Technische Universität München
TUM School of Computation,
Information and Technology

Interdisciplinary Collaboration in Computational Social Science

Angelina Voggenreiter

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz:
apl. Prof. Dr. Georg Groh

Prüfer der Dissertation:
1. Prof. Dr. Jürgen Pfeffer
2. Prof. Kenneth Joseph, Ph.D., University at Buffalo

Die Dissertation wurde am 20.04.23 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 03.07.2023 angenommen.

Abstract

Technological progress and digital developments have provided new approaches to study human emotion and behaviour. In the field of Computational Social Science (CSS), tools and methods from disciplines such as computer science, mathematics or physics are used to investigate social science problems. This dissertation introduces CSS and evaluates the value of interdisciplinary collaboration in this field. Based on three case studies, our work shows how technology can help to expand traditional social science methods and answer crucial research questions in various social science fields: Study 1 explains how automated image processing can be employed in Communication Sciences to characterize the visual discourse on climate change found on Twitter, study 2 describes how eye tracking can be utilized in Education Sciences to measure teachers' attention and study 3 describes how machine learning can be applied in Sociology to gain a deeper understanding of drug consumption. While these studies demonstrate the potential of CSS research, such as access to new datasets and analysis of large scale data, they also show the great challenges in CSS projects, such as ethical pitfalls and misuse of computational methods. We explain how interdisciplinary collaborations helped us to overcome these challenges and to expand knowledge in the Social Sciences. While highlighting the need for interdisciplinary collaboration, this dissertation also discusses common difficulties in collaborative projects and offers guidelines to successfully complete such projects.

Zusammenfassung

Technologischer Fortschritt und digitale Entwicklungen haben neue Wege eröffnet, um menschliche Emotionen und Verhaltensweisen zu untersuchen. In der Disziplin Computational Social Science (CSS) werden Tools und Methoden aus Bereichen wie Informatik, Mathematik oder Physik verwendet um sozialwissenschaftliche Probleme zu untersuchen. Diese Dissertation bietet eine Einführung in CSS und evaluiert den Wert von interdisziplinärer Kollaboration in diesem Bereich. Anhand von drei Fallstudien zeigt diese Arbeit, wie der Einsatz von Technologie dazu beitragen kann, traditionelle sozialwissenschaftliche Methoden zu erweitern und wichtige Forschungsfragen des jeweiligen Fachgebiets zu beantworten: Studie 1 erläutert, wie automatisierte Bildverarbeitung in den Kommunikationswissenschaften genutzt werden kann um den visuellen Klimawandel-Diskurs auf Twitter zu untersuchen, Studie 2 beschreibt, wie Eye Tracking in der Bildungsforschung verwendet werden kann, um die Aufmerksamkeit von Lehrern zu messen, und Studie 3 beschreibt, wie maschinelles Lernen in der Soziologie eingesetzt werden kann, um ein tieferes Verständnis von Drogenkonsum zu erlangen. Während diese Studien das Potenzial der CSS-Forschung betonen, wie beispielsweise den Zugang zu neuen Datensätzen und die Analyse großer Datenmengen, zeigen sie auch die großen Herausforderungen in CSS-Projekten, wie ethische Probleme oder die fehlerhafte Anwendung von computergestützten Methoden. Wir beschreiben, wie uns interdisziplinäre Kollaborationen in diesen Projekten geholfen haben, diese Herausforderungen zu meistern und Wissen in den Sozialwissenschaften zu erweitern. Während diese Dissertation die Notwendigkeit interdisziplinärer Zusammenarbeit betont, diskutiert sie auch häufige Schwierigkeiten in kollaborativen Projekten und bietet Richtlinien, um solche Projekte erfolgreich durchzuführen.

Publications

This thesis includes two peer-reviewed papers relevant to the examination as well as one additional paper. Table 1 gives an overview of the publications selected for the dissertation and Table 2 lists further publications.

Publications **relevant to the examination**

Title	Authors	Venue
Glowing Experience or Bad Trip? A Quantitative Analysis of User Reported Drug Experiences on Erowid.org	Mooseder, Angelina & Malik, Momin M. & Lamba, Hemank & Erowid, Earth & Thyssen, Sylvia & Pfeffer, Jürgen	Proceedings of the International AAAI Conference on Web and Social Media (ICWSM 2022), 16(1), 675-686, `https://doi.org/10.1609/icwsm.v16i1.19325` Published 06-2022
(Social) Media Logics and Visualizing Climate Change: 10 Years of #climatechange Images on Twitter	Mooseder, Angelina & Brantner, Cornelia & Zamith, Rodrigo & Pfeffer, Jürgen	Journal Social Media and Society, 9(1), `https://doi.org/10.1177/20563051231164310` Published 03-2023

Additional publications in the dissertation

Measuring Teachers' Visual Expertise Using the Gaze Relational Index Based on Real-world Eye-tracking Data and Varying Velocity Thresholds	Mooseder, Angelina* & Kosel, Christian* & Seidel, Tina & Pfeffer, Jürgen	Arxiv Report, `https://arxiv.org/abs/2304.05143` Published 04-2023

Table 1: Publications selected for the dissertation. Authors marked with an asterix contributed equally.

Further publications		
Title	Authors	Venue
What does Presence in Polarizing Environments do to Users' Behavior?	Strathern, Wienke & Mooseder, Angelina & Pfeffer, Jürgen	Proceedings of the third MEDIATE Workshop (ICWSM 2022), `https://doi.org/10.36190/2022.52` Published 06-2022
Central Figures in the Climate Change Discussion on Twitter	Kara, Anil Can & Dobrijevic, Ivana & Öztas, Emre & Mooseder, Angelina & Ghawi, Raji & Pfeffer, Jürgen	The 24th International Conference on Information Integration and Web Intelligence (iiWAS2022), 575–580, `https://doi.org/10.1007/978-3-031-21047-1_52` Published 11-2022
This Sample seems to be good enough! Assessing Coverage and Temporal Reliability of Twitter's Academic API	Pfeffer, Jürgen & Mooseder, Angelina & Hammer, Luca & Stritzel, Oliver & Garcia, David	Proceedings of the International AAAI Conference on Web and Social Media (ICWSM 2023) Accepted for Publication 11-2022
Popular and on the Rise — But Not Everywhere: COVID-19-Infographics on Twitter	Witzenberger, Benedict & Mooseder, Angelina & Pfeffer, Jürgen	5th International Data Science Conference (iDSC 2023) Accepted for Publication 04-2023

Table 2: Further publications.

Remark

In the following dissertation we will speak of "computer scientists", referring to researchers with a technical expertise, experience and education in the computational sciences and related disciplines, as well as "social scientists", referring to researchers with an expertise, experience and education in a social science discipline. This differentiation does not reflect the full spectrum of researchers background knowledge, as many researchers have knowledge in both areas. Therefore, the terms "computer scientists" and "social scientists" express a typology instead of describing a person and may even apply to the same person to different degrees.

Contents

1 Introduction to Computational Social Science

The field of Computational Social Science (CSS), which concerns applying computational methods to social science research problems, has forged new approaches to researching human behaviour and has helped to expand knowledge in various social science disciplines. At the same time, conducting studies in Computational Social Science remains challenging and requires extensive technical knowledge and a social science background. We therefore propose that an interdisciplinary collaboration between computer and social scientists is needed to solve complex research problems within CSS.

This dissertation evaluates interdisciplinary collaboration in CSS. In particular, it introduces the field of CSS, highlights the importance of interdisciplinary collaborations in this field, and critically reflects on approaches to making such collaborations successful. In so doing, this dissertation offers three main contributions. First, it presents an overview of the advantages of CSS research as well as three real world examples, in which computational methods were applied to expand knowledge in the social sciences. Thus, this dissertation answers the following question:

> **Question 1:** What opportunities do computational methods provide for social science research?

Second, this dissertations discusses barriers and difficulties in the field of CSS research. Based on three real world examples it showcases how these barriers can hinder the research process and impact the findings. Hence, this dissertation elaborates on the following research question:

> **Question 2:** What challenges arise when applying computational methods in social science research?

Third, this dissertation examines the value of interdisciplinary collaboration for CSS research. Based on three real world examples, it shows how such collaboration can help to overcome challenges of CSS and make use of its full potential. While highlighting the need for interdisciplinary collaboration, we also discuss common difficulties in collaborative projects and present best practices for successfully conducting such projects. In summary, this dissertation answers the following question:

> **Question 3:** In what ways and under what circumstances can interdisciplinary collaboration help overcome these challenges?

This dissertation is structured as follows: We begin by defining CSS and describing its application in various social science disciplines (Chapter 1). Next, we give an overview of the methodological foundations of CSS, with regard to both data collection and analysis (Chapter 2). Third, we present some of the main benefits and challenges of CSS. We then define interdisciplinary collaboration and explain how it can help tackle these challenges (Chapter 3). Afterwards, we present three case studies to exemplify these opportunities, challenges, and the importance of interdisciplinary collaboration (Chapter 4):

With Study 1 (Section 4.1), an automated analysis of the visual discourse on climate change found on Twitter, we demonstrate the potential of large scale data analysis for communication science, and the challenge of designing a CSS study to contribute directly to the social sciences. We describe how an interdisciplinary collaboration between computer and communication scientists helped us to embed the study in a social science framework and provide meaningful results for the communication science community. With Study 2 (Section 4.2), a comparison of eye tracking parameters to measure teachers' attention, we exemplify the advantages of using CSS methods to reduce research costs, time and obtrusiveness, and showcase the challenge of trusting to much in and misusing computational methods. We explain how an interdisciplinary collaboration between computer and educational scientists helped us to critically reflect on the parameters used for eye tracking research in education. With Study 3 (Section 4.3), an automated analysis of drug experience reports, we demonstrate the benefits of gaining access to new data sources and the challenges arising from the lack of ethical standards in CSS. We explain how a transdisciplinary collaboration between computer scientists and drug community experts helped us to detect ethical pitfalls and design the study in a way that protects the population under study and their interests.

We then summarize the contributions of these case studies and elaborate on the value of collaboration. Finally, we examine potential difficulties when working in interdisciplinary collaborations and suggest ground rules enabling the success of such collaborations in CSS (Section 5).

1.1 Definition

Computational Social Science (CSS) in its broadest sense can be defined as a discipline in which "computational tools and techniques are used to expand the methods and approaches used in social science research" (Abanoz, 2022). CSS should support social scientists in answering important research questions of their field by allowing access to new data sources, providing new methods for analysing data, and bringing expertise and a critical perspective to the current use of technologies and algorithms in the social sciences (Abanoz, 2022). CSS is typically applied to study human behaviour and social dynamics (Oboler, Welsh, & Cruz, 2012; Uppal & Bohon, 2017).

1.2 Areas of Application

The adoption of computational methods and tools has proven beneficial for a great variety of social science research topics. The section below offers an overview of the areas of application in CSS.

1.2.1 Sociology

Within Sociology, CSS has been employed to understand how people form relationships, groups and networks, and how they interact within these social structures. Researchers have analysed how these structures can change on an individual or social level over time or due to internal or external variables, and how social processes within these structures can influence the behaviour of individuals (Edelmann, Wolff, Montagne, & Bail, 2020).

Furthermore, computational methods have been applied to examine how people produce culture, and how they perceive cultural objects, such as movies, music, pictures, and other art objects. Similarly, researchers have studied the sociology of science by mapping scientific fields and visualizing groups within these fields as well as interactions with related disciplines or industry, investigating how consensus between scholars is formed, how knowledge is created and how scientific work is assessed, as well as evaluating processes, such as teamwork between researchers, within certain disciplines (Edelmann et al., 2020).

Additionally, CSS expanded research on the general population and specific communities by, for example, using mobile phone data to make demographic estimations of populations in certain neighbourhoods or countries, as well as collecting social media and web data to study behaviours, such as abortion or radicalization, in terms of prevalence, demographics and other related behaviours (Edelmann et al., 2020).

1.2.2 Politics

In the area of politics, CSS has been applied to study the political discourse on social media, which includes analysing the social networks of politically interested users, assessing the impact of influential actors and groups, and investigating how information flows between users with the same or different political preferences. Moreover, CSS leverages research on the influence of political events on public debates, on the popularity of certain political topics, and on processes within political discourse, such as polarization, the creation of echo chambers, and the propagation of misinformation (Edelmann et al., 2020; Haq, Braud, Kwon, & Hui, 2020). Furthermore, CSS has advanced studies of political communication on social media and its consequences for political behaviour, such as voting behaviour and elections. For example, several studies have evaluated how political stakeholders, such as parties and candidates, use platforms like Twitter to address their audience, and have analysed the impact of political campaigns on social media. Similarly, web data has been collected to gather insights into certain movements, such as the Egyptian revolution, and to study how collectives are organized on social media (Edelmann et al., 2020; Haq et al., 2020; Jungherr, 2016).

1.2.3 Psychology

In psychology, CSS has introduced new tools and systems to conduct experiments and measure psychological constructs, expanding the potential of research in many areas. In clinical psychology, for example, virtual landscapes and tasks have been created to diagnose mental disorders and cognitive impairments (Ostermann, Röer, & Tomasik, 2021), and in social psychology, a version of the famous Milgram experiment was designed with virtual avatars to study obedience (Slater et al., 2006). Furthermore, with the use of computational methods, a large variety of biopsychological devices and analysis platforms could be (further) developed to study cognitive processes and emotions (Lim, Mountstephens, & Teo, 2020; Posada-Quintero & Chon, 2020; J. Zhu, Ji, & Liu, 2019). In addition, complex computational models have been applied in cognitive, affective, educational, organizational, social and personality psychology, and neuroscience to evaluate constructs, such as mental workload, team performance, affect, cognitive ability and engagement (D'Mello, Tay, & Southwell, 2022). Likewise, in clinical psychology such models have proven their utility, for example to help better understand mental disorders and identify factors associated with them, predict the course of psychological disorders and select therapies suitable for individual patients (Bartlett, Pirrone, Javed, & Gobet, 2022; Dwyer, Falkai, & Koutsouleris, 2018).

1.2.4 Business and Economics

In Business and Economics, computational methods have helped to gain deeper insights into the complex, dynamic systems which influence teams, companies, and the economy as a whole. For example, machine learning and computational network modelling have been applied to study the structure and dynamics of (online) leadership and teams, and to identify factors associated with strong team performance (Evegroen, Schoonderbeek, & Treur, 2021; Carter, DeChurch, Braun, & Contractor, 2015; H. Zhu, Kraut, & Kittur, 2013). On a larger scale, collecting, combining and analysing huge datasets - such as scanner price data, financial transactions, satellite images, traffic sensors, news, and social media data - has been useful for modelling the current state of the economy and predicting future economic developments, such as financial crises or stock market trends (Barbaglia et al., 2021).

1.2.5 Education Studies

In education research, computational methods have been utilized to model and predict learners' performance, which is helpful for identifying students at risk, understanding factors associated with learning success and determining which kind of difficulties students might experience with the subject matter (Albreiki, Zaki, & Alashwal, 2021). Furthermore, these methods were applied to obtain deeper insights into the learning process by computationally analysing learners' behaviour and motivation, and quantitatively evaluating students' satisfaction and opinions about classes (Yunita, Santoso, & Hasibuan, 2021). In addition, as described in the psychology section above, CSS allowed the (further) development of biopsychological devices and analysis platforms to study cognitive processes and emotions, which has also proven useful for educational research. For example, eye tracking measurements were used in numerous studies for assessing students' visual attention and analysing their information processing during learning tasks to identify the effects of learners' strategies, learning difficulties, and instructional material (Lai et al., 2013; Strohmaier, MacKay, Obersteiner, & Reiss, 2020).

1.2.6 Communication Studies

The digital world has introduced new ways of communication, which certainly change to some extent how, when, why and whom we connect with. Therefore, a large part of CSS studies in communication research focus on the collection and analysis of online data to investigate online communication. Some studies are conducted from a network perspective to analyse how people organize themselves, cooperate and collaborate in online spaces, such as social media platforms or online computer games. Others rather focus on the social media content itself, by investigating the values embedded in texts, or developing methods to detect malicious content, such as hate speech, profanity, or false information (Matei & Kee, 2019).

2 Methodological Foundations of CSS

Although there is no clear consensus on what Computational Social Science comprises and what is outside of this discipline, researchers typically use keywords such as big data, social media, behavioural research, machine learning and social networks, when describing or conducting research in CSS (Purnomo, Asitah, Rosyidah, Septianto, & Firdaus, 2022; Uppal & Bohon, 2017). This chapter provides an overview of the main methodologies in CSS. As researchers are constantly adding new methods to CSS, by combining and adapting methods from other research disciplines, this list does not claim to be comprehensive, but is rather indicative of the broad variety of the research field.

2.1 Data Collection

2.1.1 Web and Social Media Data

Social media play a vital role in our society. Around 59% of the population worldwide use social media, with rates much higher in western societies. The average internet user spends more than 2.5 hours daily on social media (Dixon, 2022; Statista Research Department, 2022). Amongst the most popular social media platforms are Facebook (2.9 billion monthly active users), YouTube (2.6), Whatsapp (2.0), Instagram (1.5) and Weixin/WeChat (1.2) (We are Social, Hootsuit, & DataReportal, 2022). Besides these, there exist a great variety of smaller platforms and forums, which are frequented by niche populations. For example, the mental health forums on Beyond Blue provide opportunities for discussions amongst people struggling with mental health issues, and the website Erowid invites people to submit reports describing their experiences with drugs. Social media data has offered new possibilities to gain insights into human opinions and to analyse human behaviour and social processes. The difficulties of collecting such data depend on the individual platform. Some companies, such as the microblogging platform Twitter, offer an Application Programming Interface (API), which allows collecting large amounts of data in an easy way. However, the regulations regarding the amount and type of data which can be accessed, the costs for data collection, and the scenarios in which such data may be employed, are defined by the company and can thus change at any time (Tindall, McLevey, Koop-Monteiro, & Graham, 2022). As a result, generating replicable datasets is challenging (Lazer et al., 2020). Other platforms, especially small and traditional websites, such as forums or blogs, do not offer APIs and thus must be accessed by web scraping, a method of automatically collecting data from the web (Tindall et al., 2022). Nevertheless, there are some platforms, like TikTok, which prohibit the automated extraction of data, and use extensive mechanisms to prevent any form of web scraping. The differences in the complexity of data collection can lead to some platforms being studied more often than others (Pfeffer, Mooseder, Hammer, Stritzel, & Garcia, 2022) and thus some user groups being heavily under- or over-represented in research.

Depending on the respective website, many different types of data can be collected, such as textual data, images, mentions of users, location, or engagement metrics. When collecting social media data, it is crucial to gain knowledge about the platform under study, its' user group and the mechanisms behind it (Tindall et al., 2022). First, one should get an overview of how users interact with the platform and which kind of norms they follow. For example, Twitter offers various functionalities to engage with content, such as liking, retweeting, quoting, or replying to tweets. People employ these functionalities because they assign a specific meaning to them. For example, they might utilize the like button as a feature to express approval of the content or sympathy with the author (Bucher & Helmond, 2018). When working with such data, it is thus important to grasp the meaning they express. Similarly, researchers should try to gain an understanding of how users organize themselves on the platform before collecting data. For example, on TikTok, users have the possibility to create videos in response to other videos (so-called duets), and thus the meaning of a single video can only be fully assessed when including its counterpart in the analysis.

Second, one should consider how algorithms may influence the data and its accessibility. For example, some platforms may delete content which has received only low engagement from users after a short time interval. Therefore, it depends on the time of collection which content will be available, leading to reproducibility issues. Evaluating and detecting bias within the data imposed by algorithms may, however, not always be possible, a problem which will be discussed in section 3.2.1.

2.1.2 Survey and Experimental Data

While social media data can give insights into some aspects of human behaviour, this 'found' data is sometimes not enough to answer the research question. In these cases, data tailored specifically to the research purpose, so-called 'designed data', is needed (Abanoz, 2022). Traditionally, many social science disciplines have relied on questionnaires and lab experiments to assess humans' attitudes, opinions and emotions. Whereas recruiting participants for such studies has been quite cumbersome in the past, the introduction of several computational methods and platforms has simplified these processes: Researchers can design questionnaires and lab experiments, which can be conducted completely online, they can start marketing campaigns on social media platforms to recruit specific user groups and directly link to their study, and they can make use of crowdworking platforms to assign tasks to online participants (Watts, 2014; Zhang, Wang, Xia, Lin, & Tong, 2020).

2.1.3 Biopsychological Measurements

When studying human emotions and cognitive processes, surveys and interviews might not be the optimal research method, as they imply a high chance of self-reporting bias: Study participants could be reluctant to talk about their inner life and give untruthful or socially desired answers. Even more, the participants could be unable to answer questions about emotional and cognitive processes, as they might be not aware of the processes in their brain, or the task of experiencing a situation might interfere with the task of reporting the feelings and thought processes at the same time. To get a more objective insight into the emotions and cognitive processes of humans, a range of biopsychological measurements have been developed. For example, measuring their electrodermal activity through skin conductance response helps to detect when participants experience cognitive arousal (Posada-Quintero & Chon, 2020) and measuring their heart rate variability through electrocardiography helps to investigate their emotional responses (J. Zhu et al., 2019). While some of these measurements have a long history within the social sciences and have been used even before computers were invented, the introduction of computational methods and new technologies has brought many improvements, by enabling the development measurement devices which have a higher sensitivity, are less intrusive, and can, in some cases, even be used in mobile contexts. Furthermore, modern computational approaches have leveraged the analysis and visualization of the large amount of data created by such devices.

One such biopsychological measurement is eye tracking, which includes the detection of gaze positions and movements, as well as the determination of the pupil's diameter. While the positions and movements of the gaze can be used to assess visual attention, the pupil's diameter gives insights into emotional processing, with a larger pupil indicating that the person experiences positive feelings (Lim et al., 2020; Strohmaier et al., 2020). Most eye trackers today are video-based, non-invasive devices, which are either placed stationary in front of the person's face, e.g. on top of a computer screen, or on the person itself, e.g. via glasses or head mounted devices. These eye trackers send out infrared light, which is reflected by the cornea, and collect images of the eye with a camera. In these images, certain features can be detected, such as the position of the pupil center and the corneal reflection. Based on examples of known gaze coordinates, which can be determined in a calibration process prior to the eye tracking, these features can be mapped to a certain gaze point, often expressed in x-y-coordinates. The gaze coordinates can then be stored together with a timestamp to monitor the gaze positions over time (Stuart, 2022).

An important goal of eye tracking is to detect when a person's gaze is focussing on something (fixation) and when the gaze is shifting from one focus point to another (saccade). However, this is not so easy, as the eyes never stand completely still: Even when the eyes are focussing on a specific gaze point, they make small movements, the so-called fixational eye movements. The function of these movements is not well understood yet, but it seems like they prevent the visual percept of a stationary stimulus from fading away, like it would when the image was stabilized on the retina (Rucci, McGraw, & Krauzlis, 2016). The constant movements of an eye makes it hard to define at which point the movements are small enough to be counted as fixation and at which point they should be counted as saccade. There are many algorithms used to differentiate fixations from saccades. Dispersion-based algorithms are grounded in the idea that the eye moves over a large area during a saccade and within a small area during a fixation. For example, the Identification by Dispersion-Threshold (IDT) classifies consecutive gaze points within a certain duration window as fixation, if they are close in space, i.e. if they span an area below a dispersion threshold. Velocity-based algorithms try to make use of the fact that the velocity of the gaze movement is small during a fixation and large during a saccade. For instance, the main idea behind the Identification by Velocity-Threshold (IVT) algorithm is classifying gaze samples as fixation if their velocity is below a certain threshold parameter. In contrast, the Identification by Hidden Markov Models (IHMM) algorithm does not use a fixed velocity parameter, but a Hidden Markov Model with the velocity distributions of fixation and saccade points as states (Andersson, Larsson, Holmqvist, Stridh, & Nyström, 2017; Salvucci & Goldberg, 2000).

Figure 1 shows an example of a mobile eye tracker with two cameras mounted on glasses, one which measures eye movements (c1) and one which makes a video of the participant's surrounding (c2). At timestamp t1, the participant focuses at gaze point g1 and at timestamp t2, at gaze point g2. In a velocity-based algorithm one could measure the gaze velocity by taking the time between t1 and t2 and the angular distance between both gaze points into account (for more information, see section 4.2, fixation calculation).

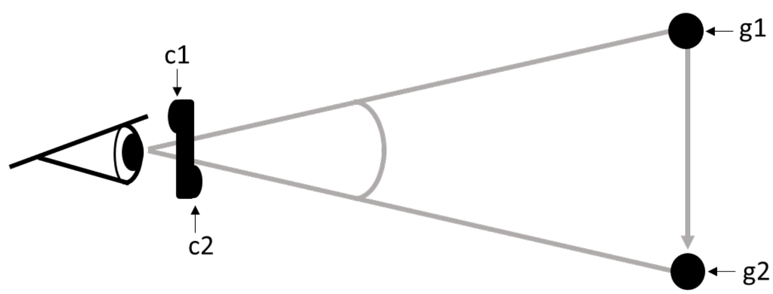

Figure 1: Mobile eye tracker mounted on glasses with two cameras (c1 and c2). The gaze moves from gaze point g1 to gaze point g2.

2.1.4 Sensing Data

Another form of data often used for CSS studies is mobile phone and sensor data. Many mobile phones include sensors such as geolocation positioning sensors (GPS), bluetooth sensors, accelerometers (measuring acceleration), magnetometers (measuring the magnetic field), and gyroscopes (measuring orientation and rotation movements). In addition, when mobile phone users make a call, cell phone providers store the call detail records (CDR), consisting of information such as a mobile user identifier, a timestamp and a geolocation estimated by using the cell phone towers (Grantz et al., 2020; Huang, Cheng, & Weibel, 2019). All of this data can give insights into users location and interaction patterns and can thus be utilized to understand human mobility patterns, perform population and refugee size estimations, model the spread of infectious diseases, optimize emergency responses and evacuation plans, and more (Grantz et al., 2020; Huang et al., 2019; Y. Wang, Li, Zhao, Feng, & Luo, 2020; Zhang et al., 2020).

Researchers can collect such data by cooperating with or buying data from mobile phone operators and similar data providers. Alternatively, they can work with study participants, who voluntarily make their own data available for research purposes, and provide them with a physical sensing device, e.g. a GPS tracker, or an electronic sensing device, e.g. an app, which collects data by making use of sensors in the mobile phone (Huang et al., 2019).

2.2 Data Analysis

2.2.1 Statistical Methods

Statistical analysis refers to a set of methods which can be used to detect trends and relationships in quantitative data.

Descriptive Statistics. The goal of descriptive statistics is to gain an understanding of the data and its distribution, which is, for example, important for detecting outliers, and to decide whether the data is suitable for a certain method of inferential statistics. The data can be described by calculating features like the minimum, maximum, mean, median, mode, standard deviation, skewness and curtosis (Zeitlin & Auerbach, 2019; Zhang et al., 2020).

Inferential Statistics. In comparison to that lies inferential statistics which aims at using information about small samples to draw conclusions about larger populations by applying rules from probability theory (Zhang et al., 2020). In so doing, inferential statistics can help to test hypotheses about relationships between variables and differences between populations. The following summary gives an overview of some of the most commonly used methods in inferential statistics (for more information see Boslaugh, 2012; Zeitlin & Auerbach, 2019):

- Two sample T-Tests: Two sample T-Tests can be used to analyse whether the central tendencies of a variable of two samples differ. While the t-test for independent samples is used for unrelated samples, the t-test for paired samples can be used for repeated measures or related data points, such as the test results of spouses or twins. The test has several presumptions, such as variables in the populations being normally distributed, measurements being independent from each other and variances being homogenous.

- Mann-Whitney U Test: The Mann-Whitney U Test can be described as a nonparametric alternative to the t-test for independent samples, as it can be used for samples which are not normally distributed and do not fulfil the presumption of variance homogeneity. The test compares the rank-sums of two samples with the hypothesis that a randomly chosen value from one sample has the same probability of being smaller than a randomly chosen value from the other sample, as being greater than it.

- Wilcoxon Signed-Rank Test: The Wilcoxon Signed-Rank Test for two samples works similarly to the Mann-Whitney U Test and can be used as a nonparametric alternative to the t-test for paired samples.
- Analysis of variance (ANOVA): The one-way ANOVA can be used to investigate whether the central tendencies of one variable of two or more independent samples differ, by comparing the variances between groups with the variances within groups. Similarly, the one-way repeated measures ANOVA can be used for paired samples and the multivariate ANOVA (MANOVA) can be used for multiple variables. ANOVAs have several presumptions, like an independent and normal distribution of residuals, and homogeneity of variances.
- Kruskal–Wallis H test: The Kruskal–Wallis H test is an extension of the Mann-Whitney U test and can be used as a nonparametric alternative to the one-way ANOVA.
- Friedman test: The Friedman test works similarly to the Kruskal–Wallis H test and can be used as a nonparametric alternative to the one-way repeated measures ANOVA.
- Correlation: The Bravais Pearson Correlation Coefficient can be calculated to test whether there exists a linear relationship between normally distributed variables on an interval or ratio scale. Similarly, the spearman rank correlation coefficient, can be described as a nonparametric alternative to the Bravais Pearson correlation coefficient, and can also be used for variables on an ordinal scale.
- Pearson chi-square test: The chi-square test can be used to assess whether two categorical variables are associated with each other. The test statistic can be calculated by creating a contingency table to analyse the co-occurrence of each value of variable x and variable y, and comparing the observed cell frequency to the expected cell frequency under an independence hypothesis. The test does not only help to reveal whether there exists a relationship between two variables, but can also give insights into what the relationship looks like, i.e. which categories occur with each other more or less often then we would expect with no relationship.
- Regression: Regression can be used to estimate the relationship between variables, i.e. how much variance in one dependent variable is explained by one or more independent variables and their interaction effects. While simple linear regression models a linear relationship between two variables, multiple linear regression models a linear relationship between multiple independent and one dependent variable. Similarly, linear logistic regression models the relationship between variables based on a logistic function and is often used for categorical dependent variables.

2.2.2 Machine Learning

The term Machine Learning describes a collection of methods to automatically identify patterns in a set of data. These patterns can then be used to divide the data into certain categories or make predictions about unknown data.

Unsupervised The goal of unsupervised machine learning is to find hidden structures in a set of unlabelled data. One common approach here is clustering, which aims at dividing the data into sets (clusters), whereby a set should consist of data points which are similar to each other.

One of the most popular clustering methods is k-means clustering. This iterative algorithm randomly distributes k cluster centers in a n-dimensional space, whereby n describes the number of features of a data point and k the number of clusters, as defined in advance by the researcher. It then calculates the distance of each data point in this space to the cluster centers and assigns each point to the cluster with the closest center. Next, it recalculates the position of cluster centers, by taking the means of all data points belonging to the cluster. It then repeats the process of assigning data points to the nearest cluster center and recalculating the cluster center, until the stopping criterion is reached (e.g. until the positions of cluster centers do not change anymore) (Joshi, 2020).

Supervised. The aim of supervised machine learning is to find relationships between input and output values in a labelled dataset and to use this knowledge to infer output values for new, unlabelled data. While regression can be used to infer numerical output values, classification is used to assign categorical labels to data. Data can be assigned to either one of two labels (=binary classification), one of multiple labels (=multiclass classification), or multiple labels (=multi-label classification) (Esposito & Esposito, 2020). There are several models for supervised machine learning (Esposito & Esposito, 2020; Joshi, 2020):

- K-Nearest-Neighbours: The k-nearest neighbour algorithm takes a data-point d as input, determines the k data points of the training set which are closest to d, and returns the averaged categorical class label of these points.
- Decision Trees: Decision Trees are binary trees, with rules on each internal node and the final categories on the leafs. The category label of a data point can be determined by following the path from the root to the leaf which fulfils all rules for the data point.
- Random forest: Random forests consists of a set of different decision trees. The final category label is determined by calculating a label based on each decision tree for the data point and aggregating the response (e.g. by choosing the label, which is received in most decision trees).
- Logistic regression: The logistic regression algorithm takes a datapoint d as input and determines the probability of d belonging to a class c, given

its input values v, based on an optimized regression model which uses a sigmoid function. As the sigmoid function returns values close to zero or close to one, this result can then be used for binary classification (e.g. by deciding that the category label should be c if the output probability is larger than 0.5).

- Naive Bayes: Naive Bayes is a probabilistic algorithm which determines the final category label of a datapoint d by calculating for each category the probability of the datapoint having values v, given that it belongs to this category, and then choosing the category which resulted in the highest probability as final class label. A special form is Gaussian Naive Bayes, which assumes that continuous features are distributed along a Gaussian (normal) distribution.

- Linear Discriminant Analysis: Linear Discriminant Analysis tries to find a linear combination of features in the training set which maximizes the distance between the class means and minimizes the variance within a class. This linear combination can then be used for a representation of the data in which the classes are best separated from each other and a suitable cut-off value in this feature space is chosen to separate the classes. The category label for a new datapoint d can then be determined by representing the datapoint in the feature space and choosing the category, in which it falls according to the cutoff value, as class label.

Neural Networks. An (artificial) neural network (ANN) is a system which resembles the functioning of a human brain and can be used for regression, classification, and clustering. A neural network consists of several connected layers, which include multiple nodes. A node can be described as a computing unit, which takes some input, and weights and combines this input based on specific rules, to calculate some output.

There exists a great variety of different types of layers which determine the rules according to which output values are created. The most common neural networks are feed forward neural networks. During their training phase, each datapoint of the training set is propagated forward through the network, meaning that the values of the datapoint are passed to the nodes of the first layer, which calculate some output values. These output values are then passed on to the nodes of the next layer and so the information travels through the network until it reaches the output layer, which calculates the output value. The output value is then compared to the real data label and the error is now propagated backwards through the network to update the node's weights. By iteratively performing forward and backward propagation, the ANN model is optimized to the training data. The output value for a new datapoint can then be predicted by passing the datapoint d through the optimized network. Another type of neural networks are convolutional neural networks, which are able to process images and can thus be used for extracting image characteristics, performing object recognition or clustering images based on visual similarity

(Joshi, 2020; Moocarme, Abdolahnejad, & Bhagwat, 2020).

2.2.3 Natural Language Processing

CSS provides methods to collect large scale datasets, for example from social media platforms, which often contain textual data. As it is in many cases almost impossible to read through this data in an acceptable amount of time, methods are needed to automatically extract meaning from text. The collection of these methods is called "Natural Language Processing" (NLP). One NLP approach is topic modelling, a technique to uncover latent themes in the text corpus, which is often realized by analysing word frequencies and co-occurences (Kherwa & Bansal, 2019). By identifying these topics, it is possible to gain an overview of the content prevalent in the corpus, to compare the popularity of themes, and to investigate whether certain topics are related to specific user attributes, events and more.

Another NLP method is sentiment analysis. The goal of sentiment analysis is to detect attitudes, opinions or emotions within a text. For example, one can study whether a text chunk expresses a positive, negative or neutral stance, whether it conveys certain emotions, such as anger or sadness, and whether it is written in a subjective or objective way (Mäntylä, Graziotin, & Kuutila, 2018; Zhang et al., 2020). There exist several approaches for analysing the sentiments within a text. Dictionary-based approaches make use of dictionaries, which contain words or expressions associated with a certain sentiment. For example, when trying to infer whether a text expresses a positive, neutral or negative stance, one could use a list for positive expressions (e.g. 'good', 'love') and one for negative ones (e.g. 'bad', 'hate'), or a wordlist with the probabilities of a word being positive or negative. When working with dictionaries, the context of the dataset should be carefully evaluated. For example, users on a website of an amusement park might use the word 'ecstasy' in a positive context (as a synonym for 'joy'), while it might be a rather neutral term on a drug website (as it describes a substance). Furthermore, by using only dictionaries, one might not be able to capture all the relevant information in the text. For this reason, many researchers combine the dictionary-based approach with a rule-based approach, meaning that they define rules about how the system should work with, for example, negations (e.g. 'not good'), degree modifiers ('extremely good'), or punctuation ('good!!!'). Others have implemented machine learning approaches to automatically define such rules based on patterns in the data, or used a combination of approaches described (Dvoynikova, Verkholyak, & Karpov, 2020).

2.2.4 Social Network Analysis

One of the main goals of computational social science is understanding human behaviour. In many cases, humans are not acting independent of others, but are embedded in complex social systems. The analysis of these systems, often called (Social) Network Analysis, can help to understand human behaviour on an individual and collective level.

A network can be described as a graph-like representation of a complex system, which focuses on capturing relational information. Typically, a network consists of nodes, which represent the actors in the network (e.g. humans), and ties, which represent the relationships between nodes (e.g. friendship relations). Network analysis can help to uncover processes in which the network was formed, by investigating attributes such as homophily (=the tendency of actors forming relationships when they are similar to each other), reciprocity (=the tendency of forming mutual relationships), transitivity (=the tendency of two actors connecting when they have a shared partner), or preferential attachment (=the tendency to form relationships with actors, who already have a high number of connections). In addition, networks can give insights into the social structure of a system and differences in the social embeddedness of actors, for example by analysing properties such as the degree distribution, and detecting components (=isolated groups of nodes) or communities (=densely connected groups of nodes) (for more information on this topic see Fu, Lou, & Boos, 2017). Figure 2 provides an example network illustrating the interactions between participants of a scout meeting. As becomes evident, the scouts in the example seem to prefer interactions with peers of the same scout troop (indicating homophily). While there is a high prefential attachment in the boys groups, as boys seem to interact mainly with one group leader, the girls in the hexagon group seem to have formed three communities, and the circle group is densely clustered, with many transitive relationships.

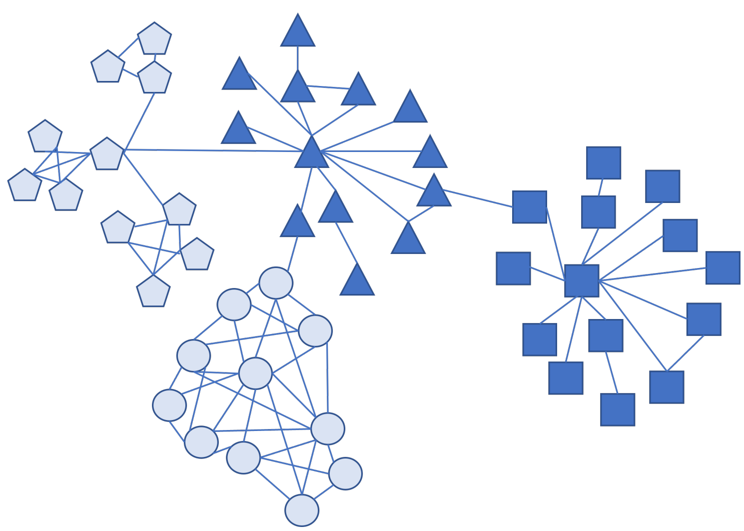

Figure 2: Example network of a scout meeting, with nodes representing scouts, ties representing interactions, shape indicating the scout group, and saturation indicating gender (light nodes correspond to girls and dark nodes to boys).

