information, besides the pure values. Tufte (1983) describes three "viewing depths", simultaneously present in data visualisations: "what is seen from a distance", the overall structure of the displayed data, "what is seen from up close", the presented data itself and "what is seen implicitly", the story, the graphic tells. So multiple layers of potentially learned information must be accounted for, when considering learning from data visualisations. The general debate about the best design for learning material is repeated on a small scale for diagrams, following similar lines of discussion. Some researchers see additional visual components similar to the concept of germane load from the Cognitive Load Theory, guiding the learner's attention. Otto Neurath (1936), for example, regarded additional "chart junk" as beneficial for understanding. Evidence that supported this claim was found for both short term (Borgo & Abdul-Rahman, 2012; Haroz et al., 2015) and long-term memory (Bateman et al., 2010). Other data visualisation experts, like Tufte (1983), argue more in line with the Cognitive Load Theory's definition of extraneous cognitive load, advocating for a strict no-nonsense approach to data visualizations, as any unnecessary elements might confuse the reader, increase the printing costs of charts and might even open the door for subtle manipulations. Some found no significant influence (Kosara & MacKinlay, 2013) or negative impact of "chart junk", depending on the type of data visualization used (Skau et al., 2015). Overall, a balance of processing speed and long-term retention (called desirable visual difficulty) needs to be found when additional, non-informative elements are presented in a data visualization (Hullman et al., 2011; Salomon & Perkins, 2005).

6.1.3 Mental Normalization

Apart from the visual design of the diagrams, one can think of other manipulations that can be used to influence learning e.g., the overall rotation. Presenting a diagram in a non-canonical rotation triggers mental normalization (Müller et al., 2021) which in turn puts a higher load on the working memory (Shepard & Metzler, 1971). First described by Shepard and Metzler (1971) as mental rotation, the effect of longer response times for higher deviations from a canonical point of view has been shown for various materials, such as box figures (Shepard & Metzler, 1971), letters (Cooper & Shepard, 1973), maps (Montello, 2010), molecules (Stieff, 2007), arbitrary shapes (Cooper, 1975), text (Koriat & Norman, 1985; Risko et al., 2014) and bar graphs (Müller et al., 2021). Besides the model of a whole stimulus being virtually rotated in working memory, other processes have been proposed, such as piece-meal rotation (Just & Carpenter, 1985; Yuille & Steiger, 1982), mental storage of a3D model (Marr, 2010; Marr & Nishihara, 1978) or an edge detection mechanism (Biederman, 1987). However, the exact process is still disputed and might even be only distinguishable with neuroimaging techniques (Gauthier et al., 2002). Therefore, we use the theory agnostic term "mental normalization" as umbrella term for all mechanisms.

As mental normalization puts a higher load on the working memory, a rotated display of data visualizations might either trigger a more intensive analysis of the stimulus (acting as a desirable difficulty) or make the material seem to be less accessible and therefore more complicated than it is (acting as a perceptual disfluency).

6.1.4 The Influence of Cognitive Offloading

The concept of cognitive offloading (Risko & Gilbert, 2016) proposes, that load on the working memory is something humans tend to avoid. They do so by using external devices or physical motion, effectively storing information in the world instead of in their heads. By storing information in the environment, humans can overcome their biological limitations, in this case the working memory capacity (Dror & Harnad, 2008). Cognitive offloading seems to boost short term task performance (Beitzel & Staley, 2015; Grinschgl et al., 2021; Kirsh, 2010) but it might hamper long term memorization. If the external device that participants used to offload is taken away, insufficient performance in follow-up testing was reported for problem solving (van Nimwegen & van Oostendorp, 2009), navigation

(Fenech et al., 2010; Gardony et al., 2013, 2015), and skill acquisition (Casner et al., 2014; Ebbatson et al., 2010). The idea that cognitive offloading might hamper long term memory formation gained some popularity under the term "Google effect" (Sparrow et al., 2011), but these findings could not be replicated (Hesselmann, 2020). Other researchers reported the adverse effect of cognitive offloading on learning in different domains (Eskritt & Ma, 2014; Henkel, 2014; Kelly & Risko, 2019a, 2019b; Pyke & LeFevre, 2011). Some results suggest, that interrupting cognitive offloading could be used as a desirable difficulty (Morgan et al., 2009, 2013). The effect of cognitive offloading on long term memorization is disputed.

Mental normalization has been shown to be a task that humans tend to load off (Risko et al., 2014). To investigate the influence of mental normalization on long term memory formation, cognitive offloading needs to be controlled for.

6.1.5 Present Research

In this paper, we report an experiment on the long-term memorization of information displayed in bar graphs, presented either upright or rotated, to investigate if mental rotation acts as desirable difficulty or perceptual disfluency. If this is the case, we would expect information from rotated diagrams to be more accurately recalled than information from unrotated diagrams. On the other hand, if mental rotation is an additional load on working memory, we would expect the opposite effect, that information from unrotated diagrams is recalled more accurate than from rotated diagrams.

To account for Tuftes (1983) multiple depths model, multiple learning targets were defined. The "from a distant" level was captured with a diagram shape recognition task (called recognition questions). To represent the "close up" level, questions on the data itself were asked. We specified two distinct conclusion that could be drawn from the data, the individual values of separate bars (called factual questions) and a height comparison between two bars (called comparison questions). The "story" level was neglected as diagrams were generated with bogus data and did not have a story to tell.

6.2 Methods

6.2.1 Participants

Ninety-one students (70 female) from the University of Tübingen, age 18-35 (M = 23.39 years, SD = 3.35 years) took part in this experiment. All participants were recruited via the volunteer platform for experiment participation of the Leibniz-Institut für Wissensmedien. All participants received 5€ in compensation for 40 minutes of their time. The study was ethically approved by the institutional review board of the Leibniz-Institut für Wissensmedien. All participants provided informed consent prior to testing. As spontaneous learning was investigated, participants were not informed of a knowledge test in the second phase of the experiment. After completion, participants were informed about the true goals of the study and were given the opportunity to withdraw their consent.

6.2.2 Apparatus

Participants were seated in front of a HP Elitebook 8530p with a $15.4''$ display with an approximate viewing distance of 60 cm. Participants were also fitted with earmuffs that had an Arduino Nano and a Gyroscope GY-521 on top of it to monitor the head movements. The gyroscope was adjusted to sit horizontal on the participants head (see Figure 6.1 for a depiction of the setup). To allow for a bit of leeway in adjustment, the gyroscope had to sit within a deviation of $\pm 1°$ (dotted lines) before the experiment started. If the participants tilted their heads more than $\pm 4°$ ($\pm$3-5° in total, accounting for the leeway of gyroscope attachment, lower lines) the screen was blacked out until participants levelled their head out again.

6.2.3 Procedure

The experiment was set up in a within-subject design and consisted of a familiarization phase and a surprise test phase (see Figure 6.2).

In the familiarization phase, participants were presented with 10 diagrams, five of which were rotated by 150° and five unrotated. Beneath each diagram

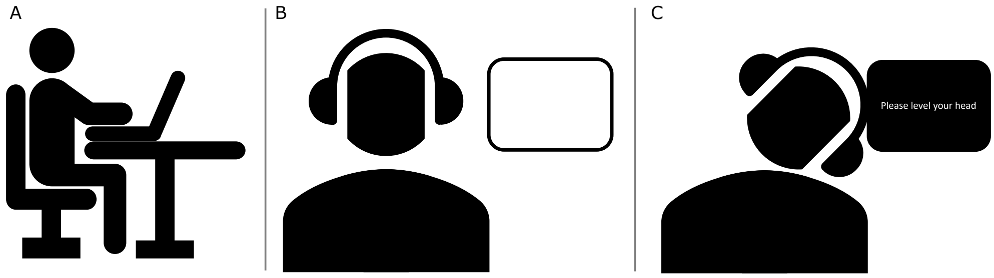

Figure 6.1: Illustration of the experiment setup (A) and a depiction of the gyroscope attachment shown as headphones (B). If the head was held level (within a ±5° margin) participants could work on the task. If the head was tilted beyond the threshold (C) the screen was blacked out and a message appeared to level the head again. This was implemented to prevent cognitive offloading of mental normalization.

(size 14.2 x 14.2 cm), participants were presented with 6 questions (font size 40) in sequence that should be answered with the information within the diagram. The diagram contained 5 bars with labels from a common topic and a y-axis with a topic-specific labelling and magnitude. The bars were labelled with words of the same topic consisting of seven letters in three syllables to control for reading speed. The questions displayed beneath the diagrams fell in two categories. "Factual" questions asked for the specific height of one of the displayed bars. "Comparison" questions asked to indicate which one of two bars was bigger/smaller. Each question type was presented three times. To control for cognitive offloading (Risko et al., 2014) effects participants were required to keep their head in an upright position. To enforce this, participants wore the gyroscope sensor on their head during the learning phase, which would trigger the screen blackening if a head tilt over 5° was detected.

After the familiarization phase, participants were asked to take off the sensor and to wait for 30 seconds until a test phase started. Participants were than shown the same questions as before, but without diagrams. The questions were displayed in the middle of the screen. Additionally, following each block of questions, a "recognition" question was presented, showing the diagram they had seen before and a bogus one with the same labels but different values. Participants should indicate, which diagram they had already seen. To avoid influences due to the random generation of the diagrams and questions, a unique set of stimuli was generated for each participant.

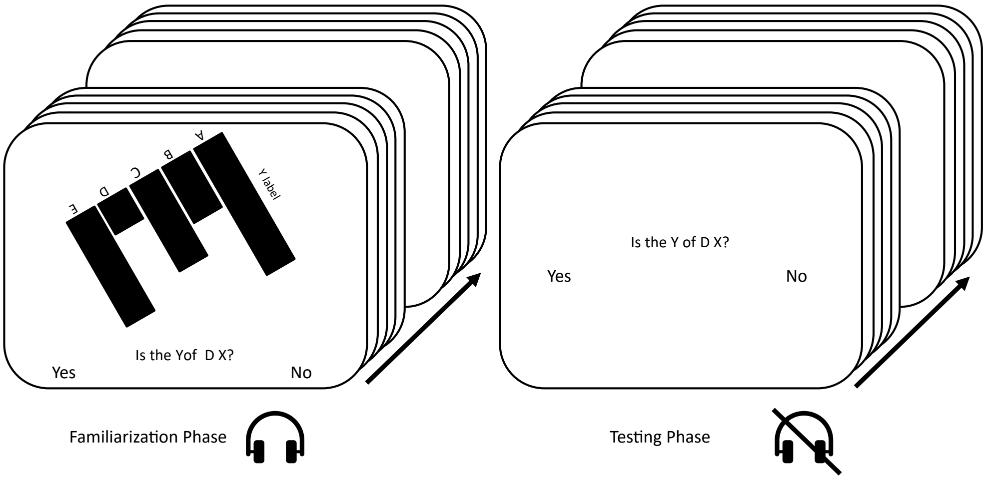

Figure 6.2: Flow diagram of study 2. In the familiarization phase participants were exposed to 10 diagrams with six question each. They needed to keep the gyroscope (here displayed as headphones) on during the learning phase. In the test phase, participants were taking the gyroscopes off and answered the same six question again plus one about the general shape of the diagram.

6.3 Results

8.37% of data points were excluded, because participants did answer these questions wrongly in the learning phase and therefore couldn't be considered as learned. We fitted two separate logistic models for factual and comparison questions in one model and recognition questions in the other model due to different data structures (see Figure 6.3 for an overview of the results). The dependent variable in both models was the correctness of the participants answers. In both models, participant and trial were set as random effects with fixed means. Accuracy rates for comparison questions were higher than accuracy rates for factual questions ($Z = 9.519,\ p < .001$). Differences for the rotation of the diagram (0°, 150°) could not be detected ($Z = 0.845$, $p = .398$). Similarly, no differences for the rotation of the diagram regarding accuracy rates in recognition questions were found ($Z = 0.801$, $p = .423$). Generally, recognition questions had the highest probability to be answered correctly (75.16%), followed by comparison questions (67.30%) and factual questions (56.01%).

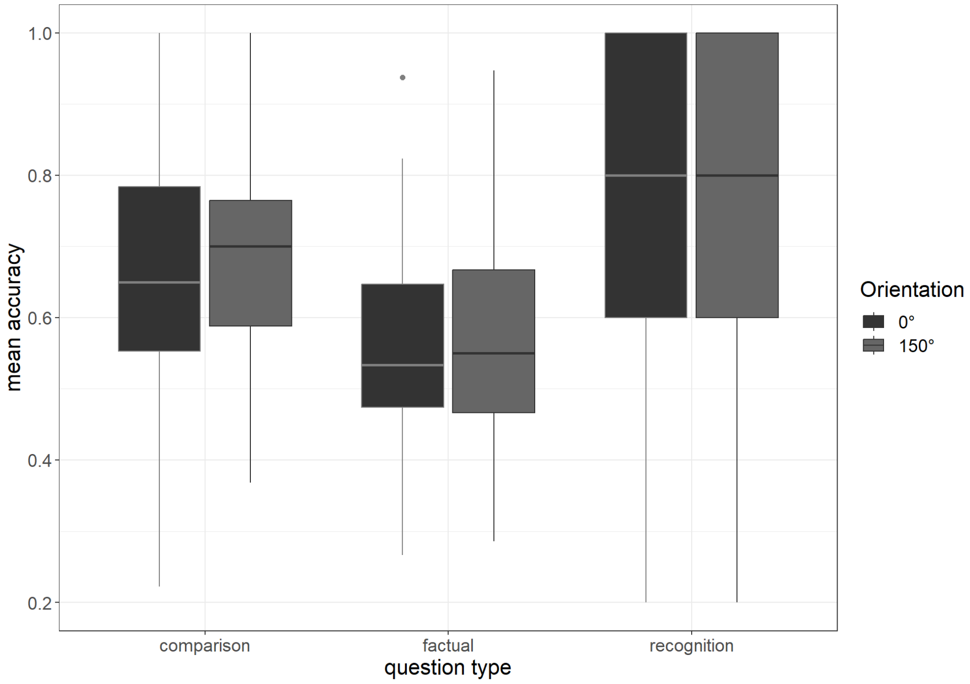

Figure 6.3: Box plot of mean accuracy of information recollection by diagram orientation and information type.

Confronted with the unconclusive results of the experiment, we conducted further investigation that was not pre-registered. We hypothesized, that mental rotation may have hampered the extraction of information. As we did not control the exposure time to the stimuli, it might have counterbalanced the adverse effect of mental rotation. Subsequently, a post-hoc analysis of the exposure times was conducted. Indeed, participants spend more time with the rotated diagram (M = 5.37 s per question) as with unrotated diagrams (M = 4.86 s per question). This difference was significant ($t(90) = -5.472$, $p < .001$). This confirmed our post-hoc hypothesis that mental normalization did hamper learning but also lead to longer self-exposure, which counterbalanced the effect.

