



HAL
open science

Small Data - Visualizing Simple Datasets for Communication and Decision Making

Pierre Dragicevic

► **To cite this version:**

Pierre Dragicevic. Small Data - Visualizing Simple Datasets for Communication and Decision Making. Human-Computer Interaction [cs.HC]. Université de Bordeaux (UB), 2023. tel-04352699v2

HAL Id: tel-04352699

<https://inria.hal.science/tel-04352699v2>

Submitted on 14 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Small Data – Visualizing Simple Datasets for Communication and Decision Making

Habilitation à Diriger les Recherches

Pierre Dragicevic

*Defended on 23 June 2023
at the University of Bordeaux*

Version: v1.2 – 6 April 2023,
with annotations added on 14 Nov 2024.

Université Bordeaux 1

Habilitation à Diriger les Recherches

Defended on 23 June 2023
at the University of Bordeaux

Small Data – Visualizing Simple Datasets for Communication and Decision Making

Pierre Dragicevic

- Reviewer* Priti Shah
Full Professor, Department of Psychology
University of Michigan
- Reviewer* Jessica Hullman
Associate Professor, McCormick School of Engineering
Northwestern University
- Reviewer* Jeffrey Heer
Full Professor, Dept. of Computer Science & Engineering
University of Washington
- Examiner* H  l  ne Sauz  on
Full Professor, Psychology
Universit   de Bordeaux
- Examiner* Guy Melan  on
Full Professor, Computer Science
Universit   de Bordeaux
- Examiner* Pascal Guitton
Professor Emeritus, Computer Science
Universit   de Bordeaux

Pierre Dragicevic

Small Data – Visualizing Simple Datasets for Communication and Decision Making

Habilitation à Diriger les Recherches, Defended on 23 June 2023

at the University of Bordeaux

Reviewers: Priti Shah, Jessica Hullman, and Jeffrey Heer

Examiners: H el ene Sauz eon, Guy Melan on, and Pascal Guitton

Universit  Bordeaux 1

 cole Doctorale de Math matiques et d'Informatique

Abstract

This habilitation thesis presents research projects that explore how data visualizations can be used to communicate quantitative facts and support decisions. The thesis identifies the umbrella topic of “small data” as the common denominator for these projects, as they all involve datasets that are smaller than what is typically found in the information visualization literature. Although much of this literature focuses on large datasets, this thesis shows that we often do not know how to best visualize even small datasets, and that interesting research questions can also arise from such datasets.

The first part of this thesis focuses on supporting rational judgments and decisions with data visualizations. It examines whether data visualization can improve reasoning with base rates, it looks at whether a cognitive bias called the attraction effect transfers to data visualizations, and discusses how to generally evaluate data visualizations for decision making.

The second part of this thesis discusses how to support effective communication with data visualizations. It begins by exploring different ways in which researchers can communicate their data to their peers: first through tabular visualizations, and then through multiverse analyses. It then reports on two studies examining how to communicate data to large audiences: a replication study on whether trivial charts can hinder truthful communication, and a study on how to convey data on humanitarian issues.

Finally, a concluding chapter brings together several of the research problems discussed here by offering perspectives on how data visualization can support rational decision-making and effective communication on humanitarian issues. It discusses how can the effective altruism movement inform research in this area, and how may immersive displays be used to connect people to invisible or distant suffering.

Résumé

Cette thèse de HDR présente des projets de recherche qui explorent comment les visualisations de données peuvent être utilisées pour communiquer des faits quantitatifs et aider à la prise de décision. La thèse identifie le thème général des “petites données” comme le dénominateur commun de ces projets, car ils impliquent tous des jeux de données plus petits que ce que l’on trouve généralement dans la littérature sur la visualisation d’information. Bien qu’une grande partie de cette littérature se concentre sur les grands jeux de données, cette thèse montre que nous ne savons souvent pas comment visualiser même les petits jeux de données, et que des questions de recherche intéressantes peuvent également être soulevées à partir de ces types de données.

La première partie de cette thèse se concentre sur l’aide aux jugements et aux décisions rationnels à l’aide de visualisations de données. Elle examine si la visualisation des données peut améliorer le raisonnement avec les fréquences de base, si un biais cognitif appelé effet d’attraction se transfère aux visualisations de données, et comment évaluer de manière générale les visualisations de données pour la prise de décision.

La deuxième partie de cette thèse s’intéresse à comment communiquer les données de manière efficace. Elle commence par explorer les différentes façons dont les chercheurs peuvent communiquer leurs données à leurs pairs : d’abord par des visualisations tabulaires, puis par des analyses multivers. Il présente ensuite deux études examinant comment communiquer des données au grand public : une étude de réplique visant à déterminer si des graphiques triviaux peuvent entraver une communication sincère, et une étude sur la manière de communiquer des données sur des questions humanitaires.

Enfin, un chapitre de perspectives rassemble plusieurs des problèmes de recherche discutés dans cette thèse en offrant des perspectives sur la façon dont la visualisation des données peut soutenir la prise de décision rationnelle et la communication efficace sur les questions humanitaires. Il examine comment le mouvement de l’altruisme efficace peut informer la recherche dans ce domaine, et comment les affichages immersifs peuvent être utilisés pour connecter les individus à des souffrances invisibles ou lointaines.

Acknowledgements

Many thanks to my students and collaborators for their joint work on the research reported here. I co-authored this research with (in alphabetical order):

Students: [Evanthia Dimara](#), [Luana Micallef](#), [Luiz Morais](#), [Charles Perin](#), and [Abhraneel Sarma](#).

Collaborators: [Nazareno Andrade](#), [Anastasia Bezerianos](#), [Fanny Chevalier](#), [Jean-Daniel Fekete](#), [Steven Franconeri](#), [Yvonne Jansen](#), [Matthew Kay](#), and [Catherine Plaisant](#).

Also many thanks to the members of my HDR committee (again in alphabetical order):

Reviewers: [Jeffrey Heer](#), [Jessica Hullman](#), and [Priti Shah](#).

Examiners: [Pascal Guitton](#), [Guy Melançon](#), and [Hélène Sauzçon](#).

I am grateful to the members of [Aviz](#) and [Potioc](#) for the many discussions and for their support.



Contents

1	Introduction	1
1.1	Purpose and Scope	1
1.1.1	Habilitation Theses	1
1.1.2	Data Visualization	2
1.1.3	Purposes of Data Visualization	3
1.1.4	Small Data	3
1.1.5	Other Work	4
1.2	Thesis Outline	6
2	Supporting Rational Judgments and Decisions	9
2.1	Reasoning with Base Rates	9
2.1.1	Introduction	10
2.1.2	First Experiment: Testing Six Visualizations	13
2.1.3	Second Experiment: Tweaking the Text	15
2.1.4	Conclusions	16
2.2	A Cognitive Bias Called The Attraction Effect	17
2.2.1	Introduction	17
2.2.2	First Experiment: Two to Three Alternatives	18
2.2.3	Second Experiment: More Alternatives	21
2.2.4	Conclusions	22
2.3	Evaluating Visualizations for Decision Making	23
2.3.1	Evaluating Elementary Multidimensional Visualizations	23
2.3.2	Using Narratives in Crowdsourcing Studies	28
2.3.3	A Taxonomy of Cognitive Biases to Study	33
2.4	General Conclusions	37
3	Supporting Effective Communication	39
3.1	Communicating Data with Tabular Visualizations	39
3.1.1	Introduction	40

3.1.2	Two examples from Jacques Bertin	42
3.1.3	Bertifier Example: Survey of Applications	44
3.1.4	Bertifier Example: User Study Data	45
3.1.5	Conclusion	45
3.2	Communicating Results from Multiverse Analyses	46
3.2.1	Introduction	47
3.2.2	Example 1 – FREQUENTIST	49
3.2.3	Example 2 – DATAVERSE	51
3.2.4	Example 3 – DANCE	52
3.2.5	Limitations and Possible Objections to EMARs	54
3.2.6	Conclusion	55
3.3	Do Trivial Charts Hinder Truthful Communication?	56
3.3.1	Introduction	56
3.3.2	First Replication of BWS’ Second Experiment	59
3.3.3	Summary of All Experiments	62
3.3.4	Conclusion	62
3.4	Communicating Humanitarian Data	63
3.4.1	Introduction	64
3.4.2	Quick Overview of the Design Space	65
3.4.3	Overview of Previous Studies	68
3.4.4	Our Own Study	71
3.4.5	Conclusion	74
4	Perspectives — Supporting Rational Decisions and Effective Communication on Humanitarian Issues	77
4.1	Introduction	77
4.2	What is Effective Altruism?	80
4.3	How Can EA Inform Visualization	81
4.4	Using Visualization and Psychology to Support EA	83
4.5	Combining Experiences with Quantitative Facts	84
4.6	Emerging Technologies	86
4.7	Conclusion	89
5	Conclusion	91
	Bibliography	95

Introduction

1

We hear about big data everywhere. I am often tempted to say that before we tackle big data, we should first solve small data. In reality, I think we should be working on both in parallel. Issues with big data – such as how to store, access and visualize large databases – are very important, pressing, and deserve attention. However, there are other data visualization problems that arise independently of dataset size – such as how to communicate data to general audiences, or how to help people make decisions based on data. These are also very important and have not received so much attention.

This habilitation thesis summarizes research projects I was involved in which focus on using data visualizations to communicate quantitative facts and support personal data-driven decisions. While trying to find a common denominator for this stream of work, I realized that many of these projects use datasets that are much smaller than what is generally found in the information visualization literature – the smallest one consists of only two numbers. Maybe this is because I like to simplify problems in order to make them more tractable and easier to comprehend. In any case, I decided that *small data* could be the umbrella topic for my thesis. However, I never consciously decided to work on small data, or identified small data as a research direction in and of itself.

I will now lay out the scope of this thesis in more detail.

1.1 Purpose and Scope

1.1.1 Habilitation Theses

Many people outside France may not know what an habilitation is. In the French academic system, the Habilitation à Diriger les Recherches (HDR) is a degree given to experienced academic researchers that allows them to officially supervise PhD students and apply for major job promotions. It is not a tenure – in France, academic researchers get tenure as soon as they are hired (although this seems to be changing). Getting the HDR requires

that the researcher writes a document summarizing their past work or a subset thereof (the thesis) and defends it in front of a jury. Researchers can still – and do – advise PhD students before that, but their supervising role is not fully recognized by the University. Researchers typically go through the HDR process about five to ten years after being hired – sometime later (it took me 16 years), and some never do it.

1.1.2 Data Visualization

The term *data visualization* is ubiquitous but academic researchers tend to prefer *information visualization*. Although several definitions of information visualization have been proposed, an especially popular one is the one by Card, Mackinlay and Shneiderman in the introductory chapter of their seminal book [KMS99]:

“Information visualization: the use of computer-supported, interactive, visual representations of data to amplify cognition.” [KMS99]

Statistical graphics, *scientific visualization* and *visual analytics* are closely related academic terms. Although they historically refer to somewhat distinct lines of research, the distinctions are subtle [Mun08; Kei+08; Kos12], and all terms are compatible with the definition above. Therefore, my co-authors and I have sometimes used *visualization* as an umbrella term for all research areas involving computer-supported visualization of data [Jan+15; DJV20]. This is consistent with many authors using visualization as a shorthand for information visualization, scientific visualization, or data visualization.

However, for people outside those academic areas, *visualization* as an isolated word can be confusing. For example, a Google Scholar search of “visualization in psychology” yields several articles about topics at the intersection of data visualization and psychology, but also completely unrelated papers about mental imagery. Before you turn this page, can you guess which of the following papers are about data visualization and which ones are about mental imagery?

1. *The Psychology of Visualization* [Mar+93]
2. *Research on Visualization in Learning and Teaching Mathematics: Emergence from Psychology* [Pre06].
3. *A History of Visualization in Psychology and Science* [PNM10]

The answers are: 1) data visualization; 2) mental imagery; 3) mental imagery. Even researchers in human-computer interaction are confused, and many think that if what you're doing is highly visual, then it's visualization. Therefore, I came to prefer the term *data visualization*. It is also a good thing that this term is broadly used outside academic research, because there seems to be no reason to use different terms within and outside academia to refer to the same concepts.

1.1.3 Purposes of Data Visualization

Bremer and Munzner [BM13] listed three reasons why people use data visualizations:

1. *To present*: to communicate data to an audience;
2. *To discover*: to learn by exploring data;
3. *To enjoy*: to consume data for personal satisfaction and pleasure.

Dimara and Stasko [DS21] recently argued for adding a fourth category:

4. *To decide*: to use data to make decisions.

This fourth category is not fundamentally separate from the second one, as people often need to learn from data in order to make data-driven decisions. However, people can learn from data without using their newly acquired knowledge to drive decisions, and conversely, visualizations that are optimized for learning and understanding are not necessarily optimized for decision making [DS21].

Over the past decades of research in data visualization, there has been a considerable amount of research on supporting data-driven exploration and discovery (2. to discover), but comparatively less so on supporting the other motivations for using data visualizations. During my career, I have been especially interested in two of them: communication (1. to present) and decision making (4. to decide). I will summarize my contributions to those two lines of research in this thesis.

1.1.4 Small Data

As I already wrote, another way in which the work reported here differs from mainstream data visualization research is by its focus on small datasets. A lot of data visualization

research focuses on visualizing large and complex datasets (e.g., large networks and large multidimensional datasets), not to mention addressing big data challenges involving heterogeneous and rapidly-growing datasets [DGG15].

Again, this line of research is really important and the focus on large datasets is largely justified. However, it is often wrongly assumed that interesting research problems in information visualization only arise when datasets are becoming large. There are two versions of this claim; One is that we already know how to visualize small datasets – just use the standard charts available in any statistical package or spreadsheet software. Another version is that very small datasets do not need to be visualized at all. In our work on trivial charts [DJ18a] (which I will summarize in [section 3.3](#)), we quoted advice from visualizations experts that is consistent with this belief:

Tufte [Tuf01] deplores trivial charts and affirms that “*tables are preferable to graphics for many small data sets*”. For Duklan and Martin [DM02], “*it makes no sense to encode only a few numbers into an overblown graphic*”. For Kelly et al [KJW05], “*graphs take up a lot of space if showing only a few data points. Hence they are best not used if there are only a few numbers to present.*”. Gillan [Gil+98] similarly suggests that “*few data points might best be presented in the body of the text*” due to the cost of “*processing the data display and integrating the information from the display and text*”.

There no empirical evidence to support these recommendations, and I also think they are misguided. Over my career I have collected several examples where visualizing a *single number* is useful. I call them *single-datum visualizations* – see [Figure 1.1](#). More examples can be found in the article by Chevalier et al. [CVG13] on concrete scales and at <http://dataphys.org/list/tags/single-datum/>.