2.2.5 Simulation

Similarly to the analysis of social systems, the simulation of these systems is a common method in CSS. Simulation describes "the process of designing a model of a real or imagined system and conducting experiments with this model [...]" (Smith, 2003, p. 1578). This method can be used to gain insights into the interactions and processes within a system, and to understand how certain variables and strategies on micro- or macro-level can effect the system as a whole (Pfeffer & Malik, 2017).

To better understand the value of simulations, it is worth to look at an example of a social system. One such example is a school class. In this system we have multiple agents, i.e. students. These agents perform actions on a micro-level, as they show individual behaviour, e.g. by studying, they interact with each other, e.g. by talking to their neighbours, and they interact with the whole system, e.g. by giving a presentation to the class. The system as a whole shows certain properties on a macro-level, e.g. the average grade and the knowledge level of the class, the current volume of the class and the atmosphere within the class (e.g. excited vs. calm). In one way, the system can affect the behaviour of individuals. Knowing the system's properties is thus important for somebody who wants to work with the system and who wants to influence individuals within the system. For example, a teacher might adapt his teaching style in a lesson according to how loud or excited a class is at the given moment. By manipulating the whole system, e.g. by decreasing the volume of the class, he can change individual behaviour, e.g. allowing students to listen and learn better. Furthermore, individual behaviour affects the system's properties. For instance, when multiple students decide to talk loudly, the overall volume of the class will increase. However, often social systems show complex patterns, which means that the overall properties of the system cannot be easily understood by looking at the sum of individual behaviour, as there exist several interaction processes within the system. For example, the volume of a class might not linearly increase with every student raising his or her voice. Instead, student A being loud might have only little effect on the system, while student B being loud can lead to students C, D, E, and F being loud as well because of friendship relations. Simulation can help to better understand these effects, and to analyse how certain changes on a micro- or macro-level can influence the system as a whole (Pfeffer & Malik, 2017). For example, a simulation of the class and its behaviour can aid a teacher in testing different intervention methods to reduce the class's volume and comparing the effects of these interventions.

Simulating a system requires two steps. First, a representation of the system, namely a model, has to be created. This model should show the main characteristics of a system (Pfeffer & Malik, 2017). For instance, the class in our example could be modelled as a network, with nodes representing students, and ties representing friendship relationships between students. Furthermore, each student node could be assessed with certain attributes, such as current volume, grade, or character traits (e.g. level of extraversion, agreeableness, ...). Nevertheless, a social system consists of myriad attributes and characteristics, many of which are unknown to the researcher or which have little or no effect on the system's properties under study. As it is impossible to represent all variables within a system, the researchers have to focus on these, which they can model and which they think of being most relevant for answering their research questions. Moreover, they may have to make certain assumptions about these features, when they are uncertain or unknown to them (Pfeffer & Malik, 2017).

In a second step, the model is used for conducting experiments, i.e. for testing how certain changes in the system on a micro- or macro level will ultimately lead to changes in the properties and/or outcomes of the system. For this, researchers often start with a basic model, consisting of multiple connected entities. They then define rules describing the behaviour of entities within the network as well as conditions under which the behaviour is activated. After this, they begin with a set of initial conditions, e.g. describing the state of each entity, and perform several iterations, in which entities can behave according to the predefined rules. Often these rules are not completely deterministic, but include a random component, thus the same initial conditions and number of iterations can lead to different output states. Finally, the researchers can evaluate the properties of the system. In a next step, they can change the set of initial conditions, or even the model itself, by adding components, relationships or entities, and repeat the process of iterations. By this, they will receive a multitude of different outcomes and will gain insights into what is possible within the model, or how certain initial states can change the model outcome (Pfeffer & Malik, 2017).

Researchers usually use one of two types of simulation for social systems, namely Agent-Based Models and system dynamics models. Agent-Based Models (ABS), in the area of social systems also often called Agent-Based Social Simulation (ABSS), simulate properties on a micro-level, which affect macro-level outcomes. Such models include several individual entities, so-called 'software agents'. Each agent takes its own decision, based on its knowledge, which is limited to its current situation, and shows individual behaviour, whereby this behaviour can change over time. Moreover, agents can interact with their environment and other agents, e.g. to pass information or ideas (Pfeffer & Malik, 2017). For instance, when generating an ABS for the class example described above, the teacher could start with a network model consisting of student nodes, a tie between students, who are friends with each other, and a volume attribute. As the first initial condition, he could set the volume of student A to 100% and all other students to 0%. Next, he could define the rule that in every iteration every student, beginning with student A, will change its volume to the average of all of his friends and his own volume, and perform 10 iterations of simulation. He could then measure the mean volume of the class, before starting with another initial condition, e.g. only student B having a volume of 100%, and performing 10 iterations. By repeatedly changing the initial conditions and comparing the outcomes, the teacher could analyse how students influence each other, or at which point it might be valuable to ask a student to remain silent. However, as becomes evident with this example, a simulation model can be a very simple representation of a system and thus the effects generated by the model might be very far away from effects in the real-world.

In contrast to ABS, system dynamics models simulate interactions between system characteristics on a macro level. One famous example of such a model is the World3 (Meadows, Meadows, & Randers, 1972). In this model complex relationships between industrialization, food production, pollution, population, and the depletion of resources on a global scale were represented to analyse possible effects of global economic growth. While this model should not be used to predict certain events, such as a collapse of the social system at a certain point of time, as the system outcomes heavily depend on the assumptions taken to create the model, it can give insights into how processes influence each other on a global level, and show possible consequences of economic growth, which have not been considered by researchers before (Pfeffer & Malik, 2017).

In summary, due to complexity and uncertainty, social systems can never be exactly represented by a simulation model and consequently, simulations can not be applied to prove real-world behaviour. They can, however, be used to analyse how a system might behave under certain circumstances with a high probability, to gain insights into the wide range of possible processes and outcomes, and to develop and advance theories on social systems (Pfeffer & Malik, 2017).

2.2.6 Visualization

Finally, one of the main methodologies of CSS is visualization. Scientific visualizations fulfill two important goals. First, they can be used by researchers to explore the data and identify patterns, trends and outliers (Foucault Welles & Meirelles, 2015). Figure 3 exemplifies the value of visualizations for the data analysis process. The figure shows three different data samples, consisting of 232 data points each. When only analysed with statistical measures, these samples would be perceived as very similar to each other, as all of them produce the same statistical measures, such as mean, standard deviation and Pearson correlation coefficient, rounded to two decimal points. However, by visualizing the data points on a graph, it becomes evident that the samples follow quite different distribution patterns.

Second, visualizations are an effective tool to communicate research results to other scientists, stakeholders and the general public (Pfeffer, 2017). While the use of visualizations can be valuable in every area of research, it is of particular importance for CSS: As CSS studies are directed to a very diverse audience, with different methodological knowledge, it is especially relevant to help readers understand the approaches used, the meaning of results as well as their limitations and implications. Images are a crucial driver in this process, as they can reduce complexity and highlight important factors. However, creating meaningful visualizations which are easily understood by the audience, is far from trivial. For this task, researchers have to choose a visualization type which matches the data and the aim of the visualization. Figure 4 gives a small insight into the huge variety of graphs, charts and diagrams. As can be seen in the picture, de-

[2] http://robertgrantstats.co.uk/drawmydata.html

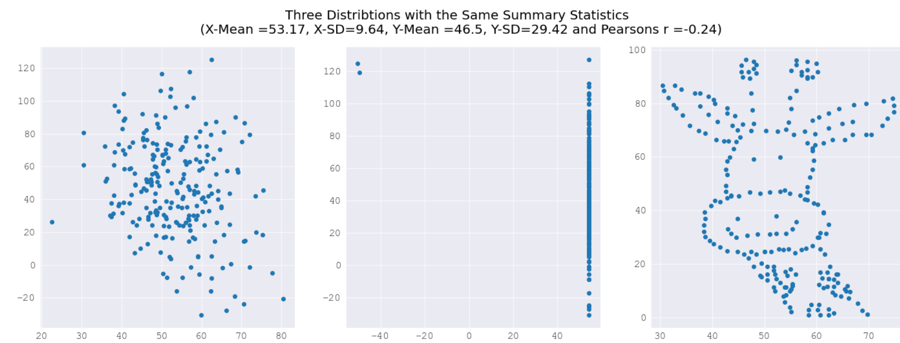

Figure 3: Example of three datasets which are different in structure, but produce the same summary statistics, rounded to two decimal points. The figure is based on own data and code, the prototype for the right subfigure was created with the tool "DrawMyData".[2]

pending on the data type and the finding which should be highlighted, different visualization types might be more appropriate than others.

Nevertheless, simply choosing a suitable image type for the research question is not enough to effectively communicate findings. Researchers also need to make sure that the image can be easily interpreted by the readers (Pfeffer, 2017). This is especially challenging in interdisciplinary research fields, like CSS, in which readers have different methodological knowledge. As Heuer, Polizzotto, Marx, and Breiter (2019) point out in their study on visualization in CSS, social and computational scientists often have different goals and consequently different visualization needs. Moreover, certain image types might be very common in one discipline, while at the same time unusual and hard to interpret for scholars of another discipline. Therefore, researchers should pay particular attention to which image types will allow their audience to quickly process the information presented. It should be noted though that the goal of presenting information in a way which is familiar to all readers and is easily embedded in their cognitive system, can lead to an oversimplification of complex structures. This is especially dangerous, when an image seems to confirm pre-existing assumptions about the world, as readers might examine the image as a metaphor for these assumptions rather than interpret it in its scientific context (Foucault Welles & Meirelles, 2015).

In addition, when designing a visualization, special attention should be given to the mapping between information and graphical elements. There exist several graphical attributes, such as position, size, color saturation (=intensity of color), color hue (=color tints) and element shape. Some of these attributes are more effective for communicating the core message than others, depending on whether a numerical or categorical variable should be shown, or in other words, whether differences in size or groups should be illustrated (Pfeffer, 2017). Figure 5 shows five different examples of visualizing the annual income of two households,

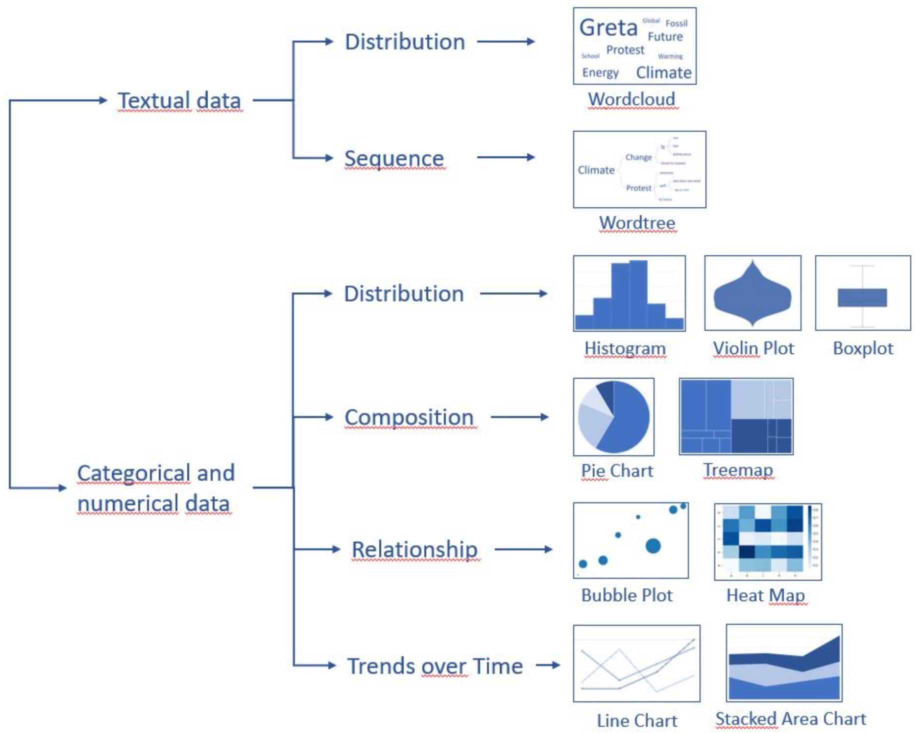

Figure 4: Examples of visualization types which can be used to present textual or categorical and numerical data.

A (=30,000\$) and B (=60,000\$). When mapping the income to position or one-dimensional size, readers can easily perceive that household B has double the income of household A. In contrast, when mapping the income to a two-dimensional size attribute, such as the circle radius, readers will find it more complicated interpreting the accurate difference between household A and B. The problem here is that humans underestimate the difference of the (two-dimensional) area between circles, as their perception focus rather lies on the difference of the (one-dimensional) radius. Consequently, they would perceive B's income as larger than A's income, but not as large as double the size (Pfeffer, 2017). Similarly, showing income by using the color saturation or hue will make it difficult for readers to interpret the differences.

As mentioned above, the interpretation ability depends on the variable which should be visualized. When using the same example to visualize a categorical difference, e.g. to show that household A belongs to the group of single parents and B belongs to the group of two-parent households, color hue would be a good attribute, as readers could see which household belongs to which class at first glance. Nevertheless, it should be considered that color hue can also be prob-

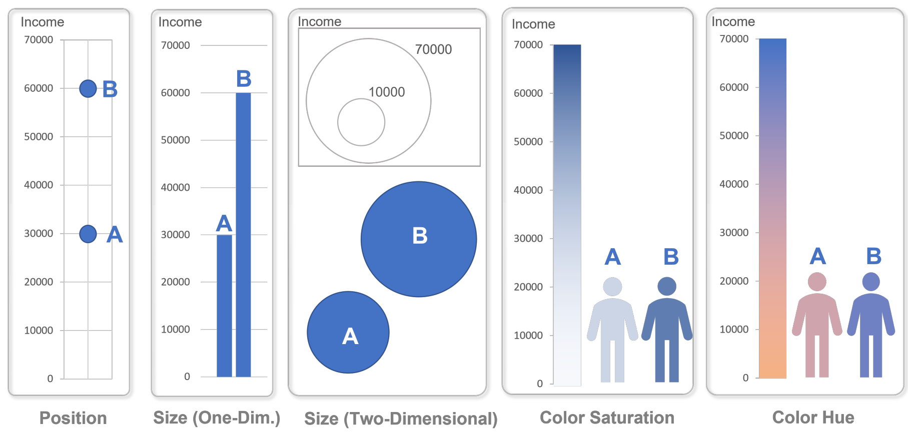

Figure 5: Household income of two households, A and B, visualized by five different graphical attributes.

lematic, e.g. when the article is printed in gray scale or for color blind readers. Furthermore, the colors might be associated by the reader with a certain meaning, depending on the context and culture of the reader. For example, the color red is often associated with 'hot' in temperature figures, with 'bad' in figures on infection risks, with 'conservative' in figures on the political landscape in the United States or with 'Social Democratic parties' in figures on the political landscape in European countries.

In summary, the process of designing powerful visualizations for CSS requires extensive knowledge of the data types and structures, the key message which should be highlighted, the visualization types and graphical attributes suitable for communicating this message, the information context, as well as the readers culture, background knowledge and perception abilities.

2.3 Methods Overview

As illustrated in this section, CSS provides a broad methodological variety by expanding traditional social science methods with new technical resources as well as computational approaches. These methods give tremendous opportunities to study social science research questions, but their application also holds some challenges.

In the next chapter, we will describe these opportunities and challenges, and will explain how collaboration can help to make use of the former and overcome the latter. We will then present three case studies, in which CSS methodologies have been applied to examine research questions in communication science, education science and sociology. Table 3 illustrates the mapping between the studies and the methods described above. Based on each study, we will demonstrate the great value of interdisciplinary collaboration for CSS research.

Study	Type	Methods
Study 1	Data Collection	Web and Social Media Data: Twitter text, images and metadata
	Data Analysis	Statistical Methods: Descriptive statistics, Pearson Chi-Squared Machine Learning: Convol. neural network, K-Means-Clustering Visualization: Donut Chart, Line Chart, Stacked Area Chart, Stacked Bar Chart, Parallel Plot
Study 2	Data Collection	Biopsychological Measurements: Eyetracking Data
	Data Analysis	Statistical Methods: Descriptive Statistics Visualization: Line charts
Study 3	Data Collection	Web and Social Media Data: Erowid text reports and metadata
	Data Analysis	Statistical Methods: Descriptive statistics, Pearson Chi-Squared Machine Learning: Classification (e.g. with Random Forest, Logistic Regression, Linear Discr. Analaysis, ...) Natural Language Processing: Sentiment Analysis Visualization: (Stacked) Bar Chart, Violin Plot, Heatmap

Table 3: Overview of the data collection and analysis methods used for the studies presented in section 4.

3 Opportunities, Challenges and the Need for Collaboration

3.1 Opportunities of CSS

Computational approaches offer new possibilities for studying social science research questions. In section 2 we described how computational methods broadened the methodological spectrum of various social science disciplines, including sociology, psychology, politics, economics, educational and communication research. In the following, we will summarize the main opportunities of computational social science.

3.1.1 Access to New Datasets

First of all, computational social science gives the opportunity to gain access to new and interesting datasets which were not available in traditional social sciences (Abanoz, 2022; Lazer et al., 2009).

New Populations. The use of social media tools allows studying populations which are typically difficult to recruit for social science studies. First, online recruitment strategies over social media platforms can simplify the process of getting in contact with hard-to-locate populations, such as people with a special job profile, people suffering from a certain medical or psychological condition, or people with specific world-views. For example, Ianelli et al. (Iannelli, Giglietto, Rossi, & Zurovac, 2020) used Facebook marketing tools to recruit participants interested in conspiracy theories. Second, the content on social media itself can be used to gain insights into the behaviour and opinions of populations, which are not only hard to locate, but might also be hard to be persuaded to take part in an interview or survey. This includes people who perform illegal or socially unaccepted behaviour, such as drug consumption, or people who fear being stigmatized, when sharing their experiences or thoughts about sensitive topics, such as sexual abuse or suicide.

For example, it is usually difficult and might be even dangerous for researchers to get in touch with extremist and terrorist groups. However, various extremist groups utilize social media for the recruitment of new participants, the promotion and discussion of various contents, mobilization and self-organisation. The collection and analysis of social media data, like ISIS related Facebook pages and Twitter accounts (Awan, 2017), or Jihadi YouTube videos (Conway & Mcinerney, 2008) can thus give insights into the motivations, narratives and activities of such groups (for more research on this topic see (Jain & Vaidya, 2021)).

New Dynamics. In the previous section we have assumed that the offline world impacts online content and that the analysis of social media data can thus give insights into the behaviour and opinions of certain populations. However, research has shown that it also works the other way round: social media are such a great part of many people's lives that own dynamics have developed on these platforms, and so online processes certainly influence the offline world, both on an individual and societal level. For example, during the Covid-19 pandemic, a large amount of false information regarding the existence and source of the virus, as well as regarding therapy and prevention measures, have been spread on social media. These so-called fake news affected people's knowledge and behavioural intentions, leading to fear, confusion, distrust in government and even violent attacks against health personnel, indicating a harm to public health (Rocha et al., 2021). By collecting and analysing web and social media data, it becomes possible to detect fake news and identify who creates them, how they spread and who is vulnerable to share these news (see, for example, Shahi, Dirkson, & Majchrzak, 2021; X. Wang, Zhang, Fan, & Zhao, 2022). The application of computational methods in the social sciences thus provides opportunities to understand the new dynamics in our increasingly digitized society and to develop recommendations for effectively responding to these dynamics.

3.1.2 Analysis of Large Scale Data

Methods from the computational sciences make it possible "to collect and analyze data with an unprecedented breadth and depth and scale" (Lazer et al., 2009, p.722). The collection of large scale datasets, e.g. consisting of web, social media, mobile phone, biopsychological or sensor based data, helps to broaden the scope of social science studies. First, such a collection enables scholars to include more research subjects in their studies than usually possible with traditional social science methodologies. For instance, when evaluating the public's opinion on a specific topic, researchers can store topic-related tweets of thousands of people within a few hours, while reaching the same sample size with traditional survey methods can take years. As there is less effort needed to collect large amounts of data with computational methods, big data scientists often do not perform random sampling, but aim at capturing all data available on a topic (Hilbert, 2015). Consequently, researchers can heavily increase the size of participants under study, which can (but not necessarily must) lead to a more diverse and representative sample. Second, by collecting large scale data, scholars can study phenomena often within a much broader time interval or at smaller time steps than possible with traditional social science methods. This allows them to analyse developments over time, detect temporal patterns and investigate the influence of certain events. For example, when analysing stress at the workplace, researchers could employ a traditional social science strategy and conduct interviews with workers, which will most probably only give insights into their current stress level. With big data techniques, in contrast, they could assess the workers' stress level via mobile heart trackers at every second of the day and thus get a much more detailed picture on stressful situations (e.g., Li et al., 2022).

The analysis of large-scale data, through methods such as machine learning and network analysis, makes it possible to study a system as a whole and to investigate complex patterns within the system. This is especially important for the social sciences as many social systems exhibit emergent behaviour, meaning that the outcome of the system cannot be easily understood when only looking at parts of the system, as the system as a whole is more than just the sum of its parts (Watts, 2014). For example, when trying to understand how a society will react to a Covid-19 regulation, it might not be enough to investigate human behaviour on an individual level and sum up the responses. Certainly, personality attributes, political preferences and knowledge about the virus might explain an individual's reaction to some degree, but this is only part of the picture. Researchers might also need to take the personal relationships and social networks into account, as opinions, emotions, and knowledge towards the virus can spread within social groups, and thus individual's responses may influence the network and the network may influence individual's responses. To summarize, CSS offers techniques to collect and analyse extremely large and highly complex social data.

3.1.3 Reduction of Research Costs, Time and Obtrusiveness

In the traditional social sciences, experiments for studying human behaviour often require a large amount of involvement from both researchers and research subjects. Field studies imply in many cases that scholars spend a large amount of time in the "field" (for example in the supermarket, or at a company), writing down the behaviour they can observe. Similarly, laboratory studies often require researchers to prepare a physical place in a way that all co founding variables are diminished, to set up an experimental situation, to invite participants to this place and to observe their behaviour or ask about their thoughts and emotions. While these methods are valuable and necessary for many research questions, computational social sciences gives the opportunity to add new techniques, which diminish the impact on researchers and study participants. This can have tremendous advantages in regards to research cost, time and obtrusiveness.

Costs and Time. In traditional experiments, a lot of money needs to be invested in researchers spending time in the laboratory or the field, searching for participants and observing human behaviour. This time can drastically be reduced by conducting studies in the online space (Watts, 2014). For example, when placing a product advertisement in a supermarket to study how many people will look at the advertisement and how many will buy the product, it can take several days until enough people have interacted with the advertisement. Therefore, scholars need to spend much time with either physically standing in the supermarket or analysing video records to assess buyers' interest and frequency. Conducting the same study on Amazon can take only minutes, as the population of potential customers is much greater on this platform, and as evaluating click times on the advertisement and the product can be fully automatized. Furthermore, taking part in a laboratory study requires a lot of effort by participants, e.g., as they need to physically visit the laboratory. Thus, to successfully implement such a study, scholars often need a high budget for motivating participants through monetary compensation. By conducting online studies, much of this effort is reduced, and thus participants are often willing to participate for a lower price. Moreover, by making use of data which participants created anyway, such as social media or mobile phone data, scholars can set the data collection costs to (almost) zero.

Obtrusiveness. In laboratory experiments, the environment and the testing devices used may put participants under stress and lead them to feel and behave differently than they would in the real world. Computational social science gives the opportunity to improve the laboratory situation by further developing existing and adding new techniques to study people's behaviour, opinions and thoughts. One common challenge is that the devices used to measure a certain variable might at the same time influence this variable, e.g. biopsychological measurement devices employed for emotion analysis, can induce uncomfortable feelings or even pain, thus confounding participants' emotions. For example, in the past, mostly mechanical eye trackers have been used, which included placing a contact lens connected to a wire physically on the participant's eyes or placing skin electrodes around their eyes. The introduction of video-based and Bluetooth technology has helped to develop eye tracking devices, which are non-invasive and mobile, and can thus be used in more realistic settings (Stuart, 2022, p. 4, 32).

Another challenge is that researchers might want to study how humans react in certain situations, for example during a fire, but these situations cannot be reproduced in a laboratory environment. When asking participants how they would react in a hypothetical situation, the artificial environment might make it difficult for participants to imagine themselves in such a situation. As a solution, virtual landscapes can be designed, allowing participants to experience a situation in an immersive way, while external variables can stll be controlled (Ostermann et al., 2021).

In addition, the feeling of being observed might lead to a tendency of giving socially desired answers or behave in a socially desired way. Shifting studies from a physical environment to an online environment can increase participants perceived anonymity and thus might reduce such desirability effects. In summary, computational social science provides the opportunity to improve the ways with which we study humans, in terms of cost, time and obtrusiveness.

3.2 Challenges of CSS

3.2.1 Unregulated Data Access

One major problem in the area of computational social science is unregulated data access. As described in section 2.1.1, a large amount of CSS studies work with web and social media data, but access to this data heavily depends on the platforms studied. Some social media companies allow researchers to access their data freely, some only allow access to researchers cooperating with the company or paying a certain amount of money, and some companies completely prohibit access to their data. Moreover, there are also differences in the technical difficulty of collecting web data, as some platforms provide easy to access APIs with good documentation or even publish certain datasets in commonly used formats, while data on other platforms is only accessible with complicated web scraping techniques.

The dependency on data providers has severe implications. First, there might arise a divide between researchers who have the money, contacts or technical knowledge to collect the data, and those who don't. Second, the fact that access might only be usable for certain researchers or might be withdrawn after some time, raises concerns of study replicability, especially as researchers might not be allowed to publish the data they have used for their study due to company regulations. Third, as companies are not obliged to reveal information about how the data is stored, which data is accessible or which data has been deleted, it may not be possible to infer whether the data collected is complete and how it might be affected by algorithms (Kirkwood, Cree, Winterstein, Nuttgens, & Sneddon, 2018; Lazer et al., 2020).

These problems do not only arise when collecting web and social media data, but also when conducting biopsychological measurements and collecting sensor data. In many cases, companies do not provide access to real raw data, but to pre-processed data, on which some noise reduction or interpolation techniques have already been conducted. This makes it difficult to compare the data acquired to datasets collected with other devices, and leads to reproducibility issues when the company decides to change their pre-processing algorithms. To summarize, data access is one of the main challenges of CSS and much technical knowledge is needed to collect data and estimate its quality, if at all possible.

3.2.2 Lack of Ethical Standards

Another challenge in computational social science is a lack of ethical and privacy standards (Lazer et al., 2020). Ethical standards are important in every area of research, but they are especially relevant where we have the tools (computational approaches) and the application domain (human behaviour, social dynamics) to do great value and great harm. When such standards are missing, scholars might be unsure of how to work with data in a responsible way, particularly when they have less experience in working with personal data.

One major ethical challenge in CSS, mostly in research with web and social media data, is gaining informed consent from study participants (van Atteveldt & Peng, 2018). Even when the legal requirements are fulfilled, meaning that the terms of use of the respective platform include the right to make data accessible for research, and that users have accepted these terms, there remains an ethical challenge, as users might not be fully aware of what kind of data is used for studies and what researchers can do with such data. Moreover, web and social media data might not only contain information about the users themselves, but also about third persons being mentioned in user posts or being shown in pictures or videos. These people might not even be aware that such content exists about them and thus have never agreed on being part of a study.

A further problem is the exposure of sensitive information. With the combination of data from various sources, it may be possible to deanonymize study subjects, or to infer delicate aspects the study participants might not even be aware of themselves(Oboler et al., 2012) For example, textual social media data can be used to predict peoples suicide risk (Ophir, Tikochinski, Asterhan, Sisso, & Reichart, 2020). Consequently, researchers have the ethical responsibility to only collect data needed for their study, to carefully consider problematic aspects of this data, to evaluate specific vulnerabilities of the population under study and to evaluate the risks of their outcomes before collecting such data. Nevertheless, it may not always be obvious which kind of problematic conclusions will be possible to derive from a certain dataset. For example, a collection of online newspaper articles does not seem like a sensitive dataset. But by investigating the times authors have published an article, and especially the times they did not, it may be possible to detect irregular absences of authors. In pandemic times, an absence of a certain timespan might indicate an infection. Thus, it may be possible to conclude which authors have been infected by the virus, and even more, which authors have been infected at the same time, leading to inferences about contact patterns and relationships. Therefore, analyzing and publishing social data requires an extensive amount of care and consideration.

Moreover, when working in such a critical domain like CSS, in which powerful techniques are applied on social data, it should be considered how the methods and results of a research project can be (mis)used by other parties, because as Sahneh et al. (2021) describe, "[...] when the application of data tools moves from one context, where ethical considerations were deemed irrelevant to the project outcome, to a different context that directly impacts human lives, the result can create harms throughout society" (p. 9). For example, generating knowledge about the ways social media platform mechanisms influence the public opinion, does not only value the public, but can also be utilized by platform providers for monetary interests. Even more, in the same way social media data can help to locate terrorist attacks or understand political discussions, it may be used to detect opponents of a party or government and influence users' political opinions (Oboler et al., 2012). Altogether, CSS projects hold many ethical challenges and, due to a lack of ethical standards in the field, they require a lot of experience and training to prevent possibly harmful practices and outcomes.

3.2.3 Need for Infrastructure

When applying computational approaches, very often some kind of technical infrastructure is required. Researchers have to gain access to data collection and analysis platforms and devices, or set up an environment in which they can write their own code and implement their own tool. They need platforms to share their data, code and results with collaborators and, ideally also with the public. When working with sensitive or large scale data, the infrastructure needs to fulfil certain criteria regarding security, performance and computing power. To make sure that others can also make use of a dataset, trained model or tool in the future, the technology developed must be documented and maintained, often beyond the project's timeline. All of these requirements constitute institutional barriers, especially for social science departments. Not every research group has the knowledge, time and budget to set up and maintain such an infrastructure and so sometimes CSS projects end before they can even get started (Lazer et al., 2020; van Atteveldt & Peng, 2018). In brief, lacking infrastructure is one of the main challenges of CSS, and technical resources and knowledge are required to overcome this barrier.

3.2.4 Low Impact for Social Sciences

While most computational social science projects aim at solving a certain social science research problem, many CSS studies fail in integrating their work in the social science context and truly creating an impact for the social science communities. Or, as Watts (2014) puts it

> [..] much of computational social science has effectively evolved in isolation from the rest of social science, largely ignoring much of what social scientists have to say about the same topics, and largely being ignored by them in return. (p.22)

One barrier in this area is a lack of serious engagement with the theories of the social science disciplines (X. Wang, Song, & Su, 2022; Theocharis & Jungherr, 2021). Basing studies on well developed theories is a crucial step to be able to fully interpret and understand the impact of data and study results (Bravo & Farjam, 2017). By simply applying a computational method on certain data, without embedding the project in the theoretical background of the field, researchers risk producing studies which are largely descriptive and do not really help in advancing understanding of the research problem. However, as frequently the case in interdisciplinary research (Cairns, Hielscher, & Light, 2020), (computer) scientists with little or no experience in the respective social science discipline might find it challenging to understand the theories and concepts of the social science field, which are based on different, often implicit, epistemologies. Even for academics more experienced in the discipline, it might be difficult to connect to related social science studies when employing new methodologies: The results of traditional studies, especially when qualitative methods have been applied, might not always be directly comparable or easily translatable to results achieved with computational, more quantitative approaches. Thus, rigorous work is needed to identify what is known about a research problem based on prior work, determine gaps within this knowledge, decide on whether the methodologies planned can help in advancing these gaps, interpret the results with taking their limitations into account, set the conclusions into context to prior work and analyze how they relate back to what is known and what has been assumed.

Another challenge in creating an impact for the social science communities is making CSS studies accessible to them. Several scholars criticise that only few CSS articles are published in social science journals and that some CSS conferences are frequented by only few social scientists (Watts, 2014; Bravo &

Farjam, 2017). Therefore, academics have to carefully consider how they can reach the social science community and contribute to the larger discussion on the research problem. In part, this is also a question of making findings understandable for the audience. Social scientists may not be familiar with computational approaches and thus special emphasis should be given to describe and explain the methods used. In doing so, scholars should take care of adapting to the language and terminology of the discipline, and create visualizations, which can be easily interpreted by the audience, but do not invite to oversimplification and false conclusions (Foucault Welles & Meirelles, 2015). It should be noted here that this does not mean the work should not serve the computer science community as well. However, we believe that both communities have to be included to make meaningful progress on the research problem. Altogether, creating an impact for the social sciences remains challenging in CSS, and a lot of domain expertise is needed to seriously engage with the theories and language of the social science field under study.

3.2.5 Misuse of and Overtrust in Computational Methods

With the goal of making computational methods applicable in the social sciences, a large variety of packages, tools and systems have been developed for researchers with less technical expertise. For example, numerous companies, like Bright Data,[3] sell pre-processed datasets from social media platforms, which come in easy-to-use formats like excel tables; various python and R packages, e.g. nltk and R Vader (Hutto & Gilbert, 2014), allow to uncover intricate concepts, such as sentiments or hate speech within large text corpora by only executing one function; and biopsychological measurement devices, such as the Tobii Eye trackers,[4] are often sold in combination with licences to analysis platforms, which allow to import the collected data, calculate key values and produce sophisticated visualizations with just a few clicks.

The promise behind these tools and packages is simple: researchers can circumvent the long and cumbersome process of writing code for data collection, training their own machine learning models, investigating how biopsychological data can be converted into numerical structures, and designing sophisticated visualization algorithms, and concentrate on the research questions they are really interested in. This development certainly has a great potential for diminishing technical barriers between disciplines and is even useful for researchers who would have the technical knowledge to produce the algorithms needed by themselves. At the same time, it holds a great danger, namely the misuse of tools and techniques based on a lack of a deeper understanding of the underlying algorithms. For example, when conducting eye tracking, many steps like converting infrared into electrical signals, calculating the pupils' positions, mapping gaze to x-y-coordinates, or classifying saccades and fixations are executed. For each of these processing steps, a variety of computational approaches exist, some of which are predetermined by the company selling the eye tracking

[3] `https://brightdata.de/`
[4] `https://www.tobii.com`

device, while others need to be chosen by the researcher in the data analysis phase. The selection of algorithms and parameters will heavily influence the end result and should thus be decided on wisely. Researchers who, however, do not have the domain or technical expertise to discern these approaches, with their potential and limitations, might choose devices or techniques unsuitable for their research problem. In particular, they might use the parameters and algorithms proposed as default option by the system, without understanding the consequences. Furthermore, in underestimating the impact of their decisions, they might fail to report the parameters and algorithms used, which makes it difficult to validate and replicate the study, a common issue which is described in more detail in section 5.1.2.

In addition, a lack of knowledge in the technical domain may lead to 'overtrust' (that is, trusting too much) in what an algorithm or tool can contribute. When employing a technology, it is pivotal to understand its limitations. For example, a machine learning model to classify hate speech might be biased towards certain minority groups and should thus not be employed in specific settings (Davidson, Bhattacharya, & Weber, 2019). However, research has shown that people with less digital literacy tend to trust in algorithmic decisions, even when they do not know what affects these decisions (Krügel, Ostermaier, & Uhl, 2022). This is especially problematic when applying technology to make decisions with societal impact, for example in a hiring process or for choosing a therapeutic treatment.

All things considered, researchers need to gain detailed information on the algorithms and tools they want to employ. However, this knowledge is not easily acquired for novices in the domain, as the literature on these approaches is often formulated in the jargon of computational science and based on the ground concepts of the field, which might not be familiar to scholars outside the computer science community (for more information on language barriers between computer and social scientists see 5.2.1). Consequently, preventing misuse and overtrust is one of the main challenges of CSS and requires extensive technical expertise.

3.3 The Need for Collaboration in CSS

The preceding sections have evaluated some of the main challenges in computational social science projects, which leads to the question of how researchers can overcome them, if at all possible. To face each of these challenges, an extensive amount of knowledge and skills from both a social and computational perspective is needed.

But can we expect computer scientists to fully comprehend the concepts of a whole new discipline, carve out the most pressing research questions, develop a theory that is well embedded into the long history of research in this area, learn to speak the language of their population under study, evaluate all ethical peculiarities which might come with working with this population, and accommodate themselves to the norms and terminology of the specific research community to present the results in a way that they benefit the social science discipline? And can we expect social scientists to acquire knowledge of computational methods to understand all the technical details, algorithms and approaches required to solve a specific research problem, critically reflect on the suitability of certain techniques and gain an overview of possible alternatives, learn programming to adapt their code to their data, and set up and maintain a computational infrastructure which they will most probably only need for this specific project?

Although computer scientists will certainly benefit from acquiring knowledge in the social sciences, and the other way round, it is unrealistic to demand that a single person should learn all the things needed to solve such CSS problems, which are usually highly complex and require extensive knowledge in multiple disciplines. Therefore, we propose that the most promising way to answer such interdisciplinary research questions and overcome the typical challenges of Computational Social Science projects is a collaboration between researchers (or experts) from both a computational and a social science domain.

This section and the following chapters elucidate the importance of interdisciplinary collaboration in CSS. First, we describe similarities and differences between computational and social scientists to illustrate what both sides can contribute to CSS research. Next, we present different forms of collaboration to reveal how researchers and experts from both domains could work together on a research problem. We then show three real-world examples of interdisciplinary collaborations in CSS, and explain the value of collaborative work in each of these projects. Furthermore, we identify factors which might lead to difficulties in collaborative projects, and provide ground rules for successful collaborations.

3.3.1 Characteristics of Computational and Social Science

To explain why both a computational and a social science perspective are essential for answering questions of Computational Social Science, it is helpful to highlight the specific characteristics of these disciplines. Social scientists and computer scientists do not only provide different skill sets and knowledge, but are usually accustomed to different ways of thinking and performing research.