6.4 Discussion

The analyses showed a significant difference between question types. Questions that concerned only one value from the diagram (factual questions) were less likely to be remembered correctly than questions that compared two values (comparison questions) and were indeed very close to guessing probability. The general shape of the diagram (recognition questions) was remembered or recalled best. In context of Tuftes (1983) multiple depths model, these findings indicate that a) indeed multiple levels of information coding seem to exist and b) that the individual data points are less significant to the onlooker than the overall shape of the diagram. As the "story" level was neglected in this study, no conclusions could be drawn on its significance for the data visualisation. Future research into this area might be fruitful.

Curiously, there was no significant difference between the two rotation conditions. This was most unexpected, as the theoretical background suggested a difference, just the direction was unclear. If mental normalization would act as a desirable difficulty (Bjork & Bjork, 2009; Diemand-Yauman et al., 2011; Paivio, 1991) or would increase the perceived difficulty (Salomon, 1984), information from rotated diagrams should have been easier to recall and thus show higher accuracy rates than information learned from unrotated diagrams. On the other hand, if easier access to the presented information frees resources for more thorough learning (Sweller & Chandler, 1994), we would expect information learned from unrotated diagrams to be more accurately recalled than information from rotated diagram.

A possible explanation for the inconclusive results might the extended exposure time on rotated stimuli trials. As exposure time was not controlled for, participants might have spent more time on rotated trials than on unrotated trials, counterbalancing the adverse effect of mental normalization costs. The results of the analysis of this post-hoc hypothesis seem to support this explanation.

Mental normalization indeed puts additional load on the working memory, but this additional load could not be fostered as a disfluency or desirable difficulty measure to support learning. The additional load seems to rather have hampered learning, which the participants seem to have mitigated by longer material exposure. Controlling for exposure time would have been beneficial for a clearer insight into this phenomenon and future research should account for that.

6.4.1 Conclusion

We investigated whether the resource-intensive process of mental normalization can be fostered for better long-term retention. Some theories point in that direction (Bjork & Bjork, 2009; Salomon, 1984). The results could not find any benefits of mental normalization on learning. A post-hoc analysis of the data rather points to the opposite. The negative effects of mental normalization on memory were mitigated by longer exposure times. This suggests that mental normalization cannot be fostered as a strategy for long-term memory retention. Further research might focus on the exact processes that make additional mental load a viable endeavour for learning, as not every additional mental load seems to have a positive effect.

6.5 Declarations

6.5.1 Conflict of Interest

The authors confirm that they have no conflict of interest to declare.

6.5.2 Funding

This research was funded by a grant of the German Federal Ministry for education and research (BMBF) 2016–2019 under grant number 01IO1616.

6.5.3 Ethical Approval

The experimental procedure was reviewed and approved by the institutional review board of the Leibniz-Institut für Wissensmedien (LEK 2018/077). All participants signed informed consent before the experiment. As participants were not informed about the test part to avoid intentional learning, they were informed afterwards about the aims of the experiment. They were given the opportunity to redact their consent.

6.5.4 Preregistration

The preregistration of this experiment can be accessed
via AsPredicted (https://aspredicted.org/68vq7.pdf)

6.5.5 Availability of Material and Data

Study material, data and analyses are available via the Open Science Framework (https://osf.io/mkf3a/?view_only=bc0dca7d97c44cdd9fa8a84853200546) A description of the Gyroscope and the respective programming code is also available (https://osf.io/qgrxp/?view_only=76459a6d9b544104a719facdeb1a9fb1).

6.5.6 Acknowledgements

I would like to thank Dr. Nora Wickelmaier for support in the pre-registration process and counseling for the data analyses. Thanks to Jana Lang for her support in data collection and to Hauke Meyerhoff for his valuable remarks on the manuscript.

Chapter 7

Discussion

In co-located collaborative work settings, equal access to information is important for efficient group work. Single Display Groupware like Multi-Touch Tables is designed to be accessed from all sides. To provide equal access to all participants, the different points of view need to be considered. Most information, however, is displayed in a distinct orientation, designed to be accessed from a certain point of view. Users with differing points of view need to engage in the costly cognitive process of mental normalization. The additional mental load might prevent equal access to the presented information. This dissertation focused on the individuals' cognitive processes (i.e., visual preception) and aimed to achieve three goals in two major lines of study.

The first line of study aimed at two goals. The first goal was to confirm the cost of mental normalization for rotated diagrams. The second goal was to explore ways to ease mental normalization costs. To allow equal access to the displayed information, the diagram should be modified to accommodate multiple points of view. This line was conducted in the manuscripts in Chapter 4 and Chapter 5. The second line investigated whether mental normalization yields benefit, e.g., facilitating long-term retention of the presented information. Chapter 6 was devoted to this question.

The confirmation of the hypothesis of an additional strain of mental normalization of data visualizations on the cognitive system was addressed with the first experiment in Chapter 4. Mental normalization costs had been shown for several other types of stimuli, including cube shapes (Shepard & Metzler, 1971), letters (Corballis & McLaren, 1984; Rüsseler et al., 2005), words (Koriat & Norman, 1985; Risko et al., 2014), molecules (Stieff, 2007), arbitrary two-dimensional shapes (Cooper, 1975), and pictures (Tarr & Pinker, 1989), but not for diagrams. By comparing response times for unrotated bar graphs and rotated bar graphs, the significant additional cost of mental normalization for the cognitive system was confirmed. Participants responded significantly slower to rotated bar graphs with co-rotating labels than to unrotated bar graph displays. This pattern was also present in the base-line conditions (bar graphs with word labels) for the other two experiments reported in Chapter 4. These results confirmed the additional costs of mental normalization of diagrams for the cognitive system. The findings were also in accordance with the above-mentioned results on other types of stimulus material. Additionally, the independent manipulation of the overall rotation of the bar graph and the separate rotation of the labels revealed the label rotation to be a major factor in response time latencies. Labels were either shown in the same rotation as the entire diagram ("co-rotating with the diagram") or, aligned upright relative to the subject, independent of the diagram's rotation ("reformatted display"). Normalization costs were significantly reduced in the latter manipulation, indicating a piece-meal mental normalization process that normalized different parts of the diagramm seperately.

The second question addressed design options to lessen the strain from mental normalization. As the first experiment revealed the label rotation as a critical factor in graph rotation, changing label characteristics would benefit rotational access the most. As a constant orientation towards the subject would not be feasible in a multi-subject setting, other approaches were tested. In the second experiment of Chapter 4, the use of icons as replacement for written labels were tested. Otto Neurath created an icon-based data visualization system as a mean to communicate important information to lesser educated classes (Neurath, 1936). He argued that the use of self-explaining icons would benefit the understandability of data visualizations. Later research was able to support this claim (Haroz et al., 2015). In the second experiment, we used the same icons as Haroz et al. (2015). The results added further support to some of the claims. Response time latencies were overall reduced in icon conditions, showing a decrease in mental normalization cost. However, despite not showing significant impact, the differences between the rotation steps remained in non-negligible effect size range, at least for the critical "rotating with the diagram" condition. Therefore, the intervention was only partially successful. Although reducing the overall response time compared to word labels, the mental normalization costs are still affecting the response time latencies. The persistence of the mental normalization effect might be explained by the "orientatedness" of the icons. While being processed faster than words, pictographs still possess an optimal point of view (Palmer et al., 1981) and therefore contribute to the access cost differences for different perspectives. Additionally, icons might only be used in certain contexts, but might not be suitable for all topics or themes.

A color manipulation was chosen as a second approach. Color cues have been shown to be processed very fast (Treisman, 1986). They have been proven to be very effective for perceptual grouping (Palmer et al., 2003), visual search (Wolfe & Horowitz, 2004, 2017) and attentional efficiency (Carter, 1982; Christ, 1975; Duncan & Humphreys, 1989). The influence of color coding of the labels to reduce mental normalization costs was tested in the third experiment of Chapter 4. Like the effect of the pictograph manipulation, the response time latencies were reduced in rotated trials with color cues to a negligible difference compared with unrotated trials. Both, pictograph, and color cues reduced the response time latencies significantly, but the color intervention seemed to be slightly more efficient.

A third approach was investigated in Chapter 5, displaying a second, redundant, but 180°rotated label. A previous experiment (Experiment 1 in Chapter 4) revealed the label rotation as a key factor for mental normalization costs. Ideally, the label would always be oriented towards the onlooker, but as there are multiple points of view on the data visualization on the multi-touch table, the implication of a rotating label is not feasible. However, a double display of unrotated and upside-down displayed labels could eliminate mental normalization costs for viewers positioned on the opposite sides of the MTT and lessen the mental burden for subjects positioned at the small sides (±90°) of the table. The effect of such a double display was unclear. The investigation of the feasibility of a redundant label display was split in two parts. The first two experiments of Chapter 5 were concerned with the impact of double label display from the perspective of basic psychological research. Literature on the influence of congruent flankers on the response time showed a beneficial effect for processing speed in reducing the response time (B. A. Eriksen & Eriksen, 1974). As the same label was displayed, we hoped to find similar results, although the effect might be diminished by the rotated display. Additionally, to the double display of labels, we also manipulated the orientation of the labels, either facing inwards or facing outwards. This was done to draw informed decisions for the application on bar graphs. For the same reason, the contrast of the labels was also manipulated.

The first two experiments of Chapter 5 yielded similar results. In both cases, the introduction of the redundant label increased response time for upright viewing conditions. However, the use of redundant labels seemed to have a leveling effect on the response times from different perspectives. Experiment 2 showed a decrease in response time differences between upright and upside-down presentation of otherwise similar stimuli conditions. We called this the "justice effect" of double label display. The other manipulations did not show a clear pattern. It seems that the label orientation (Koriat & Norman, 1985; Risko et al., 2014; Shepard & Metzler, 1971) and perceptual contrast (C.-C. Lin, 2003) have only minor influences on the response time. In the final experiment of Chapter 5, the double label display approach was applied to rotated bar graphs. The results of the first two experiments were encouraging. We were able to reproduce the balancing "justice effect" of double label displays on bar graphs, but in the applied setting, both viewing conditions (upright and upside-down) showed higher response times for double label displays. The additional label might have caused an information overload on the cognitive systems of the participants (Meyerhoff et al., 2021). It is also possible, that the differing task designs between the experiments had led to the differing outcomes: In experiment 1 and 2, the participants were asked to classify the stimulus, the third experiment contained a comparison task.

The first line of research was concerned with the mitigation of the higher demand on the cognitive system stemming from mental normalization. In a second line of research, possible benefits of mental normalization load were investigated. Previous research suggested that additional cognitive load might increase the effort put into learning by setting up additional hurdles (Diemand-Yauman et al., 2011; Mulligan, 1996; Rosner et al., 2015) or displaying the learning material as harder than it is, thus mobilizing additional cognitive resources (Geller et al., 2020). Thus, "perceptual disfluencies" might support long-term memory formation (Alter & Oppenheimer, 2009; Oppenheimer, 2008). As mental normalization puts additional load on the cognitive system (Shepard & Metzler, 1971), the utilization of mental normalization tasks as a perceptual disfluency was investigated in Chapter 6. Participants were asked to answer questions on rotated and unrotated bar graph stimuli. The questions were focused on different types of information, one could derive from the presented stimuli, such as absolute values of single bars or comparison between two bars. After a short break, they were presented with a second phase, testing the implicit learning of the first part of the experiment. The experiment did not show any positive or negative impact of mental normalization on subsequent recall performance. Additional analyses revealed a longer exposure time for rotated stimuli, which might have counteracted a general harmful effect of mental normalization on long-term memory formation. We were not able to identify a positive influence of mental normalization tasks, as they did not act as perceptual disfluency.

7.1 Implications

The results of the three presented studies hold implications for both basic and applied research. As presented in the introduction, mental normalization costs are a well-researched area of cognitive psychology. They have been shown for various types of stimuli, including words (Koriat & Norman, 1985), sentences (Risko et al., 2014), arbitrary shapes (Cooper, 1975), maps (Aretz & Wickens, 1992), and many more. Here, the first major implication for basic research derives from the first experiment in Chapter 4. The experiment aimed to show the presence of mental normalization costs. Indeed, the hypothesis of mental normalization costs for data visualizations (here: bar graphs) was confirmed. Researchers came up with many different models of the mechanism behind mental normalization (Biederman, 1987; Just & Carpenter, 1985; Marr, 2010; Marr & Nishihara, 1978; Tarr & Bülthoff, 1998; Tarr & Pinker, 1989; Xu & Franconeri, 2015; Yuille & Steiger, 1982). Some researchers proposed different mechanisms for different tasks and reported difficulties in discerning the different mechanisms without imaging procedures (Gauthier et al., 2002). Additionally, in Chapter 4, the response time difference between trials with "co-rotated" and "reformatted" labels points towards a piece-meal mental rotation process (Just & Carpenter, 1985; Xu & Franconeri, 2015; Yuille & Steiger, 1982) rather than a holistic rotation of the stimulus (Shepard & Metzler, 1971). One possible explanation lies in the structural difference between labels and bars. Labels can be described as a sequential and descriptive presentation of information, while the bars are rather pictorial and diagrammatic. This claim is supported by the results of the second experiment. In the pictograph conditions the response time differences between the unrotated baseline trials and the rotated trials was smaller for trials that showed the pictographs reformatted than for trials with co-rotating pictograph labels. However, none of the rotated trials with pictograph labels showed a significant difference in response times compared to the pictograph baseline, so the reason for a piece-meal mental rotation requires more research. A second explanation might be, that holistic rotation was mostly observed for shape information (Cooper, 1975; Shepard & Metzler, 1971), but not for stimuli that integrated multiple visual features (Hochberg & Gellman, 1977; Xu & Franconeri, 2015).

The results of the second experiment also contributed to the research on data visualization design. The use of pictographs as means to communicate complex information in a concise form was proposed by Otto Neurath (1936). This claim was supported by research of Haroz et. al. (2015), who found empirical evidence for benefits of the use of ISOTYPE-like icons in terms of processing speed and clarity of communication. The results of the second experiment in Chapter 4 are in line with the findings of Haroz and colleagues. Trials with pictographs as labels elicited an overall shorter response time latency than trials with word labels. The response time difference for co-rotating trials did however not drop below the effect size threshold that was deemed as being a negligible difference to normal, upright viewing conditions. As pictures – like words – usually have an optimal point of view (Palmer et al., 1981), the additional mental normalization cost for label rotation was probably not eradicated but just diminished. It should also be noted that we did not fully replicate the study of Haroz and colleagues (2015), but just one condition. The use of stacked pictographs instead of bars for example might have a completely different effect.

The final intervention tested in Chapter 4 was the use of color as a support or bypass for mental normalization needs. Instead of manipulating the label itself, we manipulated the color of the associated bars and used that same color scheme on beforehand shown the selection task prompts. Color coding is more versatile in its' applichation than pictographs, as pictographs can only be used for categories that can be represented by icons. Color coding has been shown to speed up visual processing tasks such as perceptual grouping (Palmer et al., 2003) or visual search (Wolfe & Horowitz, 2004, 2017). It can also help to guide attention more efficiently (Carter, 1982; Christ, 1975; Duncan & Humphreys, 1989). Indeed, the results of the third experiment in Chapter 4 supported these claims, as color coding reduced the response times below the threshold of negligibility. However, color coding might not be a feasible intervention for every type of data visualization, as the increased use of colors reduced the distance in the color space between them. Also, research shows that increasing the number of colors might hamper performance in visual processing tasks such as feature detection (Haroz & Whitney, 2012).