1.1.5 Other Work

The research reported here covers about 25% of my work over the past 10 years, and about 15% in my whole research career. Other major topics I have worked on in the past 10 years are *data visualizations beyond the desktop* (i.e., visualizations that are physical, situated, immersive, or wall-sized), and *research transparency and statistical reporting*. Before that, I have worked on animated transitions and visual tracking, new interaction and visualization techniques, input devices and motor control, and user interface prototyping tools. I will not discuss these topics, except in the last chapter on future perspectives where I will mention beyond-desktop visualizations.

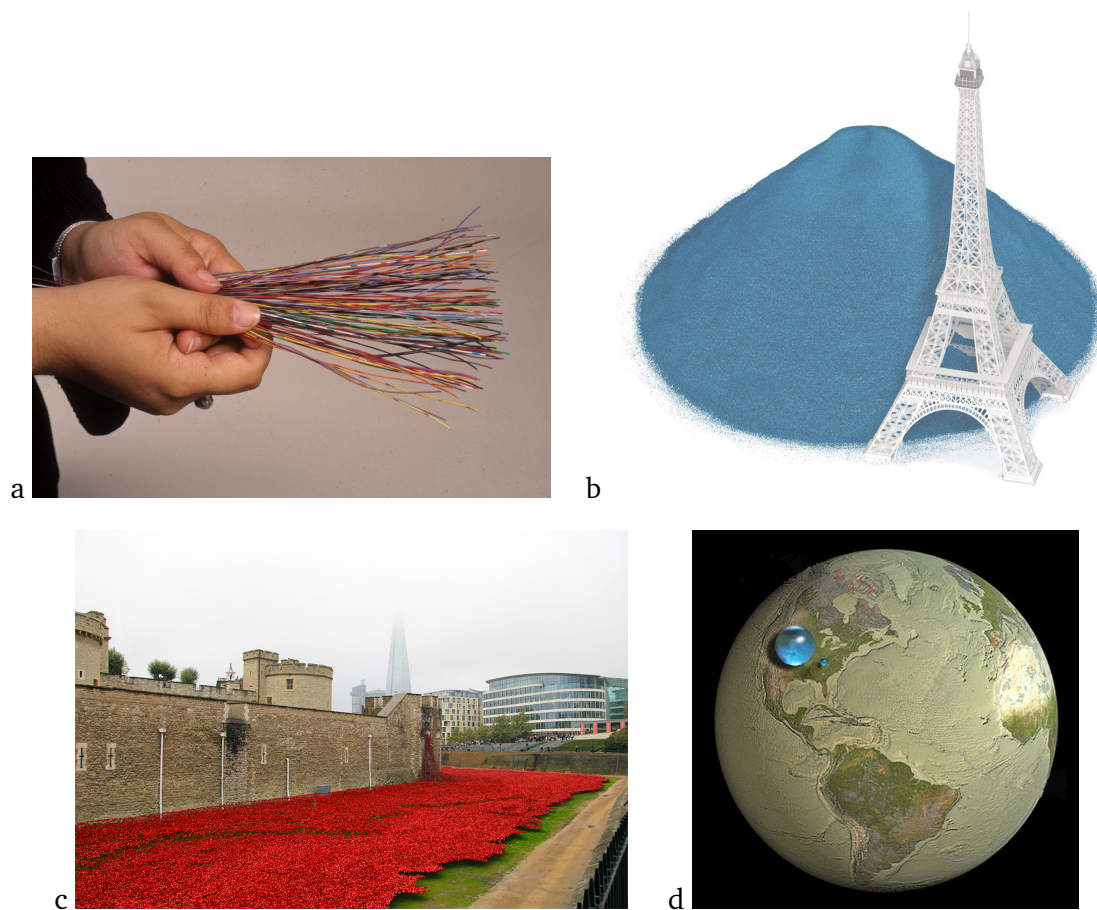


Fig. 1.1: Physical and graphical representations of single quantities: a) Bundles of wire representing the distance traveled by electricity in a nanosecond, and used by Grace Hopper to help programmers understand “*just what they’re throwing away when they throw away a millisecond*” [Plo21]; b) Number of plastic bottles purchased in a day across the world [SH19]; c) Hundreds of thousands of ceramic poppies representing WWI British military fatalities [Jan14]; d) Amount of Earth’s water [Sch19].

1.2 Thesis Outline

This report primarily consists of extended summaries of some my past articles – generally, one article by section. At many places, I have reworked the explanations to make them easier to understand, and often added new explanations. I tried to be generous with figures. To make the thesis lighter, I don't cover related work extensively, and so my contributions are sometimes not very well positioned with respect to the past literature. More comprehensive information about how each piece of work fits within the literature can be obtained by referring to the corresponding paper.

The thesis outline is as follows:

- Chapter 1. **Introduction.** The introduction you are reading now.
- Chapter 2. **Supporting Rational Judgments and Decisions.**
 1. **Reasoning with Base Rates.** We look at whether data visualization can improve how we reason with base rates, such as when we need to estimate the probability of having a disease after having received a positive test [MDF12].
 2. **A Cognitive Bias Called the Attraction Effect.** We look at whether the attraction effect – a bias whereby our choice between two options is influenced by an irrelevant (clearly undesirable) option – persists with data visualizations [DBD17c].
 3. **Evaluating Visualizations for Decision Making.** How should we evaluate data visualizations' support for judgment and decision making? We look at three different facets of this question [DBD17a; DBD17b; Dim+18].
- Chapter 3. **Supporting Effective Communication.**
 1. **Communicating Data with Tabular Visualizations.** We look at a powerful and versatile method that was used decades ago by researchers to communicate data to their peers, and explain how we adapted it to modern technology [PDF14; PFD19].
 2. **Communicating Results from Multiverse Analyses.** We look at the concept of interactive research paper where readers can change statistical analysis options and look at how fragile or robust the findings are [Dra+19].

3. **Do Trivial Charts Hinder Truthful Communication?** A 2016 study found that charts can persuade by their mere presence, simply because they look scientific. With a replication study, we look at whether this claim is justified [DJ18a].
 4. **Communicating Humanitarian Data.** We look at different ways of communicating data about human tragedies that may help the audience relate to the victims, and whether they actually work [Mor+20; Mor+21].
- Chapter 4. **Perspectives — Supporting Rational Decisions and Effective Communication on Humanitarian Issues.** We look at avenues of research that target the goals of Chapters 2 and 3, with applications to the broad area of humanitarian visualization [Dra22a]. I discuss effective altruism and immersive displays, among other things.
 - Chapter 5. **Conclusion.** We go through 17 key takeaways from this HDR thesis.

Supporting Rational Judgments and Decisions

Rationality is complicated and controversial. It is certainly wrong to think that people or their decisions can be easily categorized as rational or irrational. But I think it is also a mistake to reject the concept of rationality entirely, just because it causes a range of conceptual difficulties and can be abused. It is important that we as humans try to improve how we think, and rationality is a useful concept to that end. I will briefly get back to this issue in the conclusion of this chapter, and say more about difficulties with the concept of rationality. For now, I will refrain from even attempting to define rationality, and will instead directly dive into concrete examples.

This chapter covers two concrete examples looking at whether visualization can help people make better judgments and decisions. The first one involves reasoning with base rates, while the second one involves a choosing between multiple options that have been set up in order to elicit a cognitive bias called the attraction effect (also known as the decoy effect). This chapter will conclude with a third section looking at different aspects to consider when evaluating visualizations for their ability to support rational judgment and decision making.

Almost all of the work reported here is by Evanthia Dimara, a PhD student I co-advised between 2014 and 2017. The coming section is work by Luana Micallef, a visiting PhD student I co-advised in 2011.

2.1 Reasoning with Base Rates

People have difficulty reasoning with probabilities and are unaware of their wrong judgments, particularly in problems involving base rates. This issue has been extensively studied in psychology, but rarely in a data visualization context. In this section, we look at whether visualizations can improve reasoning with base rates. This section is based on a 2012 paper authored by Luana Micallef, myself, and Jean-Daniel Fekete [MDF12].

Supplementary material (data, code, analyses) is available at <https://osf.io/68ntk/> and <https://aviz.fr/bayes>.

2.1.1 Introduction

Both laymen and professionals have difficulty making inferences and decisions based on uncertain information [CT96; Gig+07; Hof+00]. This can have severe consequences in many domains. Physicians need to diagnose diseases based on the outcome of unreliable medical tests. Patients need to decide whether they should undertake heavy medical treatment. Wrong judgments are common and often result in overdiagnosis [WB10a], e.g., up to two thirds of breast cancers detected by mammography can be overdiagnosed [ZM04]. In other cases, patients with a positive HIV test result attempted or committed suicide before further tests turned out negative [Chi93; Gig98; Sti96]. In this domain, a crucial piece of information for effective decision making is the probability that a patient has a disease given that a test is positive.

In legal trials, juries have to convict or acquit defendants based on unreliable evidence and here too, wrong judgments abound [Koe97]. A respected professor and advisor to defense lawyers claimed on U.S. television that since only 0.1% of wife batterers murder their wives, evidence of battering should be ignored in murder trials [Goo95]. This reasoning is however fallacious, since the only important information is the probability that a husband was the murderer given that he battered his wife and she was killed.

All such scenarios involve Bayesian inference, more specifically, reasoning with *base rates*. Base rates are known to be counter-intuitive and subject to fallacious reasoning. As a concrete illustration, consider the following classic problem [Edd82]:

The probability that a woman at age 40 has breast cancer is 1%. According to the literature, the probability that the disease is detected by a mammography is 80%. The probability that the test misdetects the disease although the patient does not have it is 9.6%.

If a woman at age 40 is tested as positive, what is the probability that she indeed has breast cancer?

Out of 100 medical doctors, 95 estimated this probability to be between 70% and 80%, while the correct probability is only 7.8% [Edd82]. The probability is low because the

prevalence of the disease in the population, i.e., the *base rate*, is low. When making Bayesian inference, this information is often ignored [GH95; Hof+00], thus leading to the so-called base rate fallacy [BS07a; BS07b]. Using natural frequencies¹ instead of probabilities reduces the fallacy [CT96; GH95; HG04]. However, it is still difficult to comprehend how the different numerical quantities relate to each other².

Previously proposed solutions to improve base-rate reasoning involve educational interventions, some of which use visual representations. For example, a study has suggested that when Bayes' theorem is introduced to students through visualizations, students learn better than without a visualization [SG01]. However, prior training is not always possible. A few studies were conducted to assess the immediate benefits of visualizations, with mixed results. From them, we learn that two different but complementary types of representations may help people reason with base rates: *Euler diagrams* can convey critical information on the nested-set relations in such problems [GG01; Slo+03], while *frequency visualizations* like icon arrays may facilitate probabilistic reasoning more broadly [CT96]. In our work, we were interested in visualization designs that combine both approaches, as in Figure 2.1.

The underlying intuition that motivated this study was that hybrid visualizations showing both nested-set relations and frequencies must help people more than either design. We only knew of one study prior to 2012 which evaluated such hybrid visualizations: the study by Gary Brase [Bra09]. Brase's study did not find benefits over simpler diagrams, but the visualization designs used in this study suffer from a number of issues, making the results difficult to interpret (full discussion in [MDF12]). In addition, like most other studies, Brase's study involved a population of university students who had to participate in the study as part of their psychology course requirement, making it difficult to generalize their findings to a more diverse population. In addition, only one medical diagnosis problem was evaluated, making it difficult to generalize the findings to other problems. In our study, we addressed most of these limitations.

¹Using *natural frequencies* would mean stating “10 out of every 1,000 women” instead of giving a 1% probability. This frequency format is said to be *natural* as the denominator corresponds to the number of actual observations.

²In our example, 1% is the *base rate*, 80% is the *hit rate*, and 9.6% is the *false alarm rate*.

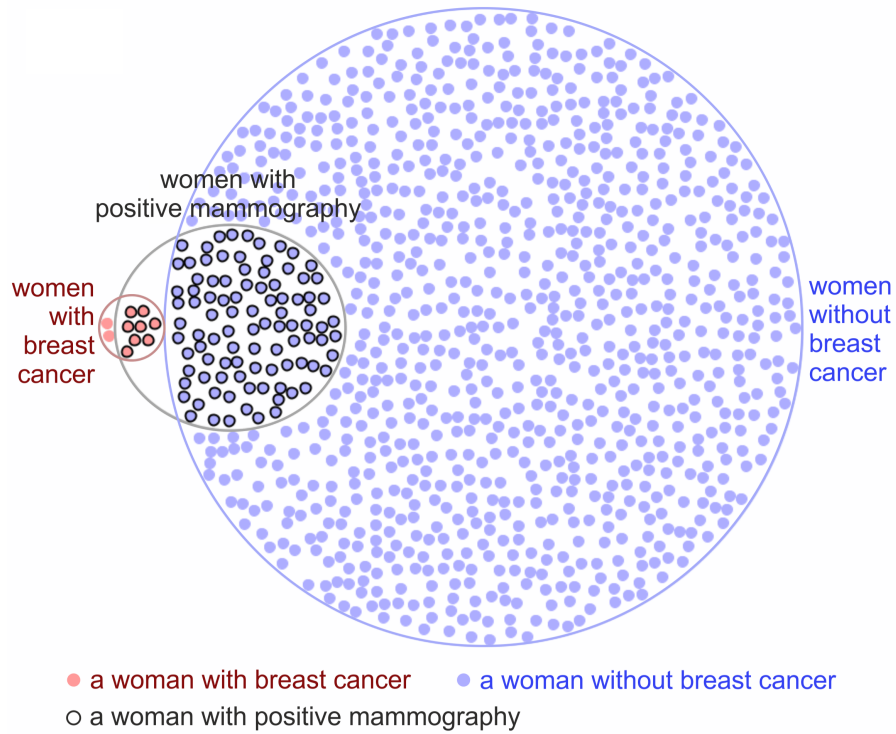


Fig. 2.1: A hybrid Euler/frequency visualization of the classic mammography problem by David M. Eddy [Edd82], described in section 2.1. Each dot is a woman; Pink dots are women with breast cancer (unknown to them) while blue dots are women without breast cancer. All women undergo a mammography. The dots with a black outline are women whose test came back positive, while all others were negative. As we saw in section 2.1, the test is such that 80% of women with breast cancer get a positive test (the black dots in the pink group). Meanwhile, among the women who do *not* have breast cancer, 9.6% get a positive test nonetheless (the black dots in the blue group). Now suppose you are a woman who just got a positive test. The only thing you learned is that you are somewhere in the black ellipse, and therefore, that the probability that you do have breast cancer is the proportion of pink dots in that group, i.e., 7.8%.

2.1.2 First Experiment: Testing Six Visualizations

We designed a set of six visualizations for base rate problems (V1–V6) which include Euler-based representations, frequency representations, and different combinations of the two (Figure 2.2).

To test these visualizations, we recruited 168 participants in a crowdsourcing platform (24 per condition). Each of them saw three base rate problems used in prior work: the *mammography* problem [Edd82] we already discussed, as well as the *cab* problem [Bar80], and the *choosing a course in economics* problem [Ajz77]. The three problems were presented on separate pages, and used the same presentation format: either text alone (V0) or text followed by a visualization (V1–V6), depending on the participant. The ordering of the problems was randomized.