Computer scientists have an expertise in technical methods (X. Wang, Song, & Su, 2022). For CSS projects, this means that they can provide access to novel and often large scale datasets, by implementing web scraping techniques to collect social media and web data, developing devices and algorithms to collect new forms of data, such as heart rate measurements or eye tracking, and designing systems to allow the conduction of online or virtual experiments. When working with data provided by third parties, they might be able to perform measures to estimate the data quality and detect potential biases and errors in pre-processed data. Even more, they can contribute methods for analysing data, and can evaluate the implications and impact of results on a technical level. In addition, computer scientists typically have the resources and skills to set up an infrastructure needed to collect, analyse and share data (Parti & Szigeti, 2021).

Furthermore, while there is a need in CSS projects to critically reflect on the use of technological devices and algorithms (Abanoz, 2022; Lazer et al., 2020), computer scientists contribute the expertise to evaluate and question computational approaches, formulate best practices regarding their application and point out potential pitfalls and biases.

Social scientists, in comparison, have an expertise in their field, which means that they can identify the essential research questions of their field, typically know the scientific state of the art regarding these questions, and can develop theoretical models, which are in line with the knowledge base of their field (Watts, 2014; X. Wang, Song, & Su, 2022; Parti & Szigeti, 2021). In regards to a CSS project, they thus have the expertise to formulate the research problem and embed it into its theoretical context. Moreover, they can evaluate the study results in the context of these theories, recognize the result's implications for their discipline and assess their impact on a societal level.

Moreover, social scientists often already have some experience with the population under study and can provide insights into their language, culture, preferences, and behavioural patterns. In CSS projects, this knowledge is extremely valuable for choosing the right data collection strategy and gaining an understanding of the data. For example, in a project on analysing anorexic behaviours on social media, a clinical psychologist might already know the social media platform and keywords frequently employed by the user group and might be able to explain specific behaviour or terminologies.

In addition, many social scientists analyse personal data on a regular basis, and thus have more training and experience to encounter the ethical challenges associated with this process (Oboler et al., 2012). For CSS projects, this expertise is very helpful to identify possibly problematic data collection, analysis and presentation strategies, and to develop mechanisms to mitigate or prevent a negative impact on the population under study.

In summary, while the expertise of computational scientists can help to mitigate problems with data access (as described in section 3.2.1), to overcome the barrier of lacking infrastructures (3.2.3), and to prevent misuse and overtrust in computational methods (3.2.5), the perspective of social scientists is valuable for facing the challenges of creating true impact for the domain of the research problem (3.2.4) and performing research, which is ethically responsible (3.2.2).

3.3.2 Combining Computational and Social Science

As described above, both the computational and social science perspectives are important for answering CSS research questions. By combining them through collaboration, researchers can encounter the challenges of computational social science and make use of its full potential.

In general, collaboration can be defined as a situation in which several people contribute their knowledge, skills, ideas, perspectives, data and tools to reach a common goal (Bramley & Ogilvie, 2021; Cairns et al., 2020). There are plentiful reasons why researchers collaborate. In many cases, academics want to gain access to knowledge or infrastructure which they cannot provide on their own. Sometimes, the goal here is not only using this knowledge for a certain research problem, but also learning from the other discipline, may it be by gaining a new perspective on research in general or getting trained in using a specific method or tool. Also, being able to distribute work to multiple parties to increase productivity, is a strong motivation for collaboration. In other cases, the collaboration is rather motivated by external factors, for example to apply for specific funds, or to increase visibility and prestige in a certain research community. Additionally, some researchers also have personal reasons for joining a collaboration, for instance because they find pleasure in working with other people (van Rijnsoever & Hessels, 2011).

In academic research projects, the degree of how strong the parties are involved in each step of the research process and how many decisions are made together, heavily varies. Consequently, there exist several definitions of collaboration. **Multidisciplinary collaboration** describes a situation in which academics with different backgrounds work on the same research problem, either sequentially or in parallel. Often the work is divided in such a way, that scholars with a specific expertise address specific aspects of the problem and present their findings in separate publications.

Interdisciplinary collaboration, in contrast, can be viewed as a more involving form of multidisciplinary collaboration, in which academics with different backgrounds work jointly on the same research problem, by integrating their perspectives and trying to create a common understanding of the problem (Barthel & Seidl, 2017; Aboelela et al., 2007). Typically, the participating parties do not simply fulfill a single task and hand it over to the other members, but they are involved in multiple project stages (Aboelela et al., 2007). The main assumption behind interdisciplinary collaboration is that the problem can not (or only with a disproportionate amount of time and resources) be solved by one discipline alone and, thus, the parties are mutually dependent on each other. Therefore, the integration of methods and solution approaches to enable mutual problem solving is of particular importance for interdisciplinary collaboration (Krause-Jüttler, Weitz, & Bork, 2022). Such an integration may not always come easily: As illustrated by Roschelle and Teasley (1995), a collaboration does not simply emerge through the coexistence of people, but requires effort from all parties to create shared knowledge.

Furthermore, **transdisciplinary collaboration**, as defined in Barthel and Seidl (2017), describes the cooperation between academic researchers and non-academic actors, such as field-experts or decision makers. In many cases, these collaborations can be extremely valuable, as non-academic actors often provide a new perspective on the research domain, have extensive knowledge of the population under study and can help to create a serious impact on a societal level. When describing the value of interdisciplinary collaboration throughout this work, we also include transdisciplinarity in the definition of interdisciplinarity, as we believe that both forms of collaboration benefit Computational Social Science in a similar way.

In conclusion, there are many different ways in which computer and social scientists (or experts) can conduct research together. We believe, that by deeply integrating the opinions, ideas, skills and knowledge of both sides, scholars can create a mutual understanding of the research problem which by far exceeds the sum of what individuals could achieve alone. Thus, for CSS projects, we recommend an interdisciplinary collaboration with a strong involvement of the parties. Certainly, such an involvement requires a lot of effort and time to allow fruitful discussions and learning processes. However, as we will show with the following three studies, this effort is rewarded with the potential to gain a deep understanding of CSS research problems and overcome some of the main challenges of CSS.

4 Case Studies

In this section, we present three case studies for inter- and transdisciplinary collaboration in Computational Social Science. Each study exemplifies a major opportunity of CSS and demonstrates how CSS can be valuable for expanding knowledge in the social sciences. At the same time, each study presents a common challenge in CSS and we explain how collaboration with scholars and experts from the social sciences has helped us in overcoming this challenge.

Study 1 was conducted in the area of Communication Sciences. In this work, machine learning techniques were used to evaluate the visual discourse on climate change on Twitter. As an example of the opportunities of CSS, this study amplifies the value of large scale data analysis, and shows how such analysis can aid in studying data spanning a much larger sample size and time frame compared to traditional social science methods. Based on this study, we describe the challenge of designing a CSS study, which has an impact on the social sciences. We explain how an interdisciplinary collaboration between computer and communication scientists helped us to embed the study in a social science framework and provide meaningful results for the communication sciences.

Study 2 lies in the area of Education and elucidates how eye tracking can be used to compare teachers' attention within a classroom. As an example for a major opportunity of CSS, it demonstrates how CSS contributes new tools which are valuable to diminish the obtrusiveness on study participants. Furthermore, we discuss the challenge of overtrust in computational methods, as this study showcases a certain lack of reflection on eye tracking algorithms in the field of education. We explain how an interdisciplinary collaboration between computer and educational scientists helped us to critically reflect on the methods applied in the field and to investigate the impact of algorithmic decisions on eye tracking results.

Study 3 lies in the area of sociology and demonstrates how machine learning methods can be applied on web data to gain a deeper sociological understanding of drug consumption. As example for a main opportunity of CSS, the study reveals how CSS can open access to new populations and variables. In addition, it exemplifies the challenge of ethical pitfalls in CSS, and we explain how a transdisciplinary collaboration between computer scientists and drug community experts helped us to design the study in an ethical way which protects the population under study and their interests.

4.1 Study 1: (Social) Media Logics and Visualizing Climate Change: 10 Years of #climatechange Images on Twitter

This publication is RELEVANT TO THE EXAMINATION.

Authors
Angelina Mooseder, Cornelia Brantner, Rodrigo Zamith, Jürgen Pfeffer

In
Journal Social Media and Society, 9(1), peer-reviewed, published on 03-2023, `https://doi.org/10.1177/20563051231164310`.

Abstract
Images have become a key vehicle for communicating climate change, especially in a visually oriented social media ecosystem. However, few studies have examined the ways in which climate change is visually communicated on those platforms. This study addresses that gap by examining more than two million images appearing alongside tweets containing #climatechange, identifying the types of images different stakeholders share and the amount of engagement those images elicit. It highlights differences in the image types that are published frequently (e.g., textual visualizations), the image types that users prefer to engage with (e.g., protest images), and the impact of bots and a cyclical communication pattern keyed to focusing events. These findings are then evaluated through a conceptual framework of media logics, which helps highlight some of the distinctions between (news) media logic and social media logic—and their emerging hybridization—within the context of climate change communication.

Contribution of Thesis Author
Theoretical conceptualization, data preprocessing, algorithmic design, discussions and implementation, data visualization, as well as manuscript writing, revision, and editing.

SAGE

(Social) Media Logics and Visualizing Climate Change: 10 Years of #climatechange Images on Twitter

Angelina Mooseder[1], Cornelia Brantner[2], Rodrigo Zamith[3] and Jürgen Pfeffer[1]

Abstract

Images have become a key vehicle for communicating climate change, especially in a visually oriented social media ecosystem. However, few studies have examined the ways in which climate change is visually communicated on those platforms. This study addresses that gap by examining more than two million images appearing alongside tweets containing #climatechange, identifying the types of images different stakeholders share and the amount of engagement those images elicit. It highlights differences in the image types that are published frequently (e.g., textual visualizations), the image types that users prefer to engage with (e.g., protest images), and the impact of bots and a cyclical communication pattern keyed to focusing events. These findings are then evaluated through a conceptual framework of media logics, which helps highlight some of the distinctions between (news) media logic and social media logic—and their emerging hybridization—within the context of climate change communication.

Keywords

Climate Change; Visual Analysis; Visual Representation; Social Media; Machine Learning; Computational Social Science; Issue Attention Cycle

Introduction

Climate change is today recognized as one of the biggest threats to human life, and its trajectory can be influenced to some extent by human action (or inaction) (IPCC, 2022). Global leaders, in response to alarms about catastrophic impacts on the Earth's environment, have made pledges to accelerate decarbonization efforts. Such pledges are meaningless without a coordinated collective response, which is especially challenging to organize because climate change 'feels' abstract to many people, does not develop in a linear fashion, requires international cooperation, and has become highly politicized (Boykoff et al., 2022; Chapman et al., 2016; IPCC, 2022). Given these challenges, it is unsurprising that effectively communicating the issue of climate change—and getting people to care about it—remains a vexing challenge for different stakeholders, especially those who seek to align public understanding of climate change with the scientific consensus (Brossard and Scheufele, 2022).

A key vehicle for communicating information about climate change is visual media. Images have become central to today's information environment as evidenced by the popularity of visuals on social media, which have become both important sources of news and general information as well as sites for meaning-making (Highfield and Leaver, 2016; Pearce et al., 2020). Images also offer a potential break from the typical critique of climate change science as being overly technical and complex—and therefore alienating—by offering the opportunity to simplify ideas into easily consumed objects that immediately draw upon and reinforce multiple associations (Pearce et al., 2019; Schäfer, 2020). In light of this, it is unsurprising that climate change stakeholders have turned to images to make an abstract and complex phenomenon more concrete (Wozniak, 2020).

While there has been some research on the use of climate change images by professional journalists and through professional media vehicles like newspapers, online news sites and broadcast news programs (see the reviews by Agin and Karlsson, 2021; Anderson, 2009; Pearce et al., 2019; Schäfer, 2012; Schäfer and Schlichting, 2014), there is far less research with regard to the use of images on social media platforms (cf. Hopke and Hestres, 2018; Pearce et al., 2019,2; Schäfer, 2020). This is

[1]School of Social Sciences and Technology, Technical University of Munich, Germany
[2]Department of Geography, Media and Communication, Karlstad University, Sweden
[3]University of Massachusetts Amherst, United States

Corresponding author:
Angelina Mooseder, School of Social Sciences and Technology, Technical University of Munich, Germany
Email: angelina.mooseder@tum.de

an important oversight because social media have become major—if not the primary—news source for large segments of people around the world (Newman et al., 2022).

One platform where images play a prominent role is Twitter, which is used by various climate change stakeholders to influence public understanding. Although it originated as a text-oriented microblogging platform, Twitter has evolved to become a major distributor of images in part because of its sociotechnical design. While Twitter's userbase only accounts for a small sliver of the world population, it is nevertheless influential among journalists, policymakers, academics, and so-called political junkies (Freelon, 2019). However, relatively little is known about how climate change images are used on Twitter (cf. Pearce et al., 2019,2; Schäfer, 2020). Instead, extrapolations are sometimes made from the larger body of literature on traditional mass media, and especially professional news media, which are governed by different value systems, processes, and sociotechnical structures (Thimm et al., 2018; Tsuriel et al., 2021; Van Dijck and Poell, 2013).

This study aims to address this shortcoming in the literature by adopting a computational approach to examine more than two million images included in climate change-related tweets over 10 years, from the time Twitter launched its image-hosting feature until the middle of 2021. In doing so, we offer an empirical contribution by identifying the kinds of stakeholders that are using Twitter to communicate images related to climate change, the types of images those stakeholders share, and the amount of engagement (e.g. likes, retweets) that those images elicit. We also offer a theoretical contribution by discussing the distinctions between (news) media logic and social media logic—and their emerging hybridization—within the context of climate change communication. Finally, we introduce a large-scale computational methodological approach and operationalize it through an unsupervised machine learning.

Literature Review

Media and Logics

In order to understand how images related to climate change might be communicated on Twitter—and how such communication might differ from traditional news media, which is the context most often examined in the literature—it is helpful to draw upon the conceptual lens of *media logic*. Media logic refers to "[t]he assumptions and processes for constructing messages within a particular medium," with the communicator's behavior being structured by the "rules or 'codes' for defining, selecting, organizing, presenting, and recognizing information as one thing rather than another" (Altheide, 2004, p. 294). Hjarvard (2008, p. 113) refines the concept by underscoring the importance of the "technological modus operandi" employed by media actors, which in turn recognizes the structuring role of technology in shaping those 'codes.' Media logic can therefore be understood as a subset of the broader conceptual framework of institutional logics (Hjarvard, 2018), and one that focuses on how media vehicles are instrumentalized for communication, and in particular, to convey information about specific issues.

A media logic lens is useful for two reasons. First, it draws attention to the fact that communicators must abide, to a certain extent, by a particular "rhythm, grammar, and format" in order to be seen as legitimate media actors producing legitimate media messages, while recognizing a certain degree of fluidity for those things (Altheide, 2004, p. 294). Second, it recognizes that technological affordances play a major role in how, and the extent to which, the communicators can abide by those codes (Altheide, 2016).

The concept of media logic is rooted in the distinct context of professional news media and has tended to focus on one-way mass media (Asp, 2014; Van Dijck and Poell, 2013). Indeed, Altheide and Snow's original formulation of the concept in 1979 arose from an evaluation of commercial television (a highly visual medium) and the efforts of media producers to efficiently (profitably) coordinate political coverage. Consequently, the formal and informal codes associated with media logic are shaped by the professional values and the media vehicles typically utilized by professional news organizations, even as they are inextricably linked to external logics (e.g., political and market logics). In other words, the logic is constrained and enabled—at least in places like the United States—by shared understandings of newsworthiness criteria, performances of neutrality, and the advancement of civic-minded ideals in order to demarcate professional boundaries, as well as by temporal and spatial restrictions such as the expected length of a newspaper article or segment on a television broadcast (Asp, 2014) and political and economic considerations (Altheide, 2004).

The rise of social media over the past two decades has required a rethinking of that concept. Van Dijck and Poell (2013) have argued that the mass (and news) media-centric conceptualization of media logic needs to be expanded into a sister concept of *social media logic*, which remixes traditional media logic across four dimensions. First, *programmability* recognizes a distinction between the scheduled editorial approach adopted by traditional media and the more unpredictable two-way crowdsourced nature of social media. Second, *popularity* points to social media's emphasis on quantifying and rewarding 'likeable' phenomena and the feedback loop generated by its technical design. Third, *connectivity* similarly highlights social media's emphasis on ‘spreadable’ content and the use of different repertoires to forge sub-communities. Finally, *datafication* involves the quantification of phenomena in ways that are both visible and invisible, as with prominently displaying the number of likes received by an image while invisibly personalizing one’s information feed based on multiple quantified inputs (Van Dijck and Poell, 2013).

In short, social media logic amplifies existing aspects of media logic and adds new ones. For example, while shareability has been recognized as an element of the newsworthiness criteria used by journalists (Harcup and O’Neill, 2017), it acts as a central component linking the four key dimensions identified by Van Dijck and Poell (2013). Moreover, social media logic permits the embrace of new communication strategies as it is not bound by the same senses of professionalism as its traditional counterpart (Tsuriel et al., 2021). Indeed, different stakeholders, from major companies to individuals users, have taken to using the meme visual format (among other amateurish visual artifacts) to communicate a mixture of messages, a practice that is both accepted and frequently used in social media exchanges (Murru and Vicari, 2021). As news media have expanded their presence on social media platforms, a contested, hybridized logic has begun to emerge (Tsuriel et al., 2021).

Climate Change Visuals in News and Social Media

Most studies of climate change have focused on traditional news media, such as newspapers and television newscasts (and their online versions), in the Global North (see the reviews by Agin and Karlsson, 2021; Schäfer and Schlichting, 2014). Three findings from that literature merit particular attention here.

First, there is a recurring observation that professional news media attention to climate change is cued by particular focusing events that draw attention for a time, but that such attention wanes shortly thereafter until the next focusing event (Djerf-Pierre, 2012; O’Neill, 2020; Schäfer and Schlichting, 2014). In other words, the pattern of climate change coverage frequently follows what Downs (1972) termed the ‘issue-attention cycle’ (see Brossard et al., 2004), leading to the critique that climate change coverage is cyclical rather than sustained. Scholars have argued that this is due in part to the politics tied to the issue and the media logic that governs professional journalism in the West (Tschötschel et al., 2020). Boykoff and colleagues’ (Boykoff et al., 2022) global longitudinal study tracking climate change coverage in traditional news media since 2004 shows that while climate change coverage has generally increased, there are still temporal peaks in coverage that confirm that there have been several climate change issue-attention cycles since 2004.

Second, professional journalistic norms that value neutrality have been found to hamper climate change coverage by giving disproportionate voice to climate change skeptics and deniers, especially in the US, thereby creating a distorted image of the causes and effects of climate change (Tschötschel et al., 2020). As social media has displaced some of the gatekeeping power previously held by professional news media (Wallace, 2018), frustrated stakeholders—both among climate skeptics/deniers and advocates/believers—have actively sought to communicate directly with their audiences by bolstering their social media presence (Pearce et al., 2019; Schäfer, 2012,2). Moreover, scholars have found that bots are being increasingly instrumentalized on social media to shape discussions around issues like climate change (Chen et al., 2021).

Third, the literature on the use of images to communicate climate change-related messages is underdeveloped in relation to the broader literature on its written counterpart. Nevertheless, the work on visual communication about climate change has produced important findings. Reviews of existing literature on the visual coverage of climate change (e.g., O'Neill and Smith, 2014; Schäfer, 2020) show that two types of images are especially common in news coverage: those depicting the consequences of climate change (e.g., extreme weather events, desertification, impacts on biodiversity, the iconic polar bear on the too-small ice floe) and those showing prominent people (e.g., politicians and celebrities). Studies also find the following types to be present, though less common: infographics (e.g., temperature curves, visualizations of the greenhouse effect), causes of climate change (e.g., power plants, traffic), nature (e.g., unsoiled habitats), solutions and opportunities for action (e.g., energy-saving lamps, alternative energy production), and protests. With regard to the pattern of coverage, the UNFCCC World Climate Conferences (COP) and the IPCC Assessment Reports have been identified as international events that trigger global media coverage of climate change (Schäfer, 2020). For example, O'Neill (2020) observed in their analysis of the use of images in five U.S. and U.K. newspapers between 2001 and 2009 that events like the release of the IPCC's fourth assessment report and COP15 drove notable increases in the coverage. They also observed changes in the kinds of images that were featured, with the latter part of the decade featuring a notable increase in climate cartoons, protest imagery, and visual synecdoches that were subverted and parodied within right-wing, climate-skeptic newspapers (O'Neill, 2020). Finally, the literature offers only some insight into the potentially distinct visual communication strategies used by different sets of stakeholders (e.g., journalists, non-governmental organizations, political actors, scientists). Interviews conducted during COP18 and COP19 by Wozniak et al. (2017) found that communicators from non-governmental organizations believed that images of protests and 'PR stunts' were particularly effective, while communicators from government delegations believed that images of political actors were particularly effective. This finding contrasts with those of some prior scholarship. Images of politicians have previously been found to be ineffective and to elicit the least positive responses from audiences, and protest images, while effective among those who express concerns about the climate, tend to generate negative responses from skeptics (Chapman et al., 2016; Corner et al., 2015; Leviston et al., 2014; Metag et al., 2016; Wang et al., 2018). In other words, it remains unclear if the image types that stakeholders think are effective really help in making the topic of climate change more accessible to the audience.

The aforementioned distinctions between media logic and social media logic raise questions about the extent to which these findings might apply to social media, and Twitter in particular. As such, the lack of a robust body of literature on the visual communication of climate change on social media hampers our understanding of the phenomenon. While there has been an increase in scholarship on how the issue of climate change is communicated on and through social media (see the reviews by Pearce et al., 2019; Schäfer, 2012), those studies have focused largely on written communication. One exception in this area is the work of León et al. (2022) who analysed 380 images that were included in top-ranked tweets about climate change over a one-year period. A key finding in relation to the present study is that the majority of images focused on people, and that the majority of identifiable people were indeed politicians and celebrities. Thus, in this regard, the results were consistent with those of traditional media. Notably, however, their study design makes it impossible to ascertain whether such communication followed the cyclical nature observed in traditional media, whether such images were typically tied to focusing events, or what kinds of stakeholders were communicating those images.

Images and Engagement on Social Media

The power of images lies in the holistic, associative, and quick way through which they are perceived as well as in their superiority over verbal material in attracting attention (Messaris and Abraham, 2001). This makes them particularly effective at articulating ideological messages (Brantner et al., 2011). In particular, the analogical quality of images make them generally perceived to be closer to reality, and they are therefore less questioned than verbal content (Brantner et al., 2011). Messaris and Abraham (2001) further suggest that people are better able to index and later recall pictorial information because of that analogical quality.

It is therefore unsurprising that visual depictions of climate change have repeatedly been found to impact not only individuals' perceptions of the importance of the issue but also their attitudes toward it (Chapman et al., 2016; Metag et al., 2016; O'Neill and Nicholson-Cole, 2009; O'Neill et al., 2013). These effects can be moderated by individual factors, such as image preferences or subjective interpretation of images, as well as content-specific factors, including the nature of the content (i.e., type and subject of the image), the amount of exposure to the content (i.e., how often it surfaces), and the context around the exposure (i.e., associated signals surrounding the image) (Metag, 2020). In other words, repeated exposure to particular kinds of images and indicators of community support for them can be presumed to increase the potential impact of the image on perceptions of and attitudes toward the issue depicted in said images (or some component of that issue). It is therefore worthwhile to examine what kinds of images tend to be engaged with the most, as social media infrastructure both uses and communicates engagement-related signals in ways that structure the user experience on the platform (Van Dijck and Poell, 2013).

Engagement on Twitter manifests primarily through the use of four different affordances. Users can 'like' a tweet, which generally indicates their support of a tweet (Bucher and Helmond, 2018). They can 'retweet' the tweet, thereby highlighting the message's "informational value" (Hwong et al., 2017, p. 481) and helping it spread. Notably, both of these forms of engagement can be executed with a simple click. However, users can also 'quote tweet' (retweet with added comment) and they may 'reply' (respond) to a tweet. The function of these latter forms is harder to discern because the act alone says nothing about the direction of user engagement (i.e., if it is supportive or dismissive). Nevertheless, the use of such affordances signals discursive engagement (Hwong et al., 2017) and can be factored into both human and algorithmic evaluations over the value of the tweet and its accompanying image (Van Dijck and Poell, 2013). In other words, the utilization of these affordances creates conditions for establishing the likelihood of exposure, and the context around the exposure, to a particular image, as well as to repeated exposure to similar images or to images disseminated by similar stakeholders. Indeed, studies show that tweets containing embedded images tend to generate higher levels of engagement than text-only tweets (Rogers, 2014 March 10). However, it remains unclear if certain image types elicit greater engagement on social media than others when it comes to the issue of climate change.

Research Questions

The conceptual framework of media logics and the issue attention cycle help draw attention to an important set of questions pertaining to *what* images are disseminated about climate change on social media, *when* those images are disseminated, *who* disseminates those images, and *how* those images are engaged with via technical affordances. However, there is limited empirical evidence for how these questions play out at the intersection of the issue of climate change, the context of social media, and the practice of visual communication. As such, we pose the following research questions:

> **RQ1**: What types of images are most common in climate change-related tweets?
>
> **RQ2**: Do climate change-related tweets featuring images follow a cyclical pattern that is tied to focusing events?
>
> **RQ3**: What kinds of stakeholders are most active in distributing climate change-related tweets containing images, and what types of images do they use?
>
> **RQ4**: What types of images tend to elicit the greatest engagement?

Methods

Data Collection

In order to address those research questions, we collected all original tweets (excluding re-tweets) that used the hashtag #climatechange and contained an image. We then downloaded all of those images based on the image URL. While there exist many hashtags related to climate change on Twitter, we focused on #climatechange to study the topic from a very general perspective, instead of targeting specific themes (e.g. by using #biodiversity), events (e.g. by using #COP21) or communities (e.g. by using #globalwarming, a hashtag rather prominent across right-wing communities) (Cann et al., 2021; Thorsen and Astrupgaard, 2021). For data collection, we used the Twitter Academic API (Pfeffer et al., 2022), which allowed us to access all historic tweets available on the platform at the time of data collection (meaning all tweets sent, which have not been removed by Twitter due to content violations or by the users themselves). Data collection took place in July 2021 and included all images posted from the time Twitter introduced its image-sharing service in August 2011 to June 2021. This produced a dataset containing 2,516,251 images.

Variables

Image Type. To categorize such a large amount of image data, we used a semi-automated, iterative approach to label images. First, we established a set of categories that we expected to be prevalent based on prior work (Dahl, 2017; DiFrancesco and Young, 2011; O'Neill et al., 2013) and that we estimated would be visually distinct. For this reason, we used only some of the categories employed in studies of traditional news media and augmented them with categories that are likely to be found on social media. For example, we found that textual content, such as screenshots of tweets, was quite common on Twitter, so we included the category 'text/quote'. We chose to adopt an iterative approach that allowed us to add new categories as we came across substantively distinct images.

Next, we employed image clustering to form groups of visually similar images. We used the VGG16 convolutional neural network* to extract 4,096 features per image. These features are specific characteristics of an image. As they are created by the neural network, we do not know exactly what each feature comprises but machines can make decisions about whether two images look similar to each other based on the features they contain.

We then performed k-means clustering, a technique that takes a set of unlabelled data points (e.g., the features in the images) and calculates the similarity between datapoints based on their vector-representation in a multidimensional space. The technique then divides the data into k subsets, called clusters, whereby similar data points are grouped into the same cluster. The number of clusters (k) must be defined in advance by the researcher. To choose an appropriate number of clusters, we had to deal with two competing interests. First, we wanted to categorize highly similar images into a single group (e.g., images showing the same object and scenery). This interest favors creating the largest number of clusters in order to take into account small and subtle differences between images. Second, the number of clusters had to be small enough to allow manual inspection and labelling of each cluster. This interest favors creating a small number of clusters to make the project feasible. We therefore settled on a k value of 5,000, which appeared to show a sufficient amount of differentiation while making manual cluster inspection feasible.

After clustering visually similar images, we manually assigned each cluster into a substantive category (e.g., 'animals'). We randomly selected a set of 100 images per cluster, which was used to represent that cluster. If the cluster size contained fewer

*https://www.tensorflow.org/api_docs/python/tf/keras/applications/vgg16 than 100 images, we evaluated all images in the cluster. In some instances, all images were substantively similar but did not fit a preexisting category and did not appear frequently enough to warrant a new category; these were coded as ‘miscellaneous.’ In other instances, the image set contained substantively distinct images and the cluster was labeled as ‘no categorization possible.’ A trained assistant manually inspected all sets and assigned a label to each cluster.

At the conclusion of this process, we aggregated those categories into a smaller set of eight conceptually distinct parent categories to facilitate the presentation of the results (see Figure 1). Sample images for each category are available in Figure 2. In the end, 12% of images could not be categorized, yielding a final dataset of 2,070,123 images spread across 2,207,543 tweets.

To validate our findings, we manually inspected 1,000 randomly selected images. 920 images were categorized correctly by the system, leading to an accuracy of 92% (Krippendorff’s alpha = 0.90). The 80 images, which were classified incorrectly by the system, showed a great visual similarity to each of the assigned categories, e.g., a picture of a diagram consisting of multiple coloured sketches was coded as ‘graph/diagram’ by our human inspection and classified as ‘cartoon’ by the system (as the features were similar to that of cartoon), and an image which depicted two persons holding a sign that demanded funding for a climate project was coded with the ’people’ category by human inspection and was classified as ‘protest’ by the system.

Stakeholder Type. To get an understanding of the most active distributors of images, we extracted all accounts that published more than 100 images associated with the hashtag #climatechange. This gave us a subset of 2,047 accounts, which we classified into different groups based on their Twitter profile description, their most recent tweets, and any readily accessible online information (e.g., their website). This classification was performed by two coders. To establish intercoder reliability, those coders double-coded 200 accounts (10% of the sample). A Krippendorff’s alpha test for categorical variables yielded a coefficient of 0.81, which exceeds the recommended minimum coefficient of 0.8 (see Riffe et al., 2019, p. 129).

Drawing on prior work and our observations, we developed a typology made up of the following stakeholder types: *advocacy actors* (e.g., non-governmental organizations, fundraisers, charitable organizations, interest groups, and individual activists), *bots* (i.e., accounts showing automated behaviour, such as publishing the same tweet multiple times in a short time), *business actors* (e.g., companies and brands, or individuals who primarily promoted their business activity or network), *journalistic actors* (e.g., news outlets and their journalists), *political actors* (e.g.,

Image Type	Description and Sub-Types
Climate Consequences	Images showing the negative impact of climate change and extreme weather events, such as hurricanes, floods, wildfires, and droughts.
Conference/Workshop	Images of conferences, plenary sessions, workshops, talks, or classrooms.
Miscellaneous	Images depicting agricultural scenes (*agriculture*); contamination and environmentally harmful substances (*pollution*); garments and attire (*clothes*); aerial views of the earth that show climate trends or like context (*earth/satellite images*); nourishing substances for humans and animals (*food*); symbolic elements, such as using an hourglass to visualize that humankind is running out of time (*symbols*); urbanized pollution and general venues (*urban areas*); and *other images* that could be consistently categorized (e.g., book covers, movie posters, flags).
Nature and Animals	Images of different kinds of wildlife, including those that are endangered or iconic (*animals*) and pristine landscapes (*nature*).
People	Images with people as the focus, such as photographs of celebrities, speakers, politicians, scientists, climate victims, and private individuals.
Protest	Images of protest marches, large gatherings of individuals engaging in collective action, and the use of signs.
Technology	Images of solar panels or wind turbines (*green technology*); oil refineries and smokestacks (*industry*); and airplanes, gasoline/electric cars, and bikes (*transportation*).
Visualizations	Images containing humorous/satirical drawings of climate change aspects and popular people, as well as memes in cartoon style (*cartoon*); graphs, charts, maps, and temperature tables (*graph/diagram*); infographics and posters (*illustration*); and memes, quotes, scientific facts, and screenshots of tweets (*text/quote*).

Figure 1. Description of image types and subtypes

Figure 2. Example images of the categories text/quote, graph/diagram, illustration, cartoon (row 1), people, conf./workshop, protest (row 2), animals, nature, climate consequences (row 3), transportation, industry, green technology, urban (row 4), symbol, earth/satellite, food (row 5), pollution, clothes, agriculture, other (row 6) (from left to right)

political or governmental institutions and projects, as well as politicians and members of their campaigns), *scientific actors* (e.g., research institutions and projects, as well as individual researchers and climate experts), and *private persons* (e.g., individuals who used their accounts mostly for personal reasons or to provide personal opinions). An *other* category was also included to account for individuals who did not clearly fit into one of the above categories.

Forty-nine of the accounts could not be classified because they were no longer on Twitter at the time of data collection or had exclusively non-English descriptions and tweets; these accounts were excluded from the analysis.

Engagement. We also collected information about four different engagement affordances—'likes,' 'retweets,' 'quote tweets,' and 'replies'—for each tweet in our dataset. These data were obtained directly from the Twitter API and represent the counts for that affordance and tweet at the time of the data collection.

Results

Salience of Image Types

The first research question asked about the types of images that were most common in climate change-related tweets.

As shown in Figure 3, just over half of the images were categorized as visualizations, with over two thirds of such images falling in a subcategory that included memes, inspirational quotes, and screenshots of quoted tweets. The second most common category was people (15% of the images), which contained photographs of politicians and celebrities, among others. The third most common category with just under one-tenth of the images, was nature and animals (9%). These images included both endangered animals as well as pristine environments. While the images are arguably distinct, they nevertheless cumulatively represent what could be lost due to climate change. The fourth most common category was conference/workshop (7%). The less-common categories were miscellaneous (5%), technologies (4%), protest (4%), and climate consequences (4%).

Pattern of Communication and Focusing Events

The second research question asked whether climate change-related tweets featuring images followed a cyclical pattern that was tied to focusing events.

There is some evidence of a cyclical pattern (see Figure 4). In particular, there are repeated spikes in both the volume and proportion of #climatechange tweets

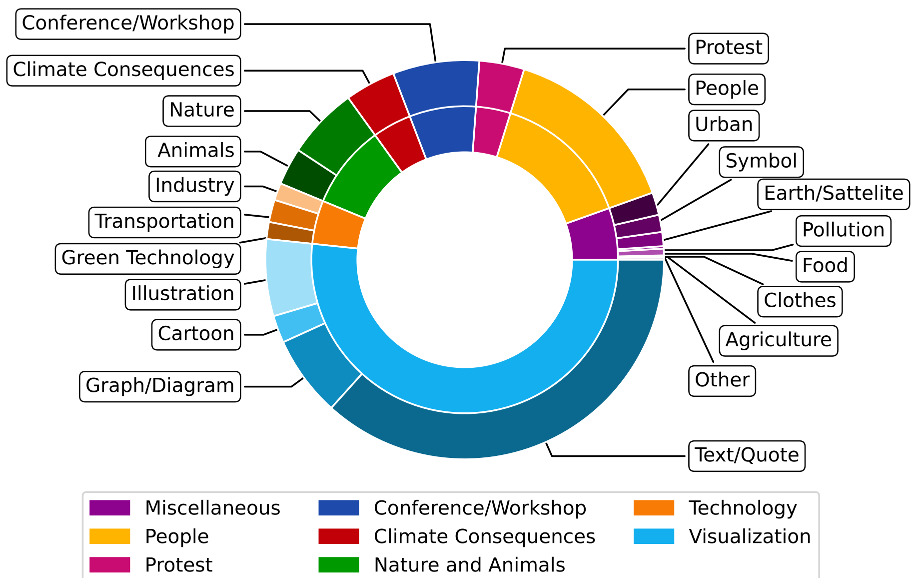

Figure 3. Distribution of images by type of image

with images, often in the latter part of each year. These peaks appeared to be linked to focusing events. For example, instances of the United Nations Climate Change Conference (COP) frequently coincided with increases in the distribution of #climatechange images. Other major policy and media events also seemed to draw attention, such as Fridays for Future and the global school strikes that followed. Notably, there is a pronounced dropoff in the volume and proportion of images starting in the beginning of 2020, which suggests that climate change may have been displaced as a key issue on Twitter as the United States began a messy transition of power and the COVID-19 pandemic took hold. The volume began to increase toward the end of 2020, but the proportion of #climatechange tweets, in relation to all English tweets, remained low in comparison to pre-Covid times.

The results also show that images have become an important vehicle for climate change communication. The proportion of #climatechange tweets containing an image, in relation to all #climatechange tweets, rose with the introduction of the image feature in 2011 and became relatively stable around 30% since 2015. A notable exception to this are the months around the One Planet summit, during roughly 50% of #climatechange tweets were published with an image.

The use of image types remained, proportionally, fairly stable over time. The most notable exception to this was visualization images. That image type gained some

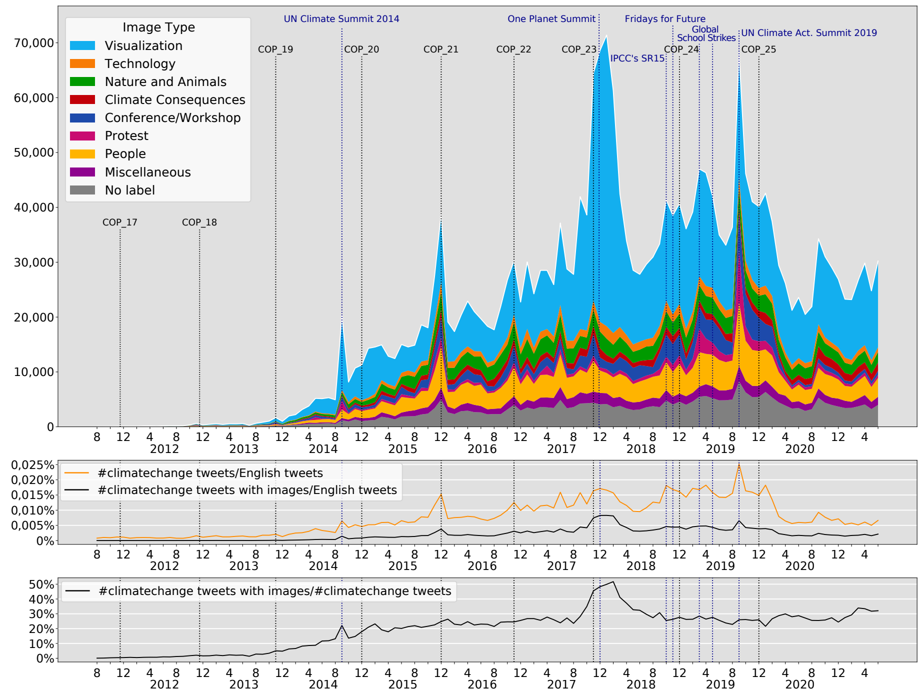

Figure 4. Amount of images associated with tweets containing the hashtag #climatechange on a monthly basis between August 2011 and June 2021 (upper chart). Proportion of the total #climatechange tweets (orange) and the #climatechange tweets with images (black) in relation to all English tweets (middle chart). Proportion of #climatechange tweets with images in relation to all #climatechange tweets (lower chart).

relative salience in the middle of 2014 and spiked in late 2017. It returned to more normal levels by early 2018, and remained the most common image type in the years that followed.