A different intervention was tested in Chapter 5. Instead of altering or replacing the word label on the bar graph stimuli, we added a second instance of the same label, but rotated it by 180°. We argued that the second label might not only reduce the need for mental normalization, as for each 180° section of the rotation one of the two displays of the label is within ±90°. Additionally, research within the flanker paradigm (B. A. Eriksen & Eriksen, 1974) implies, that the second display of the label might work like a congruent flanker, that supports faster mental processing in decision tasks. The results reported in Chapter 5 are mixed. All three experiments showed that the additional upside-down label is acting like an incongruent flanker instead of like a congruent flanker (B. A. Eriksen & Eriksen, 1974), adding additional load on the cognitive system. However, we discovered a positive effect of the additional label display: While putting additional load on the working memory for the processing for upright viewing conditions, double label displays elicited reduced response times for upside-down displays compared to single label display conditions in experiment 1 and 2. We called this the "justice effect" of double display presentation. It seems as if the redundant display helps for accessing the rotated information (M. R. Morris et al., 2006). We could not find any significant difference between outward and inward rotation of the double display (Koriat & Norman, 1985; Risko et al., 2014; Shepard & Metzler, 1971) or contrast manipulation (C.-C. Lin, 2003). In case of the application to bar graphs, the additional display might overload the information display and hamper menal processing (Meyerhoff et al., 2021).

Possible benefits of mental normalization efforts on long-term memory formation were investigated in Chapter 6. Some researchers proposed that additional load on the visual cognitive system by using so called perceptual disfluencies (Diemand-Yauman et al., 2011; Mulligan, 1996; Rosner et al., 2015) might lead to a deeper processing of the presented information (Geller et al., 2020) while others disagreed (Magreehan et al., 2016; Rhodes & Castel, 2008, 2009; Xie et al., 2018; Yue, Castel, & Bjork, 2013). The results of the experiment presented in Chapter 6 suggest that mental normalization does not yield any benefits for long-term memory formation for information presented in data visualizations. It might even be harmful.

In addition to the contributions to basic research, the reported results of this dissertation also yield implications for applied settings. The first major practical implication concerns the group interaction at the multi-touch table. The studies in this paper highlight the influence of mental normalization affordances on the information access. In the first experiment in Chapter 4, the hypothesis of additional load generated by mental normalization was confirmed for data visualizations. Therefore, the impact of mental normalization processes on equal access to information in group ware settings needs to be considered when designing and applying multi-user interfaces. Several approaches to mitigate this adverse effect were subsequently tested. The use of pictographs instead of word labels showed to be an insufficient solution to the problem of unequal information access. They elicited faster response times than word labels but didn't change the response time structure for different points of view. Therefore, pictographs should be preferred over word labels in group ware interfaces, whenever feasible, but are an inferior solution to color coding the labels. Color coding did not only speed up the visual information processing, but also providing equal access to the information in the data visualization, independent of the point of view. For reasons described earlier in the section of implications for basic research, color coding might not always be feasible, for example if a very large number of shades of color need to be employed. Besides equal group access, quick information communication regardless of the point of view might be low or zero gravity environments (Kanas and Manzey, 2008, p. 59ff), where the spatial orientation of the users cannot be assumed.

If speed of access is not a concern, a double display of word labels can be employed. The double display puts additional strain on the cognitive processing of the presented information. At the same time the justice effect might equalize the access speed for all present users. However, this trade-off needs to be thoroughly considered before implementation, as the additional load might outweigh the benefits of equal access.

We also considered potential long-term benefits that might arise from the additional computing requirements of the cognitive system caused by a divergent point of view such as better long-term memory formation. However, we couldn't find any evidence for a beneficial role of mental normalization load on learning. This does not only highlight the negative impact of divergent points of view on group ware, but also has implications for other areas of learning, such as exhibition design for museums. Information texts and data visualizations might be presented at an angle or in inaccessible areas. This could be a hindering factor for the learning goals of visitors. The results also add to a growing body of research that denies the possible benefit of perceptual fluencies for learning material design (Magreehan et al., 2016; Rhodes & Castel, 2008, 2009; Xie et al., 2018; Yue, Castel, & Bjork, 2013).

7.2 Strengths and Limitations

A considerable strength of the presented research is the compliance to several open science standards (Nosek et al., 2015). Data and material of the presented studies were made accessible via the repository of the open science framework (OSF). The study presented in Chapter 6 was also pre-registered. The experiments also adhered to common standards of good scientific practice, such as a-priori estimation of sample size requirements to detect a certain effect size (Faul et al., 2009).

The inspiration for all experiments were drawn from real-world settings and were subsequently transferred in a laboratory setting. The laboratory setting allowed for controlled experimental designs that minimized the influence of external factors. The inspiration from - and modelling of - applied settings ensured the relevance outside of basic research. This approach allowed for maximization of both internal (Campbell, 1957; H. Lin et al., 2021) and ecological validity (Kieffer, 2017) without sacrificing one for the other.

Furthermore, the design of the experiments provided some unique advantages. All experiments relied on automatic stimulus generation. All stimuli were based on parameter sets randomly drawn from pre-defined pools. This allowed for (true) randomization of the potential confounding influence of material effects on the results. In chapters 4 and 6, each participant received their own set of stimuli that only matched on the relevant factors for the experiments. Due to limitations in the online data collection process for Chapter 5, this complete randomization of study material was not feasible. In Chapter 5, however, an online data collection approach was applied. While the data for the experiments in the chapters 4 and 6 was collected in a laboratory setting in presence, the data for Chapter 5 was derived from an online sample. This approach allowed to a massive upscaling of data collection and allowed for the investigation of smaller effect sizes.

For the experiment in Chapter 6, a low-cost head tracker was invented and used. To control for the head position of the participants, an Arduino based device was developed, that used an electronic gyroscope to track the head position. Other researchers in the lab expressed their interest in using the device, but ultimately chose a different experimental design. The device and an investigation of its accuracy can be found in Appendix A.

Despite all the strengths of this project, some limitations also need to be addressed. All data was collected from student samples which might limit the applicability of the results to other populations. There is some evidence that the interpretation of data visualization needs some previous exposure and knowledge on the type of graph (Börner et al., 2016; Maltese et al., 2015). However, bar graphs were seen as widely familiar to a diverse population and as we were investigating basic cognitive processes, the results should be generalizable.

In contrast to other research on mental normalization (Koriat & Norman, 1984) stimuli rotation was not conducted in 60° but in 90° steps. This reduced resolution was chosen to reflect real-world conditions more accurately. For collaboration around a multi-touch table, group members are usually distributed around the different sides of the table, hence heightening our interest in these angles. The difference in step size might have caused less comparable results as the human visual system is tuned to find right angles (Leibowitz et al., 1955; Smith, 1962). However, the general pattern of the response time latencies matched those of previous findings (Koriat & Norman, 1984), so the loss of resolution in the mental normalization process seems to be acceptable.

The individual studies also suffered from some limitations in their design. In the first experiment in Chapter 4, two independent factors collapsed two condition cells to one, making the preferred analysis method (ANOVA) not applicable. This shortcoming was recognized before data collection, but the experimental design did not permit for a different approach. The analysis with two linear regressions instead of an ANOVA showed to be a sufficient alternative that was fitting better to the collected data. All experiments in Chapter 5 included the manipulation of contrast. All experiments showed no or negligible influence of contrast on the response times elicited by the stimuli. The reason for the lack of influence of contrast might be the non-linear nature of contrast (Beghdadi et al., 2020) and accordingly a too weak manipulation of the different contrast conditions. Contrast might still influence the readability and the potential information overload of the double display of labels in data visualizations, but the presented setup could not detect it. Further research might be needed. The study in Chapter 6 also showed some shortcomings in the design. We did not account for the influence of exposure time on the implicit learning of the data visualization information. This confounded the initial results and made them hard to interpret. Only an additional analysis that partialed out the influence of exposure time shed some light on the underlying mechanisms of mental normalization and long-term memory formation. While these insights were still valuable, the analysis was not included the pre-registered analysis plan and therefore needs to be treated with caution. A second limitation stems from the short delay between the implicit learning phase and the testing phase. Participants were only required to wait for thirty seconds in between both parts of the experiment. A bigger delay or a later follow-up phase would have granted deeper inside on the effectiveness of mental normalization as perceptual disfluency. As the experiment did not show the expected results, the impact of this shortcoming was very limited. Nevertheless, it would have had a major impact if mental normalization had acted as a perceptual disfluency and therefore needed to be addressed.

All studies presented in this dissertation use bar graphs as a data visualization. The simplicity of the bar graphs allowed for a controlled manipulation of various aspects of the data visualization. While the experimental design benefitted from the reduced design, it remains open if the findings can be generalized to other types of data visualizations. While the motivation for this dissertation was inspired by the scenario of co-located group work around a multi-touch table, a study on the effects of unequal access to information presented on the Table was not included. More than 150 ms of additional response time latency on average for non-optimal points were deemed as sufficient evidence for the hampering effect of mental normalization on the group work. There were also practical concerns to be considered. Group work includes verbal and non-verbal communication to coordinate the collective effort influence the outcome. Controlling for these factors is complicated and neglecting these processes might confound the results. Therefore, it was decided against the venture for a group experiment although the generalizability of the results of this project might be somewhat reduced.

7.3 Future Directions

The presented research could be continued in various directions. An exploratory study of group work at the multi-touch Table and the impact of optimized data visualizations would be another tempting extension of the current work. It would be interesting how the interaction between and the participation of the group members is altered by different designs. The development of a concise framework to track group interactions would be advisable. It could extend on previous works of Leilah Lyons (Roberts & Lyons, 2017; Tissenbaum et al., 2017) and the work on epistemic analysis (Andrist et al., 2015; Shaffer et al., 2009; Wooldridge et al., 2018). The potential benefits of point-of-view agnostic data visualizations could also be applied to user interface design for space flight controls. Interviews with astronauts could help to analyze the needs for directionless information displays. A fruitful strain would be the application of the present findings to a broader set of data visualizations. While this dissertation is focused on mental normalization of simple bar graphs, a wide variety of data visualizations is currently employed for information communication. Presenting more complex data could add value to the use of multi-touch tables, but additional affordances might need to be addressed. The independent rotation of labels and the plot also sparks some interesting questions. As text and plot are both visual features of the data visualization and to the best of the present knowledge, both types of information are processed on the same pathways, why then would the different parts of data visualization be treated differently? The independent rotation seems also be activated for pictograph labels despite their more pictorial qualities? What are the factors determining the use of piece-meal instead of holistic mental rotation? It would also be interesting, which features of data visualizations are remembered by participants. The experiment in Chapter 6 showed differences between the different information levels that where tested, prompting the questions what the central information is, that is communicated by a data visualizations and how different types of data visualizations differ in their central messages. A free recall experiment for previous presented data visualization might shed light on the advantages and drawbacks of different types of data visualization. Another interesting pursuit would be to further investigate ISOTYPE properties. In the present research, we built on the results of Haroz and colleagues (2015), whose results supported the claim that pictographs are beneficial for information transfer. However, the ISOTYPE system of Otto Neurath (1936) also comprised other rules for data visualization, the benefits of which is not yet supported by scientific evidence. Finally, a further investigation of the potential benefits and applicability of the "byproducts" of this project would be compelling. Both the Random stimuli generation method and the low-cost head tracker would benefit from thorough research on their potentials and limits.

7.4 Conclusion

The topic for this thesis was derived from research on group work in a co-located setting at a single display groupware, i.e, a Multi-Touch Table. The presented research was concerned with the impact of mental normalization of data visualization on the individual users' visual processing. First, it addressed the question of whether or not the requirement of mental normalization for users located at non-canonical points of view puts additional strain on these users' cognitive systems. Secondly, design options for data visualizations to lessen this additional strain were explored. Finally, the potential benefits of fostering the mental normalization costs for long-term retention were addressed.

The first line of research was concerned with the cost of mental normalization of data visualizations. It encompassed Chapter 4 and Chapter 5. Chapter 4 addressed both, the question on the occurrence of additional strain and design variations to lessen it. The additional burden of mental normalization on the cognitive system was confirmed in the first experiment: Participants reacted slower to completely rotated bar graph diagrams than for both, reformatted and unrotated diagrams. Mental normalization processes therefore do indeed put additional strain on the cognitive system. The second and third experiment in Chapter 4 explored different design options to lessen this additional burden, by employing pictographs as labels and color coding the labels and bars. Both interventions successfully reduced the overall response times to rotated diagrams, with a slightly greater reduction in the color conditions. Chapter 5 exploring more design options on their feasibility to relieve some mental normalization strain. The impact of a double display of the written labels was explored. Displaying the same label twice, but one copy being presented upside-down would limit the extend of mental normalization to 180°at the most, but might come with issues of flanker or contrast inferences. While the results of the first experiment, which focused on the double display of words where promising, the application of this design option to bar graphs showed limited use. The double display put additional strain on the cognitive system, regardless of the diagrams rotation. However, this additional load was higher for unrotated diagram conditions than for rotated displays. This resulted in a levelling of the accessibility of the data visualization for all points of view, but at the cost on a generally less accessible design. The additional strain of mental normalization processes on the cognitive system showed overall to be very robust and resilient to different design interventions.

A second line of study revolved around potential benefits of mental normalization costs for long-term memory formation. In Chapter 6, this was explored. The results suggested an adversal effect of rotated presentation of diagrams on the long-term retention of the displayed information. Thus, mental normalization processes seem not to yield benefits for long-term memory formation.

The thesis provided new insights on the impact of mental normalization demands for rotated display of data visualizations and pointed towards some design options to reduce this impact. The findings have implications for User Interfaces aiming to support collaborative co-located group work on Single Display Groupware like Multi-Touch Tables.

Appendix A

A 7 Dollar Head Tracker

A.1 Introduction

Head-tracking is a crucial procedure in medicine and cognitive psychology. In the medical field, head roll might be an indicator for several diseases and maladies, including Congenital fourth nerve palsy (Kekunnaya & Isenberg, 2014), strabism and nystagsm (Kim et al., 2004). In cognitive psychology, tracking head positions is part of eye-tracking research and offloading literature (Risko & Kingstone, 2011; Risko et al., 2014).

To track the head position several methods are commonly used. Often head roll is measured manually or semi-manually from photos (Farah et al., 2017; Guan et al., 2015) or videos (Risko et al., 2014; Zikovitz & Harris, 1999). Both methods have their drawbacks as the data is not closely linked to other collected data and additional coding by hand is needed. There are specialized instruments to assess head roll (CROM - Cervical Range-of-Motion Instrument — Inclinometer, n.d.) and some mobile eye tracking devices include gyroscopes to log head rotation (Risko & Kingstone, 2011). As mobile eye trackers are rather expensive and Inclinometers also need manual readings, we present a cheap, digital device to measure head roll in experimental conditions.

Arduino devices with gyroscope breakout boards have been used in other application settings, such as human factors and have shown reasonable accuracies (Huang et al., 2018).