Our main dependent variables were *bias* and *error*. We operationalized *bias* as $\log_{10}(p/p_e)$, with p_e being the true probability and p the participant's estimated probability. To illustrate, in the mammography problem, if a participant answers that the probability of having breast cancer is 6 out of 10, then $p = 6/10 = 0.6$, and the bias is therefore $\log_{10}(0.6/0.078) = 0.886$. The strictly positive value indicates an overestimated probability. In contrast, a negative value would have indicated an underestimation, while a value of zero corresponds to a perfectly correct answer. We operationalized *error* as the absolute value of bias, i.e., in our example, $|0.886| = 0.886$. This continuous measure of error stands in contrast with previous studies, which simply counted the proportion of perfectly correct answers, thus wasting statistical power and ignoring the extent to which people's answers are wrong.

Results for bias are in Figure 2.3. Each of the three columns show the distribution of biases for a specific base rate problem, for each of the visualization conditions (rows). Visually, the distributions differ across base rate problems but seem very similar across visualization conditions. In particular, there is no obvious sign of V1–V6 outperforming V0 (text only). This is confirmed by looking at errors: differences in mean errors are not statistically measurable, even after collapsing V1–V6 into a single condition.

Therefore, we failed to find a facilitating effect of visualization on base rate reasoning. In addition, our participants' accuracy was remarkably lower compared to many of the previous studies. For instance, considering V0 (text only), previous studies reported 35% to 72% perfectly correct answers [Bra09; CT96; GH95; Slo+03; CT96], while we only had 6% of them. We were not able to fully explain this discrepancy, but part of it might

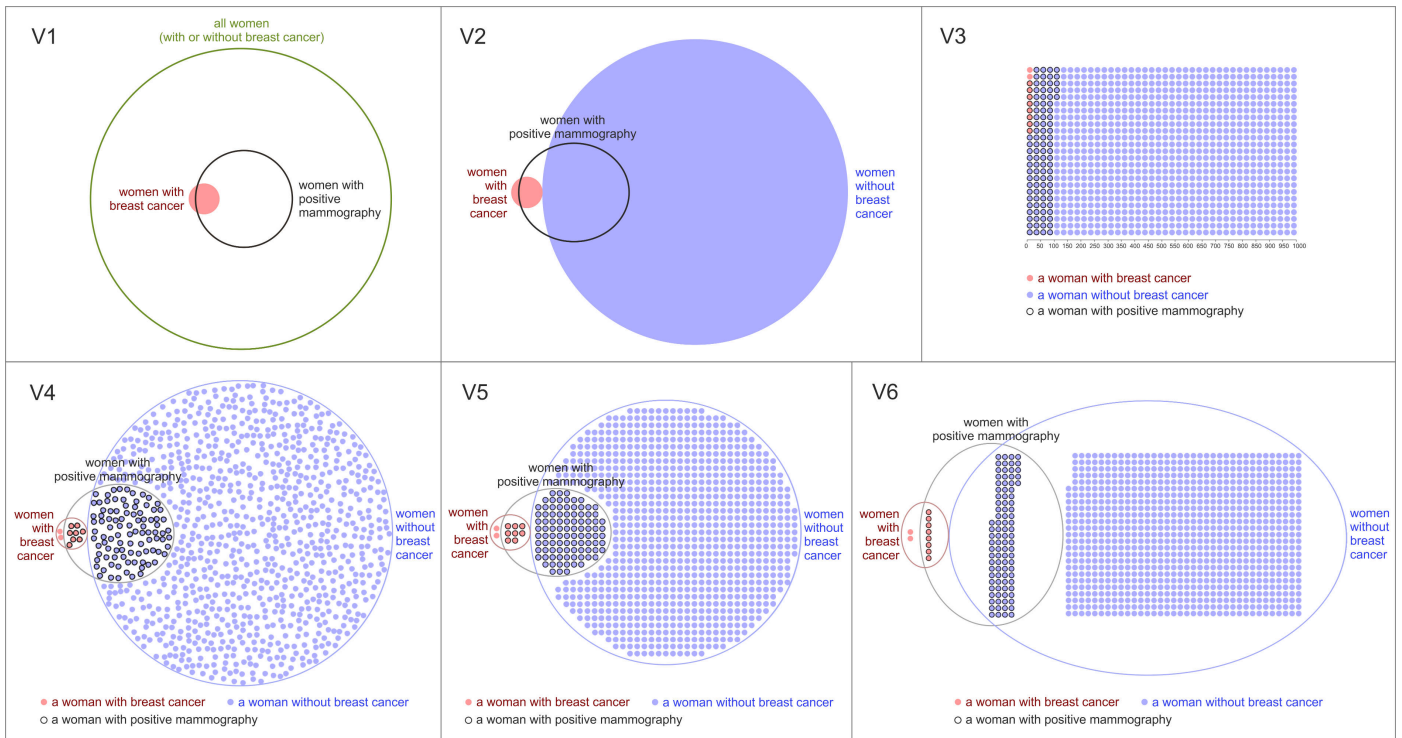


Fig. 2.2: The six visualizations evaluated in our first experiment, showing the mammography problem [Edd82].

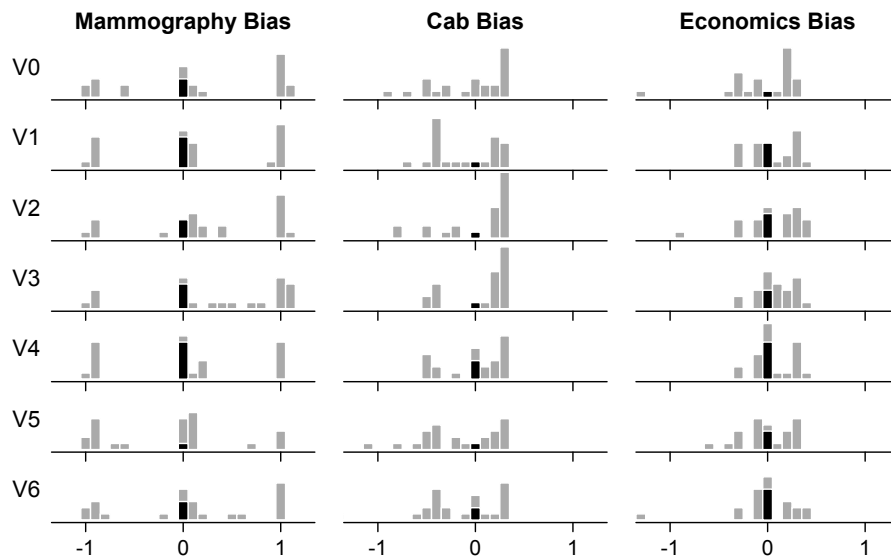


Fig. 2.3: Distributions of bias in answers per base rate problem and visualization condition ($N = 24$ each). Black bars are perfectly correct answers. A bias of -1 means an answer $10\times$ lower and a bias of 1 means $10\times$ higher.

be due to our population being more diverse than in previous studies, which generally involve highly-focused university students.

Open-text comments from participants revealed that the lack of benefit of visualizations may be partly due to people ignoring them. Various participants commented that they did not fully understand the visualization, that having two representations was confusing, or that they ignored it and that the information provided in the text was sufficient. Some doubted the credibility of the visualization and chose to trust the text. This blind focus on the numbers seems to have led many participants to miscalculate and arrive at answers that are off by an order of magnitude.

2.1.3 Second Experiment: Tweaking the Text

We ran a second experiment to confirm with a larger sample that simply adding a visualization to a base-rate problem is of little help, and to investigate alternative solutions. Since the answer to the base-rate problem is in the visualization itself (see again [Figure 2.1](#)), it should be possible to increase the chances that people find it, either by *i*) helping them make the link between the text and the visualization [[GG01](#)]; *ii*) encouraging them to search for the solution in the visualization; or *iii*) forcing them to search for the solution in the visualization.

We tested again V0 and V4, and added two alternative presentation techniques that only involved moderate modifications to the text. In the condition V4a, we added short instructions in the text that refer to the visualization. For example, after the sentence “8 of every 10 women with breast cancer will get a positive mammography”, we added “(compare the red dots that have a black border with the total number of red dots)”. This was meant to support strategy *i*) and possibly *ii*). In the condition V4b, we removed all numerical quantities from the text. For example, the previous sentence became “A large proportion of women with breast cancer will get a positive mammography”. This was meant to support strategy *iii*), and was based on the notion that an imprecise estimation can be better than a plain wrong calculation.

We tested those four conditions with 480 new participants (120 per condition), using only the mammography problem. The results for bias are reported in [Figure 2.4](#). The distributions are similar to those of the previous experiment, except for V4b which is closer to normal. Consistent with our expectations, V4b rarely yielded a perfectly correct answer (only 1 out of 120 – a participant who reported taking a screenshot and counting

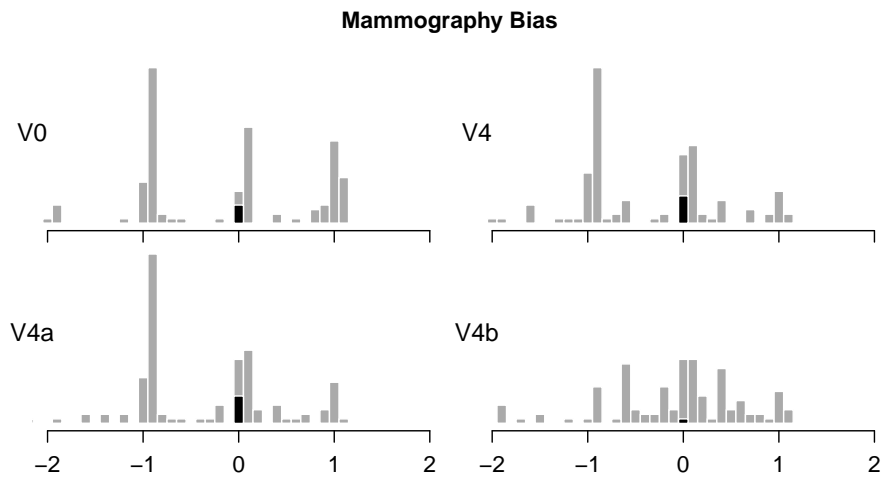


Fig. 2.4: Distributions of biases in answers to the *mammography* problem ($N = 120$ each) per presentation type. Black bars are perfectly correct answers.

the dots using MS Paint), but it yielded more accurate answers on average. The median error was 0.41, which is much lower than with V0, V4 and V4a (all 0.89). In addition, V4b yielded almost no average bias, whereas with other visual presentations subjects tended to underestimate probabilities.

2.1.4 Conclusions

In this study, we used crowdsourcing to assess the effect of six visualizations (based on Euler diagrams, glyphs and combinations of both) on base-rate reasoning using three classic problems in psychology. To our surprise, participants' accuracy was remarkably low and did not substantially improve when a visualization was provided with the text. A follow-up experiment with higher sample size confirmed that simply adding a visualization to a textual base-rate problem is of little help for crowdsourcing participants. It however revealed that removing all numbers from the text in order to force participants to perform visual estimation with the visualization substantially reduces probability estimation errors. Thus, novel representations that holistically combine text and visualizations in a way that promotes the use of visual estimation rather than numerical calculation deserve to be investigated.



Fig. 2.5: Illustration of the attraction effect in a hypothetical election: Bob has an excellent education plan, while Alice is very strong in crime control. The addition of Eve, a candidate similar but slightly inferior to Alice, raises Alice’s attractiveness as a candidate. This irrelevant option is called a decoy.

2.2 A Cognitive Bias Called The Attraction Effect

The attraction effect is a well-studied cognitive bias in decision making research, where people’s choice between two alternatives is influenced by the presence of an irrelevant (dominated) third alternative. In this section, we examine whether this cognitive bias, so far only tested with three alternatives and simple presentation formats such as numerical tables, text and pictures, also appears in visualizations. This section is based on a 2016 paper authored by Evanthia Dimara, Anastasia Bezerianos, and myself [DBD17c]. Supplementary material (data, code, analyses) is available at <https://osf.io/kfhme/> and <https://aviz.fr/decoy>.

2.2.1 Introduction

Suppose you are voting for primary elections and need to choose between candidates **Bob** and **Alice** (Figure 2.5). Bob has a solid education plan, while Alice has an excellent strategy for crime control. There is no obviously wrong choice: it all depends on how much you care about education and safety. But now suppose there is a third candidate, **Eve**. Like Alice, Eve is stronger on crime control, but her crime control plan is not as good as Alice’s. Should her presence influence your decision? Normally not, as Eve is inferior in all respects and therefore irrelevant to your choice [HPP82]. And yet countless studies have shown that adding Eve make people more likely to chose Alice in similar decision-making tasks [OP95].

This shift in preference is called the *attraction effect* (also known as the *decoy effect* and the *asymmetric dominance effect*). Studies suggest the attraction effect is quite general and robust, e.g., it occurs when people choose consumer products like beers, cars, or films [HPP82], when they gamble [Wed91], select candidates to hire [Hig96], choose a meal in a menu [HJW12], decide which suspect committed a crime [Tru12], or vote [OP95]. Even animals like hummingbirds [BHH02], bees [SWS02], and amoebae [LB11] appear to be subject to the attraction effect when selecting their food.

Like other cognitive biases, the attraction effect leads to suboptimal decisions that can have important implications in many areas such as politics and advertising. Our goal in this study was to find out whether the attraction effect also has implications for information visualization design. If someone uses a visualization to choose among several alternatives (e.g., when buying an apartment [WS92]), will the presence of inferior alternatives affect their decision? In other words, does the attraction effect transfer to visualizations?

2.2.2 First Experiment: Two to Three Alternatives

The purpose of our first experiment was to replicate the design of a standard attraction effect experiment (two alternatives plus an optional decoy, all presented in a numerical table), and test if the effect persists when alternatives are shown using a scatterplot visualization.

We reproduced the first experiment by Malkoc et al. [MHH13], which is representative of a classical attraction effect experiment. In this experiment, participants were presented a fictional scenario where they had to pick a fitness club. Each fitness club was defined by its variety and its cleanliness, both rated from -10 to +10. Some participants were given the choice between two fitness clubs (equivalent to showing only Bob and Alice in Figure 2.5), while others had to choose among three fitness clubs, one of which was a decoy (equivalent to choosing between Bob, Alice, and Eve). The attraction effect is measured by looking at the proportion of people choosing Alice when the decoy is present vs. absent. As many studies before, Malkoc et al.'s experiment used a numerical table to convey the different alternatives and their attributes (see top of Figure 2.6). We did the same, but added conditions where the same information was conveyed with scatterplots (bottom of Figure 2.6).