Active Stakeholders and Strategies

The third research question asked about the kinds of stakeholders that were most active in distributing climate change-related tweets containing images, and about the types of images they use.

To address this question, we analyzed a subsample of 1,998 accounts (see the Method section for more details) that collectively distributed more than 1 million of the labeled images. As shown in Figure 5, the most active set of stakeholders were private individuals (20%), followed by advocacy actors (19%), business actors (17%), journalistic actors (12%), scientific actors (12%), political actors (10%), bots (5%) and other actors (5%). However, the picture looks rather different in terms of the share of

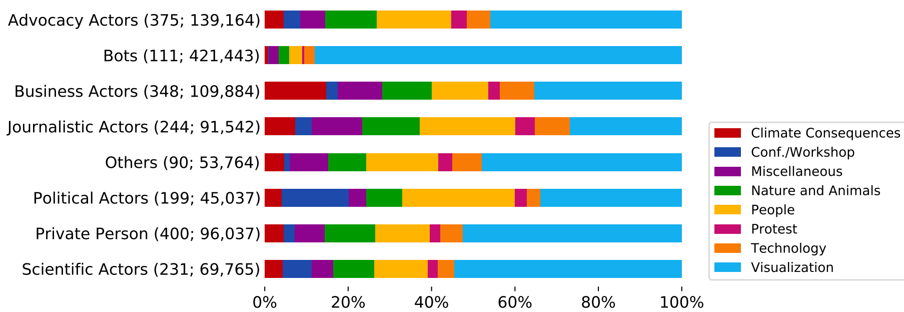

Figure 5. Number of accounts (n=1,998), number of images (n=1,026,636), and distribution of image types by the kind of stakeholder among users who shared more than 100 images using the hashtag #climatechange

images that were distributed. Even though bots only comprised 5% of the accounts, they were responsible for 41% of the images in the subsample. They were followed by advocacy actors (14% of the images), business actors (11%), private individuals (9%), journalistic actors (9%), scientific actors (7%), other actors (6%), and political actors (4%).

Figure 5 also shows some interesting differences in the types of images produced by each kind of stakeholder in the subsample. A Pearson chi-square test indicated that there was a significant relationship between stakeholder type and image category ($\chi^2(49, N = 1,026,636) = 304,202, p < 0.001$). Notably, bot accounts overwhelmingly distributed visualization images (88% of their images), especially images that incorporated text and quotes in the form of screenshotted tweets, inspirational quotes, and memes (78%). Thus, bots were clearly instrumentalized to promote some coordinated messaging but, interestingly, used images of tweets in place of the quote tweet affordance during a significant portion of the time. While scientific actors also favored visualizations overall (55% of their images), they differed by being the most likely to distribute images of graphs and diagrams (18%). In contrast, journalistic actors were the least likely to use visualizations (27%).

Political actors were relatively more likely to distribute people images (27%) as well as conference/workshop images (16%). Journalistic actors (5%) and advocacy actors (4%) were the stakeholders most likely to distribute protest images; no other stakeholder had a share in excess of 3%.

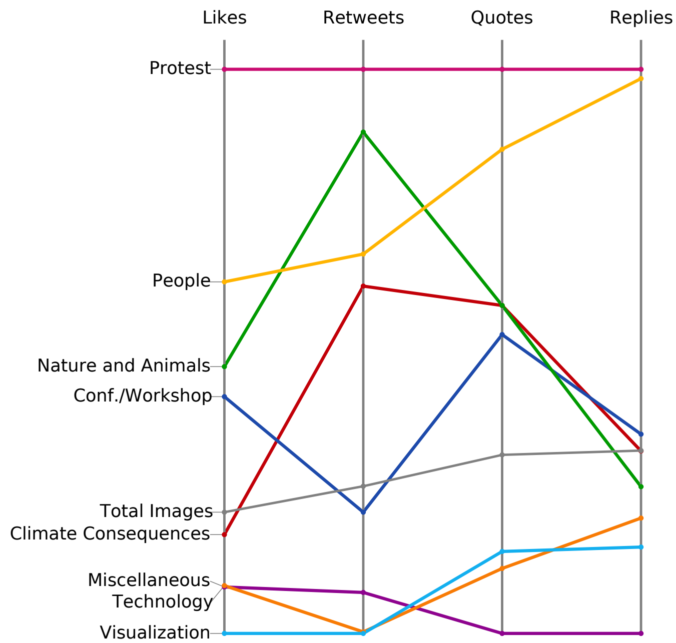

Figure 6. The mean number of likes, retweets, quote tweets, and replies per image for eight different types of images, normalized by the minimum and maximum mean value of all image types

Image Type and Engagement

The fourth research question asked about the types of images that tended to elicit the greatest engagement.

The average (mean) image in our data received 5.5 likes, 3.2 retweets, 0.2 quotes, and 0.3 replies. However, the distribution of these forms of engagement were not equal across categories. As shown in Figure 6, protest images generated the most user engagement by far (M_{likes}=10.8, $M_{retweets}$=4.5, M_{quotes}=0.3, $M_{replies}$=0.5) despite not being very common. In contrast, the most common image type, visualizations, generally received low levels of engagement (M_{likes}=4.0, $M_{retweets}$=2.7, M_{quotes}=0.2, $M_{replies}$=2.9). The second most common image type, people, received fairly high levels of engagement (M_{likes}=8.2, $M_{retweets}$=3.9, M_{quotes}=0.3, $M_{replies}$=0.5), while the third, nature and animals, attracted a high amount of retweets (M_{likes}=7.2, $M_{retweets}$=4.3, M_{quotes}=0.2, $M_{replies}$=0.3).

Discussion

Climate Change Communication

Climate change is one of the most important issues of our time (IPCC, 2022), and visual depictions of the issue matter greatly for how individuals come to perceive the issue's importance and the attitudes and affect they develop toward it (Chapman et al., 2016; Metag et al., 2016; O'Neill and Nicholson-Cole, 2009; O'Neill et al., 2013). It is therefore crucial that we better understand how the issue is communicated visually within the spaces where people gather and interact, such as on social media.

To that end, this study elicited four key empirical findings in relation to prior work. First, the most common type of image—by a large margin—was the visualization, and typically one that included some combination of some text (e.g., quote) with an image. This stands in stark contrast to prior work, which has identified images of climate change consequences and prominent individuals as being most common in traditional news media (see O'Neill and Smith, 2014; Schäfer, 2020). Second, the temporal flow in the volume of images being distributed followed a fairly cyclical pattern situated around focusing events like the COP summits. This finding is in line with prior work that examined traditional news media (O'Neill, 2020; Schäfer, 2020). Third, bots accounted for a significant share of the images communicated on Twitter, with journalistic actors and scientific actors lagging noticeably behind. While this study does not speak to individual actors' influence, it does offer further evidence that the most interested stakeholders (e.g., advocacy actors and business actors) continue to invest in a platform that affords them the opportunity to communicate directly with an audience (Pearce et al., 2019; Schäfer, 2012,2). Our findings are also in line with those of Wozniak et al. (2017), who showed that governmental communicators prefer images of political actors for promoting climate change awareness. Fourth, protest images were consistently the ones engaged with the most, which is also in line with the expectations of the non-governmental organizations (Wozniak et al., 2017). This is notable given that prior work (Chapman et al., 2016; Corner et al., 2015; Leviston et al., 2014; Metag et al., 2016; Wang et al., 2018) has raised questions about the effectiveness of such images, particularly among climate change skeptics.

These empirical findings contribute to a broader theoretical discussion regarding the logic employed in communicating the issue of climate change through images on social media. Foremost, they further illustrate that the use of social media—and Twitter in particular—involves a logic that is distinct from traditional media in some dimensions (Hjarvard, 2018; Van Dijck and Poell, 2013; see also Asp, 2014). This is no doubt due in part to the platform's technical affordances (Altheide, 2016), which promote and facilitate distribution via single-click actions and integrated hosting, which in turn reinforces the core element of shareability that characterizes social media logic. However, it is also likely due in part to the fact that social media allow a wider range of stakeholders to be media producers, which fosters a co-production of the "rhythm, grammar, and format" (Altheide, 2004, p. 294) that is considered acceptable in that networked space. Of particular note is the fact that the most common image type in this study (visualizations featuring image-text combinations) is one not seen in prior work (León et al., 2022; O'Neill and Smith, 2014; Schäfer, 2020). Traditional news media have long used pull-out quotes, which is somewhat akin to the apparent practice found here of distributing screenshots of others' tweets. However, the common practice of juxtaposing synecdoches with inspirational quotes and utilizing memes to ridicule individuals or offer social commentary suggests a grammar and format that is quite distinct (see Figure 2).

Additionally, although the genuinely digital and online multimodal 'text/quote' genre did not elicit the same engagement as regular photos, stakeholders may nevertheless favor using these aesthetically unprofessional or amateurish text-images for two reasons. First, they are easier to generate and remix in order to convey the stakeholder's perspective, especially when compared to professional photojournalistic styles. Second, they may be used to signal the stakeholder's familiarity with the distinct aesthetic vernacular that social media logic is imbued with. Visual cross-platform analysis that take specific affordances and vernaculars into account (Pearce et al., 2020) could show if this image type represents an aesthetic and logic that is specific to Twitter, or perhaps social media more broadly.

However, this study also points to continuities between the two logics, at least as they manifest through the issue of climate change. For example, the rhythm in the distribution of images appears to mirror that of traditional media, as it shows an at least partially cyclical pattern which connects to focusing events. This is in line with decades-old theorizing about the coverage of persistent issues (Downs, 1972) and the longitudinal analysis of Boykoff et al. (2022). Furthermore, many of those focusing events appeared to be pre-planned political events often orchestrated to draw media attention (e.g., climate conferences and planned demonstrations), which not only reinforces a linkage between the two logics in terms of what drives communication but also highlights the continued interlinkage with political logic (Altheide, 2004). Similarly, we see the performance of journalistic actors perhaps most closely aligning with prior scholarship on the visual communication of climate change, such as by being among the actor types most likely to distribute images of people and protests. These observations cumulatively lend themselves supporting a broader contention that the continued intertwining of news media and social media is producing a more hybridized logic (Tsuriel et al., 2021).

This theoretical contribution also helps us better understand the present state of climate change communication on Twitter, especially on the visual front, and what its future might look like. The high levels of engagement around protest imagery in particular highlight both promise and risk. While such images resonate well with people who are already climate-aware, they can reinforce cynicism and an 'us versus them' feeling in others (Chapman et al., 2016; Corner et al., 2015; Wang et al., 2018). In other words, while such images do generate engagement, that engagement may be detrimental to fostering the collective response that is necessary for tackling climate change (Leviston et al., 2014; Metag et al., 2016). In similar vein, the use of bots to promote visualized forms of text can pose a significant threat to effective climate change communication on two fronts. First, bots can be instrumentalized—as they qualitatively appear to already be—to promote polarizing messages conveyed through social media-friendly aesthetics (e.g., semi-amateur media). Second, even if the intent of such bots is not to persuade, they may nevertheless contribute noise that drowns out contributions made by scientific actors and journalists, among others, or simply create what appears to be an informational overload. Put differently, we must reckon with the growing volumes of climate change communication being generated by non-human actants that leverage both the logics and forms associated with Twitter and other social media platforms (Chen et al., 2021).

In considering this study's findings and theoretical contributions, it is important to remain mindful of the fact that it examines a single hashtag. While #climatechange is highly useful in coordinating discussion around the topic (as evidenced by the volume of tweets), it does not encompass of all related communication. Indeed, as Twitter's search and surfacing affordances have improved over the years, hashtags have perhaps become less important. This may be especially true for accounts that have large followings, as well as those that view the use of hashtags as tacky, unprofessional, or a potential branding liability. In other words, our findings are likely most reflective of actors who aimed for wider reach and may be less representative of actors who aimed to reach just their followers. We find some corroborating evidence of this in the fact that not a single fossil fuel company tweeted more than 100 images using the #climatechange. It is likely that such actors, and others, are looking to participate in (and influence) deliberations, but not through this key hashtag. It would thus be helpful for future work to further segment the coordinating affordances—and, perhaps, the key nodes—implicated in these discussions. Nevertheless, the present study does shed light on a substantial and important part of the discussion.

Figure 7. Example of algorithmic limitations: Even though the image in the middle belongs semantically to the same category as the left image (climate victims), it is visually more similar to the image on the right (politician)

Methodological Challenges

This study was based on a methodological design that is novel among studies of climate change communication. The unsupervised machine learning approach is highly useful in multiple regards. In particular, it allows for the systematic evaluation of a large volume of data—something that is becoming ever-more necessary given the sheer amount of digital communication produced on a daily basis—and allows for segmentation based on micro-features that humans might miss. This is doubly true for longitudinal work that requires larger samples in order to be representative. However, more data or the use of complicated algorithms isn't necessarily better. We see evidence of this in the limitations of this study. For example, our approach is unable to differentiate between visually similar but substantively distinct images, such as one that pairs an image of a field of solar panels with a supportive quote and one that pairs it with a critical quote. The approach we have used to automatically label images was based on purely visual characteristics and is therefore different from manual analysis of images that is able to consider the semantic meaning of images. For example, it would be interesting to investigate whether the images that have been shown to be particularly effective, namely those of people that are negatively affected by climate change (Wang et al., 2018), elicit more user engagement. However, as Figure 7 illustrates, trying to automatically differ between images of climate victims and other people, such as politicians, is an extremely difficult task. In the future, researcher could train classifiers to detect very specific subcategories, and combine these to build larger semantic categories.

Similarly, our ability to pair our findings to those from prior work is limited by differences in how categories are operationalized (and which categories can be formulated). To that end, we believe that future work in this area would benefit from hybrid approaches to content analysis (Brantner and Pfeffer, 2018; Zamith and Lewis, 2015), namely by using an approach like ours to capture a more complete range of distinct types of images and randomly sample from them in order to perform a closer analysis that allows for more nuance and considers the peculiarities of images. Such analyses could build upon the more interpretative traditions of visual communication scholarship to capture latent meanings through contextualization (Wozniak et al., 2015).

Moreover, similar to studies on multimodality in news media, future work could investigate how different modalities are combined in meaning-making within climate change communication. Our computational approach could thus be combined with the pairing of qualitative iconographic-iconological analysis and quantitative content analysis of both image and text—an approach that has already been successfully implemented in analyses of multimodal tweets (Brantner et al., 2020).

Finally, we have applied a single label classification approach to be able to differ between images based on their most prevalent theme. This restricted us though in assessing the full meaning of an image when images contained more than one category. Future work could use multilabel classification to account for images which show several categories, such as protesters in front of an industry scene.

References

Agin, S. & Karlsson, M. (2021). Mapping the field of climate change communication 1993–2018: Geographically biased, theoretically narrow, and methodologically limited. *Environmental Communication*, *15*(4), 431–446, https://doi.org/10.1080/17524032.2021.1902363.

Altheide, D. L. (2004). Media logic and political communication. *Political Communication*, *21*(3), 293–296, https://doi.org/10.1080/10584600490481307.

Altheide, D. L. (2016). Media logic. In G. Mazzoleni (Ed.), *The International Encyclopedia of Political Communication* (pp. 1–6). John Wiley & Sons.

Anderson, A. (2009). Media, Politics and Climate Change: Towards a New Research Agenda. *Sociology Compass*, *3*(2), 166–182, https://doi.org/10.1111/j.1751-9020.2008.00188.x.

Asp, K. (2014). News media logic in a new institutional perspective. *Journalism Studies*, *15*(3), 256–270, https://doi.org/10.1080/1461670X.2014.889456.

Boykoff, M., Aoyagi, M., Ballantyne, A., et al. (2022). *World newspaper coverage of climate change or global warming, 2004-2022*. Media and Climate Change Observatory Data Sets. Cooperative Institute for Research in Environmental Sciences, University of Colorado https://sciencepolicy.colorado.edu/icecaps/research/media_coverage/world/index.html.

Brantner, C., Lobinger, K., & Stehling, M. (2020). Memes against sexism? a multi-method analysis of the feminist protest hashtag #distractinglysexy and its resonance in the mainstream news media. *Convergence*, *26*(3), 674–696, https://doi.org/10.1177/1354856519827804.

Brantner, C., Lobinger, K., & Wetzstein, I. (2011). Effects of visual framing on emotional responses and evaluations of news stories about the gaza conflict 2009. *Journalism & Mass Communication Quarterly*, *88*(3), 523–540, https://doi.org/10.1177/2F107769901108800304.

Brantner, C. & Pfeffer, J. (2018). Content analysis of twitter: Big data, big studies. In S. Eldridge II & B. Franklin (Eds.), *The Routledge Handbook of Developments in Digital Journalism Studies* (pp. 79–92). Routledge.

Brossard, D. & Scheufele, D. A. (2022). The chronic growing pains of communicating science online. *Science*, *375*(6581), 613–614, https://doi.org/10.1126/science.abo0668.

Brossard, D., Shanahan, J., & McComas, K. (2004). Are issue-cycles culturally constructed? A comparison of French and American coverage of global climate change. *Mass Communication and Society*, *7*(3), 359–377, https://doi.org/10.1207/s15327825mcs0703_6.

Bucher, T. & Helmond, A. (2018). The affordances of social media platforms. In J. Burgess, A. Marwick, & T. Poell (Eds.), *The SAGE handbook of social media* (pp. 233–253). SAGE Publications.

Cann, T. J. B., Weaver, I. S., & Williams, H. T. P. (2021). Ideological biases in social sharing of online information about climate change. *PLOS ONE*, *16*(4), https://doi.org/10.1371/journal.pone.0250656.

Chapman, D. A., Corner, A., Webster, R., & Markowitz, E. M. (2016). Climate visuals: A mixed methods investigation of public perceptions of climate images in three countries. *Global Environmental Change*, *41*, 172–182, https://doi.org/10.1016/j.gloenvcha.2016.10.003.

Chen, C.-F., Shi, W., Yang, J., & Fu, H.-H. (2021). Social bots' role in climate change discussion on twitter: Measuring standpoints, topics, and interaction strategies. *Advances in Climate Change Research*, *12*(6), 913–923, https://doi.org/https://doi.org/10.1016/j.accre.2021.09.011.

Corner, A., Webster, R., & Teriete, C. (2015). *Climate visuals: Seven principles for visual climate change communication (based on international social research)*. Climate Outreach.

Dahl, T. (2017). Verbal and visual framing activity in climate change discourse: A multimodal analysis of media representations of the IPCC' 5th assessment report. In K. Flottum (Ed.), *The role of language in the climate change debate* (pp. 13–30). Routledge.

DiFrancesco, D. & Young, N. (2011). Seeing climate change: the visual construction of global warming in Canadian national print media. *Cultural Geographies*, *18*(4), 517–536, https://doi.org/10.1177/1474474010382072.

Djerf-Pierre, M. (2012). When attention drives attention: Issue dynamics in environmental news reporting over five decades. *European Journal of Communication*, *27*(3), 291–304, https://doi.org/10.1177/0267323112450820.

Downs, A. (1972). Up and down with ecology—the issue-attention cycle. *The Public Interest*, *28*, 38–51.

Freelon, D. (2019). *Tweeting Left, Right & Center: How users and attention are distributed across Twitter*. Knight Foundation.

Harcup, T. & O'Neill, D. (2017). What is news? *Journalism Studies*, *18*(12), 1470–1488, https://doi.org/10.1080/1461670X.2016.1150193.

Highfield, T. & Leaver, T. (2016). Instagrammatics and digital methods: studying visual social media, from selfies and GIFs to memes and emoji. *Communication Research and Practice*, *2*(1), 47–62, https://doi.org/10.1080/22041451.2016.1155332.

Hjarvard, S. (2008). The mediatization of society. *Nordicom Review*, *29*(2), 102–131, https://doi.org/10.1515/nor-2017-0181.

Hjarvard, S. (2018). The logics of the media and the mediatized conditions of social interaction. In C. Thimm, M. Anastasiadis, & J. Einspänner-Pflock (Eds.), *Media logic(s) revisited: Modelling the interplay between media institutions, media technology and societal change*, Transforming Communications – Studies in Cross-Media Research (pp. 63–84). Springer International Publishing.

Hopke, J. E. & Hestres, L. E. (2018). Visualizing the Paris climate talks on Twitter: Media and climate stakeholder visual social media during COP21. *Social Media + Society*, *4*(3), https://doi.org/10.1177/2056305118782687.

Hwong, Y.-L., Oliver, C., Van Kranendonk, M., Sammut, C., & Seroussi, Y. (2017). What makes you tick? The psychology of social media engagement in space science communication. *Computers in Human Behavior*, *68*, 480–492, https://doi.org/10.1016/j.chb.2016.11.068.

IPCC (2022). *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press.

Leviston, Z., Price, J., & Bishop, B. (2014). Imagining climate change: The role of implicit associations and affective psychological distancing in climate change responses. *European Journal of Social Psychology*, *44*(5), 441–454, https://doi.org/10.1002/ejsp.2050.

León, B., Negredo, S., & Erviti, M. C. (2022). Social engagement with climate change: principles for effective visual representation on social media. *Climate Policy*, *0*(0), 1–17, https://doi.org/10.1080/14693062.2022.2077292.

Messaris, P. & Abraham, L. (2001). The role of images in framing news stories. In S. Reese, O. H. Gandy Jr., & A. E. Grant (Eds.), *Framing Public Life* (pp. 215–226). Lawrence Erlbaum.

Metag, J. (2020). Climate change visuals: a review of their effects on cognition, emotion and behaviour. In D. Holmes & L. Richardson (Eds.), *Research Handbook on Communicating Climate Change* (pp. 153–160). Edward Elgar Publishing.

Metag, J., Schäfer, M. S., Füchslin, T., Barsuhn, T., & Kleinen-von Königslöw, K. (2016). Perceptions of climate change imagery: Evoked salience and self-efficacy in Germany, Switzerland, and Austria. *Science Communication*, *38*(2), 197–227, https://doi.org/10.1177/1075547016635181.

Murru, M. F. & Vicari, S. (2021). Memetising the pandemic: memes, covid-19 mundanity and political cultures. *Information, Communication & Society*, *24*(16), 2422–2441, https://doi.org/10.1080/1369118X.2021.1974518.

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., & Nielsen, R. K. (2022). *Reuters Institute Digital News Report 2022*. Reuters Institute for the Study of Journalism.

O'Neill, S. (2020). More than meets the eye: a longitudinal analysis of climate change imagery in the print media. *Climatic Change*, *163*(1), 9–26, https://doi.org/10.1007/s10584-019-02504-8.

O'Neill, S. & Nicholson-Cole, S. (2009). "Fear won't do it": Promoting positive engagement with climate change through visual and iconic representations. *Science Communication*, *30*(3), 355–379, https://doi.org/10.1177/1075547008329201.

O'Neill, S. J., Boykoff, M., Niemeyer, S., & Day, S. A. (2013). On the use of imagery for climate change engagement. *Global Environmental Change*, *23*(2), 413–421, https://doi.org/10.1016/j.gloenvcha.2012.11.006.

O'Neill, S. J. & Smith, N. (2014). Climate change and visual imagery. *WIREs Climate Change*, *5*(1), 73–87, https://doi.org/10.1002/wcc.249.

Pearce, W., Niederer, S., Özkula, S. M., & Querubín, N. S. (2019). The social media life of climate change: Platforms, publics, and future imaginaries. *WIREs Climate Change*, *10*(2), e569, https://doi.org/10.1002/wcc.569.

Pearce, W., Özkula, S. M., Greene, A. K., Teeling, L., Bansard, J. S., Omena, J. J., & Rabello, E. T. (2020). Visual cross-platform analysis: digital methods to research social media images. *Information, Communication & Society*, *23*(2), 161–180, https://doi.org/10.1080/1369118X.2018.1486871.

Pfeffer, J., Mooseder, A., Hammer, L., Stritzel, O., & Garcia, D. (2022). This sample seems to be good enough! assessing coverage and temporal reliability of twitter's academic api. arXiv:2204.02290 [cs.SI] https://arxiv.org/abs/2204.02290.

Riffe, D., Lacy, S., Watson, B., & Fico, F. (2019). *Analyzing Media Messages: Using Quantitative Content Analysis in Research*. Routledge, 4th edition.

Rogers, S. (2014, March 10). What fuels a Tweet's engagement? https://blog.twitter.com/2014/what-fuels-a-tweets-engagement.

Schäfer, M. S. (2012). Online communication on climate change and climate politics: a literature review. *Wiley Interdisciplinary Reviews: Climate Change*, *3*(6), 527–543, https://doi.org/10.1002/wcc.191.

Schäfer, M. S. (2020). News media images of climate change: reviewing the research. In D. Holmes & L. Richardson (Eds.), *Research Handbook on Communicating Climate Change* (pp. 143–152). Edward Elgar Publishing.

Schäfer, M. S. & Schlichting, I. (2014). Media representations of climate change: A meta-analysis of the research field. *Environmental Communication: A Journal of Nature and Culture*, *8*(2), 142–160, https://doi.org/10.1080/17524032.2014.914050.

Thimm, C., Anastasiadis, M., & Einspänner-Pflock, J. (2018). Media logic or media logics? an introduction to the field. In C. Thimm, M. Anastasiadis, & J. Einspänner-Pflock (Eds.), *Media logic(s) revisited: Modelling the interplay between media institutions, media technology and societal change* (pp. 1–8). Springer International Publishing.

Thorsen, S. & Astrupgaard, C. (2021). Bridging the computational and visual turn: Re-tooling visual studies with image recognition and network analysis to study online climate images. *Nordic Journal of Media Studies*, *3*(1), 141–163, https://doi.org/10.2478/njms-2021-0008.

Tschötschel, R., Schuck, A., & Wonneberger, A. (2020). Patterns of controversy and consensus in German, Canadian, and US online news on climate change. *Global Environmental Change*, *60*, https://doi.org/10.1016/j.gloenvcha.2019.101957.

Tsuriel, K., Dvir Gvirsman, S., Ziv, L., Afriat-Aviv, H., & Ivan, L. (2021). Servant of two masters: How social media editors balance between mass media logic and social media logic. *Journalism*, *22*(8), 1983–2000, https://doi.org/10.1177/1464884919849417.

Van Dijck, J. & Poell, T. (2013). Understanding social media logic. *Media and Communication*, *1*(1), 2–14, https://doi.org/10.17645/mac.v1i1.70.

Wallace, J. (2018). Modelling contemporary gatekeeping. *Digital Journalism*, *6*(3), 274–293, https://doi.org/10.1080/21670811.2017.1343648.

Wang, S., Corner, A., Chapman, D., & Markowitz, E. (2018). Public engagement with climate imagery in a changing digital landscape. *WIREs Climate Change*, *9*(2), e509, https://doi.org/10.1002/wcc.509.

Wozniak, A. (2020). Stakeholders visual representations of climate change. In D. Holmes & L. Richardson (Eds.), *Research Handbook on Communicating Climate Change* (pp. 131–142). Edward Elgar Publishing.

Wozniak, A., Lück, J., & Wessler, H. (2015). Frames, stories, and images: The advantages of a multimodal approach in comparative media content research on climate change. *Environmental Communication*, *9*(4), 469–490, https://doi.org/10.1080/17524032.2014.981559.

Wozniak, A., Wessler, H., & Lück, J. (2017). Who prevails in the visual framing contest about the United Nations climate change conferences? *Journalism Studies*, *18*(11), 1433–1452, https://doi.org/10.1080/1461670X.2015.1131129.

Zamith, R. & Lewis, S. C. (2015). Content analysis and the algorithmic coder: What computational social science means for traditional modes of media analysis. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 307–318, https://doi.org/10.1177/0002716215570576.

4.2 Study 2: Measuring Teachers' Visual Expertise Using the Gaze Relational Index Based on Real-world Eye-tracking Data and Varying Velocity Thresholds

Authors
Angelina Mooseder[5], Christian Kosel[5], Tina Seidel, Jürgen Pfeffer

In
ArXiv, `https://arxiv.org/abs/2304.05143`, Published on 04-2023

Abstract
This article adds to the understanding of teachers' visual expertise by measuring visual information processing in real-world classrooms (mobile eye-tracking) with the newly introduced Gaze Relational Index (GRI) metric, which is defined as the ratio of mean fixation duration to mean fixation number. In addition, the aim was to provide a methodological contribution to future research by showing to what extent the selected configurations (i.e. varying velocity thresholds and fixation merging) of the eye movement event detection algorithm for detecting fixations and saccades influence the results of eye-tracking studies. Our study leads to two important take-home messages: First, by following a novice-expert paradigm (2 novice teachers & 2 experienced teachers), we found that the GRI can serve as a sensitive measure of visual expertise. As hypothesized, experienced teachers' GRI was lower, suggesting that their more fine-graded organization of domain-specific knowledge allows them to fixate more rapidly and frequently in the classroom. Second, we found that the selected velocity threshold parameter alter and, in the worst case, bias the results of an eye-tracking study. Therefore, in the interest of further generalizability of the results within visual expertise research, we emphasize that it is highly important to report configurations that are relevant for the identification of eye movements.

Contribution of Thesis Author
Data preprocessing, algorithmic design, discussions and implementation, data visualization, as well as manuscript writing, revision and editing.

[5]The authors contributed equally.

Measuring Teachers' Visual Expertise Using the Gaze Relational Index Based on Real-world Eye-tracking Data and Varying Velocity Thresholds

Christian Kosel*[1], Angelina Mooseder*[2], Tina Seidel[1], Jürgen Pfeffer[2]

*These authors contributed equally

[1] Friedl Schöller Endowed Chair for Educational Psychology (Seidel),

School of Social Science and Technology, Technical University Munich, Germany

[2] Computational Social Science and Big Data (Pfeffer),

School of Social Science and Technology, Technical University Munich, Germany

{christian.kosel, angelina.mooseder, tina.seidel, juergen.pfeffer}@tum.de

Abstract

This article adds to the understanding of teachers' visual expertise by measuring visual information processing in real-world classrooms (mobile eye-tracking) with the newly introduced Gaze Relational Index (GRI) metric, which is defined as the ratio of mean fixation duration to mean fixation number. In addition, the aim was to provide a methodological contribution to future research by showing to what extent the selected configurations (i.e. varying velocity thresholds and fixation merging) of the eye movement event detection algorithm for detecting fixations and saccades influence the results of eye-tracking studies. Our study leads to two important take-home

messages: First, by following a novice-expert paradigm (2 novice teachers & 2 experienced teachers), we found that the GRI can serve as a sensitive measure of visual expertise. As hypothesized, experienced teachers' GRI was lower, suggesting that their more fine-graded organization of domain-specific knowledge allows them to fixate more rapidly and frequently in the classroom. Second, we found that the selected velocity threshold parameter alter and, in the worst case, bias the results of an eye-tracking study. Therefore, in the interest of further generalizability of the results within visual expertise research, we emphasize that it is highly important to report configurations that are relevant for the identification of eye movements.

1 Introduction

Human visual expertise in vision-intensive domains such as medicine (Gegenfurtner & Seppänen, 2013), sports (Aglioti et al., 2008), driving (Lappi et al., 2017), and aviation (Peißl et al., 2019) reflects complex cognitive and visual processing that evolves in domain experts through deliberate and consistent practice over a long period (Gegenfurtner et al., 2011). Domain experts are more skilled in the perception, interpretation, and evaluation of domain-specific visual information (Gegenfurtner, 2020). In recent years, there has been a growing interest in the visual expertise of teachers because they perceive and interpret a large amount of dynamic visual information to effectively manage the complexity of a classroom full of students. A rich body of studies has revealed that experienced and novice teachers differ markedly in their visual processing (see meta-analysis/reviews: Gegenfurtner et al., 2011; Grub et al., 2020). Besides studies based on verbal reports and think-a-loud protocols that were collected after a visual stimulus (i.e., video vignette, photograph; van Es & Sherin, 2010) was shown to the participating teachers, other studies explored teachers' visual expertise based on fine-graded data that were collected with eye-tracking devices (Kosel, Holzberger, & Seidel, 2021; McIntyre & Foulsham, 2018). Eye-tracking is an effective method to explore where, how often and how long teachers direct their visual attention (Holmqvist et al., 2015). In visual expertise research across various disciplines including teaching, two eye-tracking parameters are found to be commonly sensitive to expertise – the *number of fixations* and the *average fixation duration* (Gegenfurtner et al., 2011; Grub et al., 2020). Generally, it was found in several studies that domain experts, including experienced teachers, have more but shorter fixations, whereas domain novices have fewer but longer fixations (Grub et al., 2020; Wolff, et al., 2016). These findings underline that the processing of visual information not only varies across individuals but also across different levels of expertise (Gegenfurtner et al., 2011). Until now, however, research on visual expertise has lacked a common and single metric to capture and contrast the visual expertise of domain experts and novices. Gegenfurtner and colleagues (2020) fill this gap by introducing a novel eye-tracking metric indicative of expert visual processing, called the Gaze Relational Index (GRI). The GRI is defined as the ratio of mean fixation duration to mean fixation count. As far as we know, GRI has only been applied in two studies that used data from laboratory and stationary eye movement recordings. Gegenfurtner et al. (2020) investigated the GRI regarding 3D dynamic medical visualizations in diagnostic radiology (Gegenfurtner et al., 2020), and Grub, Biermann, Lewalter & Brünken (2022) focused on experienced and novice teachers' gazes using video vignettes (Grub et al., 2022). Studies found marginal expertise differences (Gegenfurtner et al., 2020) or no differences (Grub et al., 2022). The present study aims to go beyond laboratory-collected eye movement data (on-action) and implement the GRI into research on teachers' visual expertise by calculating the GRI for experienced and novice teachers' real-world eye gaze extracted from a mobile eye movement device (in-action).

Another challenging issue that arises in the domain of visual expertise is that the majority of eye-tracking-based studies do not implement in their reports information about how the various eye-tracking parameters are calculated, for example, which specific configurations in the eye movement event-detection algorithm to detect fixations and saccades were used. However, this is especially important as an increasing number of researchers are analyzing their gaze data with more advanced external analysis tools and scripts (Dolezalova & Popelka, 2016; Panetta et al., 2020) that are based on the raw eye-tracking data extracted from the eye tracker. We, therefore, aim to demonstrate the extent to which different configurations may affect the detection of fixations and saccades and thus also the (interpretation of) the results. The above outlined GRI is a suitable measure to investigate how different configurations affect the results of eye movement experiments, as it is a single-valued measure that allows a straightforward comparison.

1.1 Professional Vision and Visual Expertise of Teachers

Professional vision is commonly used as a conceptual framework in the field of cognitive-oriented teacher research (Goodwin, 1994; Seidel & Stürmer, 2014). The concept implies a two-step-process: (1) noticing, which describes teachers' ability to selectively direct attention to relevant events in the classroom; and (2) knowledge-based reasoning, which refers to teachers' ability to interpret these events based on their professional knowledge (Seidel & Stürmer, 2014; van Es & Sherin, 2010). Thereby, noticing and knowledge-based reasoning are not isolated processes, but interact with each other (Seidel & Stürmer, 2014). Teachers' professional knowledge drives teachers' attentional processes in a top-down process (i.e., selective attention inferred from their knowledge) and, in turn, noticing activates teachers' knowledge in form of curriculum scripts and classroom routines stored in their long-term memory (i.e., teachers can make sense of what they see) (Lachner et al., 2016). This implies that professional vision is formed primarily through consistent practice over many years in which teachers accumulate professional knowledge. It, therefore, indicates that professional vision is primarily a characteristic of experienced teachers (Berliner, 2001). Eye-tracking studies at the intersection of professional vision and visual expertise provide further evidence that teachers' visual processes change as their expertise increases (Grub et al., 2020; Kosel, Holzberger, & Seidel, 2021; van den Bogert et al., 2014). For example, it has been found that experienced teachers compared to novice teachers are able to distribute their attention more evenly across students (van den Bogert et al., 2014) and monitor a larger group of different students during teaching (Kosel, Holzberger, & Seidel, 2021). Beyond these findings especially relevant to classroom management, studies have shown that experienced teachers, similar to domain experts in other vision-intensive fields (Gegenfurtner et al., 2011; Gegenfurtner et al., 2022), have shorter but more fixations whereby domain novices have longer but fewer fixations. Since an important assumption is that fixations indicate that information is perceived and processed cognitively (Rayner, 2009), the results suggest that experts encode information more rapidly because of their more advanced and fine-graded knowledge structures that drive visual attention in a top-down process (Gegenfurtner et al., 2022).

In contrast, novices do not have this accumulated knowledge, and their attention is driven more by external and salient features of the visual stimulus in a bottom-up process (Gegenfurtner et al., 2022). The rapid information processing of experienced teachers (reflected in short fixation durations) is also consistent with Ericsson and Kintsch's (1995) theory of long-term working memory. They stated that experts increase the capacity of their working memory by building retrieval structures in their long-term memory (see also Gegenfurtner et al., 2022). The knowledge embedded in this retrieval structure is available in the working memory and enables experts to process visual information more rapidly compared to novices that have not yet fully developed a knowledge-based retrieval structure. In other words, relevant for visual expertise is not only a large amount of domain-relevant knowledge but also a superior organization of this knowledge. In addition, the ability of experts to process more information (reflected in a higher number of fixations) is related to the assumptions of Haider and Frensch's (1996) information reduction hypothesis. They argue that experts optimize the amount of information processed by separating task-relevant from task-irrelevant information. Ignoring redundant information leads to experts having more capacity in their working memory to process more relevant information. Both theories are also important parts of the Cognitive Theory of Visual Expertise (CTVE; Gegenfurtner et al., 2022) which covers further important aspects of visual expertise (e.g. parafoveally and holistic information processing). Taken together, the outlined assumptions help to understand experienced teachers' faster and more automated information processing involving less conscious effort, suggesting that experienced teachers encode and update dynamic teaching situations (with many and short fixations) more rapidly (Gegenfurtner et al., 2020; Grub et al., 2020).