A.1.1 Device Description

The device consists of an Arduino Nano V3.0, a Gyroscope GY-521 a USB cable (USB2.0 to Mini-USB, ca. 1.5m), a bread board and four male-to-male jumper wires.

The Arduino platform is an open-source project with the goal to make microcontroller programming accessible to everybody (Arduino.cc, 2021). Arduino boards come in various shapes and sizes. For this project, we decided to use the Arduino Nano board, due to its small size.

For the gyroscope we opted for the GY-521 breakout board with gyroscope and acceleration sensors (AZ-Delivery, 2021). Breakout boards are designed to ease the use of electronic sensors for laypeople. Both components are wired onto the breadboard in a layout as seen in Figure 6.3.

As a platform to mount the device on, we used acoustic earmuffs, but anything with a head band would do, such as headphones or Alice bands The breadboard with the arrangement of the electronic components is fitted to the platform with some zip ties to give it some space for placement correction on the participants' heads (see Figure A.2).

Using the Arduino IDE 1.8.13 ("Arduino IDE", 2020) and the libraries "MPU6050_tockn.h" and "Wire.h", we upload a program on the Arduino to read the x-angle of the gyroscope and send it via the serial connection (see Figure A.3). To help the initial adjustment of the gyroscope, we make use of the on-board LED of the Arduino. The script includes a code block to light the Arduinos on-board LED if the absolute x angle is greater than $1°$, helping the experimentator to adjust the device on the participants' heads at the beginning of the experiment. We were satisfied with an accuracy of $\pm 1°$, but this margin was arbitrary and can be set to the researchers liking. The baud rate was set to 9600.

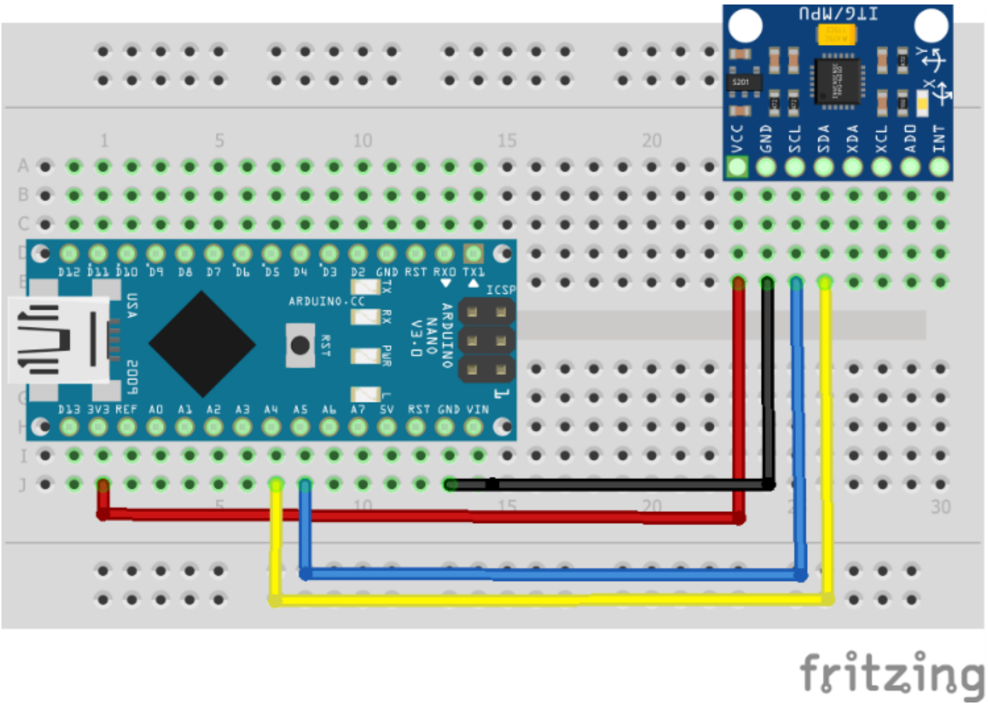

Figure A.1: Wiring on the breadboard for Arduino Nano and the Gyroscope, generated with Fritzing®.

The experiment was conducted in PsychoPy Version 1.85.1 (Peirce, 2007, 2008). To read the most recent data from the serial port, we used multi-threading approach, establishing an additional thread to read and hold the latest sensor reading, that could then be used in the main program (see Figure A.4).

To show the use and accuracy of the device we conducted two demonstrations. In demonstration 1, we show the accuracy of the head tracking gyroscope. For this, we take the breadboard with the Arduino and the Gyroscope and place it on a horizontal line displayed on a screen. In demonstration 2, the head tracker is mounted on the head of participants, who were asked to adjust their head position so that displayed lines are standing perfectly vertical in their field of view.

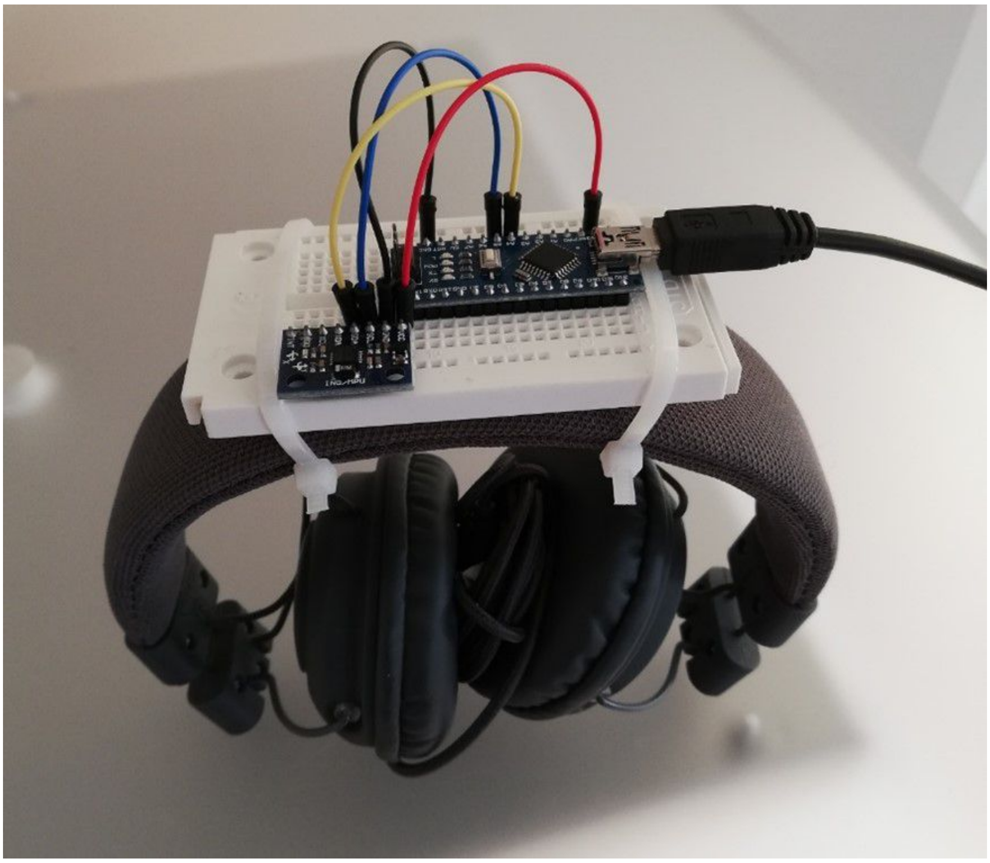

Figure A.2: The assembled head tracker.

A.2 Demonstration 1

In demonstration 1, the Gyroscope was used without subjects to show the general feasibility of using a gyroscope.

A.2.1 Methods

Apparatus

The demonstration was conducted on a HP Elitebook 8530p with PsychoPy 1.85.1. The stimuli were presented on a $23''$ monitor (Dell Panel Monitor S2340Tt). The described device for head tracking was used, but without the head mount (see Figure A.3 A).

Procedure

Thirteen horizontal lines, encompassing angles from -30° to 30° in 5° steps, were presented on the monitor in a random order. The breadboard with the mounted Arduino and gyroscope were held against the monitor. When the experimenter was satisfied with the alignment, they pressed space to record the gyroscope angle. This procedure was repeated 5 times.

A.2.2 Results

The recorded angles deviated by ±1.5° from the stimuli lines (see Figure A.3 B). It is unclear if this is due to the imprecision of the gyroscope or inaccuracy by the experimenter. The bimodal distribution of residuals (see Figure A.3 C) might suggest, that imprecision is more to blame to the experimenter than to the device itself.

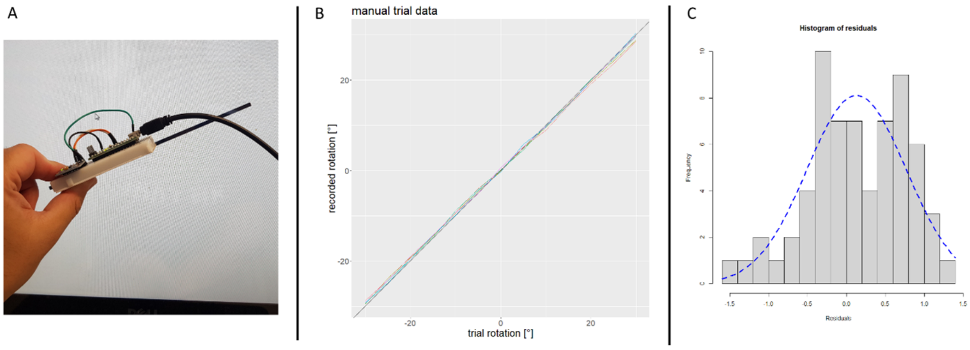

Figure A.3: A: Depiction of the demo 1 procedure. B: Line graph of the recorded angles by trial angles. C: Histogram of the residual angles (recorded angles - trial line angles).

A.3 Demonstration 2

In demonstration 2 we show an example use of the gyroscope for data collection with participants.

A.3.1 Methods

Apparatus

The head tracker from experiment 1 was mounted to the ridge of a pair of earmuffs. Stimuli were presented on an elevated

Procedure

10 participants (6 female), age 26–52 (M = 32.1, SD = 8.2) were asked to roll their head until they see a vertical line and then log the roll of their head by pressing the space bar on the keyboard. The experiment was set up on an HP Elitebook 8530p with an external monitor (Dell Panel Monitor S2340Tt) elevated to be level with the participants' heads (see Figure A.4).

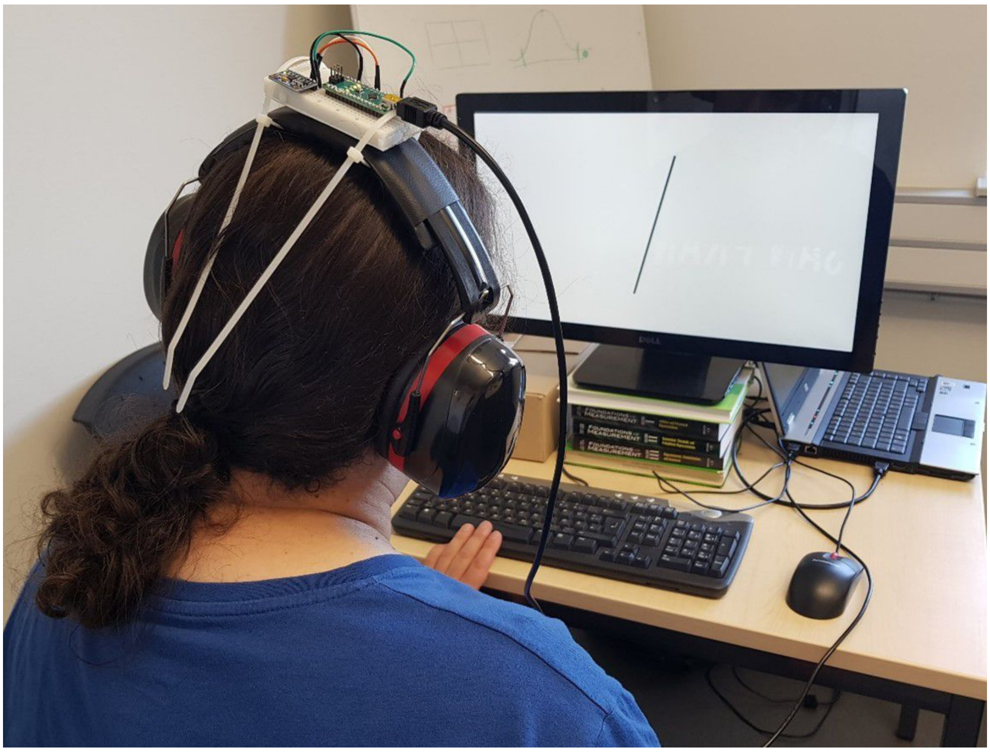

Figure A.4: Setup for demonstration 2.

A.3.2 Results

The head-mounted head-tracker showed reliable data transmission. Every time the demonstration required to record data from the head tracker, the data was tracked. Although they are not as perfectly aligned with the trial angle, all recorded data points seem to be plausible (see Figure A.5). Deviations from the trial angles seems to be rather due to human imperfection than tracking device inaccuracies – especially in combination with demonstration 1.

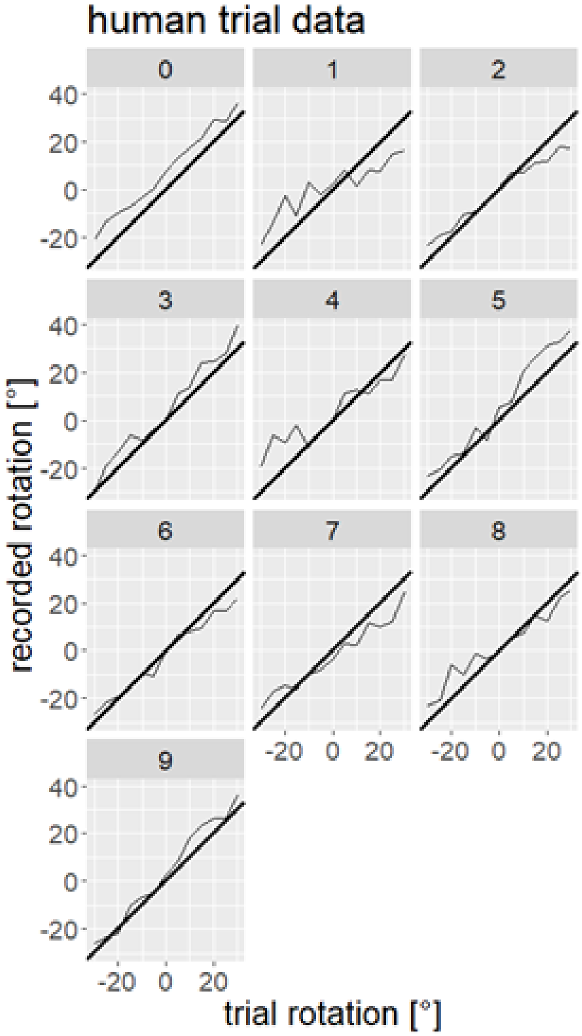

Figure A.5: head tracking data for all 10 participants.