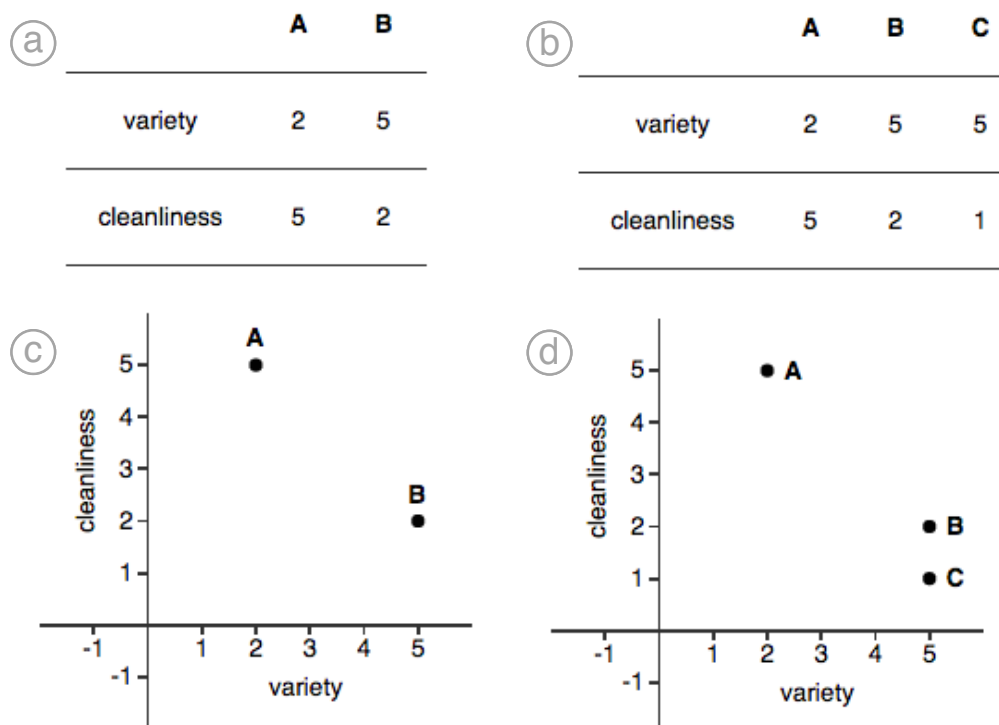


Fig. 2.6: Examples of experimental stimuli for the table (a,b) and the scatterplot (c,d) conditions. The left decision task (a,c) has no decoy, while the right decision task (b,d) includes a decoy on the *variety* attribute.

In [Figure 2.6-right](#), the decoy C is similar to (and dominated by) B. Because it is expected to make B more attractive, B is commonly referred to as *the target*. In addition, because C presumably makes B more attractive by making variety look more important as an attribute, it is conventionally referred to as a *decoy on variety*. Some participants saw a *decoy on cleanliness* instead, with A being the target (stimuli not shown on the figure); We will call this other decoy D. To balance out possible layout effects, we varied the ordering of the alternatives and the attributes in the table (an improvement over Malkoc et al.), as well as the mapping of attributes to the x and y axes of the scatterplot. In total, we used 18 stimuli, which we randomly assigned to participants (each participant only saw one stimulus). However, we only considered two independent variables in our analysis: the *decision task*, defined by the set of gyms to choose from: {A, B} (no decoy), {A, B, C} (decoy on variety), or {A, B, D} (decoy on cleanliness); and the *presentation format*: table or scatterplot.

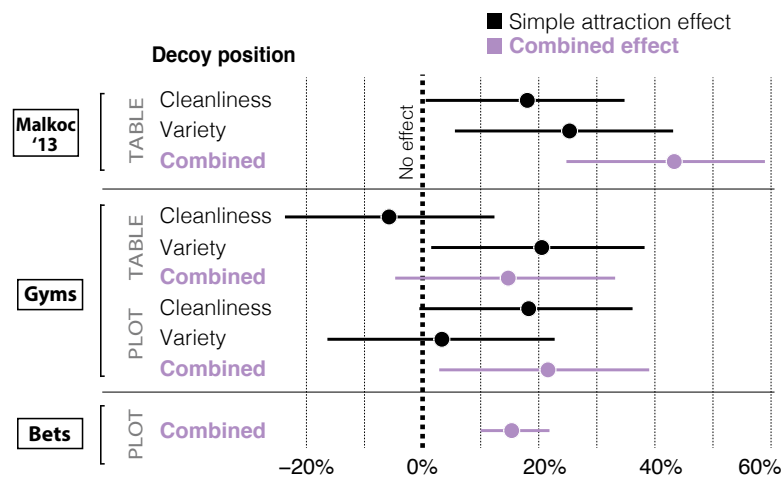


Fig. 2.7: Point estimates and 95% confidence intervals for the attraction effects in Malkoc et al. [MHH13], and in our two experiments.

We collected data from 305 participants in a crowdsourcing platform. The results are in Figure 2.7, in the middle area labeled *Gyms*. The top area labeled *Malkoc '13* reports the results from the original experiment, for comparison. The bottom area labeled *Bets* will be discussed in the next section. Back to the *Gyms* area, the top three error bars report are results for numerical tables; The top error bar indicates the difference in the percentage of participants who chose the target when there was a decoy on cleanliness vs. no decoy, with its 95% confidence interval. A positive value (to the right of the dashed line) indicates an attraction effect. Since the CI clearly crosses 0%, the result is inconclusive.³ The next error bar reports the results for the decoy on variety: here, there is evidence of an attraction effect. Finally, the purple error bar reports the *combined attraction effect*, which directly compares the two decoy conditions with each other. There is again some evidence of an attraction effect. The next three rows show the same results, but for the scatterplot condition. Again, we only see evidence for one of the two decision tasks, but when we combine both, we see evidence of an attraction effect.

Overall, we replicated the attraction effect that has been previously observed with numerical tables (although our effect is smaller than in Malkoc et al. [MHH13]), and we also found evidence of an attraction effect with scatterplots. However, using a scatterplot to visualize only three alternatives does not capture most situations where visualizations are used for decision support.

³I do not report any p-value, here or anywhere in this report, for reasons I explain in [Dra16]. I draw my inferences from confidence intervals, which convey the same information as p-values, and more.

2.2.3 Second Experiment: More Alternatives

In this second experiment, we looked at whether the attraction effect persists in more realistic situations where more than two or three alternatives are visualized.⁴ We also attempted to increase statistical power in order to get stronger evidence.

To increase statistical power, we used a within-subjects design. Such designs are less common in previous studies but have been occasionally used. Wedell [Wed91], for example, measured a clear attraction effect with numerical tables using a within-subjects procedure, where participants were given multiple decision tasks. He further increased statistical power by excluding no-decoy conditions and only measuring the combined decoy effect. We therefore decided to replicate Wedell's design.

As in Wedell's experiment, our tasks involved choosing among three lottery tickets, each defined by two attributes: the *probability* of winning, and the amount that can be won (the *prize*). Participants were presented with twenty such decision tasks in sequence. Contrary to Wedell, our experiment was incentivized for higher external validity: upon completion of all tasks, the twenty lotteries were ran by a computer and the results were used to determine a bonus reward for each participant.

We only tested scatterplots. Figure 2.8 shows two of our experimental stimuli (floating boxes added by us). In contrast to Wedell and to our previous experiment, the number of alternatives to choose from was substantially higher than three. On the left plot, the participant is asked to choose among approximately twenty lottery tickets defined by their probability to win (*x-axis*) and their prize (*y-axis*). Two tickets dominate others in terms of their expected value (A and C), so participants were expected to choose either one of them. However, both A and C have the same expected value so there is no right answer, just as with A and B in Figure 2.6 or Bob and Alice in Figure 2.5. The other dots are either decoys dominated by A or other dominated alternatives we call distractors. This stimulus was paired with another stimulus presented at another time during the sequence of 20 decision tasks (Figure 2.8-right), and where decoys were dominated by C. If participants tend to pick ticket A more often with the left stimulus than with the right one, this is evidence for an attraction effect.

To make it more difficult for participants to infer patterns in the sequence of decision tasks, we inserted irrelevant decision tasks at various positions, which were not used in

⁴Prior to this experiment, we ran another (inconclusive) experiment that is reported in full in a research report [DBD16].

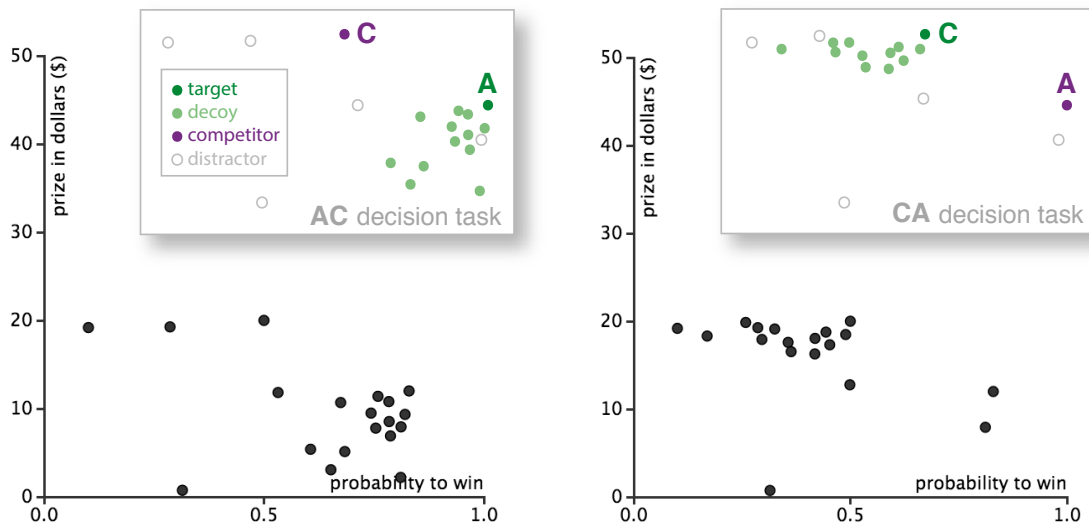


Fig. 2.8: Experimental stimuli for the two matched decision tasks *AC* and *CA* (black-and-white background images), and explanatory annotations (box overlays). Each dot is a lottery ticket.

our analyses. These tasks differed in that they had either one or three non-dominated alternatives (instead of two), and they did not exhibit an imbalance in the number of asymmetrically dominated alternatives.

We used an aggregate measure of attraction effect that uses participant’s responses to all twenty stimuli (10 paired stimuli). The result is reported at the bottom of [Figure 2.7](#), in the area labeled *Bets*. Although the effect size is likely not larger than in the previous experiment, this time the statistical evidence is remarkably strong.

2.2.4 Conclusions

In this study, we showed that the attraction effect generalizes to data visualizations. In other words, when people use scatterplots to visualize multiple alternatives to choose from, the number and position of irrelevant (dominated) alternatives may influence their choice. This shift in preferences violates basic axioms of rational choice theory [[HPP82](#)]. In addition to being the first study on the attraction effect in information visualization, our study contributes to ongoing debates in decision-making research on whether the effect generalizes to non-numerical formats [[FLB14](#); [HPP14](#); [Sim14](#); [YL14](#)], and whether it should disappear as more alternatives are added [[BLP98](#)]. However, more realistic

datasets remain to be tested, and our inconclusive results with real datasets (reported separately [DBD16]) suggest that the effects with real-world datasets may be small and hard to detect.

2.3 Evaluating Visualizations for Decision Making

So far we have summarized two studies addressing specific questions related to judgment and decision making with visualizations. But how should we evaluate visualizations for judgment and decision making more generally? Although this is a vast question, this section summarizes three contributions that bring us slightly closer to a response. The first one looks at what to measure when we want to compare elementary multidimensional visualizations for their ability to support multi-attribute choice tasks. The second one looks at whether or not we should use narratives in task instructions when we evaluate visualizations, both for analytical tasks and decision tasks. Finally, the third contribution helps us identify which cognitive biases could be interesting to study in a data visualization context.

2.3.1 Evaluating Elementary Multidimensional Visualizations

This subsection is based on a 2017 paper authored by Evanthia Dimara, Anastasia Bezerianos, and myself [DBD17a]. Supplementary material (data, code, analyses) is available at <https://aviz.fr/dm>.

Suppose we want to book a hotel, and we have access to a dataset with about one hundred hotels with a dozen attributes such as price, room size, bed size, or user ratings. We refer to tasks like these as *multi-attribute choice* tasks: tasks that consists of finding the best alternative among a fixed set, where alternatives are characterized across several attributes with known values, but where the objective function is unknown to the system (it is inside the user's head). What are the best visualizations for tasks like these?

On the one hand, a dataset like the hotel dataset is trivial to visualize: it is well-structured (a table with columns having a well-defined type), it is fairly small in visualization standards, and there is no missing or uncertain data. Many general-purpose visualization techniques exist that can visualize such a dataset, including scatterplot matrices, parallel coordinates, and tabular visualizations [KK96; OL03; PDF14]. There are also visualization

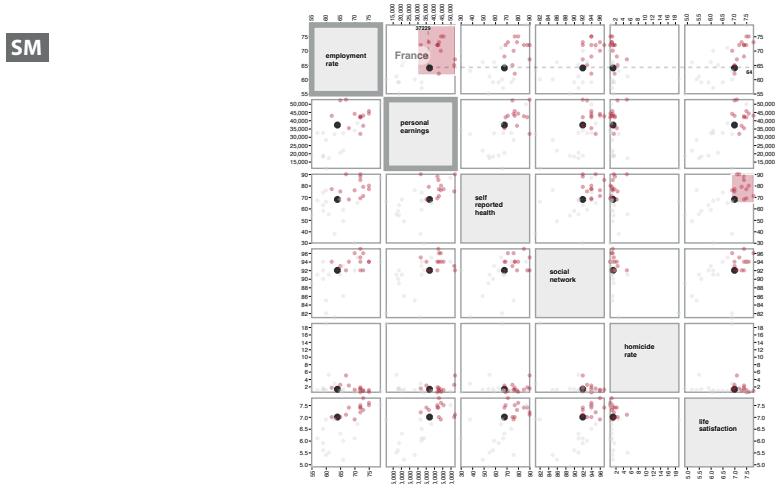
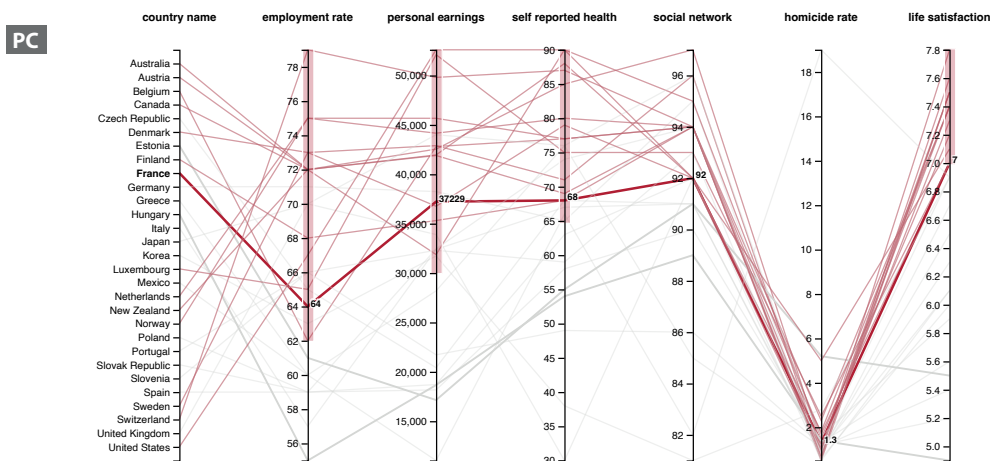
systems specifically designed for multi-attribute choice tasks, many of which use general-purpose visualizations with extra features [Paj+17; CL04; Gra+13].