To be able to capture visual expertise using a single and expertise-sensitive metric, Gegenfurtner and colleagues (2020) introduced the so-called Gaze Relation Index (GRI) in the field of visual expertise. The GRI is defined as the ratio of mean fixation duration to mean fixation count (Gegenfurtner et al., 2020). Based on the previous empirical studies (e.g., Wolff et al., 2016) and how the GRI is calculated, it can be inferred that the GRI should be higher for novice teachers than for experienced teachers. Since the GRI is still emerging in the field of visual expertise, the number of studies to date is limited. In the study by Gegenfurtner and colleagues (2020), the GRI was calculated for dynamic 3D medical visualizations. They found that the GRI was slightly, but statistically non-significant, higher for novices compared to experts (Gegenfurtner et al., 2020). In the educational context, Grub and colleagues (2022) analyzed the GRI in a standardized experimental design in which experienced and novice teachers perceived various classroom situations via short video sequences. Contrary to their hypothesis, they found no differences in the number and duration of fixations and thus no differences in GRI (Grub et al., 2022). However, as discussed by Gegenfurtner (2020), the full potential of the GRI might come to light when the experiment is situated outside the lab, using mobile eye-tracking “to mirror the full complexity of visual input that experts routinely deal with in their everyday work surroundings” (Gegenfurtner et al., 2020, p. 38). However, the number of studies that analyzed mobile eye-tracking data to explore expertise differences concerning the number and duration of fixations (the basis for the GRI) is limited (Huang et al., 2021). While the in-action study by Huang and colleagues (2021) confirms expected expertise differences regarding the two metrics, the findings of on-action eye-tracking studies are more heterogeneous (Grub et al., 2022; Kosel, Holzberger, & Seidel, 2021; van den Bogert et al., 2014; Wolff et al., 2016). One reason for this, as Gegenfurtner and colleagues (2020) described, could be that eye-tracking experiments in the laboratory cannot capture the full dynamic complexity that can only be recorded with mobile eye-tracking in teachers’ natural work environment. Thus, there is a need to further explore novice and experienced teachers’ visual expertise measured with the GRI using mobile eye-tracking data.

1.2 Classify eye movements using event-detection algorithms

Across all academic disciplines, eye-tracking-based studies rely on eye movement event-detection algorithms to analyze raw data and classify different types of eye movement, such as fixations (moments when the eye is relatively still and visual information is processed) and saccades (rapid eye movements between two or more phases of fixation). There exists a large number of different algorithms today (see for a review and evaluation of different algorithms: Andersson et al., 2017). Event-detection algorithms can be broadly grouped into dispersion- and velocity-based algorithms (Andersson et al., 2017). One of the most frequently used velocity-based algorithms for detecting fixations is the Identification by Velocity Threshold (I-VT). This algorithm uses only one parameter, the fixed velocity threshold for saccade detection where "fixations are segments of samples with point-to-point velocities below the set velocity threshold, and saccades are segments of the sample with velocities above this threshold" (Andersson et al., 2017, p. 618). The fixed and a-priori-defined velocity is most usually given in visual degrees per second (°/s). Commonly used values for the velocity threshold in lab-based eye-tracking studies range between 5 and 50°/s, using lower values for oculomotor studies and higher values for cognitive studies (Andersson et al., 2017). The I-VT algorithm is implemented in most of the recent commercial eye-tracking software like Tobii Pro (Tobii, 2022). However, since fixation is a fundamental parameter of most eye-tracking studies, outcomes depending not only on the used algorithm to separate fixations from saccades (Salvucci & Goldberg, 2000) but also on the different velocity thresholds employed for the algorithms (Andersson et al., 2017; Holmqvist et al., 2015). In other words, different velocity thresholds might produce significantly different results (Salvucci & Goldberg, 2000). The various velocity thresholds can be easily changed in most software solutions. In Tobii Pro, for example, velocity thresholds of 30°/s and 100°/s are pre-stored (30°/s = fixation filter; 100°/s attention filter). In this context, Hossain and Miléus (2016) compared different velocity thresholds for fixation identification for low-sample-rate mobile eye-trackers like the Tobii Pro Glasses 2. They point out that the IV-T fixation filter does not perform as well on mobile eye-tracker as it did on lab-based eye trackers—especially when a lot of head movements are involved in the recordings. The problem here is that head movements have an impact on velocities and many fixations would not be detected by the IV-T algorithm. They found that the default setting of 30°/s underestimates the periods during which a participant gathers information because a large proportion of smooth pursuits (eye movements in which the eyes remain fixated on a moving object) and vestibular-ocular reflex (VOR; stabilizing eye movements in the opposite direction of head movements) are classified as saccades. One way to counter this is to increase the velocity threshold on the mobile eye-tracker. Using the 100°/s attention filter would overestimate information gathering because fixations, smooth pursuits, VOR periods and 10-15% of short saccades will be classified as fixations (Hossain & Miléus, 2016). However, Hossain and Miléus (2016) found the highest precision for fixation detection in mobile eye-tracking using a velocity threshold between 90°/s and 100°/s, when head movements are involved and not compensated for with external gyroscope data. Overall, since some studies and technical reports point out that results significantly change with different velocity thresholds (Hossain & Miléus, 2016; Olsen, 2012; Salvucci & Goldberg, 2000), the selected velocity threshold must be set out in research studies to make results comparable. Most studies in the educational context (Chaudhuri et al., 2021; Cortina et al., 2015) and in other fields such as aviation (Weibel et al., 2012) where mobile eye tracking is used, however, do not specify the velocity thresholds used to detect fixations and saccades.

In addition to velocity thresholds, the so-called fixation merging is another configuration of the IV-T algorithm that needs to be addressed. The basic idea of merging fixations is that very short fixations (i.e. too short fixations do not reflect cognitive processing) are merged with the next longer fixation, which is in its vicinity (within 0.5° of visual angle) (Tobii, 2022). Merging can be set automatically in the Tobii software package (Olsen, 2012; Tobii, 2022). However, fixation merging has consequences for the classification of the number of fixations and thus for the results of fixation-based metrics such as the GRI. However, the extent of this effect has not yet been described, which makes an assessment together with different velocity thresholds relevant for future eye-tracking studies in the context of visual expertise.

1.3 The present study

In the present study, we aimed to explore teachers' visual expertise following an expert-novice paradigm. Established expertise theories and prior empirical findings point to the fact that teachers, through deliberate practice for a long period, develop visual expertise which leads to qualitatively enriched and superior ways of visually perceiving and processing information when compared to novices. Two of the expertise-sensitive eye-tracking metrics are the number of fixations and the average duration of these fixations. The introduced GRI is based on the relation between both parameters and can be used as a single value metric to assess visual expertise in vison-intense domains like teaching. However, until now the GRI is a seldom explored metric, and evidence is limited to lab-based on-action eye-tracking studies (Gegenfurtner et al., 2020; Grub et al., 2022). Therefore, the first aim was to use the GRI to measure teachers' visual expertise based on real-world gaze collected with a mobile eye-tracking device during instruction. The second aim of the present study was to investigate the impact of various velocity thresholds for eye movement identification using the Identification by Velocity Threshold (I-VT) algorithm and fixation merging on the eye-tracking parameter/GRI metric. This study is hypothesis-driven and involves two related research questions:

1. Is the gaze relational index (GRI) higher for experienced compared to novice teachers?

We aimed to explore the potential utility of GRI as an indicator of visual expertise. Based on previous findings, we hypothesized that experienced teachers use more top-down knowledge-based processing of visual information, leading to their ability to scan the visual field more rapidly.

Thus, we expected more and faster fixations among experienced teachers. Novice teachers in comparison, use more button-up salient-based processing of visual information, resulting in fewer and longer fixations. Therefore, we expect the GRI to be higher for novice teachers than for experienced teachers.

2. How do the eye-movement parameters (fixations, duration of fixations) and the GRI change…

    a) depending on the choice of velocity thresholds for eye movement detection based on the Identification by Velocity Threshold (I-VT) algorithm?
    b) depending on fixation merging?

We expected that the different velocity thresholds would lead to different results regarding the detection of fixations and saccades, thus affecting the GRI. Based on the logic behind the IV-T algorithm, we expected that the lower the selected velocity threshold, the fewer eye movements were classified as fixations. However, derived from eye-tracking protocols and studies (Andersson et al., 2017; Olsen, 2012), we hypothesize that this is not a linear process, i.e., the velocity threshold of 30°/s compared to 60°/s does not classify half of the eye movements as fixations, mainly because using higher velocity thresholds more smooth pursuit eye movements and slow saccades were identified as fixations. Furthermore, we expected that fixation merging significantly reduces the number of fixations and therefore the GRI of a participant. The extent to which outcomes differ is difficult to predict, so this research question is exploratory in nature.

2 Methods

2.1 Participants

The data were obtained from four in-service mathematics teachers (two females, two males). Each teacher gave a lesson ranging between 60 - 90 min in four different higher secondary schools (grade 9) in Germany. All participating teachers taught similar content (matrix calculus) at the time of the data collection. In addition, the sampled lesson was minimally predetermined to allow for some consistency across teachers and their individual lessons - teachers were given 5min. of their class to recap the topic and tasks of the last lesson and the remaining time to introduce a new piece of content. Two of the participating teachers are novices with an average teaching experience of 1.5 years, while the other two teachers are experienced teachers with an average teaching experience of 11 years. Teachers were between 27 and 62 years old, (M=37.25, SD=16.64). Class sizes ranged from 14 to 24 students (69 students total).

2.2 Procedure

Mobile eye-tracking recording took place during a regular class period, chosen to interfere as little as possible with the regular lesson plan. We used a Tobii Pro Glasses 2 with a temporal resolution of 100 Hz to collect eye movement data (Tobii, 2022). Before the recordings started, a calibration of the eye-tracking glasses was performed until a satisfactory calibration was achieved. All participating teachers were advised not to move their eye-tracking glasses during the recording of eye movements. After the recording, the participating teachers were interviewed through a questionnaire (assessment of the lesson, demographic data, professional experience, etc.).

2.3 Data (pre-)processing

2.3.1 Data Collection

We exported the raw data using the Tobii Lab Analysis Software (Tobii, 2022), which gave us information about the eye and gaze positions at each recording timestamp, and performed all subsequent fixation calculations in python. For each timestamp, we stored the time since the beginning of the recording in milliseconds, the pupil positions of the left and right eye at this timestamp in 3D space, the gaze points at this timestamp in 3D space and a 2D representation of the gaze points at this timestamp. The time of recordings per participant varied between 24 and 68 minutes. To control for these time differences and to limit their impact on the eye-tracking parameter, we extracted for each person all eye movements of the first 20 minutes of the recording and discarded the rest of the data for our analysis.

2.3.2 Fixation Classification Algorithm

Fixation calculation. We based our fixation calculation on the Velocity-Threshold Identification (I-VT) algorithm, as described in (Salvucci & Goldberg, 2000) and (Olsen, 2012). First, we calculated the point-to-point velocities for each pair of consecutive recording timestamps (t1,t2), by performing the following steps:

1) We calculated the timestamp of the exact point in time between t1 and t2, by taking the mean of t1 and t2.
2) We calculated the position of the left eye at the timestamp t1t2 by taking the mean of the left eye position vector at t1 and the left eye position vector at t2. We did the same for the right eye.
3) We calculated the visual angle between the left eye position at timestamp t1t2, the gaze position at t1, as well as the gaze position at t2. We did the same for the right eye. This gave us an indicator of how far the gaze has moved from timestamp t1 to timestamp t2.
4) We divided the visual angle by the time between t1 and t2 in seconds. This gave us the angular velocity of an eye movement in degrees/second at timestamp t1t2.
5) We aggregated the velocity of the left and right eye by taking the mean of both velocities. If the velocity of one eye could not be inferred (e.g. because the person had blinked with one eye at time t1 or t2 or both), we took the velocity of the other eye. If velocities of both eyes could not be inferred, the sample was declared an invalid value.
6) We calculated the gaze positions in 2D and 3D space at timestamp t1t2, by taking the mean of each gaze position on time t1 and time t2.

7) For each point, we stored the timestamp t1t2, the angular velocity v_t1t2 at this timestamp, the gaze points at t1t2, and the eye position at t1t2.
8) Next, we labeled all points with a velocity below or equal to the velocity threshold parameter as fixations and all points above the threshold as a saccade. To study the impact of this velocity parameter on eye-tracking parameters, such as duration and number of fixations, we performed our analysis with different threshold values between 10 and 150 (stepsize=10). In addition, we put special attention to the velocity threshold of 30°/sec, as this is used per default in the fixation filter of the Tobii Lab Analysis Software, and the velocity threshold of 100°/sec, as this is used per default in the attention filter of the Tobii Lab Analysis Software (Tobii, 2022).

Building fixation groups. After this, we build fixation and saccade groups by merging all consecutive points containing a fixation to a fixation group, all consecutive saccades to a saccade group, and all consecutive points with an invalid value to an invalid group. For each group, we defined the start time as the point of time between the timestamp of the first sample in this group and the timestamp of the last sample of the preceding group. Similarly, we defined the end time as the point of time between the timestamp of the last sample in this group and the timestamp of the first sample of the preceding group. We calculated the duration of the group by subtracting the start time from the end time and calculating the eye and gaze positions by taking the mean of all eye and gaze points in this group. We furthermore stored the eye movement type (fixation, saccade, invalid) as well as a counter for fixations, saccades, and invalids.

Fixation merging. We then merged fixation groups, which were divided by a saccade or invalid value, but were close in time and space. We did this using the following steps

1) For each pair of subsequent fixation groups f1,f2, we calculated the time between the end of f1 and the beginning of f2. If this time was shorter than a threshold (max_time_betw_fixations), we continued with step 2, otherwise, we continued with the

next fixation pair. We used a max_time_betw_fixations threshold of 75 milliseconds as recommended in (Olsen, 2012).

2) We calculated the visual angle between f1 and f2 by using the mean eye position of f1 and f2, the gaze position in f1, and the gaze position in f2 for the left eye. We did the same for the right eye and merged the visual angles of both eyes as described in step 5 in fixation calculation. If the overall angle was shorter than a threshold (max_angle_betw_fixations), we merged the fixation groups in the same way as merging consecutive fixations. All saccades and invalid values between f1 and f2 were thus discarded. We used a max_angle_bw_fixations threshold of 0.5 degrees, as recommended in (Olsen, 2012).

To study the impact of fixation merging on eye-tracking parameters, we performed our analysis once with and once without fixation merging.

Eye-tracking parameter calculation. We then continued to examine individual eye-tracking parameters. For each person, we calculated the number of fixations $fixnr_{Person\ x}$ as well as the mean fixation duration $fixdur_{Person\ x}$, meaning the sum of lengths of all fixations of this person, divided by the number of fixations. Furthermore, we defined the Gaze Relational Index of a person as $GRI_{Person\ x} = fixnr_{Person\ x}/fixdur_{Person\ x}$ and calculated this index for each person.

For an expert-novice contrast, we calculated the mean fixation number $fixnr_{Group\ x}$ by taking the mean of the fixation numbers of all participants in this group. Furthermore, we calculated the mean fixation duration of this group $fixdur_{Group\ x}$, by taking the mean of the mean fixation durations of all participants in this group. We then calculated the Gaze Relational Index of a group, as defined in (Gegenfurtner et al., 2020), by using the following formula: $GRI_{Group\ x} = fixnr_{Group\ x}/fixdur_{Group\ x}$.

3 Results

3.1 Differences between experienced and novice teachers' gaze relational index

The first research question examined the extent to which experienced and novice teachers differ in the GRI. Table 1 shows the eye movement parameters number of fixations, duration of fixations, and GRI using the velocity threshold of 30°/s and 100°/s, separated by expertise level. Descriptive results indicate that experienced teachers had more fixations, shorter fixation durations, and a lower GRI compared to novice teachers. Although the trend of the results has not changed, varying velocity thresholds have an impact on the eye movement parameter/GRI. For example, while the difference between expert groups in GRI is marginal at a velocity threshold of 30°/s, novice teachers' GRI is more than double that of experienced teachers at a velocity threshold of 100°/s.

Table 1

Group-based eye-tracking parameter and GRI with velocity threshold of 30/100 and no merging of fixations

	Fixation Number		Mean Fixation Duration		GRI
	M	*SE*	*M*	*SE*	
Experts (VT 30°/s)	4319.50	566.50	125.52	17.82	0.030
Novices (VT 30°/s)	3991.50	569.50	156.35	3.35	0.039
Experts (VT100°/s)	3981.50	674.50	186.79	29.51	0.047
Novices (VT100°/s)	2569.50	401.50	326.37	18.70	0.127

3.2 Impact of varying velocity threshold on the number of fixations, duration of fixations, and the GRI

The second research question examined the impact of varying velocity thresholds and fixation merging (yes/no) on eye movement parameters/GRI. Figure 1 shows the eye movement parameter/GRI for each person with each analysis. Merging categories seem to have little to no effect on the mean fixation duration, fixation number, and GRI of a person. However, as already indicated in the results of RQ1, the velocity threshold seems to have a high influence on the eye movement parameter/GRI: A higher velocity threshold leads to more samples being classified as fixation. As consecutive samples containing a fixation are merged, a higher velocity threshold implies a higher mean fixation duration. At the same time, a higher velocity threshold leads to a lower number of fixations for thresholds above 30-40°/s. At first sight, this seems to be counterintuitive, but is based on the merging of consecutive samples: For example, there could be 3 samples in the dataset, s1, s2, and s3. With a velocity threshold of 30, these would be classified as s1=fixation, s2=saccade, and s3=fixation, resulting in a fixation number of two. With a velocity threshold of 100 in contrast, s2 could be classified as fixation as well. As consecutive fixations are merged, this would result in a fixation number of one, meaning the fixation number decreases with an increasing velocity threshold. However, the above-identified relation is not linear, which means that the order of the participants in terms of their level of GRI is changed. For example, E1 has a higher GRI than N2 when using a velocity threshold of 30°/s, but a lower GRI than N2 when using a velocity threshold of 100°/s.

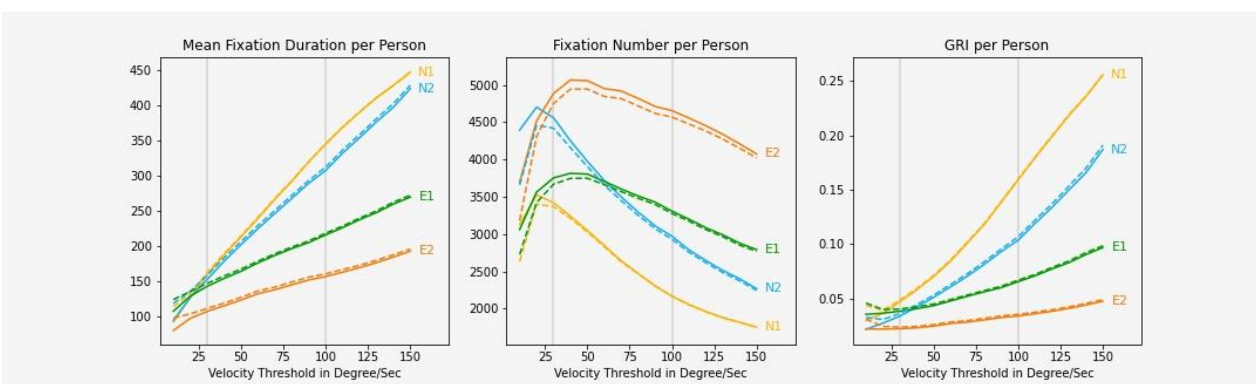

Figure 1: *Mean fixation duration, fixation number, and GRI per person for a fixation calculation with different velocity thresholds between 10 and 150 degrees per second. Solid lines represent the analysis without merging of fixation groups and dashed lines the analysis with fixation group*

merging. The velocity thresholds of 30 and 100 are marked with a grey line. E1/2 = Experienced teachers, N1/2 = Novice teachers

4 Discussion

The teaching profession heavily depends on visual information—teachers visually perceive, collect, and process information in a complex and dynamic classroom environment (Wolff et al., 2016). Over the last years, cognitively-oriented educational research found that experienced teachers develop domain-specific visual expertise which has not yet developed in novice teachers (Kosel, Holzberger, & Seidel, 2021; van den Bogert et al., 2014; Wolff et al., 2016). The present study aimed to contrast the visual expertise of experienced and novice teachers, as measured by the Gaze Relational Index (GRI), in highly dynamic real-world classroom environments using mobile eye-tracking data. Furthermore, the study explores how different configurations (varying velocity thresholds & fixation merging) of the IV-T algorithm for eye-movement classification affect the results of the study. In general, our findings correspond to the perceptual superiority of domain experts (indicated by a lower GRI) and findings suggest that different velocity thresholds for eye-movement identification significantly affected the results of our study.

4.1 Gaze Relational Index as an expertise-sensitive metric in research about teachers' visual expertise

We expected experienced teachers to process visual information faster and with more numerous fixations (characterizing the domain-specific superiority of experienced teachers in visual processing; Gegenfurtner et al., 2011), and thus need less time and effort to comprehend the complexity of classroom situations (Gegenfurtner et al., 2022). Therefore, the GRI (ratio of the mean number of fixations to the mean duration of fixations) was expected to be lower for experienced teachers than for novices. We were able to provide support for this hypothesis as we found experienced teachers have more fixations with shorter average fixation duration than novice teachers and thus, have a lower GRI compared to novice teachers. Contrary to other studies in the context of visual expertise of medical experts and novices (Gegenfurtner et al., 2020) as well as experienced and novice teachers (Grub et al., 2022), the calculated GRI in this study was more sensitive to expertise. One decisive reason for the heterogeneous results may play a role (besides more technical reasons, which will be discussed later): Compared to the outlined studies above, we have begun to step outside artificial classroom environments of laboratory setups toward the more natural conditions teachers usually face in real classrooms using mobile eye-tracking. It has been shown that eye movements in the real world generally vary more among participants (Dowiasch et al., 2020). Dowiasch and colleagues (2020) argue that this could be since mobile eye-tracking gaze recordings are generally much less restrictive than laboratory gaze recordings, allowing participants to behave more naturally. In this context, teachers often experience a much higher level of complexity in their real work environment, which is difficult to mirror in laboratory eye-tracking research. Therefore, a general transferability of results from eye-movement measurements in the laboratory to the real world seems difficult, although researchers underline to better understand visual behavior/expertise in natural environments (Dowiasch et al., 2020; Gegenfurtner et al., 2020). We have taken this step with this study and can confirm assumptions on expertise differences in the GRI. Results might indicate that experienced teachers` superior visual processing comes to the surface, especially in complex and dynamic real-life situations.

4.2 Varying velocity thresholds for eye-movement identification influence the GRI

Across all research areas, eye-tracking-based studies face the critical challenge of transforming the raw gaze signals of the eye-tracker (i.e., gaze origin and the gaze direction) into meaningful gaze parameters (i.e., fixations, saccades) (Olsen, 2012; Tobii, 2022). Not only that there are numerous algorithms for this task available, the algorithms often work with different and customizable configurations (Andersson et al., 2017). We investigated the impact of different velocity threshold settings for one of the most commonly used algorithms (IV-T: Andersson et al., 2017; Olsen, 2012) on the results of our mobile eye-tracking study (RQ2a). The results indicate that the choice of a velocity threshold influences the mean fixation duration and fixation number per person and it consequently influences the GRI per person. In addition, we identified that the selection of the velocity threshold not only influences the absolute size of the GRI but also the rank order of participants regarding their GRI. In other words, the different velocity thresholds do not have a linear effect on the number and duration of fixations and the GRI. Concerning the default fixation filter (30°/s) and attention filter (100°/s) provided by Tobii (Olsen, 2012; Tobii, 2022), the results are less influenced when interpreted on the averaged group level (experts vs. novices) than on the individual level (e.g., comparison of single participants). However, because eye-tracking studies have comparatively few participants compared to other traditional study designs (i.e., questionnaire surveys), the presumed influence on the results is all the more striking (for example, when comparing group means). From this more methodological perspective, we argue that the upcoming heterogeneity that occurs in the results of visual expertise studies (e.g., described by Klostermann & Moeinirad, 2020) regarding the number and duration of fixations of domain experts may be due not only to different study contexts (e.g., varying professional domains or tasks) but also to the choice of a specific velocity threshold. Thereto, the choice of velocity threshold should be regularly reported in publications. Furthermore, the process of fixation merging (RQ2b) did not affect our results compared to different velocity thresholds. This is probably due to the fact that in our data very few fixation groups are merged. Choosing higher parameters for the maximal time between fixations and the maximum angle between fixations would result in more fixations being merged and could consequently lead to higher differences between analyses. Since the manual settings of fixation merging are more restricted (in comparison to velocity thresholds) in current software packages (Olsen, 2012; Tobii, 2022), we assume that the influence concerning fixation merging in studies is reduced since default values are often maintained.

In sum, our study demonstrates the importance of transparently specifying configurations of algorithms for eye-movement classification in eye-tracking studies that base their interpretation on fixations and saccades. This is one step to valid, reliable, and objective measurements of eye movements in the field of visual expertise. Based on our results and in agreement with Hossain and colleagues (2016), we recommend using the 100°/s fixation filter when mobile eye tracking is used and head movements are involved.

5.2 Limitations and future directions

The present study has three main limitations that can be addressed by future research.

First, our study is limited to descriptive (group) comparisons mainly due to the small sample size. As the present results are exploratory, further research is needed to confirm these observed differences using a larger sample size.

Second, we have limited knowledge of the extent to which the GRI is related to the specific situations teachers face in the classroom. Therefore, the following considerations on this aspect must be taken into account: Our analysis showed that experienced and novice teachers differed in their visual behavior as measured by the GRI, but we know little about *how* they differed in their interpretation of what they saw. Future research should focus on a more comprehensive combination of eye-tracking and think-aloud protocols to understand how the GRI relates to the underlying instructional situations that the teacher cognitively and visually faced during the eye-tracking recording. Another way to achieve this would be to code the first-person video recorded by the mobile eye-tracking afterward to investigate the GRI, for example, in different forms of instructional interaction (frontal instruction, group work). In this context, it should also be noted that we performed a gaze-based approach (analyses are based only on the fixations) and have not integrated any Areas of Interest (AOIs)—this means that we did not take into account the distribution of attention to specific areas in the classroom. This brings up an important point that should be considered in future research. To realize the full potential of the GRI, future studies should integrate relevant AOIs in mobile eye-tracking data to analyze which areas in the classroom are being processed with a high or low GRI (for example, to analyze the GRI in relation to task-related and task-irrelevant areas). To summarize this above-described outlook, there is a further need to understand in which situations the visual expertise of experienced teachers come to the fore.

Third, we have focused on two essential parameters (velocity threshold, fixation merging) for the various configurations of the IV-T algorithm and ignored other aspects such as interpolation (filling gaps in raw eye-tracking data in which no signal was recorded) or active noise reduction (e.g., noise may be caused by imperfect system settings) (Tobii, 2022). The extent to which these (often manufacturer-specific) methods of data preprocessing influence the results (especially when studies use different eye-tracking devices) is to be clarified.

A final comment to the GRI is warranted. It should be noted that the GRI depends on the length of the recording: While the mean fixation duration should remain relatively constant over a period of time, the number of fixations increases with each recording minute, leading to a decrease in the GRI value. Therefore, we recommend analyzing participants always over the same amount of time and to state the recording time when reporting a study.

5.3 Conclusion

Our study leads to two important consequences concerning research on teachers' visual expertise: The GRI might serve as a sensitive measure of visual expertise when using mobile eye-tracking data. The lower GRI of experienced teachers indicates that they have a distinct visual behavior, which is indicative of their fine-grained domain-specific knowledge organization that is reflected in their visual expertise. Regardless of selected velocity thresholds for identifying fixations, the experienced teachers showed shorter and more fixations that resulted in a lower GRI. However, from a methodological viewpoint, the study also showed that the selected velocity threshold parameters alter and, in the worst case, bias the results of an eye-tracking study. Therefore, in the interest of further generalizability of the results within visual expertise research, researchers are encouraged to be transparent about reporting their configurations relevant to eye movement identification.

References

Aglioti, S., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature neuroscience*, *11*, 1109–1116. https://doi.org/10.1038/nn.2182

Andersson, R., Larsson, L., Holmqvist, K., Stridh, M., & Nyström, M. (2017). One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior Research Methods*, *49*(2), 616–637. https://doi.org/10.3758/s13428-016-0738-9

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, *35*(5), 463–482. https://doi.org/10.1016/S0883-0355(02)00004-6

Chaudhuri, S., Muhonen, H., Pakarinen, E., & Lerkkanen, M.-K. (2021). Teachers' Focus of Attention in First-grade Classrooms: Exploring Teachers Experiencing Less and More Stress Using Mobile Eye-tracking. *Scandinavian Journal of Educational Research*, *9891*(2), 1–17. https://doi.org/10.1080/00313831.2021.1958374

Cortina, K. S., Miller, K. F., McKenzie, R., & Epstein, A. (2015). Where Low and High Inference Data Converge: Validation of CLASS Assessment of Mathematics Instruction Using Mobile Eye Tracking with Expert and Novice Teachers. *International Journal of Science and Mathematics Education*, *13*(2), 389–403. https://doi.org/10.1007/s10763-014-9610-5

Dolezalova, J., & Popelka, S. (2016). ScanGraph: A Novel Scanpath Comparison Method Using Visualisation of Graph Cliques. *Journal of Eye Movement Research*, *9*(5), 1–13. https://doi.org/10.16910/jemr.9.4.5

Dowiasch, S., Wolf, P., & Bremmer, F. (2020). Quantitative comparison of a mobile and a stationary video-based eye-tracker. *Behavior Research Methods*, *52*(2), 667–680. https://doi.org/10.3758/s13428-019-01267-5

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245. https://doi.org/10.1037/0033-295X.102.2.211

Gegenfurtner, A. (2020). *Professional vision and visual expertise* (Habilitation Thesis). University of Regensburg, Regensburg.

Gegenfurtner, A., Boucheix, J.-M., Gruber, H., Hauser, F., Lehtinen, E., & Lowe, R. (2020). The Gaze Relational Index as a Measure of Visual Expertise, *3*, 32–40.

Gegenfurtner, A., Gruber, H., Holzberger, D., Keskin, Ö., Lehtinen, E., Seidel, T., . . . Säljö, R. (2022). Towards a Cognitive Theory of Visual Expertise: Methods of Inquiry. In.

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, *23*(4), 523–552.

Gegenfurtner, A., & Seppänen, M. (2013). Transfer of expertise: An eye tracking and think aloud study using dynamic medical visualizations. *Computers & Education*, *63*, 393–403. https://doi.org/10.1016/j.compedu.2012.12.021

Goodwin, C. (1994). Professional Vision. *American Anthropologist*, *96*(3), 606–633. https://doi.org/10.1525/aa.1994.96.3.02a00100

Grub, A.-S., Biermann, A., & Brünken, R. (2020). Process-based measurement of professional vision of (prospective) teachers in the field of classroom management: A systematic review, *12*, 75–102.

Grub, A.-S., Biermann, A., Lewalter, D., & Brünken, R. (2022). Professional Vision and the Compensatory Effect of a Minimal Instructional Intervention: A Quasi-Experimental Eye-Tracking Study With Novice and Expert Teachers. *Frontiers in Education*, *7*, 1248. https://doi.org/10.3389/feduc.2022.890690

Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2015). *Eye tracking: A comprehensive guide to methods and measures* (First published in paperback). Oxford: Oxford University Press.

Hossain, A., & Miléus, E. (2016). *Eye Movement Event Detection for Wearable Eye Trackers*. Linköping.

Huang, Y., Miller, K. F., Cortina, K. S., & Richter, D. (2021). Teachers' professional vision in action. *Zeitschrift für Pädagogische Psychologie*, *36*, 1–18. https://doi.org/10.1024/1010-0652/a000313

Klostermann, A., & Moeinirad, S. (2020). Fewer fixations of longer duration? Expert gaze behavior revisited. *German Journal of Exercise and Sport Research*, *50*(1), 146–161. https://doi.org/10.1007/s12662-019-00616-y

Kosel, C., Holzberger, D., & Seidel, T. (2021). Identifying Expert and Novice Visual Scanpath Patterns and Their Relationship to Assessing Learning-Relevant Student Characteristics. *Frontiers in Education*, *5*, 3. https://doi.org/10.3389/feduc.2020.612175

Lachner, A., Jarodzka, H., & Nückles, M. (2016). What makes an expert teacher? Investigating teachers' professional vision and discourse abilities. *Instructional Science*, *44*(3), 197–203. https://doi.org/10.1007/s11251-016-9376-y

Lappi, O., Rinkkala, P., & Pekkanen, J. (2017). Systematic Observation of an Expert Driver's Gaze Strategy—An On-Road Case Study. *Frontiers in psychology*, *8.* https://doi.org/10.3389/fpsyg.2017.00620

McIntyre, N. A., & Foulsham, T. (2018). Scanpath analysis of expertise and culture in teacher gaze in real-world classrooms. *Instructional Science*, *46*(3), 435–455. https://doi.org/10.1007/s11251-017-9445-x

Olsen, A. (2012). The Tobii I-VT Fixation Filter [Computer software]: Tobii.

Panetta, K., Wan, Q., Rajeev, S., Kaszowska, A., Gardony, A. L., Naranjo, K., . . . Agaian, S. (2020). ISeeColor: Method for Advanced Visual Analytics of Eye Tracking Data. *IEEE Access*, *8*, 52278–52287. https://doi.org/10.1109/ACCESS.2020.2980901

Peißl, S., Wickens, C. D., & Baruah, R. (2019). Eye-Tracking Measures in Aviation: A Selective Literature Review. *The International Journal of Aerospace Psychology*, *28*(3-4), 98–112. https://doi.org/10.1080/24721840.2018.1514978

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In A. T. Duchowski (Ed.), *Proceedings of the symposium on Eye tracking research & applications - ETRA '00* (pp. 71–78). New York, New York, USA: ACM Press. https://doi.org/10.1145/355017.355028

Seidel, T., & Stürmer, K. (2014). Modeling and Measuring the Structure of Professional Vision in Preservice Teachers. *American Educational Research Journal*, *51*(4), 739–771. https://doi.org/10.3102/0002831214531321

Tobii (2022). *Tobii Pro Lab User Manual v1.194.*

Van den Bogert, N., van Bruggen, J., Kostons, D., & Jochems, W. (2014). First steps into understanding teachers' visual perception of classroom events. *Teaching and Teacher Education*, *37*(1), 208–216. https://doi.org/10.1016/j.tate.2013.09.001

Van Es, E. A., & Sherin, M. G. (2010). The influence of video clubs on teachers' thinking and practice. *Journal of Mathematics Teacher Education*, *13*(2), 155–176. https://doi.org/10.1007/s10857-009-9130-3

Weibel, N., Fouse, A., Emmenegger, C., Kimmich, S., & Hutchins, E. (2012). Let's look at the cockpit. In C. H. Morimoto, H. Istance, J. B. Mulligan, P. Qvarfordt, & S. N. Spencer (Eds.), *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12* (p. 107). New York, New York, USA: ACM Press. https://doi.org/10.1145/2168556.2168573

Wolff, C. E., Jarodzka, H., van den Bogert, N., & Boshuizen, H. P. A. (2016). Teacher vision: Expert and novice teachers' perception of problematic classroom management scenes. *Instructional Science*, *44*(3), 243–265. https://doi.org/10.1007/s11251-016-9367-z

4.3 Study 3: Glowing Experience or Bad Trip? A Quantitative Analysis of User Reported Drug Experiences on Erowid.org

This publication is RELEVANT TO THE EXAMINATION.

Authors
Angelina Mooseder, Momin M. Malik, Hemank Lamba, Earth Erowid, Sylvia Thyssen, Jürgen Pfeffer

In
Proceedings of the International AAAI Conference on Web and Social Media (ICWSM 2022), 16(1), 675-686, peer-reviewed, published on 06-2022, https://doi.org/10.1609/icwsm.v16i1.19325.

Abstract
Erowid.org is a website dedicated to documenting information about psychoactive substances, with over 36,000 user-submitted drug Experience Reports. We study the potential of these reports to provide information about characteristic experiences with drugs. First, we assess different kinds of drug experiences, such as 'addiction' or 'bad trips'. We quantitatively analyze how such experiences are related to substances and user variables. Furthermore, we classify positive and negative experiences as well as reported addiction using information about the consumer, substance, context and location of the drug experience. While variables based only on objective characteristics yield poor predictive performance for subjective experiences, we find subjective user reports can help to identify new patterns and impact factors on drug experiences. In particular, we found a positive association between addiction experiences and dextromethorphan, a substance with largely unknown withdrawal effects. Our research can help to gain a deeper sociological understanding of drug consumption and to identify relationships which may have clinical relevance. Moreover, it can show how non-mainstream social media platforms can be utilized to study characteristics of human behavior and how this can be done in an ethical way in collaboration with the platform providers.

Contribution of Thesis Author
Theoretical conceptualization, data preprocessing, algorithmic design, discussions and implementation, data visualization, as well as manuscript writing, revision, and editing.

Glowing Experience or Bad Trip? A Quantitative Analysis of User Reported Drug Experiences on Erowid.org

Angelina Mooseder,[1] Momin M. Malik,[2] Hemank Lamba,[3]
Earth Erowid,[4] Sylvia Thyssen,[4] Jürgen Pfeffer[1]
[1]School of Social Sciences and Technology, Technical University of Munich
[2]Berkman Klein Center for Internet & Society at Harvard University
[3]School of Computer Science, Carnegie Mellon University
[4] Erowid.org, Grass Valley, California, United States
angelina.mooseder@tum.de

Abstract

Erowid.org is a website dedicated to documenting information about psychoactive substances, with over 36,000 user-submitted drug Experience Reports. We study the potential of these reports to provide information about characteristic experiences with drugs. First, we assess different kinds of drug experiences, such as 'addiction' or 'bad trips'. We quantitatively analyze how such experiences are related to substances and user variables. Furthermore, we classify positive and negative experiences as well as reported addiction using information about the consumer, substance, context and location of the drug experience. While variables based only on objective characteristics yield poor predictive performance for subjective experiences, we find subjective user reports can help to identify new patterns and impact factors on drug experiences. In particular, we found a positive association between addiction experiences and dextromethorphan, a substance with largely unknown withdrawal effects. Our research can help to gain a deeper sociological understanding of drug consumption and to identify relationships which may have clinical relevance. Moreover, it can show how non-mainstream social media platforms can be utilized to study characteristics of human behavior and how this can be done in an ethical way in collaboration with the platform providers.