A.3.3 Discussion

In this short paper, we proposed a cheap, lightweight Arduino-based device for head position tracking. We demonstrated reasonably high accuracy of the tracking device, making it suiTable for head position tracking in psychological experiments. The device is less expensive than a full head-mounted eye-tracking device, making it suiTable for medium to large scale testing. However, the device has to be carefully calibrated each time it is used to give accurate sensor readings. We could also not do away with the concerns of (Risko et al., 2014), who mentioned that wearing a head tracking device might alter behavior as participants are constantly reminded that they are tracked. Our experiments only covered the "roll" axis of head rotation and thus cannot be extrapolated to other head motion patterns.

A.3.4 Declarations

Ethics Approval and Consent to Participate

The experimental procedure was approved by the institutional review board of the Leibniz-Institut für Wissensmedien (LEK 2018/077). All participants signed informed consent prior to testing.

Consent for Publication

The person in Figure A.4 agreed with the publication of the image.

Availability of Data and Material

Study materials, data, and analyses are available at the Open Science Framework (https://osf.io/qgrxp/?view_only=76459a6d9b544104a719facdeb1a9fb1)

Competing Interests

The authors declared they have no conflicts of interest with respect to their authorship and the publication of this article.

Funding

Funding by the German Federal Ministry of Education and Research (BMBF) 2016-2019 under grant number 01IO1616.

References

Alibali, M. W., & DiRusso, A. A. (1999). The function of gesture in learning to count: More than keeping track. *Cognitive Development*, *14*(1), 37–56. https://doi.org/10.1016/S0885-2014(99)80017-3

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Lawrence Erlbaum Associates Publishers.

Andrist, S., Collier, W., Gleicher, M., Mutlu, B., & Shaffer, D. (2015). Look together: Analyzing gaze coordination with epistemic network analysis. *Frontiers in Psychology*, *6*, 1016. https://doi.org/10.3389/fpsyg.2015.01016

Arduino IDE. (2020).

Arduino.cc. (2021). What is Arduino.

Aretz, A. J., & Wickens, C. D. (1992). The mental rotation of map displays. *Human Performance*, *5*(4), 303–328. https://doi.org/10.1207/s15327043hup0504

Aubert, M., Lebe, R., Oktaviana, A. A., Tang, M., Burhan, B., Hamrullah, Jusdi, A., Abdullah, Hakim, B., Zhao, J.-x., Geria, I. M., Sulistyarto, P. H., Sardi, R., & Brumm, A. (2019). Earliest hunting scene in prehistoric art. *Nature*, *576*(7787), 442–445. https://doi.org/10.1038/s41586-019-1806-y

Aung, T., Niyeha, D., Shagihilu, S., Mpembeni, R., Kaganda, J., Sheffel, A., & Heidkamp, R. (2019). Optimizing data visualization for reproductive, maternal, newborn, child health, and nutrition (RMNCH&N) policymaking: Data visualization preferences and interpretation capacity among decision-makers in Tanzania. *Global Health Research and Policy*, *4*(1), 4. https://doi.org/10.1186/s41256-019-0095-1

Aydmune, Y., Introzzi, I., & Lipina, S. (2019). Inhibitory Processes Training for School-age Children: Transfer Effects. *Developmental Neuropsychology*, *44*(7), 513–542. https://doi.org/10.1080/87565641.2019.1677667

AZ-Delivery. (2021). GY-521.

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Babbage, C. (1832). Specimen of Logarithmic Tables printed with Different Coloured Inks on Variously Coloured Papers. *The Edinburgh Journal of Science*, *6*(1), 144–150.

Baddeley, A. D. (1979). Working Memory and Reading. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of Visible Language* (pp. 355–370). Springer US. https://doi.org/10.1007/978-1-4684-0994-9_21

Baddeley, A. D. (1986). *Working memory*. Clarendon Press/Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D. (2012). Working Memory: Theories, models, and controversies. , *Annual Review of Psychology*, *63*(1), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, A. D., Eysenck, M. W., & Anderson, M. C. (2009). *Memory*. Psychology Press.
OCLC: ocn237880710.

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158–173. https://doi.org/10.1037/0033-295X.105.1.158

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (pp. 47–89, Vol. 8). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Baker, D. P., & Salas, E. (1997). Principles for Measuring Teamwork: A Summary and Look Toward the Future. In *Team Performance Assessment and Measurement*. Psychology Press.

Bang, D., & Frith, C. D. (2017). Making better decisions in groups. *Royal Society Open Science*, *4*(8), 170193. https://doi.org/10.1098/rsos.170193

Bateman, S., Mandryk, R. L., Gutwin, C., Genest, A., Mcdine, D., & Brooks, C. (2010). Useful Junk ? The Effects of Visual Embellishment on Comprehension and Memorability of Charts. *Human Factors*, 2573–2582. https://doi.org/10.1145/1753326.1753716

Bause, I. M., Brich, I. R., Wesslein, A.-K., & Hesse, F. W. (2018). Using technological functions on a multi-touch table and their affordances to counteract biases and foster collaborative problem solving. *International Journal of Computer-Supported Collaborative Learning*, *13*(1), 7–33. https://doi.org/10.1007/s11412-018-9271-4

Beentjes, J. W. J. (1989). Learning from television and books: A dutch replication study based on salomon's model. *Educational Technology Research and Development*, *37*(2), 47–58. https://doi.org/10.1007/BF02298289

Beghdadi, A., Qureshi, M. A., Amirshahi, S. A., Chetouani, A., & Pedersen, M. (2020). A Critical Analysis on Perceptual Contrast and Its Use in Visual Information Analysis and Processing. *IEEE Access*, *8*, 156929–156953. https://doi.org/10.1109/ACCESS.2020.3019350

Beitzel, B. D., & Staley, R. K. (2015). The Efficacy of Using Diagrams When Solving Probability Word Problems in College. *The Journal of Experimental Education*, *83*(1), 130–145. https://doi.org/10.1080/00220973.2013.876232

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*(2), 201–210. https://doi.org/10.3758/BF03193441

Besner, D. (1983). Visual pattern recognition: Size preprocessing re-examined. *The Quarterly Journal of Experimental Psychology Section A*, *35*(1), 209–216. https://doi.org/10.1080/14640748308402126

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation : Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 12–23.

Biederman, I. (1987). Recognition-by-Components: A theory of human image understanding. *Psychological Review*, *94*(2), 115–147. https://doi.org/10.1037/0033-295X.94.2.115

Bigelow, C. (2019). Typeface features and legibility research. *Vision Research*, *165*(June 2018), 162–172. https://doi.org/10.1016/j.visres.2019.05.003

Bjork, E. L., & Bjork, R. A. (2009). Making Things Hard on Yourself, But in a Good Way: Creating Desirable Difficulties to Enhance Learning. In M. A. Gernsbacher & J. R. Pomerantz (Eds.), *Psychology and the real world* (p. 304). Worth Publishers. https://doi.org/10.1017/CBO9781107415324.004

Bodemer, D., Janssen, J., & Schnaubert, L. (2018). Group Awareness Tools for Computer-Supported Collaborative Learning. In *International Handbook of the Learning Sciences*. Routledge.

Borgo, R., & Abdul-Rahman, A. (2012). An empirical study on using visual embellishments in visualization. *Visualization and ...*, *18*(12), 2759–2768. https://doi.org/10.1109/TVCG.2012.197

Borkin, M. A., Vo, A. A., Bylinskii, Z., Isoa, P., Sunkavalli, S., Oliva, A., & Phister, H. (2013). What makes a data visualization memorable ? *IEEE Transactions on Visualization & Computer Graphics*, *19*(12), 2306–2315. https://doi.org/10.1109/TVCG.2013.234

Börner, K., Maltese, A., Balliet, R. N., & Heimlich, J. (2016). Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization*, *15*(3), 198–213. https://doi.org/10.1177/1473871615594652

Brich, I. R., Bause, I. M., Hesse, F. W., & Wesslein, A.-K. (2019). Working memory affine technological support functions improve decision performance. *Computers in Human Behavior*, *92*, 238–249. https://doi.org/10.1016/j.chb.2018.11.014

Brich, I. R., Bause, I. M., Hesse, F. W., & Wesslein, A.-K. (2021). How spatial information structuring in an interactive technological environment affects decision performance under working memory load. *Computers in Human Behavior*, *123*, 106860. https://doi.org/10.1016/j.chb.2021.106860

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. https://doi.org/10.31234/OSF.IO/D6NU5

Buchsbaum, B. R. (2013). The role of consciousness in the phonological loop: Hidden in plain sight. *Frontiers in Psychology*, *4*, 496. https://doi.org/10.3389/fpsyg.2013.00496

Buisine, S., Besacier, G., Aoussat, A., & Vernier, F. D. (2012). How do interactive tabletop systems influence collaboration? *Computers in Human Behavior*, *28*(1), 49–59. https://doi.org/10.1016/j.chb.2011.08.010

Bundesen, C., & Larsen, A. (1975). Visual transformation of size. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(3), 214–220. https://doi.org/10.1037/0096-1523.1.3.214

Burnett, S. A. (1986). Sex-related differences in spatial ability: Are they trivial? *American Psychologist*, *41*(9), 1012–1014. https://doi.org/10.1037/0003-066X.41.9.1012

Byrne, M. D. (2002). Reading Vertical Text: Rotated VS. marquee. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *46*(17), 1633–1635. https://doi.org/10.1177/154193120204601722

Cahill, M. C., & Carter, R. C. (1976). Color Code Size for Searching Displays of Different Density. *Human Factors: The Journal of Human Factors and Ergonomics Society*, *18*(3), 273–280. https://doi.org/10.1177/001872087601800308

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, *54*, 297–312. https://doi.org/10.1037/h0040950

Carayannis, E. G., & Campbell, D. F. (2009). 'Mode 3' and 'Quadruple Helix': Toward a 21st century fractal innovation ecosystem. *International Journal of Technology Management*, *46*(3-4), 201–234. https://doi.org/10.1504/IJTM.2009.023374

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations Still improve students' learning from text. *Educational Psychology Review*, *14*(1), 5–26. https://doi.org/10.1023/A:1013176309260

Carter, R. C. (1982). Visual search with color. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(1), 127–136. https://doi.org/10.1037/0096-1523.8.1.127

Casner, S. M., Geven, R. W., Recker, M. P., & Schooler, J. W. (2014). The Retention of Manual Flying Skills in the Automated Cockpit. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *56*(8), 1506–1516. https://doi.org/10.1177/0018720814535628

Cattaneo, Z., Postma, A., & Vecchi, T. (2007). The picture superiority effect in working memory for spatial and temporal order. *Psychologia*, *50*, 102–109.

Chernoff, H. (1973). The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, *68*(342), 361–368. https://doi.org/10.1080/01621459.1973.10482434

Chokron, S., Kazandjian, S., & De Agostini, M. (2009). Effects of Reading Direction on Visuospatial Organization : A Critical Review. In G. Aikaterini & K. Mylonas (Eds.), *Quod Erat Demonstrandum: From Herodotus' ethnographic journeys to cross-cultural research: Proceedings from the 18th International Congress of the International Association for Cross-Cultural Psychology.*

Christ, R. E. (1975). Review and analysis of color coding research for visual displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *17*(6), 542–570. https://doi.org/10.1177/001872087501700602

Christen, M., Brugger, P., & Fabrikant, S. I. (2021). Susceptibility of domain experts to color manipulation indicate a need for design principles in data visualization. *PLOS ONE*, *16*(2), e0246479. https://doi.org/10.1371/journal.pone.0246479

Chvátil, V. (2015). Codenames.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, *13*(2), 259–294. https://doi.org/10.1016/0364-0213(89)90008-6

Cleveland, W. S., & McGill, R. (1984). Graphical Perception: Theory, Experimentatio, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, *79*(387), 531–554.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. Erlbaum.

Collins, D. W., & Kimura, D. (1997). A large sex difference on a two-dimensional mental rotation task. *Behavioral Neuroscience*, *111*(4), 845–849. https://doi.org/10.1037/0735-7044.111.4.845

Conrad, J., Shah, A. H., Divino, C. M., Schluender, S., Gurland, B., Shlasko, E., & Szold, A. (2006). The role of mental rotation and memory scanning on the performance of laparoscopic skills: A study on the effect of camera rotational angle. *Surgical Endoscopy and Other Interventional Techniques*, *20*(3), 504–510. https://doi.org/10.1007/s00464-005-0363-7

Conrad, R. (1964). Acoustic Confusions in Immediate Memory. *British Journal of Psychology*, *55*(1), 75–84. https://doi.org/10.1111/j.2044-8295.1964.tb00899.x

Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology*, *7*(1), 20–43. https://doi.org/10.1016/0010-0285(75)90003-1

Cooper, L. A., & Podgorny, P. (1976). Mental transformations and visual comparison processes: Effects of complexity and similarity. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(4), 503–514. https://doi.org/10.1037/0096-1523.2.4.503

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. *Visual Information Processing: Proceedings*, *6*, 555. https://doi.org/10.1111/j.1467-9280.1993.tb00468.x

Corballis, M. C., & McLaren, R. (1984). Winding one's Ps and Qs: Mental rotation and mirror-image discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(2), 318–327. https://doi.org/10.1037/0096-1523.10.2.318

Costa, A., Silva, J., Pinheiro-Chagas, P., Krinzinger, H., Lonnemann, J., Willmes, K., Wood, G., & Haase, V. (2011). A Hand Full of Numbers: A Role for Offloading in Arithmetics Learning? *Frontiers in Psychology*, *2*.

Cowan, N. (1999). An Embedded-Processes Model of Working Memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 62–101). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.006

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Dambacher, M., Haffke, P., Groß, D., & Hübner, R. (2016). Graphs versus numbers: How information format affects risk aversion in gambling. *Judgment and Decision Making*, *11*(3), 223–242.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, *15*(6), 254–262. https://doi.org/10.1016/j.tics.2011.04.003

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the: Effects of disfluency on educational outcomes. *Cognition*, *118*(1), 111–115. https://doi.org/10.1016/j.cognition.2010.09.012

Dietz, P., & Leigh, D. (2001). DiamondTouch. *Proceedings of the 14th annual ACM symposium on User interface software and technology - UIST '01*, (January 2001), 219. https://doi.org/10.1145/502348.502389

Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science*, *8*(4), 302–307. https://doi.org/10.1111/j.1467-9280.1997.tb00442.x

Dror, I., & Harnad, S. (2008). Offloading Cognition onto Cognitive Technology. In I. Dror & S. Harnad (Eds.), *Cognition Distributed: How Cogitive Technology Extends Our Minds*. John Benjamins.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity contents. *Perception & psychophysics*, *54*(6), 716–732.

Ebbatson, M., Harris, D., Huddlestone, J., & Sears, R. (2010). The relationship between manual handling performance and recent flying experience in air transport pilots. *Ergonomics*, *53*(2), 268–277. https://doi.org/10.1080/00140130903342349

Eitel, A., Scheiter, K., & Schüler, A. (2012). The time course of information extraction from instructional diagrams. *Perceptual and Motor Skills*, *115*(3), 677–701. https://doi.org/10.2466/22.23.PMS.115.6.677-701

Ellis, R., Allport, D. A., Humphreys, G. W., & Collis, J. (1989). Varieties of Object Constancy. *The Quarterly Journal of Experimental Psychology Section A*, *41*(4), 775–796. https://doi.org/10.1080/14640748908402393

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics*, *25*(4), 249–263. https://doi.org/10.3758/BF03198804

Eskritt, M., & Ma, S. (2014). Intentional forgetting: Note-taking as a naturalistic example. *Memory & Cognition*, *42*(2), 237–246. https://doi.org/10.3758/s13421-013-0362-1

Evergreen, S. D. H. (2017). *Effective data visualization*. Sage.