But surprisingly, there is very little empirical data to help us decide which tool to use. Most existing studies are either qualitative studies without a comparison baseline [Yi+05; Gra+13; ATS95; Paj+17; PF00; AAJ03; AA03], or use elementary analytic tasks such as value retrieval [Kua+12; WS92] or correlation estimation [LMV10; Yi+05]. Low-level analytic tasks like these are important but they are not decision tasks. How do we evaluate visualization tools for their ability to support multi-attribute choice?

To answer this, we identified a range of subjective and objective metrics and used them to compare three elementary multidimensional visualizations in a within-subject study (see Figure 2.9). We first explained the techniques to the participants and gave them low-level analytic tasks (value retrieval, range, correlation). Such tasks made sure that participants were fully trained and allowed us to ensure that they properly understood the techniques when they performed the choice tasks, thus eliminating potential confounds (e.g., a technique yielding poor decisions because participants did not know how to use it). A dual evaluation may also uncover potentially interesting interactions between a technique's ability to support analytic tasks and its ability to support decision tasks.

After performing all analytic tasks with all techniques, participants were told they would now use the techniques to make a personal choice. They were asked to imagine planning their vacations and looking for the ideal holiday package. We used three synthetically generated datasets containing 100 data cases each (holiday packages) and 8 dimensions: price per person (euro/day), hotel quality, archaeological interest, landscape interest, night life interest, security level, sport activity level, and kids friendly. Except for price, all dimensions were ratings from 0 to 100. The meaning of each attribute was explained to the participants. Participants performed three such decision tasks, each time with a different technique and a different dataset (fixed ordering for datasets, counterbalanced ordering for techniques).

There is no simple way to define the accuracy of a multi-attribute choice task, given its subjective nature. Although Pareto dominance is one such measure [DBD17c], participants are unlikely to select a dominated alternative given the number of alternatives and attributes. We thus decided to use as an indicative measure of accuracy the *consistency between the choice made by a participant and their self-reported choice criteria*.



TV

country name	employment rate	personal earnings	self reported health	social network	homicide rate	life satisfaction
Australia	72	42550	85	97	1.2	7.4
Austria	72	41904	69	94	0.5	7.5
Belgium	82	42811	77	94	1.7	7
Canada	72	41961	88	92	1.8	7.4
Czech Republic	65	20424	68	91	0.9	6.3
Denmark	73	42904	71	96	0.9	7.8
Estonia	61	17145	55	91	5.2	5.5
Finland	88	35319	68	94	2.3	7.4
France	64	37229	88	92	1.3	7
Germany	71	38251	65	95	0.8	6.7
Greece	60	28900	76	85	1.1	5.4
Hungary	55	18667	54	89	1.4	4.9
Italy	57	32404	67	91	1	6.1
Japan	70	33900	30	92	0.5	6.1
Korea	63	31733	38	81	2.8	6.9
Luxembourg	85	52110	75	93	2.5	7
Mexico	90	11020	66	82	19	6.9
Netherlands	75	45671	77	94	1.1	7.5
New Zealand	72	31878	90	95	1.5	7.2
Norway	75	44164	80	94	0.6	7.6
Poland	59	18172	58	90	1.3	5.6
Portugal	66	21722	49	86	1.2	5.2
Slovak Republic	59	18719	63	92	1.6	5.9
Slovenia	66	32308	59	93	0.6	6
Spain	59	32454	74	94	0.9	6.5
Sweden	73	36766	79	92	1	7.3
Switzerland	79	49810	87	94	0.7	7.5
United Kingdom	70	44008	76	96	1.2	6.9
United States	67	52607	90	92	5	7.1

Fig. 2.9: The three visualizations we evaluated: parallel coordinates (PC), scatterplot matrix (SM), and tabular visualization (TV). All three visualizations supported basic interactions for highlighting individual data cases, selecting ranges of attribute values, and reordering dimensions. Here we see a training dataset of country indicators, with 29 data cases and 6 dimensions. The datasets used in the experimental tasks (both analytic and decision) had 100 data cases and 8 dimensions.

Thus before the first decision task, participants were asked to rate how *important* each of the 8 holiday package attributes was to them, on a scale from 0 to 10. They were also asked to indicate the *direction* of their preference, i.e., whether they prefer the attribute to be high or low. For example, a holiday package with lots of physical activity can be perceived as desirable by an athletic person but undesirable by someone with limited mobility. As preferences may evolve during the session, the questionnaire was administered before and after each decision task (4× total).

We used participants' preference ratings to roughly estimate how desirable each alternative *should be* using a normalized weighted sum approach [Tri13], resulting in a desirability score of $d=1$ for the “best” alternative in the dataset and $d=0$ for the worst alternative (details in our paper [DBD17a]). Desirability scores can be computed using the preferences elicited either before the decision task (d_{pre}), or after the task (d_{post}). Since preferences can evolve while exploring options, d_{post} may seem more indicative of the “true” desirability. However, a participant may also update their preferences *after* the choice was made, e.g., as a way of rationalizing their choice. Thus, we consider both d_{pre} and d_{post} and define the *accuracy* of a decision task as $a = \max(d_{pre}, d_{post})$, with d_{pre} and d_{post} being the desirability scores of the chosen alternative.

This decision accuracy score is an approximation and is not meant to capture decision quality perfectly. The elicited criteria may not be completely reliable, and cannot fully capture the complexity of real choice criteria (i.e., someone may want an attribute to be neither too high nor too low). However, if a visualization happens to be misleading or particularly hard to use, we should expect participants to make choices that are clearly inconsistent with their criteria, thus yielding an abnormally low decision accuracy.

After each decision, participants were also asked to evaluate the choice they just made on four subjective dimensions: *satisfaction* with their choice, *confidence* in their choice, the *easiness* with which they made their choice, and their *attachment* to their choice (measured by asking them if they would switch to the choice proposed by a hypothetical recommender system). Finally, at the end of the experiment, participants were asked which technique they prefer overall for a) analytic tasks, and b) decision tasks. Participants were timed on all tasks.

We ran a lab experiment with 21 volunteers having different levels of reported expertise in data visualization (range 2–9 on a 0–10 scale, average 6). The experiment lasted 1 hour and a half on average. As we analyze multiple dependent variables, we are subject to statistical multiplicity issues and any finding should be taken as tentative. With that

said, our paper explicitly distinguishes between planned and unplanned analyses, and so in contrast to typical exploratory analyses, there is no hidden multiplicity [Ber07].

Overall, we found the three visualizations to be comparable on most metrics, with perhaps a slight advantage for tabular visualizations. More specifically, all three techniques yielded close-to-perfect accuracy in analytic tasks, confirming that participants were sufficiently trained. There were however large differences in completion times: the scatterplot matrix was slowest for value retrieval and range tasks, but by far the fastest in correlation tasks. This result matches previous intuitions and findings on visualization for analytic tasks.

In terms of actual decision-making tasks, we found our techniques to be comparable across most metrics, with a slight speed advantage for the tabular visualization. Participants also preferred the tabular visualization over the scatterplot matrix overall. However, they reported being more attached to choices they made with the scatterplot matrix on average, a result that needs to be confirmed by further studies. If the effect is real, one explanation could be that the scatterplot matrix's better support for overview tasks (confirmed by our results with the correlation task) made participants more confident that they did not miss a particularly interesting alternative.

Our accuracy metrics for decision support showed a much larger variability in responses than analytic tasks. Perhaps as a result, many of our metrics were not sensitive enough to capture differences across conditions that likely exist [Coh94]. Additional work is needed on establishing more sensitive metrics of choice quality. Nevertheless, it is interesting that tabular visualizations allowed participants to reach decisions of similar accuracy faster. Thus, although decision time is rarely central when assessing decision support, it could be used as a tie-breaker when visualizations achieve similar decision accuracy.

Again, all these findings are tentative and need to be confirmed in separate experiments. In the meantime, with the data we have so far, tabular visualizations seem to be a compelling choice for supporting multi-attribute choice tasks, despite the low attention they have received in the literature on multidimensional visualization [PDF14]. Nevertheless, several decision-support visualization tools use a tabular layout [CL04; Gra+13; Paj+17], and our experiment provides preliminary empirical justification for this design choice.

2.3.2 Using Narratives in Crowdsourcing Studies

This subsection is based on a 2017 paper authored by Evanthia Dimara, Anastasia Bezerianos, and myself [DBD17b]. Supplementary material (data, code, analyses) is available at <https://aviz.fr/narratives>.

Crowdsourcing enables rapid evaluation of visualizations [HB10; KCS08; Bou+12; Boy+14], including for judgment and decision making tasks [MDF12; DBD17c]. However, engaging contributors and obtaining high-quality responses is hard [Hul11; ECD14]. To address this, visualization studies often use task instructions with narrative elements in order to engage participants, help simulate the real use of a system, and elicit a more representative user behavior. A minimalist form of narrative consists in attributing a meaning to the datasets – for example, to evaluate HomeFinder, Williamson et al. [WS92] used questions such as “*what neighborhood has the most expensive houses?*”. Narratives can also invite participants to imagine themselves in a hypothetical decision situation. For example, Yi et al. [Yi+05] evaluated their Dust & Magnet system by asking participants to choose a cereal brand.

Unsurprisingly, hypothetical decision scenarios are common when evaluating visualization tools for decision support. For example, Aseniero et al. [Ase+15] evaluated a tool designed for software release plans by instructing participants to imagine taking the role of a project manager identifying an optimal plan. In order to evaluate a tool for preferential choices, Bautista and Carenini [BC08] immersed participants in shopping scenarios and put them in a situation of finding a hotel. Similarly, Daradkeh et al. [DCM13] asked participants to make hypothetical investments. Finally, the very study presented in the previous subsection used a hypothetical scenario where participants had to choose a holiday package.

Narratives are commonly used when evaluating visualizations, but the real effects of adding narrative elements to task instructions are unclear, in particular in crowdsourcing settings where incentives vary across people, and attention and motivation are hard to control for. For example, what would be the difference between instructing participants to identify the data point with the minimum X value, and instructing them to imagine they are trying to find the cheapest available house? Both versions are equivalent at the task level and consist of finding an extremum on a particular data dimension. The second version is possibly more salient and engaging, and with a context that is easy to understand, characteristics linked to good crowdsourcing performance [Hul11; MW10].

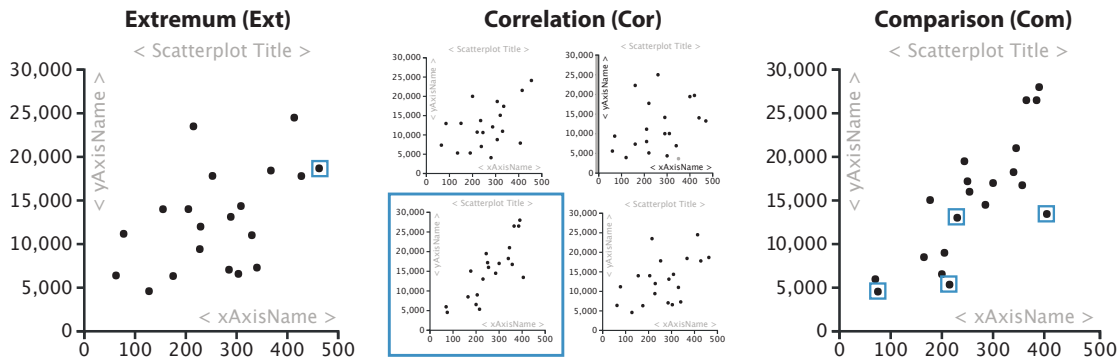


Fig. 2.10: The three tasks used in our experiment, with correct answers in blue (annotations ours). Plot labels depended on the condition – for example, scatterplot axes were labeled [X, Y] for ABS, and [size (m²), price (\$)] in the narrative conditions. In the Ext and Com tasks, the scatterplots supported basic interactions: hovering over a data point highlighted it, displayed horizontal and vertical projection lines, and overlaid the data point’s X and Y values on the axes.

However, the first version is more succinct and less demanding in terms of time and patience, aspects that have also been emphasized in crowdsourcing guidelines [ECD14].

To help address this knowledge gap, we ran a study to find out if crowdsourcing participants perform differently in simple visualization tasks when task instructions are enhanced with narrative components. We gave them three basic visualization tasks inspired from previous task taxonomies [AES05; WL90; RM90], that were analytical in nature but could be mapped to decision tasks:

- An **extremum** task (**Ext**), where participants had to find the data point with highest value according to the X dimension (leftmost scatterplot in Figure 2.10).
- A **correlation** task (**Cor**), where participants had to find the scatterplot with the highest correlation among four different ones (second panel in Figure 2.10).
- A **comparison** task (**Com**), where participants had to compare data points across their two dimensions simultaneously (third panel in Figure 2.10). The task required finding a data point without any “competitor”, a competitor being defined as a data point that has both larger X and smaller Y . There were four possible correct answers.