1 Introduction

In the past decade, social media research has primarily focused on large platforms like Twitter or Facebook to analyze social phenomena. This has lead to over-studied platforms in which challenges of generalizability are well-known (Ruths and Pfeffer 2014). In contrast, studying unique, small online communities can provide new insights into the breadth of human behavior. In particular, it can offer the opportunity to examine less publicly discussed aspects of human experience, such as drug[1] consumption.

Around the world, the consumption of drugs and the number of available substances has risen dramatically over the last decade and since 2005, around 950 new psychoactive substances have been identified worldwide (United Nations 2020). The great complexity of the world's drug landscape poses new challenges for social workers, governmental institutions, medical personnel, and drug consumers themselves (Arillotta et al. 2020; D'Agnone 2015; Schifano 2020). It is important to detect adverse reactions to substances, understand under which circumstances they arise and implement meaningful harm reduction approaches, as well as to identify positive effects of specific drugs and understand their therapeutic potential. To achieve these goals, information is needed about why, when and in which context certain drugs are used and which effects they produce in different settings.

One website providing such information is Erowid.org. For more than 25 years, the site has been collecting and curating information about (often illicit) drugs, serving people who use these substances, as well as family members, clinicians, educators, policy makers, and the curious general public. Of particular interest are the Experience Reports published on Erowid: in more than 36,000 reports, users[2] described what happened when they consumed one or more of over 800 different substances. Analyzing these reports gives a unique opportunity to study drug use from the consumer's perspective.

In cooperation with the Erowid staff members, we investigate the subjective experiences of drug consumers in a large-scale, quantitative manner. We first give a description of the dataset (Section 3) and the user base (Section 4). We then identify how various drugs and user variables are linked to characteristics of the drug experience (Section 5). Furthermore, we test whether the subjective outcome of a drug experience is predictable, using information about the drug, consumer and situational factors (Section 6). Based on a dataset of 36,711 user Experience Reports we find that:

- **Age plays a significant role in the motivation for and evaluation of drug consumption.** Younger people report more about bad and difficult experiences with drugs, while older people report more about using drugs for medical purposes.
- **Males and females differ in their drug experiences.** Females report more about using drugs for medical pur-

[1]We refer to 'drug' as a term for the wide range of unapproved recreational, steroidal, performance enhancing, sedative, or other bio-active chemical technologies and the disapproved use of approved pharmacological agents of any variety.

[2]We refer to (Erowid) users as term for people, who submit Experience Reports to Erowid.

poses, while at the same time report more about health problems and addiction associated with drugs.

- **The outcomes of reported drug experiences are linked to the substances consumed.** LSD is associated with negative experiences, while MDMA is associated with positive experiences.
- **Reported health problems and addiction are linked to specific substances.** The data revealed understudied patterns, such as an association between DXM and addiction, which may have clinical relevance.
- **Although drugs and situational factors are correlated to the outcome of an experience, they do not yield enough information to predict whether an experience will be positive, negative, or associated with addiction.**

We then discuss the benefits and limitations of using subjective drug Experience Reports found online. Our research presents new approaches for psychopharmacological research and can help to obtain a deeper sociological understanding of drug consumption.

2 Background and Related Work

2.1 Analysis of Drug Consumption

The criminalization of substance abuse makes it difficult to accurately study drug consumption. Researchers often rely on surveys of focus groups (Substance Abuse and Mental Health Services Administration 2019; European Monitoring Centre for Drugs and Drug Addiction 2020), which suggests a high risk of reporting bias: respondents may not feel comfortable expressing their opinion about drug use and may conceal information about behavior which is illegal or not socially accepted. While collections of anonymous, self-reported experiences with drug usage are by no means a representative sample and in fact are likely a highly biased sample, we assume that their content is relatively free of self-reporting bias. Therefore, we use information substance consumers have voluntarily revealed in online communities to contribute new knowledge about drug consumption.

2.2 Online Drug Communities

Erowid.org was founded in 1995, early in Web history, but it is not the only drug-related Internet community. There are other communities that allow users to discuss substance related topics,[3], exchange recipes and recommendations for substance use,[4] and publish information about the assumed contents of substances they have received.[5] Many of these forums have the aim to offer accurate information about substances, share experiences of both positive and negative effects, give advice or warnings about dosages, and provide support for users experiencing negative reactions (Soussan and Kjellgren 2014; Rolando and Beccaria 2019). By reading such information, users try to gain new knowledge about substances and to minimize their risk of experiencing adverse effects (Norman, Grace, and Lloyd 2014; Duxbury 2018; Bilgrei 2019; Berning and Hardon 2016).

[3]https://drugs-forum.com/

[4]https://bluelight.org/

[5]https://pillreports.net/

The analysis of such communities has been used for detecting new trends on the drug market (Arillotta et al. 2020; Schifano 2020; Rhumorbarbe et al. 2019), finding common drug-drug combinations (Chary, Yi, and Manini 2018), and understanding the use of specific substances (Andersson and Kjellgren 2017; Bonson, Buckholtz, and Murphy 1996).

2.3 Erowid.org

The website Erowid.org has become a valuable resource for researchers seeking to gain information about the compounds, dosages, classification and effects of drugs, especially for substances with no or little medical documentation (Karila et al. 2016; Stanciu, Penders, and Rouse 2016). The organization Erowid Center has used the website to recruit participants for surveys on visitor demographics and experiences with certain substances[6] as well as for surveys on specific drugs conducted in collaboration with academic researchers (Pal et al. 2013; Baggott et al. 2011, 2010; Gamma et al. 2005).

The experience reports published on Erowid have been qualitatively analyzed to gain an understanding of the use of synthetic cannabinoids and Kratom (Swogger et al. 2015; Newman et al. 2016). Furthermore, these reports have been used for anthropological case studies, in which the authors visualized patterns, such as substance co-use, common dosages of specific drugs or wordpair-substance relationships (Krieg et al. 2016; Krieg, Berning, and Hardon 2017). However, there still exist many more use cases on which such data could be applied. The aim of this paper is to understand the potential of subjective drug consumption reports to provide information about the characteristic experiences of these drugs.

2.4 Ethical Considerations

We are conducting research on open (though anonymous or pseudonymous) admissions of sometimes illegal behavior, and we have a responsibility to not put together data in such a way that it would deanonymize individuals and potentially put them at risk. When submitting reports, users explicitly grant permission to use their reports for scientific research. Therefore, care is already taken in the writing and editing process to omit or obscure identifiable details. Our data contains no personal identifiable information (PII). Nothing in our models tries to fill in or infer information from individual reports that could help identify the authors.

Second, the manufacture, distribution, possession and/or consumption of many of the substances described in Erowid Experience Reports are illegal in many or most jurisdictions. By studying these reports in a nonjudgmental way, we risk abetting and condoning behavior that could be illegal and that many consider immoral. However, there is a long precedent in sociology and anthropology of studying such behavior; for drugs specifically, we work within the educational and *harm reduction* frameworks (Marlatt 1996) supported by Erowid and by many biomedical researchers, social workers, and medical practitioners. Harm reduction is an evidence-based approach built on more than three decades of

[6]https://www.erowid.org/general/survey/

empirical research (Stone and Shirley-Beavan 2018), and it prioritizes providing accurate, judgement-free information, safe environments for drug consumption, addiction treatment, and decriminalization of drug usage and markets, all of which lead to less harm from drug consumption. The current study obtained the explicit consent of Erowid Center before downloading reports off their website. Furthermore, we recognize the labor and expertise of the Erowid staff members through co-authorship.

3 Data

3.1 Erowid

Erowid is viewed by advocates of drug policy reform and harm reduction as one of the most important resources on drugs (Jarnow 2016). It describes itself as

> "a member-supported organization providing access to reliable, non-judgmental information about psychoactive plants, chemicals, and related issues. We work with academic, medical, and experiential experts to develop and publish new resources, as well as to improve and increase access to already existing resources. We also strive to ensure that these resources are maintained and preserved as a historical record for the future." [7]

Much of the site is devoted to psychoactive substance 'vaults', subsections containing extensive information about various psychoactive drugs. The 'experience vaults',[8] a collection of narratives about consuming psychoactive substances (and/or practicing psychoactive methods, such as fasting or meditation), largely submitted by site readers (with additional Experience Reports republished and compiled from other sources), is particularly fascinating.

Over a hundred thousand experience reports have been submitted to Erowid, about 35% of which have been published. Reports pass through a review process, with each report being reviewed by two out of a few dozen trained volunteers who read and rate the submissions and pass them on to senior reviewers for a possible publication (Erowid, Thyssen, and Erowid 2018; Witt 2015; Erowid and Erowid 2006). During the review process, each report is tagged with 'primary' categories (= type of experience, such as 'Bad Trip') and 'secondary' categories (= situational factors, such as 'Nature/Outdoors') (Erowid, Thyssen, and Erowid 2018). The standardized format of Experience Reports, with fields where users input substance name, dosage, and form of consumption (pill, smoked, etc.), as well as the quality control through Erowid reviewers make them particularly valuable compared to similar sites.

3.2 Data Collection

Following the rules of spidering given on the Erowid site,[9] we contacted Erowid and received permission to crawl the site. We then collected 36,778 html pages between 2021-02-16 and 2021-04-07. As 65 of these pages did not follow the formatting standards of Experience Reports, they were omitted, leaving a dataset of 36,713 reports. Each report generally consisted of an identification number, title, text, publication date, author information, such as pseudonym, age,[10] gender and weight, substance information, dosage information, number of views, year of experience, and category labels assigned by the Erowid team.

3.3 Data Cleaning

User and Report Information. We cleaned and regularized the data extensively. We manually inspected all reports with an unusual author age under 13 or over 70. In cases where the report showed indications that the age information was wrong (eg. the age being four-digit, or the activities not matching the age, such as a 12-year-old being the driver of a car), we replaced the age with a missing value. We converted the user weight information, which was given in pounds (lb), kilograms (kg), or stones (st), into pounds. We replaced unrealistic low (<70 lb) and high (>500 lb) weights with a missing value. The inspection of reports with unrealistic weights revealed two reports about a cat and a dog, which had unintentionally consumed drugs. These reports were excluded from the dataset, resulting in a set of 36,711 reports. Reports on Erowid were not assigned with a unique user ID, but only with a pseudonym. Each user could have multiple pseudonyms and one pseudonym could be assigned to multiple users. Hence, it is not possible to evaluate the exact number of users and their submission history. When assessing user demographics, we counted each report as one user to obtain an upper estimation of demographics.

Categories and Context. To assess the characteristic experience of a report, we stored all 'primary' categories, which were assigned to a report by Erowid reviewers. To assess the context and location of a drug experience, we used 'secondary' category labels and context labels which were also assigned to reports by Erowid reviewers: We created the feature 'Party', which is 1 if a report contains the label 'Large Party', 'Club/Bar', 'Rave/Dance Event' or 'Festival/Large Crowd'. We constructed the feature 'Therapeutic', which is 1 if the report contains the label 'Therapeutic Intent or Outcome' or 'Therapeutic Session'. We furthermore added binary features based on the labels 'Workplace', 'School', 'Public Space', 'Nature/Outdoors', 'Guides/Sitters', 'Alone' and 'Multi-Day Experience'.

Substance Information. For each report, we used the dosage field as the source of substance information, as this provided additional information about dosage amount and consumption method. The distribution of the top 100 drugs in our dataset can be found in section A of an appendix we published online.[11] The dataset comprises reports of about 845 distinct substances (674 when grouped into larger classes). However, only 44% of these substances were included in more than ten reports. We decided for all analyses, where the drug was an independent variable (see Section 4.2, 5.2, 6), to focus only on reports about the ten most popular

[7] https://www.erowid.org/general/about/about.shtml

[8] https://erowid.org/experiences/

[9] https://erowid.org/general/about/about_archives1.shtml

[10] The age field was included in 2009.

[11] https://mediatum.ub.tum.de/1639245

Drug	Reports	Drug	Reports
Cannabis	7,274 (20%)	DXM	843 (2%)
MDMA	2,406 (7%)	Ketamine	780 (2%)
Salvia Div.	2,319 (6%)	Cocaine	776 (2%)
LSD	2,263 (6%)	P. cubensis	657 (2%)
DMT	943 (3%)	Meth.	553 (2%)

Table 1: Top Ten Drugs used for the current study with number of reports and their percentage in the dataset. Reports describing the use of multiple substances were counted once for each drug.

Drug	Perc	Dosage distribution
Cannabis	35%	15.0 - 142k (M=486, SD=3,029)
MDMA	76%	0.50 - 160k (M=284, SD=3,810)
Sal. Div.	75%	0.03 - 500k (M=1,932, SD=12,246)
LSD	87%	10.0 - 5k (M=250, SD=279)
DMT	68%	0.08 - 1k (M=51, SD=67)
DXM	80%	0.60 - 510k (M=18,971, SD=67,538)
Ketamine	45%	0.16 - 260k (M=986, SD=13,929)
Cocaine	14%	0.12 - 11k (M=1,350, SD=1,568)
P. cub.	81%	1.75 - 170k (M=4,015, SD=7,707)
Meth.	20%	5.00 - 25k (M=668, SD=2,719)

Table 2: Top Ten Drugs used for the current study with percentage of reports where dosage information was given and distribution of dosage information. LSD is reported in ug, all other drugs in mg.

drugs. The ten most reported substances were 'Cannabis', 'MDMA', 'Salvia Divinorum', 'LSD', 'Mushrooms', 'Alcohol – Beer/Wine', 'Tobacco – Cigarettes', 'DMT', 'DXM' and 'Ketamine'. As 'Mushrooms' is a superclass for a great variety of psychedelic mushrooms, we decided to focus instead on the more specific class 'Mushrooms - p. cubensis', which was the 15th most reported substance. Since both 'Alcohol - Beer Wine' and 'Tobacco - Cigarettes' were used in more than 90% of cases in combination with other drugs, we decided not to include them among examined categories. Instead, we considered the 11th and 12th most reported substances, 'cocaine', and 'amphetamines'. As amphetamines is again a superclass, we included the most common amphetamine 'methamphetamine' (ranked 20 of the most reported substances). We then stored for each report whether it would contain one or more of our top ten drugs: We created one indicator feature for each drug (e.g. 'mdma'). Each feature was assigned 1 if the respective report contained the drug and 0 otherwise. Furthermore, we constructed one indicator feature for each combination of drugs (e.g. 'mdma-dxm' or 'mdma-dxm-dmt'). As there were 10 drugs, we had 1024 possible combinations, although many combinations did not appear in the data set. From including only combinations with at least one report, our dataset yielded 142 drug and drug combination features.

Table 1 shows the resulting single substances, the number of reports per substance and the percentage of all reports in the dataset. In our data, 15,861 reports include at least one of these top ten drugs, which accounts for 43% of our dataset.

Dosage Information. After extracting the substance information for each report, we were also interested in the dosage amount. However, extracting dosages proved to be a challenging task for several reasons. First, inconsistent and incommensurable measures were used: grams and ounces, but also 'tablets', 'bowls', 'shots', 'lines', 'cookies', 'drops' and more. Since the same volume or weight can have different concentrations of an active substance, even with domain expertise the standardization of these dosage amounts would not be possible. Second, the dosage was often described in approximations such as 'multiple' or 'some' [unit]. Third, there were numerous methods to consume a certain drug (e.g. smoking, drinking,...) and it was not always possible to generate rules about how a dosage with one consumption method can be converted in a dosage with another consumption method. Fourth, there were numerous forms of substances per drug with varying levels of the active constituent and therefore the dosage amount had to be adapted to the substance form. Finally, there was a huge amount of missing data for dosages. For example, more than 60% of cocaine reports did not include dosage information. To gain at least a rough estimation of dosages, we selected for each of our top ten drugs a reference consumption method, unit, and substance form. Using the information about common doses given on various websites,[12] we created a set of heuristics per drug about how to convert dosages of other units, consumption methods or substance forms. For smoked drugs, we generally converted a 'bowl' to the height of one common dose and estimated a 'hit' to be one third of a bowl. All data, for which we could not infer any rules about either unit, consumption method or substance form, was set to a missing value and ignored for all dosage analyses. The heuristics and an overview of the most common terminologies used can be found in section B in the online appendix.[11] Table 2 shows for each drug the percentage of reports where dosage information was given and the distribution of dosages in these reports. It should be noted that the maximum dosage values can be quite high, as some users reported the amount of drugs they had taken over several days or together with other consumers.

3.4 Dataset Overview

Table 3 gives an overview of the dataset, regarding the number of reports and substances, distribution of number of substances used per report, as well as distribution of report views, drug experience year, user age, weight and gender.

4 Trends on Erowid

In this section, we analyze the whole dataset of 36,711 reports to identify 1) which drug trends are prevalent on Erowid and 2) what kind of users report to Erowid. Following Paul et al.'s (2016) work on drugs-forum.com, we compare these trends to national and international estimations of drug usage to obtain a basic understanding of the data source and the group of Erowid contributors.

[12] https://www.erowid.org/, https://www.trippingly.net/, https://drugs.tripsit.me, https://dancesafe.org

Variable	Statistics
Reports	36,711
Substances	845
Drugs p. r.	1-13 (M=1.62, SD=1, P=100%)
Views	74-777k (M=15,711, SD = 25k, P=100%)
Year	1848 - 2021 (M=2,007, SD=6, P=99%)
Age	7-80 (M = 25, SD=9, P=33%)
Weight	70-500 (M=162, SD=37, P=93%)
Gender	Male=29,052, Fem.=5,449, Not Spec.= 2,210

Table 3: Dataset Overview. P stands for the percentage of reports containing that information.

4.1 Drug Popularity

Table 1 shows the 10 most common substances (as defined in section 3.3) and their distribution in the dataset.

First, we find that *Erowid contributors show a higher interest in substances outside the most commonly used drugs*. On a global level, the estimated distribution of substance use is quite skewed towards cannabis: According to the UN's estimations, 71% of past-year 'drug' users, have consumed cannabis, which makes it by far the most consumed substance worldwide, excluding alcohol and tobacco. Other commonly used substances are MDMA/ecstasy (21%), amphetamine and methamphetamine (10%) as well as cocaine (7%) (United Nations 2020). We find that the most common drugs are also popular on Erowid, although the distribution of drugs seems to be more balanced in our data: Comprising 20% of reports, cannabis is the most prevalent drug on Erowid and also MDMA (7%), methamphetamine (2%) and cocaine show a comparatively high prevalence (2%). The fact that these percentages are not higher indicates that the Erowid users also show a high interest in other substances, which are likely less prevalent on a global level.

Second, we find that *this interest is especially strong regarding psychedelic substances*. Seven of the top ten reported drugs, namely MDMA, *Salvia divinorum*, LSD, DMT, DXM, ketamine and *Psilocybe cubensis* are substances which are categorized as hallucinogens by the North American National Survey on Drug Use and Health (RTI International 2019). While in the USA the prevalence of hallucinogens is quite low (1% of all past month substance use) in comparison to alcohol (85%), marijuana (17%) and cocaine (1%) (Substance Abuse and Mental Health Services Administration 2019), this is not the case among Erowid Experience Reports: 25% of reports contain at least one of these seven hallucinogens. Therefore, the group of Erowid contributors likely does not correspond to global or even North American drug consumption but rather to consumption by a population with a strong interest in psychedelic experiences.

4.2 Demographic Trends

Figure 1 shows the age and gender distribution over all reports. Note that 'gender' is coded only as 'male', 'female', or 'not specified'; if users select the option "non-binary/other", "prefer not to answer" or remain with the default option ("- - -") when asked for their gender, this will be displayed as 'not specified' in the downloaded dataset.

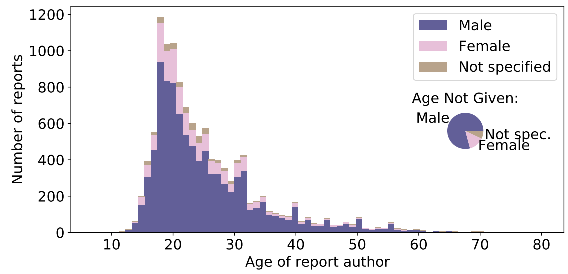

Figure 1: Stacked bar chart representing the age and gender distribution over all reports.

First, we find that *users on Erowid seem to have very similar age trends as estimated for global drug consumption*. Surveys conducted worldwide show that drug consumption is more popular among younger people, with peak levels between 18 and 25 years (United Nations 2018). This trend is also prevalent in our data. Among the third of reports that contained a valid self-report of age, at the time of experience users were on average 25 years old, with 54% of users being between 18 and 25 years old (min$=7$, max$=80$, MD$=22$, IQA$=19-29$). The distribution rises between ages 17 and 18, as the number of reports from 18 years olds is twice as high as the number from 17 years olds. Possible explanations are consumption differences, as drugs may be less accessible or interesting for minors, reporting bias, as users may feel more confident to report their age when over 18, reviewer bias, as some reviewers are less inclined to publish reports written by minors or quality-of-report bias, as writing quality or data content may be lower with juvenile authors, making reports by these authors less likely to be published.

Second, we find that *the group of Erowid contributors is highly skewed towards males*. Both in North America[13] and the European Union (European Monitoring Centre for Drugs and Drug Addiction 2020), drug consumption is reported more frequently by men than women, with around 60% of those reporting drug use in the EU being male and around 40% being female. The 'gender' imbalance on Erowid is even greater: of all reports where gender is specified, 84% are reported as 'male'. Potentially, males are more interested in or willing to write about their own drug experiences or find the website more interesting than females.

5 Characteristic Drug Experiences

In this section, we 1) analyze which characteristics of a drug experience are described by categories given on Erowid and 2) identify associations between these characteristics and individual drugs as well as user variables.

5.1 Category Overview

Here we analyze which characteristics of a drug experience are expressed within a certain category. We first give a short description of each category and their distribution over all

[13] Data from https://www.datafiles.samhsa.gov/

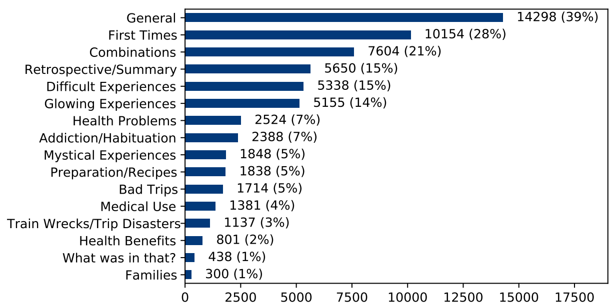

Figure 2: Number and percentage of reports per category in the dataset.

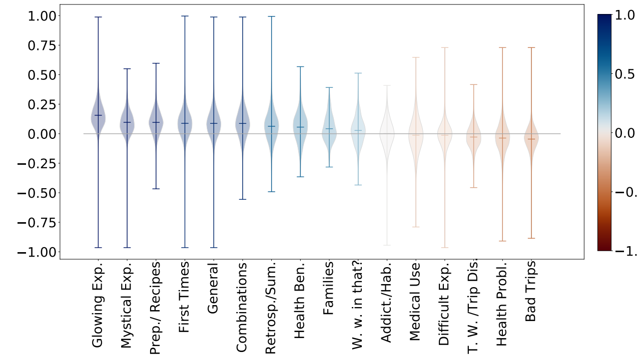

Figure 3: Distribution of sentiments per category. Lines mark the minimum, mean and maximum value. Color corresponds to the average sentiment score.

36,711 reports. We then analyze the sentiments expressed in each category as well as category co-occurrence.

Description. During the review process conducted by Erowid, each report is assigned by the Erowid team one or more of 15 primary category labels. Figure 2 presents the distribution of each category in the dataset.

The labels *'Bad Trips'*, *'Train Wrecks/Trip Disasters'* and *'Difficult Experiences'* were assigned to reports about complicated, not entirely positive experiences during drug consumption, often caused by pharmacological reactions. In contrast, the label *'Glowing Experiences'* was assigned to joyful experiences with drugs and the category *'Mystical Experiences'* was dedicated to reports about psychoactive substance induced transcendent encounters.

The label *'Medical Use'* was attached to reports on the consumption of substances for medical reasons, while the label *'Health Problems'* was assigned to reports about medical issues in general. In both sections, users often reported adverse drug effects. In contrast, the *'Health Benefits'* label was assigned to user experiences in which a substance helped to overcome certain health issues. The label *'Addiction/Habituation'* was assigned to reports about drug dependence, and many of these reports were not about a specific event, but rather a longer period of time. The label *'Retrospective/Summary'* was assigned to reports written in hindsight or over a longer period of time. The category *'What was in that?'* was used for reports in which users suspected a discrepancy between the ingredients they thought their drug would include and the real composition of the drug.

'General' was a cumulative category and had no specific meaning. The label *'First Times'* was assigned to reports about a person's first experience with a substance, and the label *'Combinations'* was assigned when the consumption of more than one substance was the main topic of the report. The category *'Preparation/Recipes'* was assigned to reports with a strong focus on the form of consumption. When reading the Experience Reports, it became evident that users on Erowid often reveal a high curiosity and experiment with new ways of consuming substances. The category *'Families'* was assigned to a great variety of reports, with the common factor of family members being involved. This included reports about persons consuming drugs in the presence of or together with family members, about users thinking of family during the drug experience, about facing drug addiction with the help of family members, and more.

Sentiments Expressed per Category. To assess the sentiments described in a category, we used VADER (Hutto and Gilbert 2014), a Python package which allows the calculation of sentiments expressed in a text and even includes the presence of negations, degree modifiers and more. We excluded the words 'like', 'ecstasy' and 'funny' from the Vader dictionary, as they had different connotations in a drug specific context. We then lower cased the texts, calculated the Vader compound sentiment score (-1=fully negative, 1=fully positive) for each sentence in a report and stored the average compound score per report. Figure 3 illustrates the distribution of sentiment scores per category.

First, we find that *labels assigned by the Erowid team correspond to the sentiments expressed in the categories*. As expected, reports in categories about negative drug experiences, such as 'Bad Trips', 'Train Wrecks/Trip Disasters' and 'Difficult Experiences', show on average rather negative sentiments, while reports in the category 'Glowing experiences' show on average rather positive sentiments. Moreover, the category 'Health Problems' has a rather negative average sentiment score, while 'Health Benefits' has a slightly positive average sentiment score. These results can be seen as a validation of Erowid's labelling process.

Second, we find that *much positivity is expressed in categories about new and mystical drug experiences*. The categories 'Mystical Experiences', 'Preparation/Recipes', 'Combinations' and 'First Times' show on average positive sentiments. Research on similar platforms suggests that online community members sometimes show characteristics of so called 'psychonauts': Their drug consumption is mainly motivated by curiosity about new substances and their possible applications, as well as the goal to gain knowledge about the inner self and the mysteries of life (Rolando and Beccaria 2019). The fact that reports in these four categories show a high amount of positive sentiments supports the hypothesis that Erowid contributors are also often interested in learning about new substances, combinations and their preparation, and value the knowledge gained by spiritual experiences.

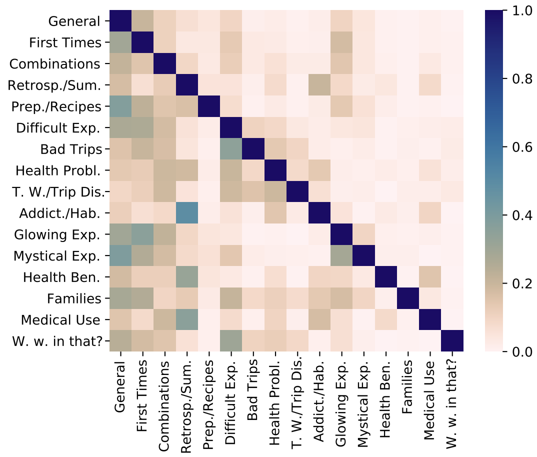

Figure 4: Category co-occurrences. The color in a cell (i, j) represents the proportion of reports in category j, which are also assigned with category i (i.e., co-occurrence frequencies are row-normalized).

Category Co-ocurrence. To analyze category similarity we calculated the label co-occurrence in all reports, which is shown in Fig 4. First, we find that *categories, which we would expect to have a high topical overlap, show a high co-occurrence*. For example, there is a considerable co-occurrence between the categories 'Difficult Experiences', 'Bad Trips', 'Health Problems' and 'Train Wrecks/Trip Disasters'. Moreover, there is a considerable overlap between 'Addiction/Habituation', 'Health Benefits' and 'Medical Use' with the category 'Retrospective/Summary'. A qualitative inspection of these reports showed that many reports in the first three categories were not written about a specific event, but rather a longer period of time. Furthermore, the genre 'What was in that?' often occurs in combination with 'Difficult Experiences', which is not surprising, as unexpected drug effects may indeed lead to troublesome drug experiences.

Second, we find that *new and mystical experiences co-occur with joyful experiences*. There is a rather high overlap between 'Mystical' and 'Glowing' experiences. In line with the results of the sentiment scores, this suggests that Erowid contributors interpret transcendent encounters often as positive, valuable and/or joyful experiences. Moreover, we find a rather high co-occurrence between 'First Times' and 'Glowing Experiences'. Possible reasons for this are that Erowid users may find it joyful to experiment with new substances (corresponding to the description of psychonauts), that there may be a selection effect in that people who have positive initial experiences are more likely to submit reports about it to Erowid, and/or that there may be a causal relationship in that first experiences indeed are more frequently positive.

In summary, we have shown that categories on Erowid describe a great variety of characteristic experiences. Categories range from positive drug experiences (e.g. 'Glowing Experiences') to negative experiences (e.g. 'Bad Trips') and even include special topics (e.g. 'Addiction/Habituation'). These results provide us with a better understanding of categories and help to interpret associations between categories and drugs or user variables.

5.2 Drug-Category Associations

In this section, we analyze the associations between categories and the top ten substances in our dataset to determine whether characteristic drug experiences are linked to specific substances.

Methods. To examine the correlation between drug and category, we measured the chi-square distance, which is the difference between the observed cell frequency and the expected cell frequency under an independence hypothesis (the product of the row and column marginals), normalized by expected cell frequency. We limited this analysis to only the 15,861 reports which contained at least one of the top ten substances.

Results. There was a significant association between drug and category ($\chi^2(135, N = 32,796) = 5,924, p < 0.01$). Figure 5 shows the chi-square distance for each drug-category pair. The exact chi-square distances and contributions can be found in section C of the online appendix.[11]

First, we find that the *subjective outcomes of drug experiences are linked to the substance consumed*. Key Findings (=findings with very high or low chi-square distances in comparison to values in the same row and column) include:

- LSD correlates with negative experiences ('Bad Trips', 'Train Wrecks/Trip Disasters').
- MDMA correlates with positive experiences ('Glowing Experiences').
- DMT, *Salvia divinorum* and *Psilocybe cubensis* correlate with 'Mystical Experiences'.

The type of experience may be induced by the pharmacological effects of the drug. For example, LSD often produces effects of paranoia and anxiety, MDMA often leads to euphoria, and DMT, *Salvia divinorum* and *Psilocybe cubensis* are known to produce hallucinogenic experiences. However, it is surprising that such correlations can be found in the data, as all of these substances also produce other pharmacological effects, which could lead to both positive and negative experiences. Furthermore, the effects of a psychoactive substance may also be influenced by the dosage, consumer variables, and the setting. Therefore we suggest that not only the biological effects play a role, but also the drug consumers expectations of these effects. For example, when users read that MDMA leads to 'Glowing Experiences', they may expect to have a positive experience with MDMA, and put themselves in a position to enjoy such a positive experience (e.g. by meeting friends, dancing,..). This could then in turn positively influence the possibility of having a positive experience and interpreting it as positive.

Second, we find that *reported health problems and addiction are linked to substances*. Key Findings include:

- Methamphetamine and cocaine correlate with 'Addiction/Habituation'.

- *Salvia divinorum*, cannabis and LSD negatively correlate with ‘Addiction/Habituation’.
- DXM correlates with ‘Health Problems’ and weakly with ‘Addiction/Habituation’.

The first two relations are in line with existing research, as both methamphetamine and cocaine are known to have a high potential for addiction, while for *Salvia divinorum*, cannabis and LSD only weak to no withdrawal symptoms are known. Therefore, the Erowid data might help to compare drugs regarding their addiction potential or even detect worrisome relationships between certain substances and medical issues. For example, our data suggests that DXM may be related to health problems and addiction, which should be investigated through further research (see Section 7.2). Such an approach could be especially beneficial for new substances, with not much clinical data available.

Third, we find that *specific usage patterns are linked to substances*. Key Findings include:

- Methamphetamine and cocaine correlate with ‘Retrospective/Summary’.
- MDMA correlates with ‘What was in that?’.

The first relation suggests that users write more often about methamphetamine and cocaine in retrospect, than they do with other drugs. This can be explained by the high addiction potential of these drugs and the tendency to use these substances repeatedly over a period of time, which may lead consumers to write about this period instead of a specific recent event. In addition, there are certainly reasons for the second relation: MDMA is the active substance in most ecstasy tablets, which vary in content and concentrations. Consequently, users often under- or overestimate the MDMA concentrations or fail to detect dangerous additional substances (Vrolijk et al. 2017). While this particular connection is already known to users and medical professionals, it illustrates a real-world trend reflected in aggregate patterns within Erowid Experience Reports.

In summary, our data reveals that the subjective outcomes of drug experiences, reported health problems and drug usage patterns are linked to specific substances. From a sociological perspective, this can help us better understand the motivations, expectations and usage patterns that drug consumers have with certain substances. From a medical perspective, these results reveal understudied patterns, such as the association between DXM and addiction, which may have clinical importance.

5.3 Associations between Category and User Variables

In this section, we analyze the associations between categories and user variables to find out whether characteristic drug experiences are linked to consumers’ age and gender.

Methods. To examine the relation between age and category, we selected all reports with age information given ($n = 11,993$), assigned an age group to each report (‘Under 18’, ‘18-25’, ‘26-40’, ‘Over 40’) and performed a chi-square test of independence. Furthermore, again using chi-square distance, we analyzed the relationship between gender and

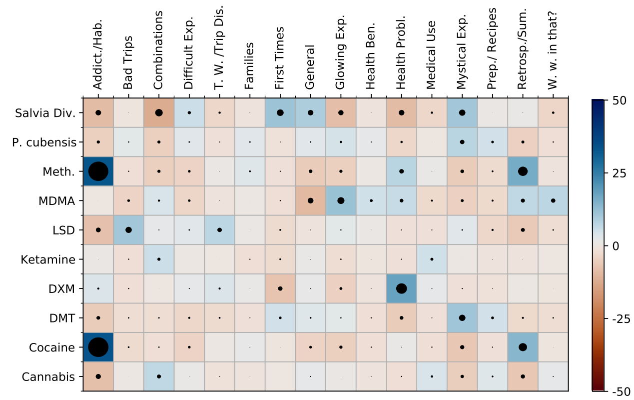

Figure 5: Chi-square distances of drug-category pairs. Positive cell values (blue) represent a positive correlation and negative cell values (red) a negative correlation between the given drug and the given category. The colors and circle sizes are proportional to the chi-squared distance; for example, the Addiction/Meth. pair has a distance of 33, Combinations/Salvia Div. of -12 and Bad Trips/Cocaine of -3.

category on all reports where gender information was specified ($n = 34,501$).

As there is a bias towards young, male users among Erowid contributors, the report numbers per group differ (‘Under 18’: 1,245, ‘18-25’: 6,488, ‘26-40’: 3,374, ‘Over 40’: 886, Male: 2,9052, Female: 5,449). However, as the chi-square test compares the observed cell frequencies with the expected cell frequencies and the expected cell frequency is not below 5 in more than 20% of cases, the differences in group size do not limit the significance of our results.

Results. Figure 6 shows the chi-square distance for each age-category pair. There was a significant relationship between age group and category ($\chi^2(45, N = 20,758) = 689, p < 0.01$). The exact chi-square distances and contributions can be found in section C of the online appendix.[11]

First, we find that *older people share more long time experiences*. It is not surprising that people above 25 have positive correlations to ‘Retrospective/Summary’ and ‘Preparation/Recipes’ as well as a weak positive correlation to the category ‘Addiction/Habituation’. Due to a higher age they had more years in which they could have experienced drug consumption; therefore, they should be able to share more drug consumption insights and to write more retrospective reports than younger people. Furthermore, they have a higher probability of having experienced an addiction and are therefore more able to report about it.

Second, we find that *older people report more about using drugs for medical purposes*. The two older groups showed positive correlations with ‘Medical Use’, while the two younger groups showed negative correlations with it. One explanation for this is, that older people likely experience more medical issues than younger people. Another is, that they might be more motivated to contribute data about the medical use of certain substances (rather than writing entertaining stories of recreational use).

Third, we find that *younger people report more about negative experiences*. The two younger groups showed positive correlations with ‘Bad Trips’ and ‘Difficult Experiences’, and people below 18 also had more reports in the category ‘Train Wrecks/Trip Disasters’ than expected under an independence null hypothesis. There are several potential explanations for this: First, younger people may have less experience with taking psychoactive drugs than older drug consumers. Therefore, they probably know less about their own limits, may take higher doses than appropriate, have a lower tolerance to drugs, take these substances in less ideal settings, and are more likely to be overwhelmed by the pharmacological effects of drugs. In addition, younger people may have different motivations for (reporting) drug consumption than older people, and therefore choose a different kind of context, substance, and dosage, which may lead to a higher chance of having a negative experience.

Furthermore, there was a significant association between gender and category ($\chi^2(15, N = 59,109) = 794, p < 0.01$). We find that *females report more about using drugs for medical purposes, while at the same time report more about health problems and addiction in relation to drugs*. One possible explanation for this is that females may focus on the harm reduction approach and submit reports to warn others about the addictive potential and health consequences of (medical) substances. However, the percentage of females is lower in our data than in other studies; therefore, it is also possible that only a certain kind of female drug consumer, namely women interested in health related issues, report to Erowid. Another explanation could be that males and females might differ in their motivations for drug consumption, such that women more often take drugs for medical reasons or because of drug dependence. This is in line with existing research suggesting females report using psychoactive drugs to help with anxiety or to feel better more often than males (Kettner, Mason, and Kuypers 2019).