Fairchild, M. D. (2013). *Color Appearance Models* (1st ed.). Wiley. https://doi.org/10.1002/9781118653128

Farah, M. L., Santinello, M., Carvalho, L. E., & Barcellos, R. (2017). A simple method to measure head tilt using iPhone®. *Investigative Ophthalmology & Visual Science*, *58*(8), 2933.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149–60. https://doi.org/10.3758/BRM.41.4.1149

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175–91.

Fechner, G. T. (1859). *Über ein psychophysisches Grundgesetz und dessen Beziehung zur Schätzung von Sterngrößen*. Hirzel.

Fenech, E. P., Drews, F. A., & Bakdash, J. Z. (2010). The Effects of Acoustic Turn-by-turn Navigation on Wayfinding. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*(23), 1926–1930. https://doi.org/10.1177/154193121005402305

Fisher, D. M. (2014). Distinguishing between taskwork and teamwork planning in teams: Relations with coordination and interpersonal processes. *Journal of Applied Psychology*, *99*(3), 423–436. https://doi.org/10.1037/a0034625

Fisher, R. (1935). *The design of experiments*. Hafner Press.

Fraunhofer-Institut für integrierte Schaltungen. (2022). mp3.

Frege, G. (1918). Der Gedanke. *Beiträge Zur Philosophie des Deutschen Idealismus*, *2*, 58–77.

Gardony, A. L., Brunyé, T. T., Mahoney, C. R., & Taylor, H. A. (2013). How Navigational Aids Impair Spatial Memory: Evidence for Divided Attention. *Spatial Cognition & Computation*, *13*(4), 319–350. https://doi.org/10.1080/13875868.2013.792821

Gardony, A. L., Brunyé, T. T., & Taylor, H. A. (2015). Navigational Aids and Spatial Memory Impairment: The Role of Divided Attention. *Spatial Cognition & Computation*, *15*(4), 246–284. https://doi.org/10.1080/13875868.2015.1059432

Gauthier, I., Hayward, W. G., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (2002). BOLD activity during mental rotation and viewpoint-dependent object recognition. *Neuron*, *34*(1), 161–71. https://doi.org/10.1016/s0896-6273(02)00622-0

Gelb, I. (n.d.). Sumerian Language. *Encyclopedia Britannica.*

Geller, J., Davis, S. D., & Peterson, D. J. (2020). Sans Forgetica is not desirable for learning. *Memory*, *28*(8), 957–967. https://doi.org/10.1080/09658211.2020.1797096

Goldberg, J. (2016). It Takes A Village To Determine The Origins Of An African Proverb. *NPR.*

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Grinschgl, S., Papenmeier, F., & Meyerhoff, H. S. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology*, *74*(9), 1477–1496. https://doi.org/10.1177/17470218211008060

Grondin, S. (2016). *Psychology of Perception.* Springer International Publishing. https://doi.org/10.1007/978-3-319-31791-5

Guan, X., Fan, G., Wu, X., Zeng, Y., Su, H., Gu, G., Zhou, Q., Gu, X., Zhang, H., & He, S. (2015). Photographic measurement of head and cervical posture when viewing mobile phone: A pilot study. *European Spine Journal*, *24*(12), 2892–2898. https://doi.org/10.1007/s00586-015-4143-3

Hall, E. T. (1963). A System for the Notation of Proxemic Behavior. *American Anthropologist*, *65*(5), 1003–1026. https://doi.org/10.1525/aa.1963.65.5.02a00020

Haroz, S., Kosara, R., & Franconeri, S. L. (2015). ISOTYPE Visualization. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 1191–1200. https://doi.org/10.1145/2702123.2702275

Haroz, S., & Whitney, D. (2012). How capacity limits of attention influence information visualization effectiveness. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2402–2410. https://doi.org/10.1109/TVCG.2012.233

Hayward, W. G., & Williams, P. (2000). Viewpoint dependence and object discriminability. *Psychological Science*, *11*(1), 7–12. https://doi.org/10.1111/1467-9280.00207

Heath, C., & Luff, P. (1992). Collaboration and controlCrisis management and multimedia technology in London Underground Line Control Rooms. *Computer Supported Cooperative Work (CSCW)*, *1*(1), 69–94. https://doi.org/10.1007/BF00752451

Heil, M., & Jansen-Osmann, P. (2008). Sex differences in mental rotation with polygons of different complexity: Do men utilize holistic processes whereas women prefer piecemeal ones? *Quarterly journal of experimental psychology (2006)*, *61*(5), 683–9. https://doi.org/10.1080/17470210701822967

Henkel, L. A. (2014). Point-and-Shoot Memories: The Influence of Taking Photos on Memory for a Museum Tour. *Psychological Science*, *25*(2), 396–402. https://doi.org/10.1177/0956797613504438

Henninger, F., Shevchenko, Y., Mertens, U., Kieslich, P. J., & Hilbig, B. E. (2019). Lab.js: A free, open, online study builder. https://doi.org/10.31234/OSF.IO/FQR49

Hertzog, C., & Rypma, B. (1991). Age differences in components of mental-rotation task performance. *Bulletin of the Psychonomic Society*, *29*(2), 209–212. https://doi.org/10.3758/BF03335237

Hesselmann, G. (2020). No conclusive evidence that difficult general knowledge questions cause a "Google Stroop effect". A replication study. *PeerJ*, *8*. https://doi.org/10.7717/peerj.10325

Higgins, S. E., Mercier, E. M., Burd, E., & Hatch, A. (2011). Multi-touch tables and the relationship with collaborative classroom pedagogies: A synthetic review. *International Journal of Computer-Supported Collaborative Learning*, *6*(4), 515–538. https://doi.org/10.1007/s11412-011-9131-y

Higgins, S. E., Mercier, E. M., Burd, E., & Joyce-Gibbons, A. (2012). Multi-touch tables and collaborative learning. *British Journal of Educational Technology*, *43*(6), 1041–1054. https://doi.org/10.1111/j.1467-8535.2011.01259.x

Hintzman, D. L., O'Dell, C. S., & Arndt, D. R. (1981). Orientation in cognitive maps. *Cognitive Psychology*, *13*(2), 149–206. https://doi.org/10.1016/0010-0285(81)90007-4

Hochberg, J., & Gellman, L. (1977). The effect of landmark features on mental rotation times. *Memory & Cognition*, *5*(1), 23–26. https://doi.org/10.3758/BF03209187

Holsanova, J., Holmberg, N., & Holmqvist, K. (2009). Reading information graphics: The role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology*, *23*(9), 1215–1226. https://doi.org/10.1002/acp.1525

Hou, F., Lu, Z.-L., Bex, P., & Reynaud, A. (2021). Editorial: The Contrast Sensitivity Function: From Laboratory to Clinic. *Frontiers in Neuroscience*, *15*.

Howell, W. C., & Kraft, C. L. (1959). *Size, Blur, and Contrast as Variables Affecting the Legibility of Alpha-numeric Symbols on Radar-type Displays* (tech. rep.). Aerospace Medical Laboratory, Wright Air Development Center, Air Research and Development Command, United States Air Force.
https://play.google.com/store/books/details?id=_6wnAAAAMAAJ.

Huang, S.-M., Gi, P.-Y., & Wu, P.-Y. (2018). A Design of Wearable Goniometers for Measuring Carpal Angular Movement. *2018 IEEE International Conference on Advanced Manufacturing (ICAM)*, 1–4. https://doi.org/10.1109/AMCON.2018.8615058

Hullman, J., Adar, E., & Shah, P. (2011). Benefitting InfoVis with visual difficulties. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2213–2222. https://doi.org/10.1109/TVCG.2011.175

IJsfontein. (2018). Beschützt die Ausstellung!

Janis, I. L. (1972). *Victims of Groupthink: A psychological study of foreign-policy decisions and fiascoes*. Mifflin Company.

Järvelä, S., Kirschner, P. A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., Koivuniemi, M., & Järvenoja, H. (2015). Enhancing socially shared regulation in collaborative learning groups: Designing for CSCL regulation tools. *Educational Technology Research and Development*, *63*(1), 125–142. https://doi.org/10.1007/s11423-014-9358-1

Jenkins, J. R., Neale, D. C., & Deno, S. L. (1967). Differential memory for picture and word stimuli 1. *Journal of Educational Psychology*, *88*(5), 303–307.

Jolicoeur, P. (1987). A size-congruency effect in memory for visual shape. *Memory & cognition*, *15*(6), 531–43. https://doi.org/10.3758/bf03198388

Jolicoeur, P. (1990). Orientation congruency effects on the identification of disoriented shapes. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(2), 351–364. https://doi.org/10.1037/0096-1523.16.2.351

Jolicoeur, P., & Besner, D. (1987). Additivity and interaction between size ratio and response category in the comparison of size-discrepant shapes. *Journal of experimental psychology. Human perception and performance*, *13*(3), 478–87. https://doi.org/10.1037//0096-1523.13.3.478

Jones, B., & Anuza, T. (1982). Effects of sex, handedness, stimulus and visual field on “Mental rotation”. *Cortex*, *18*(4), 501–514. https://doi.org/10.1016/S0010-9452(82)80049-X

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological review*, *92*(2), 137–172.

Kanas, N., & Manzey, D. (2008). *Space psychology and psychiatry*. Springer Dodrecht.
https://doi.org/10.1007/978-1-4020-6770-9.

Kekunnaya, R., & Isenberg, S. (2014). Effect of strabismus surgery on torticollis caused by congenital superior oblique palsy in young children. *Indian Journal of Ophthalmology*, *62*(3), 322. https://doi.org/10.4103/0301-4738.116468

Kelly, M. O., & Risko, E. F. (2019a). Offloading memory: Serial position effects. *Psychonomic Bulletin & Review*, *26*(4), 1347–1353. https://doi.org/10.3758/s13423-019-01615-8

Kelly, M. O., & Risko, E. F. (2019b). The isolation effect when offloading memory. *Journal of Applied Research in Memory and Cognition*, *8*(4), 471–480. https://doi.org/10.1037/h0101842

Kharrufa, A., Leat, D., & Olivier, P. (2010). Digital mysteries: Designing for learning at the tabletop. *ACM International Conference on Interactive Tabletops and Surfaces - ITS '10*, 197. https://doi.org/10.1145/1936652.1936689

Khooshabeh, P., Hegarty, M., & Shipley, T. F. (2013). Individual Differences in Mental Rotation. *Experimental Psychology*, *60*(3), 164–171. https://doi.org/10.1027/1618-3169/a000184

Kieffer, S. (2017). ECOVAL: Ecological Validity of Cues and Representative Design in User Experience Evaluations. *AIS Transactions on Human-Computer Interaction*, *9*(2), 149–172.

Kieras, D. E. (2016). *A Summary of the EPIC Cognitive Architecture* (S. E. F. Chipman, Ed.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199842193.013.003

Kim, D.-S., Coats, D. K., McCreery, K. M., Paysse, E. A., & Wilhelmus, K. R. (2004). Accuracy of clinical estimation of abnormal head postures. *Binocular vision & strabismus quarterly*, *19*(1), 21–24.

Kirsh, D. (2010). Thinking with external representations. *AI and Society*, *25*(4), 441–454. https://doi.org/10.1007/s00146-010-0272-8

Kolers, P. A. (1968). The recognition of geometrically transformed text. *Perception & Psychophysics*, *3*(1-B), 57–64. https://doi.org/10.3758/BF03212713

Kombartzky, U., Ploetzner, R., Schlag, S., & Metz, B. (2010). Developing and evaluating a strategy for learning from animations. *Learning and Instruction*, *20*(5), 424–433. https://doi.org/10.1016/j.learninstruc.2009.05.002

Kopp, B., Mattler, U., & Rist, F. (1994). Selective attention and response competition in schizophrenic patients. *Psychiatry Research*, *53*(2), 129–139. https://doi.org/10.1016/0165-1781(94)90104-X

Koriat, A., & Norman, J. (1984). What is rotated in mental rotation? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(3), 421–434. https://doi.org/10.1037/0278-7393.10.3.421

Koriat, A., & Norman, J. (1985). Reading rotated words. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(4), 490–508. https://doi.org/10.1037/0096-1523.11.4.490

Kornell, N., & Vaughn, K. E. (2016). How Retrieval Attempts Affect Learning. In *Psychology of Learning and Motivation* (pp. 183–215, Vol. 65). Elsevier. https://doi.org/10.1016/bs.plm.2016.03.003

Kosara, R., & MacKinlay, J. (2013). Storytelling: The next step for visualization. *Computer*, *46*(5), 44–50. https://doi.org/10.1109/MC.2013.36

Kretschmer, K. (1962). *Die italienischen Portolane des Mittelalters: ein Beitrag zur Geschichte der Kartographie und Nautik*. G. Olms.

Kruger, R., Carpendale, S., Scott, S. D., & Greenberg, S. (2004). Roles of orientation in tabletop collaboration: Comprehension, coordination and communication. *Computer Supported Cooperative Work*, *13*(5-6), 501–537. https://doi.org/10.1007/s10606-004-5062-8

Kupke, S., Rahn, J., Müller, T., & Oestermeier, U. (2019). Ideas to Market 3.0: Entwicklung eines toolgestützten Vorgehensmodells zur interdisziplinären und branchenübergreifenden Generierung von Verwertungsoptionen.

LaFortune, J., & Macuga, K. L. (2018). Learning movements from a virtual instructor: Effects of spatial orientation, immersion, and expertise. *Journal of Experimental Psychology: Applied*, *24*(4), 521–533. https://doi.org/10.1037/xap0000189

Larkin, J. H., & Simon, H. A. (1987). Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science*, *11*(1), 65–100. https://doi.org/10.1016/S0364-0213(87)80026-5

Latané, B., & Nida, S. (1981). Ten years of research on group size and helping. *Psychological Bulletin*, *89*(2), 308–324. https://doi.org/10.1037/0033-2909.89.2.308

Lawson, R., Humphreys, G. W., & Watson, D. G. (1994). Object recognition under sequential viewing conditions: Evidence for viewpoint-specific recognition procedures. *Perception*, *23*(5), 595–614. https://doi.org/10.1068/p230595

Lee, M. D., Reilly, R. E., & Butavicius, M. A. (2003). An Empirical Evaluation of Chernoff Faces , Star Glyphs , and Spatial Visualizations for Binary Data. *Proceeding APVis '03 Proceedings of the Asia-Pacific symposium on Information visualisation - Volume 24*, *24*, 1–10.

Legge, G. E., Parish, D. H., Luebker, A., & Wurm, L. H. (1990). Psychophysics of reading XI Comparing color contrast and luminance contrast. *Journal of the Optical Society of America A*, *7*(10), 2002. https://doi.org/10.1364/josaa.7.002002

Legge, G. E., Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading-I. Normal Vision. *Vision Research*, *25*(2), 239–252.