All participants did all three tasks in sequence, but saw different instructions depending on the condition they were randomly assigned to. Each task came with five different sets of instructions, broken down into two pages (see Table 2.1 for an overview of all textual instructions):

Condition	Task	Page 1	Page 2
ABS	Ext	You will be asked to answer a few questions about data. In the next page you will see many data points displayed in a diagram.	Which is the data point with the largest value of X?
	Cor	Now you will see four diagrams with data points. You will be asked to compare them.	In which of these four diagrams is Y most related to X?
	Com	Now you will see one of the previous diagrams again. You will be asked a question that requires identifying "competitors". In our case, a data point is a competitor of another data point if it has both larger X and smaller Y.	Select a data point that has no competitor.
SEM	Ext	You will be asked to answer a few questions about houses. In the next page you will see many houses displayed in a diagram.	Which is the biggest house?
	Cor	Now you will see four diagrams with houses. You will be asked to compare them.	In which of these four diagrams is price most related to size?
	Com	Now you will see one of the previous diagrams again. You will be asked a question that requires identifying "competitors". A house is a competitor of another house if it is both bigger and cheaper.	Select a house that has no competitor.
AN-NAR	Ext	You will be asked to answer a few questions about houses. Imagine that you are a real estate analyst and you need to understand the house market. You focus on extremely rich customers who seek to buy a house that is as big as possible. In the next page you will see the houses currently on the market, displayed in a diagram.	Given what you read, which house would be the most attractive to your customers?
	Cor	Now you want to focus on regular customers who are not necessarily very rich. You want to investigate how reliable some real estate agencies are. You will see four diagrams with houses. Each diagram shows the houses proposed by a different agency. An agency that sets arbitrary prices is NOT reliable. While in a reliable agency, price is very related to size.	Given what you read, which of these four real estate agencies is the most reliable?
	Com	Now you will see one of the previous diagrams again. It shows the houses offered by the best agency. You need to report on their best deals. A good deal is a house that has no "competitor". A house is a competitor of another house if it is both bigger and cheaper.	Given what you read, select a house that is a good deal.
DM-NAR	Ext	You will be asked to make a few decisions about houses. Imagine you are moving to a new city and you need to buy a house. You are extremely rich and you want your house to be as big as possible. In the next page you will see the houses currently on the market, displayed in a diagram.	Given what you read, which house would you buy?
	Cor	You don't have as much money as you initially thought. So before buying a house, you need to find a reliable real estate agency. You will see four diagrams with houses. Each diagram shows the houses proposed by a different agency. An agency that sets arbitrary prices is NOT reliable. While in a reliable agency, price is very related to size.	Given what you read, which of these four real estate agencies would you choose?
	Com	Now you will see one of the previous diagrams again. It shows the houses offered by the best agency. You will finally get to choose your house. A good choice is a house that has no "competitor". A house is a competitor of another house if it is both bigger and cheaper.	Given what you read, which house would you buy?

Tab. 2.1: The text instructions in each experiment condition. The condition O-NAR (not listed here) had the same page 1 as AN-NAR, and the same page 2 as SEM.

- In the **abstract** (ABS) condition, the dataset had no specific meaning. Both page 1 and page 2 used abstract wordings. An example of question was “*Which is the data point with the largest value of X?*”.
- In the **simple semantics** (SEM) condition, data points were houses and dimensions were price and size. Questions were of the type “*Which is the biggest house?*”.
- In the **analytic narrative** (AN-NAR) condition, page 1 contained a narrative that asked participants to put themselves in the situation of a real estate analyst and to find answers to analytical questions. An example of question on page 2 was “*Given what you read, which house would be the most attractive to your customers?*”.
- In the **decision-making narrative** (DM-NAR) condition, page 1 had a a narrative that asked participants to put themselves in the situation of a house buyer and, given some criteria and constraints, to make choices. Questions were of the type “*Given what you read, which house would you buy?*”.
- In the **optional narrative** (O-NAR) condition, page 1 was identical to AN-NAR and page 2 was identical to SEM. An example of question was “*Which is the biggest house?*”. In contrast with AN-NAR, participants did not have to read the narrative on page 1 to be able to answer the question.

To assess the merits of the different instructions, we used both objective performance metrics and self-reported impressions. In terms of objective metrics, we used a normalized measure of *accuracy* ranging from 0 to 1 for each of the three tasks. To see if task instructions affect the degree to which participants pay attention to the tasks, we also measured *in-task attention* by testing their ability to recall the options presented to them in the correlation task, after all tasks were completed. To further assess attention fatigue, we measured participant’s *post-task attention* through an instructional manipulation check [OMD09] at the end of the experiment. As for the subjective metrics, we asked participants to report their *confidence* in their answer and the *easiness* with which they arrived at their answer for each of the three tasks, and after all three tasks were completed, the overall *enjoyability* of the experiment and the *usefulness* of the visualizations for completing the tasks.

We recruited 405 crowdsourcing workers (about 80 per condition). As with the previous study, the strength of our findings is somehow weakened by the use of multiple dependent variables, but there is no hidden statistical multiplicity in our planned analyses.

We expected that providing task context in the form of data semantics or backstory narratives would improve the overall quality of responses. However, adding semantics to the data (i.e., using SEM instead of ABS) did not seem to help much, and some of

the narratives we used might have even harmed performance. In particular, it appears that participants who were given the decision-making narrative (DM-NAR) performed less accurately on average than those who were only given minimal context (SEM) or no context at all (ABS). As for in-task attention, participants exhibited a better recall of the correlation task when given minimal semantics (SEM) than no context (ABS), suggesting they were paying more attention. However, adding a backstory narrative on top of that (AN-NAR or DM-NAR) decreased their recall, bringing it to the same level as ABS. For post-task attention, the results were mostly inconclusive.

Why do narratives sometimes seem to harm performance? As we mentioned, crowdworkers generally appreciate succinct instructions [ECD14]. An otherwise simple task can appear more demanding in attention and time if it requires reading a long (in crowdsourcing standards) piece of text beforehand. Also, experienced contributors are generally used to performing abstract and mechanical tasks since these abound on crowdsourcing platforms. The fairly good performances we observed for ABS do suggest contributors were overall able to understand the context-less tasks and willing to carry them out.

Despite our results on performance, we have strong evidence that adding data semantics improves subjective experience on a range of metrics (confidence, perceived easiness, enjoyability, and perceived usefulness of the visualization). With that said, our backstory narratives did not yield measurable subjective benefits compared to data semantics alone (SEM). Thus, even though crowdsourcing contributors appreciate working with meaningful data, they may not be particularly interested in more elaborate narratives and may prefer to focus on carrying out their task.

Overall, our study provides compelling reasons for incorporating data semantics in crowdsourced evaluations of visualizations, i.e., stating what the datasets and their dimensions mean. However, it also uncovered what could be termed a “double-edged sword effect” of narratives – thus it seems safer to use elaborate narratives parsimoniously, unless there are clear reasons to do so. When evaluating domain-specific and decision-support visualization systems, avoiding narratives is difficult; For example, if we want to evaluate a system meant to help customers choose a car, plain instructions such as “select the best car” may lack external validity, and it may seem more suitable to provide a backstory narrative to help participants imagine a real car purchase situation. Our study suggests that such narratives will lead to lower performance, but if efforts are made to keep the narrative short, the decrease should be relatively small and an acceptable price to pay for better external validity. Much more empirical work is needed to find out how to best design narratives, a question this study does not address.

2.3.3 A Taxonomy of Cognitive Biases to Study

Despite a growing interest in cognitive biases in information visualization, studies remain rare. Meanwhile, most experimental tasks studied in the cognitive bias research use textual representations. In order to bridge this gap between cognitive psychology and visualization research, we did an extensive review and analysis of the cognitive bias literature and derived a taxonomy of cognitive biases targeted to information visualization researchers. In this taxonomy of 156 cognitive biases, biases are classified by user task, instead of by proposals for psychological explanations of why biases occur, as is commonly the case for cognitive bias taxonomies.

This subsection is based on a 2018 paper by Evanthia Dimara, Steven Franconeri, Catherine Plaisant, Anastasia Bezerianos, and myself [Dim+18].

Our taxonomy is reproduced in Table 2.3. Each row lists a particular cognitive bias. Biases are organized into six different types of tasks, plus a category “other”. These tasks correspond to experimental tasks used in cognitive bias experiments, but they also map to real-world judgment and decision tasks:

- **Estimation:** in estimation tasks, people are asked to assess the value of a quantity. For example, a person may need to estimate the likelihood of theft to decide whether to insure their car, or their future financial needs in order to choose a retirement plan.
- **Decision:** decision tasks are any task involving the selection of one over several alternative options – for example, choosing a car to buy or a university to apply to.
- **Hypothesis assessment:** this category refers to any task involving an investigation of whether one or more propositions are true or false.
- **Causal attribution:** this category refers to any task involving an assessment of causality. For example, someone may need to decide whether a person had a car accident because the road was in bad condition, or because that person is not a good driver.
- **Recall:** this category includes all systematic biases that have been experimentally observed when participants were asked to recall or recognize previous material.
- **Opinion reporting:** this category includes all systematic biases that have been experimentally observed when participants were asked to answer questions regarding their beliefs or opinions on political, moral, or social issues.

To help group biases by similarity within each task category, we additionally categorized biases into *flavors* (second column in Table 2.3), which roughly capture the kind of cognitive phenomenon behind each bias. The flavors we identified are: **Association**,

where cognition is biased by associative connections between information items; **Baseline**, where cognition is biased by a comparison with (what is perceived as) a baseline; **Inertia**, where cognition is biased by the prospect of changing the current state; **Outcome**, where cognition is biased by how well something fits an expected or desired outcome; and **Self perspective**, where cognition is biased by a self-oriented view point.

In the taxonomy table, each bias is additionally described by one representative peer-reviewed paper (column *Ref*), the extent to which the bias might be relevant to visualization research (column *Relevance to InfoVis*), and a short description (column *Short description*). The color code for relevance to visualization is explained in [Table 2.2](#):

Tab. 2.2: Color code for relevance to visualization.

8	Evidence for the alleviation of the cognitive bias in visualization
7	Evidence for the existence of the cognitive bias in visualization
6	Studied in visualization, but no clear evidence of existence or alleviation
5	Discussed in visualization research as relevant, but not yet studied
4	Not discussed in visualization but likely relevant
3	Probably relevant to visualization
2	Potentially relevant to visualization
1	Relevance to visualization currently unclear

This taxonomy is not meant to be authoritative and is not claiming to be perfect, if only because many of its aspects (e.g., categorization in tasks and flavors, relevance to visualization) are the result of the authors' subjective judgment. The goal is only to give a convenient overview of the rich set of known cognitive biases for visualization researchers who may be interested in studying them, and some representative references as a starting point. We hope this taxonomy will inspire future studies and ultimately help visualization designers anticipate – and possibly alleviate – limitations in human judgment. Please note that this taxonomy is from 2018 and has not been updated – a few relevant studies have been published in information visualization since then.

Tab. 2.3: Our task-based taxonomy of cognitive biases.

#	Flavor	Cognitive bias	Ref	Relevance to InfoVis	Short description	
TASK: ESTIMATION						
1	Association	Availability bias	[TK73]	5 [DDB14; CH17]	Events more probable if easy to remember	
2		Conjunction fallacy	[TK83]	5 [ZC07]	Specific outcomes more probable than general	
3		Empathy gap	[Loe05]	1	Estimations affected by not recognizing the role of current emotional state	
4		Time-saving bias	[Sve08]	4	Overestimate time saved when increasing speed	
5	Baseline	Anchoring effect	[FB11]	7 [Cho+18; VZS17b]	Estimation affected by first piece of information	
6		Base rate fallacy	[BS07a]	6 [MDF12; Kha+15]	Ignore base rate probability of general population	
7		Dunning-Kruger effect	[KD99]	5 [VZS17a]	Low-ability people overestimate their performance (opposite for high-ability)	
8		Gambler's fallacy	[TK75]	4	Current outcome that is more frequent will be less frequent in future	
9		Hard-easy effect	[LF77]	3	Overconfidence for hard tasks, underconfidence for easy	
10		Hot-hand fallacy	[GVT85]	5 [CH17]	Current outcome that is more frequent will be more frequent in future	
11		Insensitivity to sample size	[TK75]	5 [ZC07; ED15]	Estimate probability ignoring sample size	
12		Regressive bias	[Att53]	4	Overestimate high probabilities, underestimate low ones	
13		Subadditivity effect	[TK94]	4	Overall probability less than the probabilities of the parts	
14		Weber-Fechner law	[CF14]	6 [Har+14]	Failure to perceive small differences in large quantities	
15	Inertia	Conservatism	[PE66]	7 [ZBK15]	New information insufficiently updates probability estimates	
16		Exaggerated expectation	[WK85]	4	Exaggerating evidence to fit a conclusion	
17	Outcome	Illusion of validity	[TK75]	5 [ZC07]	Overconfidence in judgment based on intuition and anecdotes	
18		Impact bias	[SS04]	1	Predict future emotional reactions as more intense	
19		Outcome bias	[BH88]	2	Evaluate decision maker only by choice outcome	
20		Planning fallacy	[BGP10]	5 [DJ14]	Overoptimistic task completion predictions, especially for self	
21		Restraint bias	[NHP09]	1	Overestimate of ability to resist temptation	
22		Sexual overperception bias	[Has03]	1	Over or underestimate of romantic interest from others	
23		Curse of knowledge	[KH02]	7 [XWF17]	Experts assume that novices have same knowledge	
24		Extrinsic incentives bias	[Hea99]	1	Others have extrinsic motivations (e.g.money), self are intrinsic (e.g.learning)	
25		False consensus effect	[RGH77]	2	Overestimate the agreement of others with own opinions	
26		Illusion of control	[Tho99]	3	Overestimation of one's influence on an external event	
27	Self-perspective	Illusion of transparency	[SG03]	1	Overestimate insight of others into own mental state, and vice versa	
28		Naive cynicism	[KG99]	2	Predict that the others will be more egocentrically biased	
29		Optimism bias	[Wei80]	4	Positive outcomes more probable for oneself than others	
30		Out-group homogeneity bias	[PR82]	4	Estimate out-group will be more homogenous than in-group members	
31		Pessimism bias	[Sha+07]	4	Positive outcomes less probable for oneself than others	
32		Spotlight effect	[GMS00]	1	Overestimate probability that people notice one's appearance/behavior	
33		Worse-than-average effect	[Kru99]	3	Underestimate own achievements relative to others in difficult tasks	
TASK: DECISION						
34	Association	Ambiguity effect	[RB90]	4	Choices affected by their association with unknown outcomes	
35		Authority bias	[Mil63]	1	Choices affected by their association with authority	
36		Automation bias	[MRO12]	5 [Sac+16]	Choices affected by their association with an automated system	
37		Framing effect	[TK81]	5 [ZC07; ED15]	Choices affected by whether alternatives are presented as gains or losses	
38		Hyperbolic discounting	[Tha81]	4	Choices affected by small short-term rewards	
39		Identifiable victim effect	[JL97]	6 [Boy+17]	Donation choices affected by whether victims are identifiable	
40		Loss aversion	[Tha+97]	7 [ZBK15; Das+15]	Choices affected by whether alternatives are gains or losses	
41		Neglect of probability	[Sun02]	4	Choices affected by disregard of probability	
42		Pseudocertainty effect	[TK86]	5 [ZC07]	Choices affected by whether some alternatives are framed as certain	
43		Zero-risk bias	[BGK93]	4	Choices affected by alternatives with complete risk elimination	
44	Baseline	Attraction effect	[HPP82]	8 [Dim+19; DBD17c]	Choices affected by irrelevant dominated alternatives	
45		Ballot names bias	[Woo+11]	6 [Woo+11]	Voting choices affected by the order of candidate names	
46		Cheerleader effect	[WV14]	1	Choices affected by whether people are in a group	
47		Compromise effect	[Sim89]	5 [DBD17c]	Choices affected if presented as extreme or average alternatives	
48		Denomination effect	[RS09]	4	Choices affected by whether the total amount comes for smaller currency bills	
49		Disposition effect	[WC98]	4	Selling choices affected by initial and not current value	
50		Distinction bias	[HZ04]	4	Choices affected by how many are the alternatives	
51		Less is better effect	[Hse98]	4	Choices affected if presented separately or juxtaposed	
52		Money illusion	[SDT97]	4	Choices affected by nominal monetary values	
53		Phantom effect	[PW00]	4	Choices affected by dominant but unavailable alternatives	
54	Inertia	Endowment effect	[MG15]	4	Choices affected by ownership of alternatives	
55		Escalation of commitment	[Sta76]	4	Choices affected by continued commitment to suboptimal outcome	
56		Functional fixedness	[GB05]	4	Choices of object use affected by the traditional way of use	
57		Mere-exposure effect	[Zaj01]	7 [Das+17; ED15; Kap+16]	Choices affected by familiarity (repeated exposure)	
58		Simmelweis reflex	[Mor+13]	3	Choices of medical practices affected by former established norms	
59		Shared information bias	[Bak10]	4	Group choices affected by sharing only known information	
60		Status quo bias	[SZ88]	5 [OPR16]	Choices affected by the urge to avoid a change, even when better expected value	
61		Well traveled road effect	[JJ82]	4	Travel route choices affected by road familiarity	
62		Outcome	Reactance	[Bre81]	4	Choices affected by the urge to do the opposite of what someone wants you to do
63			IKEA effect	[NMA12]	5 [SB17]	Choices affected by whether alternatives involved self-effort
64	Self-perspective	Not invented here	[AP15]	4	Choices affected by alternatives of origin external to an organization	
65		Reactive devaluation	[RS91]	4	Choices affected by whether alternatives allegedly originated with an antagonist	
66		Social comparison bias	[GST10]	1	Hiring choices affected by own competences	
TASK: HYPOTHESIS ASSESSMENT						
67	Association	Illusory truth effect	[HGT77]	4	Statement considered true after repeated exposure to it	
68		Rhyme as reason effect	[MT00]	1	Statement more likely true if it rhymes	
69		Barnum effect	[Tob+88]	2	High accuracy ratings for vague and general statements	
70	Outcome	Belief bias	[EBP83]	3	Hypothesis true if conclusion is believable	
71		Clustering illusion	[KT72]	5 [VZS17a; ED15]	Seeing patterns in noise, e.g. clusters in a dot field	
72		Confirmation bias	[Mah77]	5 [ED15]	Favor reasoning or information that confirms preferred hypothesis	
73		Congruence bias	[Was60]	4	Seeking confirmation of preferred hypothesis, but not for alternatives	
74		Experimenter effect	[RR97]	3	Subconsciously influence study participants to confirm a hypothesis	
75		Illusory correlation	[CC69]	5 [ZC07]	Perceived relationship between variables that does not exist	
76		Information bias	[BBH88]	4	Seek additional information irrelevant to a hypothesis or action	
77		Pareidolia	[VFP11]	4	Seeing faces in noise, e.g. in your toast	