In summary, our data reveals gender and age play a significant role in the motivation for and interpretation of drug consumption. While younger people report more about negative drug experiences, older people and females report more about health related aspects like medical use and addiction.

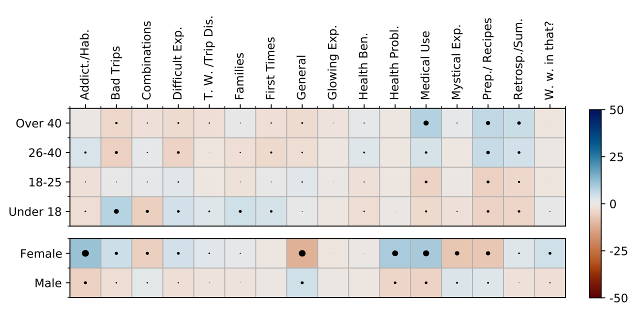

Figure 6: Chi-square distances of age group-category pairs and gender-category pairs. Legend and circle dimensions correspond to figure 5.

6 Predictability of Drug Experience Outcome

Drug use is a complex interplay of pharmacological and psychological processes. The substances consumed as well as the dosage, the consumers psychological and physical status, the location and many more factors may affect the outcome of a drug experience. While consumers may take precautions to increase their probability of having a positive drug experience and avoid having a negative experience or becoming addicted, their success seems to be unpredictable.

Surprisingly, we found regularities with all these different drug experiences: For example, LSD seems to be comparatively highly associated with negative drug experiences, while MDMA seems to be comparatively highly associated with positive ones. Furthermore, there are many variables, like dosages or settings, which we have not analyzed yet and which may reveal further patterns. Are these regularities strong enough to predict the outcome of a drug experience?

In this section we use information about the substances a user has consumed, the dosage amount, user demographics, context, and location during the experience to classify whether or not an experience will be 1) ‘glowing’ (=positive), 2) ‘difficult’ (=negative) and 3) related to addiction. The goal here is not ‘prediction’ per se, but ‘predictability’ (Martin et al. 2016). In other words, we want to use predictive performance as a goodness-of-fit measure to see the maximum amount of variance an explanatory model might aspire to explain. This helps to gain deeper insights into the level of complexity drug use encompasses: If the outcomes are predictable, this would show that there are specific patterns leading to a positive, negative or addiction experience and that users therefore can influence the outcome of their drug experience by making specific choices about the substance and setting.

6.1 Models

Labels. We created three classification tasks: For the first task, we chose the variable ‘Glowing Experience’ (0/1) as label. We treated this label as a marker for whether a report was about a positive experience, as this category had the highest average sentiment score. For the second task, we chose ‘Difficult Experiences’ (0/1) as label. We treated this label as a marker for whether a report was about a rather negative experience, as this was one of the four most negative categories as evaluated by the sentiment analysis and had more than twice as many reports than the other three most negative categories. For the third task, we chose ‘Addiction/Habituation’ (0/1) as label. While this category is not the same as a clinical diagnosis of addiction, exploring this label provides valuable information about experiences that domain experts classify as being about dependency issues.

Data. For each classification task, we excluded all reports which were not about at least one of the top ten substances or which had no consumer weight or gender information included. For the experience tasks, we excluded all reports which were about one of the top ten drugs, but had no information about the dosage of this drug. This drastically decreased the dataset. Table 4 shows the dataset sizes for all models. The datasets were moderately unbalanced for the experience tasks (1=19%,0=81%) and highly unbalanced for the addiction model (1=5%, 0=95%).

Features and Feature Selection. Table 4 shows the feature sets we used for each model. The construction of each feature is described in section 3.3. As described in section 5.1, reports about addiction were often written as summaries of a longer time period and therefore included no dosage information. Consequently, we did not include dosage amount as a feature for the addiction model.

From these sets of features, we chose all variables which had the strongest relationship with the target variable using Lasso regression. For each model, we split the dataset into a training set and a test set with a ratio of 4:1. We then assessed the optimal regularization parameter α for the lasso regression using Pythons LassoCV[14] on the training set. We chose 10-fold cross validation with α values from 0.001 to 10 with a 0.01 step size. We then included Lasso regression with the particular alpha as a feature selection part in a scikit pipeline.[14] The selected alpha values are shown in table 4; the resulting features for each model can be found in section D in the online appendix.[11]

Classifiers. We used six different linear and nonlinear classifiers: A random forest; logistic regression, with 'liblinear' as a solver; Linear Discriminant Analysis; k-nearest-neighbors; scikit's decision tree, an optimized classification and regression tree (CART) algorithm; and a Gaussian Naive Bayes.[14] Each model was included as a classifier in a scikitlearn pipeline as second element after the feature selection part, and evaluated using the standard parameters given in scikit-learn (version 0.23.2).[14]

6.2 Model Evaluation

We compared each model to the majority vote model. As the datasets were highly imbalanced, accuracy alone would be not informative for assessing the models performance. Therefore, we used the Geometric Mean (GMean), which is defined as $\sqrt{sensitvity * specitivity}$. The GMean is low, when either the prediction performance for the majority class or the prediction performance for the minority class (or both) is low. We were searching for a model, which would at least preserve the accuracy of the majority vote model, but had a better performance regarding the GMean.

6.3 Results

Table 5 presents the GMean, as well as the accuracy and sensitivity for each classification task and model. We find that given our data and models *the outcome of a drug experience is not predictable*. For each classification task, the best performing model reached a GMean less than 0.5. These low GMean values were caused by a low sensitivity: Only 10% of reports, which were about a 'glowing' experience, could be detect as such by our model. The same held for difficult experiences (sensitivity of 0.4%) and addiction experiences (sensitivity of 14%).

The results indicate that the relationship between drug consumption and subjective outcome is highly complex and has a low degree of predictability. Hence consumers cannot simply choose a specific drug, dosage and setting, in order to ensure a positive experience and to prevent a difficult or addiction experience. These findings emphasize the danger in drug consumption: We have seen that specific drug experiences are associated with the drugs themselves, age groups and gender. Consumers can therefore assess their vulnerability and even influence their probability of having a specific experience by choosing a particular drug. But in the end, the outcome of drug consumption is still somewhat unpredictable and therefore remains a risk.

[14] https://scikit-learn.org/stable/

7 Discussion and Conclusion

The Erowid Experience Report collection gave us the unique possibility to get a consumer perspective on drug consumption. We have shown that this data can reveal valuable information about the relationship between drug consumption variables and the characteristics of a drug experience.

Our research shed light on the subjective evaluation of drug experiences. We found that negative drug experiences are more prevalent for younger people and LSD users, while positive experiences occur more often with MDMA or first time consumption. Moreover, we gained deeper insights into health consequences of drug use: We found that reported health problems and addiction are linked to specific substances and that females report more often about these topics. Finally, our research highlighted the risk of drug consumption: Even when consumers could control the substance, dosage and situational factors, it is unpredictable, whether their experience will be joyful, difficult or associated with addiction.

7.1 Limitations

Although subjective experience reports can reveal fascinating patterns, they should be analyzed with care. First, we do not know how the sample of drug consumers, whose experiences are published on Erowid, compares to larger populations of drug consumers. As we have shown, a higher percentage of Erowid contributors are male and focus more on psychedelic drugs than drug consumers identified by national studies. They are also likely more engaged in systematic exploration with and reflection on psychoactive substances. In addition, there might be a selection bias in the data, e.g. as reviewers might favour reports about unusual substances, and/or a reporting bias, e.g. as contributors may only write about topics, substances and effects which they define as interesting.

Second, our synthesized metadata may contain errors, especially the drug dosages. Standardizing drug dosages is a challenging task which requires extensive domain knowledge. Even if users reported dosages in standardized ways, it remains unclear whether this information is correct, as users are not always aware of the specific content and composition of the substance consumed. For example, Vrolijk et. al (2017) compared user-generated online information on ecstasy tablets to information from the validated Dutch Drugs Information and Monitoring System (DIMS) Database, and found that users tend to overestimate MDMA concentration

Label	Glowing or Not	Difficult or Not	Addiction or Not
Dataset	7,414 Reports (1=19%,0=81%)	7,414 Reports (1=19%, 0=81%)	14,113 Reports (1=5%, 0=95%)
Features	- Drugs & drug combinations (142) - Context/location (9) - Author gender (1) - Author weight (1) - Drug dosages (10)	- Drugs & drug combinations (142) - Context/location (9) - Author gender (1) - Author Weight (1) - Drug dosages (10)	- Drugs & drug combinations (142) - Context/location (9) - Author gender (1) - Author Weight (1)
Feat. Sel. Model	Lasso (α=0.001)	Lasso (α=0.991)	Lasso (α=0.001)

Table 4: Size of dataset, with balance of positive and negative reports in parentheses, features and feature selection model for the three classification tasks.

Model	Majority Vote	Random Forest	Logistic Regression	Linear Disc. Analysis	K-Neighbors Classifier	Decision Tree	Gaussian NB
Glowing or Not	0.0 (0.81, 0.0)	0.38 (0.76, 0.156)	0.13 (0.81, 0.017)	**0.31** **(0.81, 0.101)**	0.28 (0.76, 0.087)	0.38 (0.74, 0.16)	0.38 (0.31, 0.885)
Difficult or Not	0.0 (0.81, 0.0)	**0.06** **(0.81, 0.004)**	0.0 (0.81, 0.0)	**0.06** **(0.81, 0.004)**	0.0 (0.81, 0.0)	0.0 (0.81, 0.0)	0.18 (0.8, 0.033)
Addiction or Not	0.0 (0.95, 0.0)	0.41 (0.94, 0.174)	0.19 (0.95, 0.038)	0.7 (0.93, 0.515)	0.32 (0.93, 0.106)	**0.37** **(0.95, 0.136)**	0.7 (0.55, 0.917)

Table 5: GMean for each classification task and model, with accuracy and sensitivity in parentheses. The best performing model (highest GMean with same accuracy as majority vote model) for each classification task is bolded.

and, in 15.3% of cases, provided dangerously wrong information. The existence of the 'What was in that?' category confirms information gaps among Erowid users as well.

Third, the prediction results are bound to the data, classifiers and parameters used. Although we tried a wide range of models, further research may find others which reveal better results and may show other factors, with which consumers could control the outcome of their drug experience.

Fourth, it should be emphasized that Erowid has collected Experience Reports for more than 25 years and even extracted some older reports from books and journals, dating back as early as 1848. Over this time span the substance availability, drug composition and drug consumption patterns have certainly changed, which may impact our results. Further research could study temporal trends on Erowid to gain more information on the history of drug consumption.

7.2 Implications and Future Work

The increasing complexity of the world's drug landscape has brought new challenges for drug consumers, medical personnel, social workers, institutions and researchers (Arillotta et al. 2020; D'Agnone 2015; Schifano 2020). Quantitatively analyzing experience reports can help to shed light on drug-category associations, which need to be more in the focus of research. In our study for example, we found a positive association between 'Addiction/Habituation' and dextromethorphan (DXM), a substance which is legally available over-the-counter in the United States.[15] While the withdrawal symptoms of DXM are largely unknown,[15] there have been sporadic clinical reports about patients suffering from DXM dependence (Miller 2005; Mutschler et al. 2010). Further research could investigate more deeply the relationships we have found, for example by manually inspecting the reports and conducting psychopharmacological studies. Moreover, the chi-square approach we used in this paper can be applied to other substances and topics, and help to generate new hypotheses for research.

In addition, studying Erowid Experience Reports can help to obtain a deeper sociological understanding of drug use. Our results indicate age and gender play a significant role in the motivation for and interpretation of drug consumption. This should allow further research to investigate whether there are also demographic differences in the choices drug consumers make, for example regarding the strength of drug dosages or the setting for drug consumption.

Finally, Erowid's labeling and categorization process allows the analysis of a great variety of topics, which were beyond the scope of this paper. Many of the categories described, such as 'Families' or 'Mystical Experiences', show a great potential for further research, as they can provide new insights into how people experience drug consumption.

[15]https://www.drugabuse.gov/drug-topics/commonly-used-drugs-charts

References

Andersson, M.; and Kjellgren, A. 2017. The slippery slope of flubromazolam: Experiences of a novel psychoactive benzodiazepine as discussed on a Swedish online forum. *Nordic Studies on Alcohol and Drugs*, 34(3): 217–229.

Arillotta, D.; Schifano, F.; Napoletano, F.; Zangani, C.; Gilgar, L.; Guirguis, A.; Corkery, J. M.; Aguglia, E.; and Vento, A. 2020. Novel opioids: Systematic web crawling within the e-psychonauts' scenario. *Frontiers in Neuroscience*, 14: 149.

Baggott, M.; Coyle, J.; Erowid, E.; Erowid, F.; and Robertson, L. 2011. Abnormal visual experiences in individuals with histories of hallucinogen use: A web-based questionnaire. *Drug and Alcohol Dependence*, 114(1): 61–67.

Baggott, M. J.; Erowid, E.; Erowid, F.; Galloway, G. P.; and Mendelson, J. 2010. Use patterns and self-reported effects of Salvia divinorum: An internet-based survey. *Drug and Alcohol Dependence*, 111(3): 250–256.

Berning, M.; and Hardon, A. 2016. Educated guesses and other ways to address the pharmacological uncertainty of designer drugs: An exploratory study of experimentation through an online drug forum. *Contemporary Drug Problems*, 43(3): 277–292.

Bilgrei, O. R. 2019. Community-consumerism: Negotiating risk in online drug communities. *Sociology of Health & Illness*, 41(5): 852–866.

Bonson, K. R.; Buckholtz, J. W.; and Murphy, D. L. 1996. Chronic administration of serotonergic antidepressants attenuates the subjective effects of LSD in Humans. *Neuropsychopharmacology*, 14(6): 425–436.

Chary, M.; Yi, D.; and Manini, A. F. 2018. Candyflipping and other combinations: Identifying drug-drug combinations from an online forum. *Frontiers in Psychiatry*, 9: 135.

D'Agnone, O. 2015. What have we learned and what can we do about NPS? *Drugs and Alcohol Today*, 15(1): 28–37.

Duxbury, S. W. 2018. Information creation on online drug forums: How drug use becomes moral on the margins of science. *Current Sociology*, 66(3): 431–448.

Erowid, E.; and Erowid, F. 2006. The Value of Experience: Erowid's Collection of First-Person Psychoactive Reports. *Erowid Extracts*, 10: 14–19.

Erowid, F.; Thyssen, S.; and Erowid, E. 2018. Mountains of Experience: The process of reviewing 112,000 first-person reports about the use of psychoactives. *Erowid Extracts*, 30: 6–7.

European Monitoring Centre for Drugs and Drug Addiction. 2020. *European Drug Report 2020: Trends and Developments*. Publications Office of the European Union, Luxembourg.

Gamma, A.; Jerome, L.; Liechti, M. E.; and Sumnall, H. R. 2005. Is ecstasy perceived to be safe? A critical survey. *Drug and Alcohol Dependence*, 77(2): 185–193.

Hutto, C.; and Gilbert, E. 2014. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*.

Jarnow, J. 2016. Why Psychedelic History Matters. *VolteFace*.

Karila, L.; Billieux, J.; Benyamina, A.; Lancon, C.; and Cottencin, O. 2016. The effects and risks associated to mephedrone and methylone in humans: A review of the preliminary evidences. *Brain Research Bulletin*, 126, Part 1: 61–67.

Kettner, H.; Mason, N. L.; and Kuypers, K. P. C. 2019. Motives for classical and novel psychoactive substances use in psychedelic polydrug Users. *Contemporary Drug Problems*, 46(3): 304–320.

Krieg, L. J.; Berning, M.; Colombo, G.; Azzi, M.; and Hardon, A. 2016. Visualising Erowid: A data driven anthropological research on drugs. Chemical Youth, University of Amsterdam.

Krieg, L. J.; Berning, M.; and Hardon, A. 2017. Anthropology with algorithms? *Medicine Anthropology Theory*, 4(3).

Marlatt, G. A. 1996. Harm reduction: Come as you are. *Addictive Behaviors*, 21(6): 779–788.

Martin, T.; Hofman, J. M.; Sharma, A.; Anderson, A.; and Watts, D. J. 2016. Exploring limits to prediction in complex social systems. In *Proceedings of the 25th International Conference on World Wide Web*, 683–694.

Miller, S. 2005. Dextromethorphan psychosis, dependence and physical withdrawal. *Addiction Biology*, 10(4): 325–327.

Mutschler, J.; Koopmann, A.; Grosshans, M.; Hermann, D.; Mann, K.; and Kiefer, F. 2010. Dextromethorphan withdrawal and dependence syndrome. *Deutsches Arzteblatt international*, 107(30): 537–540.

Newman, M.; Denton, G.; Walker, T.; and Grewal, J. 2016. The experience of using synthetic cannabinoids: A qualitative analysis of online user self-reports. *European Psychiatry*, 33: S309–S310.

Norman, J.; Grace, S.; and Lloyd, C. 2014. Legal high groups on the internet — The creation of new organized deviant groups? *Drugs: Education, Prevention and Policy*, 21(1): 14–23.

Pal, R.; Balt, S.; Erowid, E.; Erowid, F.; Baggott, M.; Mendelson, J.; and Galloway, G. 2013. Ketamine is associated with lower urinary tract signs and symptoms. *Drug and Alcohol Dependence*, 132(1–2): 189–194.

Paul, M. J.; Chisolm, M. S.; Johnson, M. W.; Vandrey, R. G.; and Dredze, M. 2016. Assessing the validity of online drug forums as a source for estimating demographic and temporal trends in drug use. *Journal of Addiction Medicine*, 10(5): 324–330.

Rhumorbarbe, D.; Morelato, M.; Staehli, L.; Roux, C.; Jaquet-Chiffelle, D.-O.; Rossy, Q.; and Esseiva, P. 2019. Monitoring new psychoactive substances: Exploring the contribution of an online discussion forum. *Internat. Journal of Drug Policy*, 73: 273–280.

Rolando, S.; and Beccaria, F. 2019. "The junkie abuses, the psychonaut learns": A qualitative analysis of an online drug forum community. *Drugs and Alcohol Today*, 19(4): 282–294.

RTI International. 2019. *2018 National Survey on Drug Use and Health Public Use File Codebook*. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration.

Ruths, D.; and Pfeffer, J. 2014. Social Media for Large Studies of Behavior. *Science*, 346(6213): 1063–1064.

Schifano, F. 2020. Analyzing the open/deep web to better understand the new/novel psychoactive substances. *Brain Sciences*, 10(3): 146.

Soussan, C.; and Kjellgren, A. 2014. Harm reduction and knowledge exchange—a qualitative analysis of drug-related Internet discussion forums. *Harm Reduction Journal*, 11(1): 25.

Stanciu, C. N.; Penders, T. M.; and Rouse, E. M. 2016. Recreational use of dextromethorphan, "Robotripping"—A brief review. *The American Journal on Addictions*, 25(5): 374–377.

Stone, K.; and Shirley-Beavan, S. 2018. The Global State of Harm Reduction 2018. Technical report, Harm Reduction International, London.

Substance Abuse and Mental Health Services Administration. 2019. *Key Substance Use and Mental Health Indicators in the United States*. HS Publ. No. PEP19-5068, NSDUH Series H-54.

Swogger, M. T.; Hart, E.; Erowid, F.; Erowid, E.; Trabold, N.; Yee, K.; Parkhurst, K. A.; Priddy, B. M.; and Walsh, Z. 2015. Experiences of Kratom Users: A Qualitative Analysis. *Journal of Psychoactive Drugs*, 47(5): 360–367.

United Nations. 2018. *World Drug Report 2018*. United Nations publication, Sales No. E.18.XI.9.

United Nations. 2020. *World Drug Report 2020*. United Nations publication, Sales No. E.20.XI.6.

Vrolijk, R. Q.; Brunt, T. M.; Vreeker, A.; and Niesink, R. J. M. 2017. Is online information on ecstasy tablet content safe? *Addiction*, 112(1): 94–100.

Witt, E. 2015. The trip planners: The unusual couple behind an online encyclopedia of psychoactive substances. *The New Yorker*,, November 16, 2015.

5 Discussion

In the preceding chapter, we have presented three CSS studies, which were conducted as collaborative projects between computer scientists and researchers or experts from the social sciences. In the following, we report on our experiences with these inter- and trans-disciplinary collaborations. In particular, we describe one major opportunity per study and show how computational methods allowed us to advance knowledge of social science research questions. In addition, we elaborate on one challenge per study and we explain how interdisciplinary collaboration has helped us in overcoming this challenge. Table 4 gives an overview of the opportunities, challenges and value of collaboration per study, with reference to the sections in which each opportunity and challenge has been introduced. After highlighting the value of interdisciplinary collaboration, we discuss possible difficulties in collaborative projects and suggest ground rules to enable the success of these projects.

Study	Type	Example
Study 1	Opportunity	Analysis of large scale data (3.1.2)
	Challenge	Generating impact for social science (3.2.4)
	Collaboration	Connect to social science theories
		Translate between computational to social science
		Communicate results in the language of the audience
Study 2	Opportunity	Reduction of research cost, time, obtrusiveness (3.1.3)
	Challenge	Overtrust in computational methods (3.2.5)
	Collaboration	Critically reflect on algorithms
Study 3	Opportunity	Access to new data sources (3.1.1)
	Challenge	Lack of ethical standards (3.2.2)
	Collaboration	Detect ethical pitfalls
		Protect community interests

Table 4: Overview of the opportunities, challenges and value of collaboration presented by the studies in Section 4.

5.1 Summary and Contributions

5.1.1 Study 1: Creating Impact for Social Science Research

Opportunity. The study presented in section 4.1 exemplifies how Computational Social Science helps to evaluate data on an unprecedented scale (as described in 3.1.2) and better understand the dynamics of large social systems. Traditional social science studies on visual climate change communication have mainly focused on manually analysing images (Dahl, 2017; DiFrancesco & Young, 2011). These methods have worked well for studying traditional news media, because comparatively few images on climate change have been published in newspapers and magazines. For social media platforms, however, on which not only journalists, but all kinds of users contribute images to the climate change discussion, a manual content analysis of images published within a decent time frame is almost unfeasible to conduct.

By applying methods from the computational sciences, namely an algorithm to collect Twitter data, a neural network to extract image properties and a k-means clustering approach to group similar images, we could create a large scale dataset of all #climatechange related images on Twitter, and automatically categorize these images to analyse the distribution, publishing patterns as well as stakeholder and user preferences for image groups. In particular, these computational approaches did not only allow us to study the whole population of Twitter users contributing to the discussion on #climatechange, but also analyse the whole timeframe of visual climate change communication on Twitter, starting with the introduction of the image feature in 2011. Therefore, CSS has helped us to characterize the visual climate change discourse on Twitter and obtain insights into the communication dynamics happening on social media, which would not have been possible with traditional social science approaches. As we have shown with this study, a great potential of CSS lies in the collection and analysis of big data, as it enables researchers to study large and complex social systems.

Challenge. One common challenge within CSS, which could be exemplified by this study, is designing a study, which has an impact on the social science discipline in which the research problem is located. As described in section 3.2.4, studies approaching social science problems with the aid of computational methods sometimes fail to integrate their research into the social science discipline as they do not sufficiently engage with the theories and language of the field.

To make sure to create impact on the communication sciences with our work, we had to overcome several barriers. First, we had to base our study on the existing theories on communication on social media and climate change. Identifying the relevant literature, understanding the concepts related to our work, determining gaps within the body of research, identifying how we could fill these gaps and contribute to this research, and developing a theory well embedded in the body of literature required extensive knowledge in the field of communication science.

Second, we had to make our findings comparable to prior work on visual climate change communication on traditional and social media. However, translating between traditional social science studies and computational social science studies has proven incredibly difficult. Both fields employed very different methodologies, leading to differences of how and in what context findings could be interpreted. For example, many of the existing studies on visual climate change communication involved qualitative content analyses of images, and thus the image categories within these studies were created based on semantic meaning. In contrast to that, our study was based on a quantitative approach, and thus we had to make sure that the image categories were visually different from each other to make them identifiable by an automated system. Therefore, we could not simply adopt image categories proposed by prior work in the communication sciences, to allow for easy comparisons of the distribution of categories across studies, but we needed a social and a computer science perspectives to create meaningful results for the social sciences, while being bounded to the restrictions of the computational approach.

Third, we had to present our findings in a way that they would be easily accessible and understandable by the community of researchers working on climate change communication in traditional and social media. Therefore, we had to adapt our terminology to the communication science jargon, identify publication outlets consumed by this community, create figures which would be easily understandable with less technical knowledge and put a special focus on well explaining the methodologies used with their impacts and limitations. To fulfil these goals, both technical knowledge as well as experience with the communication science concepts and jargon was required.

Collaboration. In the study presented, an interdisciplinary collaboration between communication and computational sciences was the key factor to overcome these challenges. First, the social science perspective helped us to obtain a deep understanding of the basic concepts of communication sciences, the theories in the field and the research problem, while the computer science perspective provided an extensive overview of the technical possibilities to tackle this problem, as well as a critical reflection of these technologies. In joint discussions, we could gain insights into each others disciplines, create a shared knowledge of the research problem, develop a theory well embedded in the body of literature, and define an approach which would make use of technological innovation, while creating findings valuable for the field.

The collaboration especially helped us in translating between both disciplines and finding a balance between what is computationally possible, and what is meaningful and valuable for the social science domain. For example, we created image categories in an iterative approach, by integrating each others perspective and opinions in every round of iteration. First, we started with a collection of categories identified in prior studies on climate change communication in traditional media. We then evaluated these categories from a technical perspective and split or aggregated categories based on what would be visually identifiable by an automated system. Moreover, we combined our experience of communication and social media to add new categories we expected to be prevalent on Twitter (in comparison to traditional media). We then performed several rounds of coding clusters based on these categories, while adding new categories as we came across substantively distinct images. This was followed by multiple discussions integrating feedback from both disciplines to split and aggregate categories. Although the technical approach restricted us in some ways, the interdisciplinary collaboration helped us tremendously in reaching a categorization scheme which would be computationally feasible and which we found to be well connected to prior work and meaningful from a communication science perspective.

Moreover, by combining our knowledge of communication and computer science, we could identify aspects of our study which could be difficult to understand for a communication science audience. By vividly discussing these aspects, visualizing findings and describing in our own words what we would interpret from such illustrations as well as explaining and asking questions about concepts and methodologies, we could approach these topics from both perspectives and formulate them in the language of the communication science field. Furthermore, with the experience of the communication science experts, we could choose and publish within a journal well known in the field, to make our work easily accessible for researchers studying climate change communication. In summary, the study presented in section 4.1 has shown that interdisciplinary collaboration in CSS is extremely valuable to bridge barriers between computational and social sciences. Particularly, it has proven useful to translate computer science methodologies into the domain of the research problem and to ensure that the research outcomes benefit knowledge creation within the social science community.

5.1.2 Study 2: Gaining a Critical Data Perspective

Opportunity. The study presented in section 4.2 demonstrates how computational methods can improve the ways in which we study human behaviour and cognitive processes (as described in section 3.1.3). In particular, by making use of technological innovation and sophisticated algorithms, we could assess the visual attention of teachers in a non-obtrusive way.

Assessing teachers' visual attention (defined in patterns of gaze locations over time) in a classroom setting with traditional social science methods is almost impossible. Due to conflicting mental processes and unconscious behaviour, study participants would not be able to report their visual attention while performing lectures. Consequently, all forms of subjective measurements would not be applicable to this research scenario. At the same time, due to the speed of eye movements, external human observers would be unable to precisely locate and record teachers' visual attention, and so most forms of objective measurement methods from the traditional social sciences would be unsuitable as well. While there exist several other objective measurements to study humans' visual attention, such as electrical or mechanical eye tracking methods making use of contact lenses, they are either unsuitable for a mobile context or rather invasive, inducing discomfort and even pain for participants.

Fortunately, the utilization of technological progress and computational methods for social science research has led to the development of mobile, video-based eye tracking systems, such as the one used in study 4.2. This system allowed us to collect and store teachers' gaze locations in a very precise way, while they moved within their natural work environment and performed common tasks within the classroom. After this, the application of multiple algorithms helped us to preprocess this data, calculate attention measurements, and evaluate attention differences between teacher novices and experts. Besides the advantages of mobility and unobtrusiveness, the collection and evaluation of such eyetracking data could be conducted by a small research team and was therefore comparably cost- and time-friendly. As we have exemplified with this study, computational technologies have opened completely new ways of getting insights into cognitive processes, and thus a great benefit of CSS lies in the introduction of new research methodologies for the social sciences, reducing research costs, time and obtrusiveness.

Challenge. The study has highlighted one of the main challenges in CSS, namely the misuse of and overtrust in computational methods (see also 3.2.5). On the one hand, the development of user-friendly eye tracking systems and analysis platforms has provided great opportunities for the social sciences in general, and the education science in specific. Educational scientists with less technical knowledge can now import, analyse and visualize eye tracking results in an easy way, without having to develop their own algorithms. On the other hand, the use of such systems still requires certain technical expertise and the unreflected application of algorithms and default parameters can quickly result in undesirable side effects.

In general, there seems to be a lack of critical reflection on the technological processes behind eye tracking. Researchers throughout most educational studies did not report details on the data analysis process, e.g. which thresholds were selected, and did not justify their decision for the use of certain algorithms and parameters. When reviewing the literature, one can almost obtain the impression that most scholars in this discipline were not aware of the options they could choose within the analysis process or did not understand the difference between certain algorithms and parameters, as well as the impact these could have on their findings. In contrast, there seemed to exist a general trust in the respective eye tracking system and the expectation that the analysis platform would by default deliver accurate measurements of attention.

Collaboration. The collaboration between computational and social scientists helped us not only to design a study which would use eye tracking to answer a social science question, but to critically reflect on the methods and algorithms used during this process.

In the beginning of our collaboration, we put much emphasis on understanding each other's disciplines and the ways in which we would normally do research. After one of the educational scientists provided background knowledge on the domain of eye tracking studies in education and the research problem, including basic concepts, terminology and tools, we collaboratively went through the typical research process in this domain and critically discussed each step of this process. This helped us in creating a mutual understanding of the concepts, difficulties and potential in each other disciplines and reflecting our typical approaches.

Through these discussions, we could very soon identify topics which needed more attention from both a computational and social science point of view, such as the decision for certain parameters and algorithms for fixation calculation.

While expert knowledge in the social sciences helped us to evaluate the current use of eye tracking parameters and algorithms within the educational context, technical expertise helped us to calculate fixations and compare the effects of different thresholds. By this we could show that algorithms and parameters can influence eye tracking results. Based on these findings, we combined our knowledge to assess how this would impact the interpretation, both on a technical and social level, and propose suggestions on the future use of eye tracking in the context of education research. In conclusion, the study presented in section 4.2 emphasizes the importance of interdisciplinary collaboration, as such collaboration helps to detect possible pitfalls in the application of technological approaches and define best practices on the use of CSS methods.

5.1.3 Study 3: Overcoming Ethical Challenges

Opportunity. The study presented in Section 4.3 has exemplified how computational methods can help to expand knowledge in social sciences by providing access to new data sources (as described in 3.1.1). In particular, by applying the computational data collection method "webscraping" to a platform of interest for sociology and psychology, we could gain insights into the perspectives of a hard-to-reach population and investigate new variables, i.e. substances. First, in most countries, the consumption of drugs is illegal or at least considered immoral for many substances. Therefore, it is usually very difficult for researchers to get in contact with people who consume drugs, as they may fear being criminalized or stigmatized. In the past, researchers have put tremendous effort in creating data sources to study drug consumption: They analysed the compounds in waster-water to identify drug consumption patterns in local areas (Zuccato, Chiabrando, Castiglioni, Bagnati, & Fanelli, 2008); they observed the amount of drug packages confiscated at customs and border forces to estimate the amount of substances currently available in the country (European Monitoring Centre for Drugs and Drug Addiction, 2019b, p. 13); they performed pharmacological analyses with substances voluntarily submitted to centres, in which consumers can get their drugs tested, to obtain insights into the compounds of substances currently in circulation (Brunt & Niesink, 2011); they used statics about drug-related clinical incidents, treatments and deaths to detect possibly dangerous trends in drug consumption (European Monitoring Centre for Drugs and Drug Addiction, 2020, p. 59, 65; European Monitoring Centre for Drugs and Drug Addiction, 2019a, p.9); and they conducted surveys assessing the volume of drug consumption within the general public, which very likely impose underreporting bias (European Monitoring Centre for Drugs and Drug Addiction, 2019a, p. 6). Most of these approaches do not take the perspective of the consumer themselves into account and the few which do, create problems with finding participants who are willing to talk about their consumption.

By collecting data from the Erowid website, we could consider the perspectives of people who voluntarily reported about their experiences with drugs and who had an intrinsic motivation to participate in building knowledge about the topic. For example, the authors shared their reasons for taking drugs in general or choosing specific substances or a specific setting. In addition, they described their thoughts and emotions during the consumption process and reported on the effects they experienced as positive and negative. Although the data is certainly not representative of all drug consumers worldwide, it gave insights into the thoughts, feelings and behaviour of a population, hardly any other data source in this area can provide.

Moreover, by collecting the Erowid dataset we could provide information about substances, which may not be in the focus of current drug consumption research. The incredibly wide range of different substances available worldwide makes it hard to keep updated on the current trends in drug consumption (Rhumorbarbe et al., 2019). As there is little data available on substances abroad the most commonly used drugs, many studies and surveys concentrate only on analysing the most popular substances, such as cannabis or MDMA. This can result in specific substances being understudied, and dangerous trends being undetected. The Erowid experience vault includes reports on more than 800 different substances, which provides the opportunity to also investigate less commonly used drugs. By collecting and analysing these reports, we could show, for example, a positive association between addiction and dextromethorphan, a substance which is mentioned only in few clinical reports on withdrawal symptoms. As this paper has shown, CSS has the potential to open up new research data for social science fields, in regards to prior hard-to-access populations and variables.

Challenge. As described in section 3.2.2, the field of CSS currently lacks ethical and privacy standards. The study presented in section 4.3 shows the multitude of ethical considerations which need to be taken within a CSS project, and highlights the challenge of performing CSS research in a way which does not create harm for the population under study or society in general.

First, like in many studies evaluating online data, one aspect we had to consider was gaining informed consent from the users whose reports we analysed. Second, as we were studying behaviour which is illegal or considered immoral in many countries, we had to make sure not to take a position in which we would promote such behaviour. At the same time, since we were working with a population which has been stigmatized because of such behaviour, we had to make sure that our study would not lead to further stigmatization of the population in general and individuals in specific, and would not impose any other disadvantages, like legal measures, on the study participants.

Collaboration. The multidisciplinary collaboration with drug community experts from the Erowid organization helped us to detect ethical pitfalls and to design the study in a way that the population under study and its interests were protected.

The first ethical challenge consisted of gathering the informed consent of the people whose reports we analysed. It has to be noted that the website Erowid.org is not a communication platform, on which users can interact with each other, but a platform on which users actively submit reports. These reports are then reviewed by the Erowid team and published if they meet certain quality standards. When submitting reports, users have to agree that these reports are processed for scientific research and publication. Thus, we could assume that users on Erowid.org, in contrast to many other social media platforms, share data with the goal of contributing to knowledge in the field. Consequently, the ethical challenge shifted from gaining consent to use reports for research purposes (which was fulfilled by the design of the submission process), to making sure to use the data for research which also benefits the community. By collaborating with experts from the Erowid organization, we could obtain insights into the research questions and topics discussed within the community. In addition, by including their feedback on our research goals, and later on findings, we could ensure that community interests were fulfilled by our study.

Furthermore, we had to make sure to study the population of drug consumers without either condoning or stigmatizing their behaviour, as well as without presenting results in a way that could lead to disadvantages for study participants. By combining our technical knowledge with steps taken by the Erowid reviewer team, we could work with anonymized data and make sure to only present aggregated results in the publication, which would prevent deanonymization of individuals by third parties. Additionally, by following the harm-reduction framework proposed by the social sciences, we could analyse the drug consumption behaviour in a judgmental-free way. Furthermore, the Erowid team helped us to identify language which would be perceived as stigmatizing by the community and replace it with alternative expressions or further definitions. For example, while the term 'user' is applied often in social media research, as it relates to a person using a specific platform or medium, it might be considered inappropriate or derogatory in the drug community context, as it may also describe people 'using' drugs. We therefore included a definition of 'user' in the publication to ensure that people submitting drug experience reports would be characterized in a neutral way. In summary, the study presented in section 4.3 highlights the great value of collaboration between computer and social scientists/experts to combat the ethical challenges often prevalent in CSS projects, and perform research in a responsible way.

5.2 Towards Successful Interdisciplinary Collaborations

5.2.1 Difficulties in Interdisciplinary Projects

While inter- and multidisciplinary collaboration can help to encounter the challenges of CSS and make use of its full potential, the difficulties of collaborative research should not be underestimated.

Language Barriers The parties involved in a collaborative project are most likely accustomed to the unique language of their discipline. Therefore, they might use different jargon to describe the same things, or the same terminology to describe different things. Furthermore, some studies, concepts, standards and phenomena might be well known in one discipline and not at all in another. This is in particular troublesome, when certain information belongs to the unknown known, so individuals might not be aware that this information is part of their background knowledge, that their current actions and decisions are influenced by this information and that other team members might not have this information. Consequently, the team members might not see the need to explain specific vocabulary and concepts or might not be sure how to interpret the statements of the other parties and misinterpret terms, and thus might have troubles understanding each other (Sahneh et al., 2021; Cairns et al., 2020; Parti & Szigeti, 2021). These language barriers can have tremendous consequences for collaborative projects. For example, in interviews with social and computer/data scientists about their experience with collaborations between these fields, one participant reported:

> There are difficulties along the way because these are two very separate disciplines. And there has to be a translator between them to facilitate that communication . . . but because they don't understand each other, the collaboration terminates itself at some point. (Parti & Szigeti, 2021, p. 8)

Conflicting Goals In interdisciplinary collaborations, the parties involved might have very different research goals. While computer scientists typically have a strong focus on the technical process and aim at developing innovative methods and solutions, social scientists are usually more interested in the findings of this process, and their value to solve a social research problem (Sahneh et al., 2021). These goals might lead to different expectations on how each party can contribute to the project, how much time and effort should be invested in certain steps of the research process, how much details should be reported on these steps when presenting the findings to a larger audience, and which journals or conferences should be chosen for publication (Cairns et al., 2020; Kirkwood et al., 2018). If these expectations are not integrated and contributors are not given the opportunity to pursue their own research goals, feelings of frustration may arise. For example, several computer scientists in collaborative projects reported the impression that they only fulfill a service role to provide results, without their methodological innovations being appreciated (Sahneh et al., 2021). Likewise, some social scientists gained the impression of only being invited to collaborate because of funding requirements, without being valued for what they can contribute to the research process (Barthel & Seidl, 2017; Viseu, 2015). The main problem here seems to be that the partners are not fully included in the project and not allowed to contribute their perspectives. In interdisciplinary collaborations, there often exists unspoken hierarchies, which define whose knowledge is valued higher and whose opinions count more than others (Cairns et al., 2020). Viseu (2015) describes this imbalance as follows:

> Integration is also deeply asymmetrical. The social sciences (often a single social scientist) are typically brought in after the project

> has taken shape. This asymmetry is present in every aspect of integration — from power to personnel numbers, funding, knowledge production and, ultimately, independence — but remains hidden in mundane interactions that dictate what counts as a valid social-science activity and who gets to define it. (p. 291)

As a consequence, unmet expectations can result in tensions and open conflicts, as well as team members withdrawing from the project.