Legge, G. E., Rubin, G. S., & Luebker, A. (1987). Psychophysics of reading-V. The role of contrast in normal vision. *Vision Research*, *27*(7), 1165–1177. https://doi.org/10.1016/0042-6989(87)90028-9

Legge, G. E., Rubin, G. S., Pelli, D. G., & Schleske, M. M. (1985). Psychophysics of reading - II. low vision. *Vision Research*, *25*(2), 253–266.

Leibowitz, H. W., Myers, N. A., & Grant, D. A. (1955). Radial Localization of a Single Stimulus as a Function of Luminance and Duration of Exposure*. *Journal of the Optical Society of America*, *45*(2), 76. https://doi.org/10.1364/JOSA.45.000076

Lin, C.-C. (2003). Effects of contrast ratio and text color on visual performance with TFT-LCD. *International Journal of Industrial Ergonomics*, *31*(2), 65–72. https://doi.org/10.1016/S0169-8141(02)00175-0

Lin, H., Werner, K. M., & Inzlicht, M. (2021). Promises and Perils of Experimentation: The Mutual-Internal-Validity Problem. *Perspectives on Psychological Science*, *16*(4), 854–863. https://doi.org/10.1177/1745691620974773

Lin, S., Fortuna, J., Kulkarni, C., Stone, M., & Heer, J. (2013). Selecting semantically-resonant colors for data visualization. *Computer Graphics Forum*, *32*(3 PART4), 401–410. https://doi.org/10.1111/cgf.12127

Lindgren, M., Stenberg, G., & Rosén, I. (1996). Effects of nicotine in visual attention tasks. *Human Psychopharmacology: Clinical and Experimental*, *11*(1), 47–51. https://doi.org/10.1002/(SICI)1099-1077(199601)11:1{$⟨$}47::AID-HUP743{$⟩$}3.3.CO;2-T

Lleras, A., Chu, H., & Buetti, S. (2017). Can we “apply” the findings of Forster and Lavie (2008)? On the generalizability of attentional capture effects under varying levels of perceptual load. *Journal of Experimental Psychology: Applied*, *23*(2), 158–179. https://doi.org/10.1037/xap0000116

Logie, R. H., & Pearson, D. G. (1997). The Inner Eye and the Inner Scribe of Visuo-spatial Working Memory: Evidence from Developmental Fractionation. *European Journal of Cognitive Psychology*, *9*(3), 241–257. https://doi.org/10.1080/713752559

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. https://doi.org/10.1016/j.tics.2013.06.006

Lyons, L. (2009). Designing Opportunistic User Interfaces to Support a Collaborative Museum Exhibit. *Proceedings of the 9th International Conference on Computer Supported Collaborative Learning - Volume 1*, 375–384.

Magreehan, D. A., Serra, M. J., Schwartz, N. H., & Narciss, S. (2016). Further boundary conditions for the effects of perceptual disfluency on judgments of learning. *Metacognition and Learning*, *11*(1), 35–56. https://doi.org/10.1007/s11409-015-9147-1

Maisto, A. A., & Queen, D. E. (1992). Memory for pictorial information and the picture superiority effect. *Educational Gerontology*, *18*(2), 213–223. https://doi.org/10.1080/0360127920180207

Maltese, A. V., Harsh, J. A., & Svetina, D. (2015). Data Visualization Literacy: Investigating Data Interpretation Along the Novice—Expert Continuum. *Journal of College Science Teaching*, *45*(1), 84–90.

Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences*, *9*(6), 296–305. https://doi.org/10.1016/j.tics.2005.04.010

Marr, D. (2010). *Vision : A computational investigation into the human representation and processing of visual information*. MIT Press.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *200*(1140), 269–294. https://doi.org/10.1098/rspb.1978.0020

Mateescu, M., Pimmer, C., Zahn, C., Klinkhammer, D., & Reiterer, H. (2021). Collaboration on large interactive displays: A systematic review. *Human-Computer Interaction*, *36*(3), 243–277. https://doi.org/10.1080/07370024.2019.1697697

Mausfeld, R. (2011). 12. Color Perception: From Grassmann Codes to a Dual Code for Object and Illumination Colors. In *Color Vision* (pp. 219–250). De Gruyter. https://doi.org/10.1515/9783110806984.219

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning*, *41*, 31–48.

McBride, D. M., & Dosher, B. A. (2002). A comparison of conscious and automatic memory processes for picture and word stimuli: A process dissociation analysis. *Consciousness and Cognition*, *11*(3), 423–460. https://doi.org/10.1016/S1053-8100(02)00007-7

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, *7*(7), 293–299. https://doi.org/10.1016/S1364-6613(03)00134-7

McCrudden, M. T., & Rapp, D. N. (2017). How Visual Displays Affect Cognitive Processing. *Educational Psychology Review*, *29*(3), 623–639. https://doi.org/10.1007/s10648-015-9342-2

McGrath, J. E. (1984). *Groups: Interaction and Performance* (1st Edition). Prentice-Hall, Inc.

Meyerhoff, H. S., Jardine, N., Stieff, M., Hegarty, M., & Franconeri, S. (2021). Visual ZIP files: Viewers beat capacity limits by compressing redundant features across objects. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(1), 103–115. https://doi.org/10.1037/xhp0000879

Meyerhoff, H. S., Papenmeier, F., & Huff, M. (2017). Studying visual attention using the multiple object tracking paradigm: A tutorial review. *Attention, Perception, & Psychophysics*, *79*(5), 1255–1274. https://doi.org/10.3758/s13414-017-1338-1

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The psychological review*, *64*, 81–97.

Milner, A. D., & Goodale, M. A. (1993). Visual pathways to perception and action. *Progress in Brain Research*, *95*, 317–337. https://doi.org/10.1016/s0079-6123(08)60379-9

Montello, D. R. (1991). Spatial orientation and the angularity of urban routes. *Environment and Behavior*, *23*(1), 47–69. https://doi.org/10.1177/0013916591231003

Montello, D. R. (2010). You are where? The function and frustration of You-Are-Here (YAH) maps. *Spatial Cognition & Computation*, *10*(2-3), 94–104. https://doi.org/10.1080/13875860903585323

Morgan, P. L., Patrick, J., & Tiley, L. (2013). Improving the effectiveness of an interruption lag by inducing a memory-based strategy. *Acta Psychologica*, *142*(1), 87–95. https://doi.org/10.1016/j.actpsy.2012.09.003

Morgan, P. L., Patrick, J., Waldron, S. M., King, S. L., & Patrick, T. (2009). Improving memory after interruption: Exploiting soft constraints and manipulating information access cost. *Journal of Experimental Psychology: Applied*, *15*(4), 291–306. https://doi.org/10.1037/a0018008

Moritz, J. (2019). *The influence of representation control on task performance and learning* [Doctoral dissertation, Universität Tübingen]. Tübingen.

Moritz, J., Meyerhoff, H. S., & Schwan, S. (2020). Control over spatial representation format enhances information extraction but prevents long-term learning. *Journal of Educational Psychology*, *112*(1), 148–165. https://doi.org/10.1037/edu0000364

Morris, C. J., Ebert, D. S., & Rheingans, P. (1999). An Experimental Analysis of the Pre-Attentiveness of Features in Chernoff Faces. *Proceedings Applied Imagery Pattern Recognition*, 12–17. https://doi.org/10.1117/12.384865

Morris, M. R., Paepcke, A., Winograd, T., & Stamberger, J. (2006). TeamTag: Exploring centralized versus replicated controls for co-located tabletop groupware. *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*, 1273. https://doi.org/10.1145/1124772.1124964

Mullane, J. C., Corkum, P. V., Klein, R. M., & McLaughlin, E. (2009). Interference Control in Children with and without ADHD: A Systematic Review of Flanker and Simon Task Performance. *Child Neuropsychology*, *15*(4), 321–342. https://doi.org/10.1080/09297040802348028

Müller, T., Hesse, F. W., & Meyerhoff, H. S. (2021). Two people, one graph: The effect of rotated viewpoints on accessibility of data visualizations. *Cognitive Research: Principles and Implications*, *6*(1), 31. https://doi.org/10.1186/s41235-021-00297-y

Mulligan, N. W. (1996). The effects of perceptual interference at encoding on implicit memory, explicit memory, and memory for source. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1067–1087. https://doi.org/10.1037/0278-7393.22.5.1067

Neurath, O. (1936). *International picture language : The first rules of isotype.* Kegan Paul, Trench, Trubner and Co.

Neurath, O., & Odgen, C. K. (1937). *Basic Isotype*. Kegan Paul, Trench Trubner and Co.

Norman, D. A., & Draper, S. W. (1986). *User Centered System Design; New Perspectives on Human-Computer Interaction*. L. Erlbaum Associates Inc.

Norris, D., & Kalm, K. (2021). Chunking and data compression in verbal short-term memory. *Cognition*, *208*, 104534. https://doi.org/10.1016/j.cognition.2020.104534

Norris, D., Kalm, K., & Hall, J. (2020). Chunking and redintegration in verbal short-term memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *46*(5), 872–893. https://doi.org/10.1037/xlm0000762

Norton, E. S., Kovelman, I., & Petitto, L.-A. (2007). Are There Separate Neural Systems for Spelling? New Insights into the Role of Rules and Memory in Spelling from Functional Magnetic Resonance Imaging. *Mind, Brain, and Education*, *1*(1), 48–59. https://doi.org/10.1111/j.1751-228X.2007.00005.x

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., … Yarkoni, T. (2015). Promoting an open research culture. *Science (New York, N.Y.)*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Nothelfer, C., Gleicher, M., & Franconeri, S. L. (2017). Redundant encoding strengthens segmentation and grouping in visual displays of data. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(9), 1667–1676. https://doi.org/10.1037/xhp0000314

O'Donell, B. M., Barraza, J. F., & Colombo, E. M. (2010). The effect of chromatic and luminance information on reaction times. *Visual Neuroscience*, *27*(3-4), 119–129. https://doi.org/10.1017/S0952523810000143

Ohnishi, M., Otsukuni, T., Takahashi, A., Sugiyama, M., Hirakimoto, M., Ogawa, S., Suzuki, A., Oshima, Y., Sheu, C. F., & Oda, K. (2020). Effects of luminance contrast and character size on reading speed. *Vision Research*, *166*(August 2019), 52–59. https://doi.org/10.1016/j.visres.2019.09.010

OpenAI. (2022). ChatGPT: Optimizing Language Models for Dialogue.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237–241. https://doi.org/10.1016/j.tics.2008.02.014

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *45*(3), 255–287. https://doi.org/10.1037/h0084295

Palmer, S. E., Brooks, J. L., & Nelson, R. (2003). When does grouping happen? *Acta Psychologica*, *114*(3), 311–330. https://doi.org/10.1016/j.actpsy.2003.06.003

Palmer, S. E., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. D. Baddeley (Eds.), *Attention and Performance IX* (pp. 135–151). Erlbaum.

Paterson, D. G., & Tinker, M. A. (1929). Studies of typographical factors influencing speed of reading. II. Size of type. *Journal of Applied Psychology*, *13*(2), 120–130. https://doi.org/10.1037/h0074167

Paterson, D. G., & Tinker, M. A. (1931a). Studies of typographical factors influencing speed of reading X. Style of type face. *Journal of Applied Psychology*, *15*(3), 241–247. https://doi.org/10.1037/h0074261

Paterson, D. G., & Tinker, M. A. (1931b). Studies of typographical factors influencing speed of reading. VI. Black type versus white type. *The Journal of Applied Psychology*, *15*(3), 241–247.

Pegado, F., Nakamura, K., & Hannagan, T. (2014). How does literacy break mirror invariance in the visual system? *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00703

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*(January), 1–8. https://doi.org/10.3389/neuro.11.010.2008

Peissig, J. J., & Tarr, M. J. (2007). Visual object recognition: Do we know more now than we did 20 years ago? *Annual Review of Psychology*, *58*(1), 75–96. https://doi.org/10.1146/annurev.psych.58.102904.190114

Peters, M., Lehmann, W., Takahira, S., Takeuchi, Y., & Jordan, K. (2006). Mental rotation test performance in four cross-cultural samples (N = 3367): Overall sex differences and the role of academic program in performance. *Cortex*, *42*(7), 1005–1014. https://doi.org/10.1016/S0010-9452(08)70206-5

Phillips, L. D., & Phillips, M. C. (1993). Faciliated Work Groups: Theory and Practice. *Journal of the Operational Research Society*, *44*(6), 533–549. https://doi.org/10.1057/jors.1993.96

Popper, K. (1978). Three Worlds.

Pyke, A. A., & LeFevre, J.-A. (2011). Calculator use need not undermine direct-access ability: The roles of retrieval, calculation, and calculator use in the acquisition of arithmetic facts. *Journal of Educational Psychology*, *103*(3), 607–616. https://doi.org/10.1037/a0023291

Quaiser-Pohl, C. (2003). The mental cutting test ”Schnitte” and the picture rotation test - Two new measures to assess spatial ability. *International Journal of Testing*, *3*(3), 219–231. https://doi.org/10.1207/S15327574IJT0303_2

Quispel, A., Maes, A., & Schilperoord, J. (2016). Graph and chart aesthetics for experts and laymen in design: The role of familiarity and perceived ease of use. *Information Visualization*, *15*(3), 238–252. https://doi.org/10.1177/1473871615606478

Rafal, R., Gershberg, F., Egly, R., Ivry, R., Kingstone, A., & Ro, T. (1996). Response channel activation and the lateral prefrontal cortex. *Neuropsychologia*, *34*(12), 1197–1202. https://doi.org/10.1016/0028-3932(96)00045-0

Ramakers, R., Vanacken, D., Luyten, K., Coninx, K., & Schöning, J. (2012). Carpus. *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology - UIST '12*, 35. https://doi.org/10.1145/2380116.2380123

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience? *Personality and social psychology review : an official journal of the Society for Personality and Social Psychology, Inc*, *8*, 364–82. https://doi.org/10.1207/s15327957pspr0804_3

Rensink, R. A. (2014). On the Prospects for a Science of Visualization. In *Handbook of Human Centric Visualization* (pp. 147–175). Springer New York. https://doi.org/10.1007/978-1-4614-7485-2_6

Rensink, R. A. (2017). The nature of correlation perception in scatterplots. *Psychonomic Bulletin & Review*, *24*, 776–797. https://doi.org/10.3758/s13423-016-1174-7

Rensink, R. A., & Baldridge, G. (2010). The perception of correlation in scatterplots. *Computer Graphics Forum*, *29*(3), 1203–1210. https://doi.org/10.1111/j.1467-8659.2009.01694.x

Reupert, A., Straussner, S. L., Weimand, B., & Maybery, D. (2022). It Takes a Village to Raise a Child: Understanding and Expanding the Concept of the "Village". *Frontiers in Public Health*, *10*.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*(4), 615–625. https://doi.org/10.1037/a0013684

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, *16*(3), 550–554. https://doi.org/10.3758/PBR.16.3.550

Richlan, F., Kronbichler, M., & Wimmer, H. (2011). Meta-analyzing brain dysfunctions in dyslexic children and adults. *NeuroImage*, *56*(3), 1735–1742. https://doi.org/10.1016/j.neuroimage.2011.02.040

Risko, E. F., & Gilbert, S. J. (2016). Cognitive Offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Risko, E. F., & Kingstone, A. (2011). Eyes wide shut: Implied social presence, eye tracking and attention. *Attention, Perception, & Psychophysics*, *73*(2), 291–296. https://doi.org/10.3758/s13414-010-0042-1

Risko, E. F., Medimorec, S., Chisholm, J., & Kingstone, A. (2014). Rotating with rotated text: A natural behavior approach to investigating cognitive offloading. *Cognitive Science*, *38*(3), 537–564. https://doi.org/10.1111/cogs.12087

Roberts, J., & Lyons, L. (2017). Scoring qualitative informal learning dialogue: The squild method for measuring museum learning talk. *Computer-Supported Collaborative Learning Conference, CSCL*, *1*, 487–494.