#	Flavor	Cognitive bias	Ref	Relevance to InfoVis	Short description
TASK: CAUSAL ATTRIBUTION					
78		Group attribution error	[AM85]	2	Group traits extrapolated to an individual member
79		Hostile attribution bias	[GHW92]	1	Ambiguous intents read as hostile
80		Illusion of external agency	[Gil+00]	2	Positive outcomes attributed to mysterious external agents
81		Just-world hypothesis	[HB05]	2	What goes around comes around
82		System justification	[JPC02]	3	Inertial bias for unfair systems (e.g. slavery)
83		Actor-observer asymmetry	[MN75]	3	Failures of others due to behavior or personality, own failures due to situation
84		Defensive attribution hypothesis	[Sha70]	1	Failure or mishap of others judged by own similarity with the actor
85		Egocentric bias	[RS+79]	3	Own contribution overestimated
86	Self-perspective	Fundamental attribution error	[GM95]	1	Failures of others due to behavior or personality, own failures due to situation
87		In-group favoritism	[TD81]	1	Success and positive traits for ingroup members over outgroup
88		Self-serving bias	[CS99]	2	Own achievement attributed to behavior or personality, failures to situation
89		Ultimate attribution error	[Pet79]	3	Failures of outgroup due to behavior or personality, ingroup failures due to situation
TASK: RECALL					
90		Childhood amnesia	[UN93]	1	Harder to recall event details before certain age
91		Cryptomnesia	[BM89]	3	Memory mistaken for imagination, inspiration (e.g. unintentional plagiarism)
92		Cue-dependent forgetting	[TuL74]	4	Failure to recall information without memory cues
93		Digital amnesia	[SLW11]	3	Less likely to remember easily searchable information
94		Duration neglect	[FK93]	2	Recall unpleasant experiences according to intensity, ignoring duration
95		Fading affect bias	[WS09]	1	Emotion of unpleasant events fades, but pleasant does not
96		False memory	[LP74]	1	Imagination mistaken for a memory
97		Humor effect	[SLH10]	3	Easier to recall humorous items
98		Leveling and sharpening	[KGP00]	4	Recall sharpens some features, weakens others
99		Levels-of-processing effect	[CT75]	4	Easier to recall result of deep level analysis
100		Misinformation effect	[PKF09]	3	Recall colored by new information
101		Modality effect	[GVPO8]	4	Easier to recall items presented auditorily than visually
102		Mood-congruent memory	[MMS95]	1	Recall biased toward mood-congruent memories
103	Association	Next-in-line effect	[Bre73]	2	Failure to recall words of previous speaker in turns speaking
104		Part-list cueing effect	[Sla68]	4	Harder to recall material after reexposure to subset
105		Picture superiority effect	[MD02]	4	Easier to recall images (symbolic representations) than words
106		Positivity effect	[MC05]	1	Easier to recall positive events than negative
107		Processing difficulty effect	[OM85]	4	Easier to recall information which was hard to comprehend
108		Reminiscence bump	[JP96]	1	Easier to recall events from adolescence and early adulthood
109		Source confusion	[HC04]	2	Memory distorted after hearing people speak about a situation
110		Spacing effect	[Gre89]	3	Easier to recall information from spaced than massed exposures
111		Suffix effect	[MCP71]	2	Recency effect diminished by an irrelevant sound at list end
112		Suggestibility	[LJ89]	1	Ideas suggested by a questioner mistaken for memory
113		Telescoping effect	[JCM06]	3	Temporal displacement of an event
114		Testing effect	[McD+07]	3	Recall tests lead to better memory than recognition tests
115		Tip of the tongue phenomenon	[Bro91]	1	Recall parts of an item but not the whole
116		Verbatim effect	[Pop+08]	1	Easier to recall gist than verbatim wording
117		Zeigarnik effect	[Ein+03]	1	Easier to recall interrupted tasks than completed
118		Bizarreness effect	[McD86]	4	Easier to recall bizarre items
119	Baseline	List-length effect	[CR03]	4	Harder to recall items from longer lists
120		Serial-positioning effect	[Mur62]	4	Best recall first (primacy) and last (recency) items in a series
121		Von Restorff effect	[PWA98]	4	Distinct items are better remembered
122	Inertia	Continued influence effect	[JS94]	5 [Wal+17]	Recall only first information even after correction
123		Choice-supportive bias	[MJ00]	2	Recall past choices as better than they were
124	Outcome	Hindsight bias	[FB75]	5 [ZC07]	Recall past predictions as more accurate after seeing the outcome
125		Rosy retrospection	[Mit+97]	1	Remember past overly positively
126		Cross-race effect	[TKB04]	1	More difficultly distinguishing people of outgroup race
127	Self-perspective	Self-generation effect	[CH89]	4	Self-generated content is easier to recall than if simply read
128		Self-reference effect	[RKK77]	1	Easier to recall self-related information
TASK: OPINION REPORTING					
129		Halo effect	[NW77]	2	Personality trait ascription affected by overall attractiveness
130	Association	Moral credential effect	[MM01]	1	Non-prejudice credentials allow prejudicial statements
131		Negativity bias	[Fis80]	3	Social judgments affected more by negative than positive information
132	Baseline	Focusing effect	[SK98]	4	Beliefs based on the most pronounced part of given information
133		Backfire effect	[NR10]	3	Prior beliefs stronger when correction attempted
134	Inertia	Omission bias	[SMB91]	3	Moral blame affected by whether the harm was due to inaction
135		Bandwagon effect	[NCG93]	2	Beliefs affected by opinions of others
136		Moral luck	[Cus08]	1	Moral blame depends on event outcome, not just intent and action
137	Outcome	Social desirability bias	[CM60]	2	Respond in questionnaires in a socially approved manner
138		Stereotyping	[Cor+02]	3	Assuming characteristics of group member
139		Women are wonderful effect	[RG04]	1	Associate more positive characteristics to women
140		Anthropocentric thinking	[Col+17]	1	Humans are the center of the universe
141		Anthropomorphism	[WCE10]	5 [Boy+17]	Humans as analogical base for reasoning about non-human life and processes
142		Ben Franklin effect	[JL69]	1	Opinion of others is affected by one's behavior towards them
143		Bias blind spot	[PK07]	2	Belief that biases are more prevalent in others than oneself
144	Self-perspective	Illusion of asymmetric insight	[Pro+01]	1	Belief that one knows more about others than others know about oneself
145		Illusory superiority	[DMH89]	2	Personality traits favorable to oneself over others
146		Naive realism	[HC54]	1	Belief we experience objects in our world objectively
147		Third-person effect	[Ant+15]	2	Others more vulnerable to mass media messages than oneself
148		Trait ascription bias	[PR06]	2	Own traits are variable, others are predictable
149		Zero-sum bias	[RBW15]	1	Belief that one's gain is another one's loss
TASK: OTHER					
150	Association	Attentional bias	[Jan+10]	3	People extract or process information in a weighted manner
151		Risk compensation	[Hed00]	3	Risk tolerance based on constant risk, not minimization
152	Baseline	Surrogation	[CHT13]	3	Metrics or models overtake what they were constructed to measure
153		Unit bias	[GRD06]	1	People eat more food from bigger containers
154	Outcome	Ostrich effect	[KLS09]	4	Avoiding negative information

2.4 General Conclusions

Although judgment and decision making is a huge topic in psychology, behavioral economics and marketing research, for a long time there has been very little work on it in visualization. This chapter summarized five research papers I was involved in that advanced the research area a tiny bit. These papers date back from 2012–2018 and we see more work on this now, some of which has been perhaps inspired by the work we did. For example, our study on base-rate reasoning [MDF12] has inspired follow-up studies in visualization [Ott+15; Kha+15; KBH18], and several papers exploring new biases in visualization [HWS21; Wal+19; Law+21; Xio+22] cite our bias taxonomy [Dim+18].

A lot remains to be done, both in terms of understanding how cognitive biases interact with visualizations, and knowing how to best evaluate visualizations for their ability to support rational judgment and decision making. This chapter already mentioned a few elements to consider in evaluations, including what to measure. I have not discussed a key aspect of measurement, which is the choice of *ground truth*. For example, in our study on base-rate reasoning, we (and previous work before us) consider that there is one correct answer to the mammography problem, and it is the one obtained by applying Bayes' rule: this answer, 7.8%, was our ground truth, and it allowed us to define an error metric. Ground truth relates to the notion of rationality which, as I mentioned in this chapter's introduction, is complicated. I will take up this discussion here, since this is an important issue that has not really been discussed in the papers I summarized.

Defining a ground truth can be easy in math-oriented problems like the mammography problem, but is often harder in tasks that more closely capture real-life judgments and choices. For example, in our study of multidimensional visualizations, the tasks we studied (multi-attribute choice tasks) have a subjective element that makes it hard to define a ground truth: there seems to be no right or wrong answer to a question asking participants to report which holiday package they personally like the best. However, it remains possible to identify choices that are more reasonable than others, especially if we have information about participants' choice criteria. Similarly, in our experiment on the attraction effect, choosing between the cleaner gym and the gym that has more variety is a matter of personal preference (Figure 2.6), and there is no ground truth. However, if we know that a choice was influenced by irrelevant information (the presence of a third, undesirable gym), it can be considered as sub-optimal, in part because it opens people to manipulation. In those cases as in many other cases in cognitive bias research, the

ground truth is not a correct answer but a *normative theory*, i.e., a set of rules or axioms that constrain what can be considered a rational judgment or choice.

In an experiment, the choice of a normative theory can seem perfectly reasonable at first, but turn out to be flawed upon further examination. For example, a classic framing study puts participants in a hypothetical scenario requiring them to choose whether to undergo a high-risk surgery [Mox+03]. To some participants, the doctor says that 90% of patients are alive after five years, and to others, the doctor says that 10% are dead after five years. The ground truth is that participants should be insensitive to the way information is framed, because it is the same information. However, Gigerenzer [Gig18] argued that in real-life communication, framing is often used to convey extra information, that intelligent listeners intuitively pick up. For example, U.S. doctors fear making direct recommendations due to a law that facilitates malpractice suits, and indirect recommendations can take place through framing [Gig18]. Thus, in the classic experiment I mentioned, it could be that participants' answers deviate from the ground truth not because participants are irrational, but because the ground truth is wrong.

If we think back about the attraction effect study, it is possible that the presence of a dominated alternative (and more generally, the way dominated alternatives are distributed) can give useful information. For example, if a real estate investor sees a cluster of similar houses that are all dominated, it could mean that this type of house represents a larger market, and therefore that the non-dominated house that is similar to them (the target) is truly a better choice. However, situations also exist where the number and position of dominated alternatives are clearly irrelevant and where a preference for the target would be irrational. This was the case for our experiment involving real bets, and for which we could argue that any deviation from the normative theory was irrational. Our qualitative data does indicate that the vast majority of our participants were unable to rationalize their choices based on where the dominated alternatives were located. Nevertheless, in future studies on judgment and decision making, and particularly studies of cognitive biases, it seems important to carefully justify the choice of normative theory, and discuss possible cases where this normative theory might be wrong in ecological situations.

Supporting Effective Communication

While visualizing data can help people make decisions for themselves, it is also important that insights gained from data can be efficiently conveyed to other people, to inform their own decisions. In most of the use case scenarios I have covered in the previous chapter, people explore data by themselves (e.g., a list of houses to buy, or a list of holiday plans) and draw conclusions or take action based on what they saw. Here, I focus on cases where someone (for example, a scientist or a journalist) wants to visualize data they are familiar with to convey it to others: this is the area of communicative visualization.

In this chapter, I present four research projects in which I was involved that have to do with communicative visualization (and also with small data). The first two contribute new *tools and techniques* to help researchers communicate data to *other researchers*. The other two projects are *studies* that examine how to communicate data to a *general audience*.