Lack of Time One of the most critical barriers in interdisciplinary collaboration projects is the factor of time (Cairns et al., 2020). In such projects, a lot of time and effort is needed to bridge cultural barriers, share ideas and opinions, look into each others perspectives, create a common understanding of the topic, solve conflicting goals and expectations, and explain domain-specific concepts, methodologies and terminology throughout the research process (Krause-Jüttler et al., 2022; Sahneh et al., 2021). While this is also important in disciplinary collaborations to some degree, the time and effort required for interdisciplinary projects is usually higher, as the collaboration partners have a higher cognitive distance between each other (van Rijnsoever & Hessels, 2011). Consequently, for every stage of the research process in which multiple collaboration partners are involved, an extra amount of time needs to be invested to allow for the integration of different perspectives. Otherwise, scholars may feel pressured, as reported by a collaboration partner within interdisciplinary sustainability research (Cairns et al., 2020):

> Actually for me that's a really exciting space to be in, developing methodologies and designing projects, but I think it did feel stressful and frustrating in the context I felt that we didn't really have the time for it. So that's what was frustrating. (p. 1716)

Due to conflicting priorities, researchers might not always have the time to allow for learning and knowledge building across disciplines and exploring each others perspectives. As a result, they may try to silently steer the project in the direction of their own research goals, while keeping interactions with the other parties involved on a shallow level. Such a case of interdisciplinary collaboration was reported by Verouden, van der Sanden, and Aarts (2016), and one of the project members interviewed explained the problem as follows:

> The difficult thing is that, without pointing fingers, our board differs from theirs in many ways ... they have entirely different disciplinary concerns and tasks, and this difference surfaced in the process. And yes, one can try to coordinate that ... But I also think that there are some real differences that are not easily bridgeable. Putting all this effort into trying is not worth the energy ... It seems better to just continue under our own flags. (p. 280)

This strategy, however, can ultimately cause conflicts within the research team (Verouden et al., 2016), and may lead to researchers feeling dissatisfied by the project outcome. Therefore, the time and effort needed for interdisciplinary collaboration should not be underestimated and may constitute a great barrier for many or most research projects.

Academic Gatekeeping Another difficulty for interdisciplinary research is the disciplinary organisation of universities, research institutes and publication outlets. In many universities, researchers are strongly integrated within their units, and have problems finding and selecting partners of other disciplines willing to collaborate (Lazer et al., 2020; Parti & Szigeti, 2021; van Atteveldt & Peng, 2018). Moreover, department focused budgeting as well as administrative concerns hinder close collaboration (Lazer et al., 2020; Krause-Jüttler et al., 2022; Cairns et al., 2020). In addition to that, interdisciplinary work is sometimes more difficult to publish, as publication outlets tend to reject work with approaches which are unknown or unconventional within the discipline (Parti & Szigeti, 2021). For example, one scientist interviewed in a survey on interdisciplinary collaboration between data and social scientists described his experience with such projects as follows:

> A transdisciplinary approach is one of those things that everyone endorses, and almost no one does. . . . Universities remain set up in disciplines [. . .]. Journals remain set up in disciplines . . . you know, the Journal of Psychology, the Journal of Geography and research councils . . . as set up in disciplines and we've applied for grants, and we've been told it's not social science enough. (Parti & Szigeti, 2021, p. 11)

Even more, researchers have few incentives to cooperate with scholars of other fields (Lazer et al., 2020). As van Rijnsoever and Hessels (2011) have shown in their study on research collaborations, while disciplinary collaborations are rewarding for researchers' careers, this is not the case for interdisciplinary work. Part of the problem may be that often fundings are allocated within disciplines and performance evaluations at universities underappreciate interdisciplinary efforts (Lazer et al., 2020; van Rijnsoever & Hessels, 2011). In summary, like Lazer et al. (2020) criticize in their article on CSS, "Collaboration is often not encouraged, and too often is discouraged" (p. 1060).

5.2.2 Guidelines for Interdisciplinary Collaboration

To increase the effectiveness of interdisciplinary collaborations and resolve the difficulties which may arise in such projects, we suggest observing the following practices.

Planning for Impact. When starting an interdisciplinary collaboration, it should be made sure that all partners can benefit from the collaboration and that the results are fruitful for both sides. Often, a large motivation for computer scientists to take part in a CSS project is to see their technical innovation applied in a real world setting and to know that this method will solve a social research problem. Adding up to this, the main motivation for many social scientists is to expand knowledge on crucial research questions within their field. Therefore, it has to be made sure that the study can be embedded in the theoretical context of the social science discipline and that the results are interpretable from a social perspective (Bravo & Farjam, 2017). Furthermore, the study, especially the methodologies and analysis results, have to be presented in such a way that they are comprehensible for other social scientists in the respective research community, and should be published in a way that the community is able to find them.

Nevertheless, only applying a common computational method on a social science problem might not be enough for a computational scientist to justify participation in a research project. As van Atteveldt and Peng (2018) state, "collaboration requires research that is innovative and challenging to both sides, and in many cases what we [social scientists] need is a good programmer to help us gather, clean, analyze, and visualize data rather than a computer scientist to invent a new algorithm" (p. 87). However, as the experience with the case studies described above has shown, even when it seems in the beginning like computer scientists are only needed to fulfil a certain data analysis task, they can contribute much more to the project than initially anticipated, when allowed to share their perspectives. For example, they may detect pitfalls and biases in the current application of methods in the research domain, or they may find patterns in the data structure, which require more innovative algorithms. Likewise, only including social scientists in the project for specific activities, such as creating a questionnaire or observing the adherence to ethical guidelines, does most probably not satisfy their expectations for a fruitful research project (Viseu, 2015; Barthel & Seidl, 2017). Instead, by fully integrating them in the research process, they can provide much more to the project, sometimes even in unexpected ways. For example, they may give insights into the language of the population under study, which could be a critical factor for collecting web data, or they can provide deep context to the research problem, which allows designing sophisticated theories. Therefore, we believe, when computer and social scientists truly collaborate, meaning that they work together in the whole research pipeline, instead of one fulfilling a service role for the other, both parties will be able to integrate their own perspectives, broaden their horizons, and create approaches and findings which are valuable for their own domain AND elevate the projects overall success.

To ensure that the research project creates value for all parties involved, some time should be invested in the beginning of the collaboration to plan the research process, integrate the individual aims and expectations, and create a clear common vision. This planning phase should include, for example, the following steps (Sahneh et al., 2021):

- Outline the overall research goal

- Allow each researcher to define their goals and expectations
- Determine outputs which benefit all parties involved
- Describe milestones and measurements of success
- Define roles, responsibilities and tasks within the project
- Plan time for exchange of knowledge, perspectives and feedback

Openness In some collaborative projects, team members follow their own research agenda and meet only to update each other on the progress they have made with their tasks. Such forms of research projects can work efficiently, but they neglect the main potential of collaboration, namely the exchange of perspectives, ideas and opinions to create a shared understanding of the research problem, which exceeds the sum of what all individual parties can contribute. To make sure that this potential is fully exploited, discussions between team members should be embraced rather than avoided. For example, one member of an interdisciplinary project in sustainability research described his experience with collaboration as follows:

> It feels like it's supposed to be slightly chaotic you know? ... You need people coming from different perspectives with strongly held views to challenge each other in order to bring out the most of it and carry on. (Cairns et al., 2020, p. 1717)

To ensure that all team members have the opportunity to learn from each other and challenge their perspectives, we suggest complying with the following rules during the research process (Sahneh et al., 2021):

- Allow time for discussing conflicting perspectives.
- Show respect for each others opinions and experiences.
- Take time to explain ideas in a comprehensible way. Adapt to the language of the recipient and explain domain terminology.
- Ask questions and repeat others' perspectives in own words to make sure that they were understood correctly.

Certainly, these rules will only work if the researchers are open to new perspectives and have the social skills to discuss and communicate ideas. Furthermore, they require a lot of time and effort from the team members. But this effort is rewarded with new learnings which help to raise the knowledge of the research problem to a new level. As an example, a computer scientist working on a project with social scientists described how beneficial this openness can be for advancing the research goals:

> I developed a framework that I thought I would own, and I only intended to offer others to apply it for their purposes. But suddenly it got taken over by others, requesting changes and interpreting things differently. Initially I said 'oh no – we cannot use it like this'; 'oh no – it's not meant like this'. I had to learn to let go and open up and be collaborative. [...] I found through this experience that the framework has much more potential, which I would have most likely not been aware of without this kind of interdisciplinary research work. (Siebert, Siebers, Vallejos, & Nilsson, 2020, p.714)

Involvement in All Steps. Similarly, to ensure that the joint knowledge is used in its full potential, all parties should ideally be involved in all steps of the research pipeline. With this, the team members can provide their expertise, increase the shared understanding of the research problem, critically reflect on planned tools and techniques, and identify possible pitfalls. Certainly, there exists a great variety of CSS research projects and each project follows its own plan and structure. As an example of how both computer and social scientists can be involved in the research pipeline, we will describe one data science project, in which social media data should be collected to study human behaviour. In such a project, the following pipeline, including contributions from both partners, could be established:

- Research planning: The social scientists identify the essential research questions of their field, and develop theoretical models, which are in line with the knowledge base of their field. The computer scientists provide suggestions on how these research questions could be answered by making use of technologies for data collection and analysis. Both parties decide on how the research goals can be fulfilled, and on which outputs should be generated by the project.

- Data Collection: When deciding on a social media platform, the social scientists give insights into what kind of platform is often used by the population under study and the computer scientists provide knowledge about how and what kind of data can be collected from this platform. When planning the data collection strategy, the social scientists suggest the timeframe which should be studied or the search keywords which are often used by the population, while the computer scientists analyse the characteristics of the platform, e.g. regarding algorithms which might influence the data presented. Whereas the social scientists elaborate on the constructs that should be analysed (e.g. user engagement), both parties provide ideas on how these constructs can be measured (e.g. with likes) given the data available on the platform. Lastly, the computer scientists perform the data collection.

- Data Preprocessing: The computer scientists bring the data in a readable format, and present a first overview of the database. They detect unusual values and outliers in the data, while the social scientists help with the interpretation of these values and with the decision of which data should be excluded from the study. When the computer scientists develop strategies to aggregate data and to convert the data in a consistent format, the social scientists consult on which data can be treated as synonymous (e.g. in Study 4.3 the terms 'cannabis' and 'grass') or how data can be converted in other formats.

- Data Analysis: Based on the background knowledge provided by the social scientists about the research questions, the population under study and the constructs which should be measured, the computer scientists develop and perform analysis strategies. During this step, both parties stay in close contact, so that the computer scientists can ask for clarification on the research topic and the social scientists can early articulate pitfalls or further considerations based on the analysis performed. For example, when performing sentiment analysis, the social scientists provide insights on how the sentiment dictionary should be adapted to fit to the terminology of the population under study.

- Interpretation: The computer scientists describe the methodologies applied, and explain what kind of conclusions can be drawn from the results presented. The social scientists embed the results in the theory and research background, to interpret the findings. The social and computer scientists work together on visualizing and presenting the results in a way that is comprehensible for an audience with different research backgrounds.

Communication. Understanding each other can often be a challenge in collaboration between social and computer scientists (Kirkwood et al., 2018). Therefore, a first step in every collaboration project should be to find a common language and explain the terminology used in the disciplines. Furthermore, all researchers involved should make sure that they understand each other's goals and the requirements which need to be fulfilled so that all parties can benefit from the project and its results (Beck, Meinecke, Matsuyama, & Lee, 2017). As there are certainly differences in approaches and background knowledge, the scholars should always try to describe their techniques in a comprehensible way, also stating possible advantages, pitfalls, and limitations. Throughout the project, an open communication should be upheld, allowing to ask questions when clarifications are needed (Sahneh et al., 2021).

5.2.3 Future Work on Interdisciplinary Collaboration

While the guidelines described above give insights into how researchers can increase the success of their collaboration projects, there remain challenges which cannot be solved by the researchers alone, but require changes in external structures. As criticized in many articles (Lazer et al., 2020; van Rijnsoever & Hessels, 2011; Parti & Szigeti, 2021; Cairns et al., 2020), there exist several structures in universities, publication outlets and funding institutes, which do not support or even hinder interdisciplinary research. Although numerous initiatives have been developed to encourage interdisciplinary collaboration, there is a need for more research on how collaborations can be supported effectively. In particular, future work could 1) investigate, which methods, trainings, and events help researchers to get in contact with scholars of other disciplines and perform collaborations, 2) evaluate the value and practicability of alternative performance evaluations for researchers which appreciate interdisciplinary efforts to the same extent as disciplinary work, 3) identify factors which facilitate the acceptance of interdisciplinary work by publication outlets, 4) study, how fundings could be distributed to support interdisciplinarity and 5) analyse which incentives motivate scientists to collaborate with experts outside of their discipline.

5.3 Conclusion

In summary, this dissertation has highlighted the advantages of Computational Social Science. Based on three studies, we have shown, that computational methods provide great opportunities for studying social science research problems. Not only do these methods allow access to new datasets and enable researchers to analyse large scale data, but they also contribute ways in which human behaviour can be investigated with less costs, time and obtrusiveness.

Nevertheless, as shown by this dissertation and exemplified within the three case studies, the application of computational methods in social science research also holds many challenges. While social scientists may have difficulties to gain access to data and infrastructure, and might misuse or trust too much in computational methods, computer scientists might find it challenging to design studies which generate real impact for the domain of the research problem, and to combat the ethical problems which arise when researching human behaviour.

Finally, in this dissertation, we have evaluated the potential of collaborations between social and computer scientists and/or experts for CSS research. Based on three studies we have shown that such collaboration can help to integrate different perspectives on the research problem and to overcome the challenges arising in CSS projects. Furthermore, with the experiences derived from our projects and reports on other collaborations, we described common difficulties in interdisciplinary projects and provided guidelines for making collaboration in CSS successful. We conclude that interdisciplinary collaboration presents some challenges but it is of great value to the field of Computational Social Science.

References

Abanoz, E. (Ed.). (2022). *Opportunities and Challenges for Computational Social Science Methods*. IGI Global. Retrieved from http://dx.doi.org/10.4018/978-1-7998-8553-5

Aboelela, S. W., Larson, E., Bakken, S., Carrasquillo, O., Formicola, A., Glied, S. A., ... Gebbie, K. M. (2007). Defining Interdisciplinary Research: Conclusions from a Critical Review of the Literature. *Health Services Research*, *42*(1 Pt. 1), 329–346. Retrieved 2023-01-19, from http://dx.doi.org/10.1111/j.1475-6773.2006.00621.x

Albreiki, B., Zaki, N., & Alashwal, H. (2021, September). A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques. *Education Sciences*, *11*(9), 552. Retrieved from http://dx.doi.org/10.3390/educsci11090552

Andersson, R., Larsson, L., Holmqvist, K., Stridh, M., & Nyström, M. (2017). One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior Research Methods*, *49*(2), 616–637. Retrieved from http://dx.doi.org/10.3758/s13428-016-0738-9

Awan, I. (2017, April). Cyber-Extremism: Isis and the Power of Social Media. *Society*, *54*(2), 138–149. Retrieved from http://dx.doi.org/10.1007/s12115-017-0114-0

Barbaglia, L., Consoli, S., Manzan, S., Reforgiato Recupero, D., Saisana, M., & Tiozzo Pezzoli, L. (2021). Data Science Technologies in Economics and Finance: A Gentle Walk-In. In S. Consoli, D. Reforgiato Recupero, & M. Saisana (Eds.), *Data Science for Economics and Finance* (pp. 1–17). Cham: Springer International Publishing. Retrieved from http://dx.doi.org/10.1007/978-3-030-66891-4_1

Barthel, R., & Seidl, R. (2017). Interdisciplinary Collaboration between Natural and Social Sciences – Status and Trends Exemplified in Groundwater Research. *PLOS ONE*, *12*(1), e0170754. Retrieved from http://dx.doi.org/10.1371/journal.pone.0170754

Bartlett, L. K., Pirrone, A., Javed, N., & Gobet, F. (2022). Computational Scientific Discovery in Psychology. *Perspectives on Psychological Science*, *18*(1), 178—189. Retrieved from http://dx.doi.org/10.1177/17456916221091833

Beck, S. J., Meinecke, A. L., Matsuyama, Y., & Lee, C.-C. (2017). Initiating and Maintaining Collaborations and Facilitating Understanding in Interdisciplinary Group Research. *Small Group Research*, *48*(5), 532–543. Retrieved from http://dx.doi.org/10.1177/1046496417721746

Boslaugh, S. (2012). *Statistics in a nutshell* (2nd ed.). Beijing, Farnham: O'Reilly Media.

Bramley, A., & Ogilvie, L. (2021). *Research Collaboration: A step-by-step guide to success*. IOP Publishing. Retrieved from http://dx.doi.org/10.1088/978-0-7503-2727-5

Bravo, G., & Farjam, M. (2017). Prospects and Challenges for the Computational Social Sciences. *Journal of Universal Computer Science*, *23*(11), 1057–1069.

Brunt, T. M., & Niesink, R. J. (2011). The Drug Information and Monitoring System (DIMS) in the Netherlands: Implementation, results, and international comparison. *Drug Testing and Analysis*, *3*(9), 621–634. Retrieved from http://dx.doi.org/10.1002/dta.323

Bucher, T., & Helmond, A. (2018). The Affordances of Social Media Platforms. In *The SAGE Handbook of Social Media* (pp. 233–253). SAGE Publications Ltd. Retrieved from http://dx.doi.org/10.4135/9781473984066.n14

Cairns, R., Hielscher, S., & Light, A. (2020). Collaboration, creativity, conflict and chaos: doing interdisciplinary sustainability research. *Sustainability Science*, *15*(6), 1711–1721. Retrieved from http://dx.doi.org/10.1007/s11625-020-00784-z

Carter, D. R., DeChurch, L. A., Braun, M. T., & Contractor, N. S. (2015). Social network approaches to leadership: An integrative conceptual review. *Journal of Applied Psychology*, *100*(3), 597–622. Retrieved from http://dx.doi.org/10.1037/a0038922

Conway, M., & Mcinerney, L. (2008). Jihadi Video and Auto-Radicalisation: Evidence from an Exploratory YouTube Study. In *Proceedings of the 1st European Conference on Intelligence and Security Informatics* (pp. 108–118). Berlin, Heidelberg: Springer-Verlag. Retrieved from http://dx.doi.org/10.1007/978-3-540-89900-6_13

Dahl, T. (2017). Verbal and Visual Framing Activity in Climate Change Discourse: A Multimodal Analysis of Media Representations of the IPCC' 5th Assessment Report. In K. Flottum (Ed.), *The Role of Language in the Climate Change Debate* (1st ed., p. 13-30). Routledge.

Davidson, T., Bhattacharya, D., & Weber, I. (2019). Racial bias in hate speech and abusive language detection datasets. In *Proceedings of the third workshop on abusive language online.* Association for Computational Linguistics. Retrieved from http://dx.doi.org/10.18653/v1/w19-3504

DiFrancesco, D., & Young, N. (2011). Seeing climate change: the visual construction of global warming in Canadian national print media. *Cultural Geographies*, *18*(4), 517–536. Retrieved from http://dx.doi.org/10.1177/1474474010382072

Dixon, S. (2022, August). *Average daily time spent on social media worldwide 2012-2022.* Retrieved 2022-10-18, from https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/

Dvoynikova, A., Verkholyak, O., & Karpov, A. (2020). Analytical Review of Methods for Identifying Emotions in Text Data. In *Proceedings of the III International Conference on Language Engineering and Applied Linguistics (PRLEAL-2019)* (pp. 8–21). CEUR-WS.

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology*, *14*(1), 91–118. Retrieved from http://dx.doi.org/

10.1146/annurev-clinpsy-032816-045037

D'Mello, S. K., Tay, L., & Southwell, R. (2022). Psychological Measurement in the Information Age: Machine-Learned Computational Models. *Current Directions in Psychological Science*, *31*(1), 76–87. Retrieved from http://dx.doi.org/10.1177/09637214211056906

Edelmann, A., Wolff, T., Montagne, D., & Bail, C. A. (2020). Computational Social Science and Sociology. *Annual Review of Sociology*, *46*(1), 61–81. Retrieved from http://dx.doi.org/10.1146/annurev-soc-121919-054621

Esposito, D., & Esposito, F. (2020). *Introducing Machine Learning* (1st ed.). USA: Microsoft Press.

European Monitoring Centre for Drugs and Drug Addiction. (2019a). *Estimating the size of the main illicit retail drug markets in Europe: an update.* Luxembourg: Publications Office of the European Union. Retrieved from https://data.europa.eu/doi/10.2810/468358

European Monitoring Centre for Drugs and Drug Addiction. (2019b). *EU Drug Markets Report 2019.* Luxembourg: Publications Office of the European Union.

European Monitoring Centre for Drugs and Drug Addiction. (2020). *European Drug Report 2020: Trends and Developments.* Luxembourg: Publications Office of the European Union.

Evegroen, S. A., Schoonderbeek, Y., & Treur, J. (2021). The effect of emergent team roles on team performance: A computational network model. In *Computational collective intelligence* (pp. 111–125). Springer International Publishing. Retrieved from http://dx.doi.org/10.1007/978-3-030-88081-1_9

Foucault Welles, B., & Meirelles, I. (2015). Visualizing Computational Social Science: The Multiple Lives of a Complex Image. *Science Communication*, *37*(1), 34–58. Retrieved from http://dx.doi.org/10.1177/1075547014556540

Fu, X., Lou, J.-D., & Boos, M. (2017). *Social Network Analysis: Interdisciplinary Approaches and Case Studies* (1st ed.). CRC Press. Retrieved from https://doi.org/10.1201/9781315369594

Grantz, K. H., Meredith, H. R., Cummings, D. A. T., Metcalf, C. J. E., Grenfell, B. T., Giles, J. R., … Wesolowski, A. (2020, September). The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nature Communications*, *11*(1), 4961. Retrieved from http://dx.doi.org/10.1038/s41467-020-18190-5

Haq, E. U., Braud, T., Kwon, Y. D., & Hui, P. (2020). A Survey on Computational Politics. *IEEE Access*, *8*, 197379–197406. Retrieved from http://dx.doi.org/10.1109/ACCESS.2020.3034983

Heuer, H., Polizzotto, A., Marx, F., & Breiter, A. (2019). Visualization Needs in Computational Social Sciences. In *Proceedings of Mensch und Computer 2019* (pp. 463–468). Hamburg, Germany: ACM. Retrieved from http://dx.doi.org/10.1145/3340764.3344440

Hilbert, M. (2015). ICT4ICTD: Computational Social Science for Digital Development. In *2015 48th Hawaii International Conference on System Sciences* (pp. 2145–2157). HI, USA: IEEE. Retrieved from `http://dx.doi.org/10.1109/HICSS.2015.258`

Huang, H., Cheng, Y., & Weibel, R. (2019). Transport mode detection based on mobile phone network data: A systematic review. *Transportation Research Part C: Emerging Technologies*, *101*, 297–312. Retrieved from `http://dx.doi.org/10.1016/j.trc.2019.02.008`

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media, ICWSM 2014*, *8*(1), 216–225. Retrieved from `https://doi.org/10.1609/icwsm.v8i1.14550`

Iannelli, L., Giglietto, F., Rossi, L., & Zurovac, E. (2020). Facebook Digital Traces for Survey Research: Assessing the Efficiency and Effectiveness of a Facebook Ad–Based Procedure for Recruiting Online Survey Respondents in Niche and Difficult-to-Reach Populations. *Social Science Computer Review*, *38*(4), 462–476. Retrieved from `http://dx.doi.org/10.1177/0894439318816638`

Jain, P. N., & Vaidya, A. S. (2021). Analysis of Social Media Based on Terrorism — A Review. *Vietnam Journal of Computer Science*, *08*(01), 1–21. Retrieved from `http://dx.doi.org/10.1142/S2196888821300015`

Joshi, A. V. (2020). *Machine Learning and Artificial Intelligence.* Cham: Springer International Publishing. Retrieved from `http://dx.doi.org/10.1007/978-3-030-26622-6`

Jungherr, A. (2016). Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics*, *13*(1), 72–91. Retrieved from `http://dx.doi.org/10.1080/19331681.2015.1132401`

Kherwa, P., & Bansal, P. (2019). Topic modeling: A comprehensive review. *EAI Endorsed Transactions on Scalable Information Systems*, *7*(24). Retrieved from `http://dx.doi.org/10.4108/eai.13-7-2018.159623`

Kirkwood, S., Cree, V., Winterstein, D., Nuttgens, A., & Sneddon, J. (2018). Encountering #Feminism on Twitter: Reflections on a Research Collaboration between Social Scientists and Computer Scientists. *Sociological Research Online*, *23*(4), 763–779. Retrieved from `http://dx.doi.org/10.1177/1360780418781615`

Krause-Jüttler, G., Weitz, J., & Bork, U. (2022). Interdisciplinary Collaborations in Digital Health Research: Mixed Methods Case Study. *JMIR Human Factors*, *9*(2), e36579. Retrieved from `http://dx.doi.org/10.2196/36579`

Krügel, S., Ostermaier, A., & Uhl, M. (2022). Zombies in the loop? humans trust untrustworthy AI-advisors for ethical decisions. *Philosophy & Technology*, *35*(1). Retrieved from `http://dx.doi.org/10.1007/s13347-022-00511-9`

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., ... Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, *10*, 90–115.

Retrieved from http://dx.doi.org/10.1016/j.edurev.2013.10.001

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., ... Van Alstyne, M. (2009). Computational Social Science. *Science*, *323*(5915), 721–723. Retrieved from http://www.jstor.org/stable/20403004

Lazer, D., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., ... Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science*, *369*(6507), 1060–1062. Retrieved from http://dx.doi.org/10.1126/science.aaz8170

Li, X., Zhu, W., Sui, X., Zhang, A., Chi, L., & Lv, L. (2022). Assessing workplace stress among nurses using heart rate variability analysis with wearable ECG device–a pilot study. *Frontiers in Public Health*, *9*. Retrieved from http://dx.doi.org/10.3389/fpubh.2021.810577

Lim, J. Z., Mountstephens, J., & Teo, J. (2020). Emotion Recognition Using Eye-Tracking: Taxonomy, Review and Current Challenges. *Sensors*, *20*(8), 2384. Retrieved from http://dx.doi.org/10.3390/s20082384

Matei, S. A., & Kee, K. F. (2019). Computational communication research. *WIREs Data Mining and Knowledge Discovery*, *9*(4). Retrieved from http://dx.doi.org/https://doi.org/10.1002/widm.1304

Meadows, D. H., Meadows, D., & Randers, J. (1972). *Limits to growth.* New York, NY: Universe Books.

Moocarme, M., Abdolahnejad, M., & Bhagwat, R. (2020). *The Deep Learning with Keras Workshop - Second Edition.* Packt Publishing.

Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, *27*, 16–32. Retrieved from http://dx.doi.org/10.1016/j.cosrev.2017.10.002

Oboler, A., Welsh, K., & Cruz, L. (2012). The danger of big data: Social media as computational social science. *First Monday*, *17*(7). Retrieved from http://dx.doi.org/10.5210/fm.v17i7.3993

Ophir, Y., Tikochinski, R., Asterhan, C. S. C., Sisso, I., & Reichart, R. (2020). Deep neural networks detect suicide risk from textual facebook posts. *Scientific Reports*, *10*(1), 16685. Retrieved 2022-11-03, from http://dx.doi.org/10.1038/s41598-020-73917-0

Ostermann, T., Röer, J. P., & Tomasik, M. J. (2021). Digitalization in psychology: A bit of challenge and a byte of success. *Patterns*, *2*(10), 100334. Retrieved from http://dx.doi.org/10.1016/j.patter.2021.100334

Parti, K., & Szigeti, A. (2021). The future of interdisciplinary research in the digital era: Obstacles and perspectives of collaboration in social and data sciences - an empirical study. *Cogent Social Sciences*, *7*(1). Retrieved from http://dx.doi.org/10.1080/23311886.2021.1970880

Pfeffer, J. (2017). Visualization of political networks. In J. N. Victor, A. H. Montgomery, & M. Lubell (Eds.), *The oxford handbook of political networks.* Oxford University Press. Retrieved from http://dx.doi.org/10.1093/oxfordhb/9780190228217.013.13

Pfeffer, J., & Malik, M. M. (2017). Simulating the dynamics of socio-economic systems. In B. Hollstein, W. Matiaske, & K. Schnapp (Eds.), *Networked governance* (pp. 143–161). Springer International Publishing. Retrieved from http://dx.doi.org/10.1007/978-3-319-50386-8_9

Pfeffer, J., Mooseder, A., Hammer, L., Stritzel, O., & Garcia, D. (2022). This sample seems to be good enough! assessing coverage and temporal reliability of twitter's academic api. *Arxiv*. (Preprint on Arxiv: https://arxiv.org/abs/2204.02290)

Posada-Quintero, H. F., & Chon, K. H. (2020). Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review. *Sensors*, *20*(2), 479. Retrieved from http://dx.doi.org/10.3390/s20020479

Purnomo, A., Asitah, N., Rosyidah, E., Septianto, A., & Firdaus, M. (2022). Mapping of Computational Social Science Research Themes: A Two-Decade Review. In V. S. Reddy, V. K. Prasad, D. N. Mallikarjuna Rao, & S. C. Satapathy (Eds.), *Intelligent Systems and Sustainable Computing* (Vol. 289, pp. 617–625). Singapore: Springer Nature Singapore. Retrieved from http://dx.doi.org/10.1007/978-981-19-0011-2_55

Rhumorbarbe, D., Morelato, M., Staehli, L., Roux, C., Jaquet-Chiffelle, D.-O., Rossy, Q., & Esseiva, P. (2019). Monitoring new psychoactive substances: Exploring the contribution of an online discussion forum. *Internat. Journal of Drug Policy*, *73*, 273–280.

Rocha, Y. M., de Moura, G. A., Desidério, G. A., de Oliveira, C. H., Lourenço, F. D., & de Figueiredo Nicolete, L. D. (2021). The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review. *Journal of Public Health*. Retrieved from http://dx.doi.org/10.1007/s10389-021-01658-z

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In *Computer supported collaborative learning* (pp. 69–97). Springer Berlin Heidelberg. Retrieved from http://dx.doi.org/10.1007/978-3-642-85098-1_5

Rucci, M., McGraw, P. V., & Krauzlis, R. J. (2016, January). Fixational eye movements and perception. *Vision Research*, *118*, 1–4. Retrieved from http://dx.doi.org/10.1016/j.visres.2015.12.001

Sahneh, F., Balk, M. A., Kisley, M., Chan, C.-k., Fox, M., Nord, B., ... Merchant, N. (2021). Ten simple rules to cultivate transdisciplinary collaboration in data science. *PLOS Computational Biology*, *17*(5), e1008879. Retrieved from http://dx.doi.org/10.1371/journal.pcbi.1008879

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the symposium on Eye tracking research & applications - ETRA '00* (pp. 71–78). Palm Beach Gardens, Florida, United States: ACM Press. Retrieved from http://dx.doi.org/10.1145/355017.355028

Shahi, G. K., Dirkson, A., & Majchrzak, T. A. (2021). An exploratory study of COVID-19 misinformation on Twitter. *Online Social Networks and Media*, *22*, 100104. Retrieved from http://dx.doi.org/10.1016/j.osnem.2020

.100104
Siebert, P., Siebers, P.-O., Vallejos, E. P., & Nilsson, T. (2020). Driving complementarity in interdisciplinary research: a reflection. *International Journal of Social Research Methodology*, *23*(6), 711–718. Retrieved from http://dx.doi.org/10.1080/13645579.2020.1743545
Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., ... Sanchez-Vives, M. V. (2006). A Virtual Reprise of the Stanley Milgram Obedience Experiments. *PLoS ONE*, *1*(1), e39. Retrieved from http://dx.doi.org/10.1371/journal.pone.0000039
Smith, R. D. (2003). Simulation. In A. Ralston, E. Reilly, & D. Hemmendinger (Eds.), *Encyclopedia of computer science* (4th ed., p. 1578–1587). GBR: John Wiley and Sons Ltd.
Statista Research Department. (2022, September). *Worldwide digital population July 2022.* Retrieved 2022-10-18, from https://www.statista.com/statistics/617136/digital-population-worldwide/
Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: A systematic literature review. *Educational Studies in Mathematics*, *104*(2), 147–200. Retrieved from http://dx.doi.org/10.1007/s10649-020-09948-1
Stuart, S. (Ed.). (2022). *Eye Tracking: Background, Methods, and Applications* (Vol. 183). New York, NY: Springer US. Retrieved from http://dx.doi.org/10.1007/978-1-0716-2391-6
Theocharis, Y., & Jungherr, A. (2021). Computational social science and the study of political communication. *Political Communication*, *38*(1-2), 1–22. Retrieved from http://dx.doi.org/10.1080/10584609.2020.1833121
Tindall, D., McLevey, J., Koop-Monteiro, Y., & Graham, A. (2022). Big data, computational social science, and other recent innovations in social network analysis. *Canadian Review of Sociology/Revue canadienne de sociologie*, *59*(2), 271–288. Retrieved from http://dx.doi.org/10.1111/cars.12377
Uppal, A., & Bohon, S. (2017). Thoughts on the Vulnerability of the Field of Computational Social Science. In *Proceedings of the 2017 International Conference of The Computational Social Science Society of the Americas* (pp. 1–5). Santa Fe NM USA: ACM. Retrieved from http://dx.doi.org/10.1145/3145574.3145580
van Atteveldt, W., & Peng, T.-Q. (2018). When Communication Meets Computation: Opportunities, Challenges, and Pitfalls in Computational Communication Science. *Communication Methods and Measures*, *12*(2-3), 81–92. Retrieved from http://dx.doi.org/10.1080/19312458.2018.1458084
van Rijnsoever, F. J., & Hessels, L. K. (2011). Factors associated with disciplinary and interdisciplinary research collaboration. *Research Policy*, *40*(3), 463–472. Retrieved from http://dx.doi.org/10.1016/j.respol.2010.11.001
Verouden, N. W., van der Sanden, M. C., & Aarts, N. (2016). Silence in Interdisciplinary Research Collaboration: Not Everything Said is Relevant,

Not Everything Relevant is Said. *Science as Culture*, *25*(2), 264–288. Retrieved from http://dx.doi.org/10.1080/09505431.2016.1141191

Viseu, A. (2015). Integration of social science into research is crucial. *Nature*, *525*(7569), 291–291. Retrieved from http://dx.doi.org/10.1038/525291a

Wang, X., Song, Y., & Su, Y. (2022). Less Fragmented but Highly Centralized: A Bibliometric Analysis of Research in Computational Social Science. *Social Science Computer Review*. Retrieved from http://dx.doi.org/10.1177/08944393211058112

Wang, X., Zhang, M., Fan, W., & Zhao, K. (2022). Understanding the spread of COVID-19 misinformation on social media: The effects of topics and a political leader's nudge. *Journal of the Association for Information Science and Technology*, *73*(5), 726–737. Retrieved from http://dx.doi.org/10.1002/asi.24576

Wang, Y., Li, J., Zhao, X., Feng, G., & Luo, X. (2020). Using Mobile Phone Data for Emergency Management: a Systematic Literature Review. *Information Systems Frontiers*, *22*(6), 1539–1559.

Watts, D. (2014). Computational Social Science: Exciting Progress and Future Challenges. In *Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2013 Symposium.* Washington: The National Academies Press. Retrieved from http://dx.doi.org/https://doi.org/10.17226/18558

We are Social, Hootsuit, & DataReportal. (2022, January). *Most popular social networks worldwide as of January 2022, ranked by number of monthly active users (in millions).* Retrieved 2022-10-18, from https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Yunita, A., Santoso, H. B., & Hasibuan, Z. A. (2021). Research Review on Big Data Usage for Learning Analytics and Educational Data Mining: A Way Forward to Develop an Intelligent Automation System. *Journal of Physics: Conference Series*, *1898*(012044). Retrieved from http://dx.doi.org/10.1088/1742-6596/1898/1/012044

Zeitlin, W., & Auerbach, C. (2019). *Basic Statistics for the Behavioral and Social Sciences Using R.* Oxford, UNITED STATES: Oxford University Press, Incorporated.

Zhang, J., Wang, W., Xia, F., Lin, Y.-R., & Tong, H. (2020). Data-driven Computational Social Science: A Survey. *Big Data Research*, *21*, 100145. Retrieved from http://dx.doi.org/10.1016/j.bdr.2020.100145

Zhu, H., Kraut, R. E., & Kittur, A. (2013). Effectiveness of shared leadership in wikipedia. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*(6), 1021–1043. Retrieved from http://dx.doi.org/10.1177/0018720813515704

Zhu, J., Ji, L., & Liu, C. (2019). Heart rate variability monitoring for emotion and disorders of emotion. *Physiological Measurement*, *40*(6), 064004. Retrieved from http://dx.doi.org/10.1088/1361-6579/ab1887

Zuccato, E., Chiabrando, C., Castiglioni, S., Bagnati, R., & Fanelli, R. (2008).

Estimating Community Drug Abuse by Wastewater Analysis. *Environmental Health Perspectives*, *116*(8), 1027–1032. Retrieved from http://dx.doi.org/10.1289/ehp.11022