Robertson, N. (1961). Modern Museum Is Startled by Matisse Picture. *The New York Times*, 45.

Rogers, Y., Lim, Y.-K., Hazlewood, W. R., & Marshall, P. (2009). Equal opportunities: Do shareable interfaces promote more group participation than single user displays? *Human–Computer Interaction*, *24*(1-2), 79–116. https://doi.org/10.1080/07370020902739379

Rosner, T. M., Davis, H., & Milliken, B. (2015). Perceptual blurring and recognition memory: A desirable difficulty effect revealed. *Acta Psychologica*, *160*, 11–22. https://doi.org/10.1016/j.actpsy.2015.06.006

Rubin, G. S., & Legge, G. E. (1989). Psychophysics of reading. VI-The role of contrast in low vision. *Vision Research*, *29*(1), 79–91. https://doi.org/10.1016/0042-6989(89)90175-2

Rudnicky, A. I., & Kolers, P. A. (1984). Size and case of type as stimuli in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(2), 231–249. https://doi.org/10.1037/0096-1523.10.2.231

Rüsseler, J., Scholz, J., Jordan, K., & Quaiser-Pohl, C. (2005). Mental rotation of letters, pictures, and three-dimensional objects in German dyslexic children. *Child Neuropsychology*, *11*(6), 497–512. https://doi.org/10.1080/09297040490920168

Salomon, G. (1984). Television is "easy" and print is "tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology*, *76*(4), 647–658. https://doi.org/10.1037/0022-0663.76.4.647

Salomon, G., & Perkins, D. (2005). Do Technologies Make Us Smarter? Intellectual Amplification With, Of, and Through Technology. In *Intelligence and technology: The impact of tools on the nature and development of human abilities* (pp. 71–86). Lawrence Erlbaum Associates Publishers.

Salwei, A. M., & de Diego-Lázaro, B. (2021). Does Language Make a Difference? A Study of Language Dominance and Inhibitory Control. *Frontiers in Psychology*, *12*.

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal*, *45*(2), 185–213. https://doi.org/10.1006/ijhc.1996.0048

Scheiter, K., Schubert, C., & Schüler, A. (2018). Self-regulated learning from illustrated text: Eye movement modelling to support use and regulation of cognitive processes during learning from multimedia. *British Journal of Educational Psychology*, *88*(1), 80–94. https://doi.org/10.1111/bjep.12175

Schnotz, W. (2001). Wissenserwerb mit Multimedia. *Unterrichtswissenschaft*, *29*(4), 92–318. https://doi.org/10.25656/01:7717

Schnotz, W. (2002). Towards an integrated view of learning from text and visual displays. *Educational Psychology Review*, *14*(1), 101–120. https://doi.org/10.1023/A:1013136727916

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, *13*(2), 141–156. https://doi.org/10.1016/S0959-4752(02)00017-8

Schwab, F., Hennighausen, C., Adler, D. C., & Carolus, A. (2018). Television Is Still "Easy" and Print Is Still "Tough"? More Than 30 Years of Research on the Amount of Invested Mental Effort. *Frontiers in Psychology*, *9*, 1098. https://doi.org/10.3389/fpsyg.2018.01098

Scott, S. D., Sheelagh, M., Carpendale, T., & Inkpen, K. M. (2004). Territoriality in collaborative tabletop workspaces. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, (December 2014), 294–303. https://doi.org/10.1145/1031607.1031655

Semrud-Clikeman, M., Fine, J. G., Bledsoe, J., & Zhu, D. C. (2012). Gender differences in brain activation on a mental rotation task. *International Journal of Neuroscience*, *122*(10), 590–597. https://doi.org/10.3109/00207454.2012.693999

Shaffer, D., Hatfield, D., Svarovsky, G., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A., & Mislevy, R. (2009). Epistemic Network Analysis: A Prototype for 21st-Century Assessment of Learning. *International Journal of Learning and Media*, *1*, 33–53. https://doi.org/10.1162/ijlm.2009.0013

Shen, C., Ryall, K., Forlines, C., Esenther, A., Vernier, F. D., Everitt, K., Wu, M., Wigdor, D., Morris, M. R., Hancock, M., & Tse, E. (2006). Informing the design of direct-touch tabletops. *IEEE Computer Graphics and Applications*, *26*(5), 36–46. https://doi.org/10.1109/MCG.2006.109

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703. https://doi.org/10.1126/science.171.3972.701

Sherif, M., & Sherif, C. W. (1965). Ingroup and intergroup relations. In J. O. Whittaker (Ed.), *Introduction to Psychology.* Saunders. https://brocku.ca/MeadProject/Sherif/Sherif_1965f.html.

Shor, R. E. (1971). Symbol processing speed differences and symbol interference effects in a variety of concept domains. *The Journal of General Psychology*, *85*(2), 187–205. https://doi.org/10.1080/00221309.1971.9920672

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *69*(1), 99. https://doi.org/10.2307/1884852

Skau, D., Harrison, L., & Kosara, R. (2015). An Evaluation of the Impact of Visual Embellishments in Bar Charts. *34*(3).

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, *4*(6), 592–604. https://doi.org/10.1037/0278-7393.4.6.592

Sloan, L. L. (1969). Variation of acuity with luminance in ocular diseases and anomalies. *Documenta Ophthalmologica*, *26*(1), 384–393. https://doi.org/10.1007/BF00943999

Sloan, L. L. (1977). *Reading Aids for the Partially Sighted: A Systematic Classification and Procedure for Prescribing*. Williams & Wilkins.

Smith, S. L. (1962). Angular estimation. *Journal of Applied Psychology*, *46*(4), 240–246. https://doi.org/10.1037/h0044445

Sobkow, A., Fulawka, K., Tomczak, P., Zjawiony, P., & Traczyk, J. (2019). Does mental number line training work? The effects of cognitive training on real-life mathematics, numeracy, and decision making. *Journal of Experimental Psychology: Applied*, *25*(3), 372–385. https://doi.org/10.1037/xap0000207

Song, H., & Schwarz, N. (2008). If It's Hard to Read, It's Hard to Do: Processing Fluency Affects Effort Prediction and Motivation. *Psychological Science*, *19*(10), 986–988. https://doi.org/10.1111/j.1467-9280.2008.02189.x

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. *Science*, *333*(6043), 776–778. https://doi.org/10.1126/science.1207745

Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology*, *25*(2), 207–222. https://doi.org/10.1080/14640747308400340

Stasser, G., & Birchmeier, Z. (2003). Group creativity and collective choice. In *Group creativity: Innovation through collaboration* (pp. 85–109). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195147308.003.0005

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, *48*(6), 1467–1478. https://doi.org/10.1037/0022-3514.48.6.1467

Stewart, J., Bederson, B. B., & Druin, A. (1999). Single display groupware. *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99*, (February 2013), 286–293. https://doi.org/10.1145/302979.303064

Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction*, *17*(2), 219–234. https://doi.org/10.1016/j.learninstruc.2007.01.012

Sutherland, S. W. (1989). The Forgotten Research of Miles Albert Tinker. *Journal of Visual Literacy*, *9*(1), 10–25. https://doi.org/10.1080/23796529.1989.11674437

Sweller, J., & Chandler, P. (1994). Why Some Material is Difficult to Learn. *Cognition and Instruction*, *12*(3), 185–233. https://doi.org/10.1207/s1532690xci1203_1

Tang, A., Tory, M., Po, B., Neumann, P., & Carpendale, S. (2006). Collaborative coupling over tabletop displays. *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*, (January 2006), 1181. https://doi.org/10.1145/1124772.1124950

Tapley, S. M., & Bryden, M. P. (1977). An investigation of sex differences in spatial ability: Mental rotation of three-dimensional objects. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *31*(3), 122–130. https://doi.org/10.1037/h0081655

Tarampi, M. R., Heydari, N., & Hegarty, M. (2016). A tale of two types of perspective taking. *Psychological Science*, *27*(11), 1507–1516. https://doi.org/10.1177/0956797616667459

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *PsychonomicBulletin & Review*, *2*(1), 55–82.

Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, *67*(1-2), 1–20. https://doi.org/10.1016/S0010-0277(98)00026-2

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, *21*(2), 233–282. https://doi.org/10.1016/0010-0285(89)90009-1

Tetlock, P. E., Skitka, L., & Boettger, R. (1989). Social and cognitive strategies for coping with accountability: Conformity, complexity, and bolstering. *Journal of Personality and Social Psychology*, *57*, 632–640. https://doi.org/10.1037/0022-3514.57.4.632

Tinker, M. A. (1956). Effects of angular alignment upon readability of print. *Journal of Educational Psychology*, *47*, 358–363. https://doi.org/10.1037/h0044504

Tinker, M. A., & Paterson, D. G. (1931). Studies of typographical factors influencing speed of reading. VII. Variations in color of print and background. *Journal of Applied Psychology*, *15*(5), 471–479. https://doi.org/10.1037/h0076001

Tissenbaum, M., Berland, M., & Lyons, L. (2017). DCLM framework: Understanding collaboration in open-ended tabletop learning environments. *International Journal of Computer-Supported Collaborative Learning*, *12*(1), 35–64.

Treisman, A. M. (1986). Features and objects in visual processing. *Scientific American*, *255*(5), 114B–125.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Tse, E., Histon, J., Scott, S. D., & Greenberg, S. (2004). Avoiding interference. *Proceedings of the 2004 ACM conference on Computer supported cooperative work - CSCW '04*, (January), 252. https://doi.org/10.1145/1031607.1031647

Tufte, E. R. (1983). *The visual display of quantitative information*. Graphics Pr.

Ulrich, R., Prislan, L., & Miller, J. (2021). A bimodal extension of the Eriksen flanker task. *Attention, Perception, & Psychophysics*, *83*(2), 790–799. https://doi.org/10.3758/s13414-020-02150-8

Valensi, G. (1939). *Procédé de télévision en couleurs* (FR841335A).

van Nes, F. L., & Jacobs, J. C. (1981). The effect of contrast on letter and word recognition. *IPO annual progress report*, *16*, 72–80.

van Nimwegen, C., & van Oostendorp, H. (2009). The questionable impact of an assisting interface on performance in transfer situations. *International Journal of Industrial Ergonomics*, *39*(3), 501–508. https://doi.org/10.1016/j.ergon.2008.10.008

Varriale, V., van der Molen, M. W., & De Pascalis, V. (2018). Mental rotation and fluid intelligence: A brain potential analysis. *Intelligence*, *69*, 146–157. https://doi.org/10.1016/j.intell.2018.05.007

Ward, G. (2001). A critique of the working memory model. In *Working memory in perspective* (pp. 219–239). Psychology Press. https://doi.org/10.4324/9780203194157

Weber, W., & Wenzel, A. (2013). Interaktive Infografiken: Standortbestimmung und Definition. In W. Weber, M. Burmester, & R. Tille (Eds.), *Interaktive Infografiken* (pp. 3–23). Springer. https://doi.org/10.1007/978-3-642-15453-9_1

Weiser, M. (1999). The computer for the 21st century. *ACM SIGMOBILE Mobile Computing and Communications Review*, *3*(3), 3–11. https://doi.org/10.1145/329124.329126

Westendorf, L., Shaer, O., Varsanyi, P., van der Meulen, H., & Kun, A. L. (2017). Understanding Collaborative Decision Making Around a Large-Scale Interactive Tabletop. *Proceedings of the ACM on Human-Computer Interaction*, *1*(CSCW), 1–21. https://doi.org/10.1145/3134745

Whitehouse, A. J. O., Maybery, M. T., & Durkin, K. (2006). The development of the picture-superiority effect. *British Journal of Developmental Psychology*, *24*(4), 767–773. https://doi.org/10.1348/026151005X74153

Wigdor, D., & Balakrishnan, R. (2005). Empirical investigation into the effect of orientation on text readability in tabletop displays. *ECSCW 2005 - Proceedings of the 9th European Conference on Computer-Supported Cooperative Work*, (September), 205–224. https://doi.org/10.1007/1-4020-4023-7_11

Wilkinson, L., & Wills, G. (2011). *The grammar of graphics* (2. ed., softcover reprint of the hardcover 2. ed., 2005). Springer.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*(6), 495–501. https://doi.org/10.1038/nrn1411

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 1–8. https://doi.org/10.1038/s41562-017-0058

Wooldridge, A. R., Carayon, P., Shaffer, D. W., & Eagan, B. (2018). Quantifying the Qualitative with Epistemic Network Analysis: A Human Factors Case Study of Task-Allocation Communication in a Primary Care Team. *IISE transactions on healthcare systems engineering*, *8*(1), 72–82. https://doi.org/10.1080/24725579.2017.1418769

Xie, H., Zhou, Z., & Liu, Q. (2018). Null Effects of Perceptual Disfluency on Learning Outcomes in a Text-Based Educational Context: A Meta-analysis. *Educational Psychology Review*, *30*(3), 745–771. https://doi.org/10.1007/s10648-018-9442-x

Xu, Y., & Franconeri, S. L. (2015). Capacity for visual features in mental rotation. *Psychological Science*, *26*(8), 1241–1251. https://doi.org/10.1177/0956797615585002

Yu, D. (2010). Comparing reading speed for horizontal and vertical English text. *Journal of Vision*, 17.

Yue, C. L., Bjork, E. L., & Bjork, R. A. (2013). Reducing verbal redundancy in multimedia learning: An undesired desirable difficulty? *Journal of Educational Psychology*, *105*(2), 266–277. https://doi.org/10.1037/a0031971

Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is-and is not-a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory and Cognition*, *41*(2), 229–241. https://doi.org/10.3758/s13421-012-0255-8

Yuille, J. C., & Steiger, J. H. (1982). Nonholistic processing in mental rotation: Some suggestive evidence. *Perception & Psychophysics*, *31*(3), 201–209.

Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory & Cognition*, *13*(3), 193–201. https://doi.org/10.3758/BF03197681

Zikovitz, D. C., & Harris, L. R. (1999). Head tilt during driving. *Ergonomics*, *42*(5), 740–746.

Zwinger, S., & Zeiller, M. (2016). Interactive Infographics in German Online Newspapers. *Proceedings of the 9th Forum Media Technology 2016*, 11.