3.1 Communicating Data with Tabular Visualizations

A tabular visualization is a table whose cells contain colors and shapes instead of (or in addition to) numerals. Tabular visualizations are quite versatile – for example, we saw that they can be used to support data-driven decision making (subsection 2.3.1), but they can also be effective communication tools. In this section, I discuss a particular family of tabular visualizations designed for scientific communication that date back to the 1970s, and describe the work I and my co-authors did to bring them up-to-date and to make them known to and available to a larger audience.

This section is based on a 2014 paper authored by Charles Périn, Jean-Daniel Fekete, and myself [PDF14]. It also borrows from a 2019 paper [PFD19] and photo material (<https://www.aviz.fr/slowdata>) by the same authors. Supplementary material is available at <https://aviz.fr/bertifier>, and the Bertifier tool is available as a web app at https://aviz.fr/bertifier_app/.

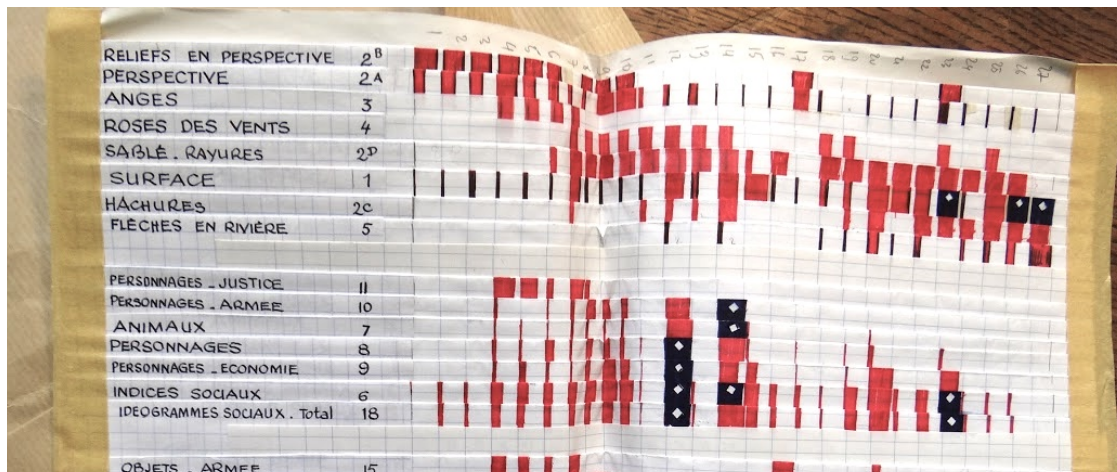


Fig. 3.1: Part of a Bertin matrix handcrafted by the Laboratoire de Graphique (c. 1970–1980, *Archives Nationales*). This matrix is like a regular table where numbers have been replaced by visual shapes, and rows and columns were ordered in a way that reveals patterns. A common approach was to draw an initial matrix from numerical data, photocopy it, cut it along either rows or columns, reorder it, photocopy it again, and start over. Once the result was deemed satisfactory, a clean and annotated version of the matrix was produced for publication. The process could take weeks.

3.1.1 Introduction

Jacques Bertin was a French cartographer known for his theoretical contributions to data visualization. His most influential work is his 1967 monograph “*La Sémiologie Graphique*” [Ber67], where he lays out a theoretical framework for visualizing data that influenced areas such as cartography, social sciences and information visualization. Much less known but equally important is his later work from 1975, “*La Graphique et le traitement graphique de l’information*” [Ber75; Ber81]. This book, which is our focus, details a method for visually processing and communicating numerical tables that was meant to be effective, generic and accessible to any scientist and researcher [Pal03]. It was based on four simple steps: *i*) compile empirical data into a two-dimensional table (called *matrix*), *ii*) represent cell values visually, *iii*) group similar rows and columns to reveal patterns, and *iv*) annotate the end result for communication purposes.

Bertin devised and refined his matrix method after years of work with researchers such as geographers, agricultural economists, ethnologists, and historians [Pal03]. In the 1970s, he founded a “Laboratoire de Graphique” whose main purpose was to help social scientists from his institution turn their data into insightful figures to be used in scientific publications. However, all figures were hand-crafted, and his method required

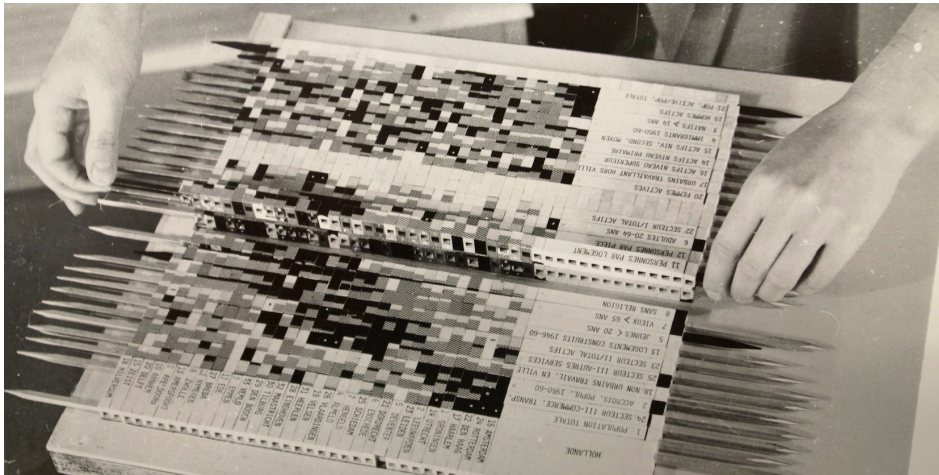


Fig. 3.2: A reorderable matrix (c. 1970–1980, photo Serge Bonin, Archives Nationales). Reordering Bertin matrices was a crucial but also time-consuming part of the process. In 1968, Bertin and his colleagues designed a mechanical device to facilitate reordering. It consisted in separate plastic cells, and a rod mechanism for reordering them across either rows or columns. The device underwent successive refinements up to the 80s, was sold to other French labs, and even used in schools [PFD19].

using physical material and devices that involved tedious manipulations—often weeks of work (see Figures 3.1 and 3.2). There has been several computer adaptations of Bertin’s method [BC94; Sii99; Saw96], but they are limited in functionality, feature only rudimentary interactions, and the end results are not optimized for communication. In our paper [PDF14], we addressed this gap by contributing:

- A list of requirements for a faithful computer adaptation of Bertin’s method derived from an examination of his writings and artifacts created by his lab,
- An extensive review of existing computer systems for tabular visualization with their features and limitations,
- BERTIFIER (www.bertifier.com), a web-based tabular visualization authoring system that implements Bertin’s method,
- *Crossets*, a novel interaction technique for creating, manipulating and fine-tuning tabular visualizations,
- Algorithm adaptations to support user-driven matrix ordering,
- A user study suggesting that users from different backgrounds can use BERTIFIER to help them visualize and understand their own tabular data.

This section focuses on reporting examples of Bertin matrices made with and without BERTIFIER, in order to illustrate their usefulness and their expressive power.

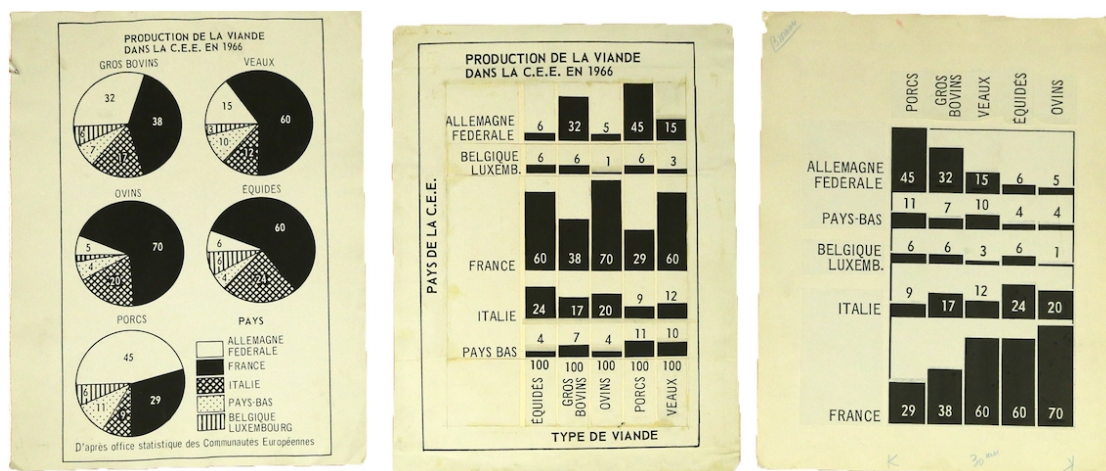


Fig. 3.3: Meat production per country shown as pie charts (left), matrix (middle), and reordered matrix (right). Drafts for the book *La Graphique* [Ber75], Archives Nationales.

3.1.2 Two examples from Jacques Bertin

In Figure 3.3, taken from *La Graphique* [Ber75], Bertin uses a simple example to illustrate why matrices are an effective representation for data. On the left, pie charts show the contribution of different countries in the production of different types of meat. Each pie is a meat type. While common, this type of design makes it difficult to understand the data, leading Bertin to qualify it as “useless”. The middle chart is a matrix visualization. With such a visualization, the way production is distributed across countries (rows) and meat types (columns) becomes easier to read. However, since countries and meat types do not have a natural order, many other matrices can be produced, including the one on the right, which is a better choice because it better shows general trends and patterns.

Figure 3.4 shows a more complex matrix, with a larger and more heterogeneous dataset. It is one of the earliest applications of Bertin’s method to research data [VLB67]. The visualization lists all war ships (228 in total) owned by France and Great Britain in 1947. Each column is a ship. The top row shows whether each ship is French (black rectangle) or British (white rectangle). The bottom row shows the number of cannons for each ship – the taller the black bar, the more cannons. The middle area encodes each ship’s year of construction – the lower the small black rectangle the more recent. This particular matrix is sorted by year of construction. It can be seen that the rate at which ships were built increased between 1700 and 1747, both for French and British ships (since the proportion of black/white rectangles on the first row doesn’t clearly change over time).

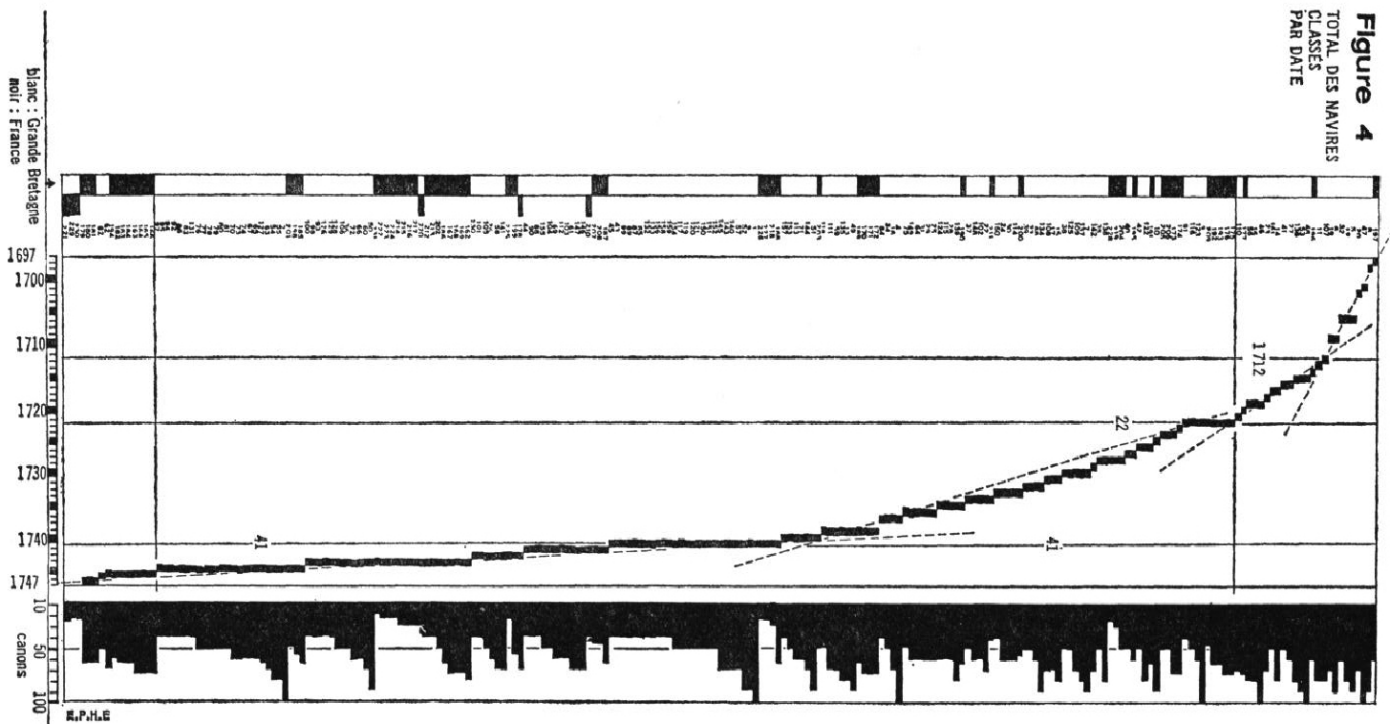


Fig. 3.4: Bertin matrix from 1967, showing information about the 227 ships (each ship is a column) owned by the French and British navy in 1947 [VLB67]. The original figure has a vertical aspect ratio but has been rotated by 90° here to save space.

The paper [VLB67] contains three other figures where the matrix is sorted differently to reveal other insights. This visualization manages to convey a fair amount of raw data in a small space – Bertin described his procedure as “*simplifying without destroying*”: raw data must be shown as is, without aggregation or complex processing.

The four navy figures were among the case studies we used when we developed and tested our tool BERTIFIER. We successfully reproduced them with the tool (see <https://aviz.fr/Bertifier/Examples>). Since the navy dataset was not available, we reconstructed it from Wikipedia articles.

COMPUTER ADAPTATIONS OF BERTIN'S METHOD		SOFTWARE			USER INTERFACE					FEATURES																						
		YEAR	DEVELOPED IN BERTIN'S LAB	SOFTWARE TESTED IN BERTIN'S LAB	UI DEGREE OF COMPATIBILITY	UI SPATIAL DIRECTNESS	UI TEMPORAL DIRECTNESS	SUBJECTIVE USABILITY	UI CONSISTENCY	ANIMATED TRANSITIONS	ENCODING			LAYOUT				COMMUNIC.														
		1980	1990	2000	2010	AVAILABILITY	UI DEGREE OF COMPATIBILITY	UI SPATIAL DIRECTNESS	UI TEMPORAL DIRECTNESS	SUBJECTIVE USABILITY	UI CONSISTENCY	ANIMATED TRANSITIONS	DATA TYPES SUPPORTED	CHANGE ROW/COL ENCODING	SHAPE ORIENTATION	SPECIFY HEADERS	ROW/COL AUTO REORDERING	ROW/COL MANUAL REORDERING	ROW/COL RESIZING	ROW/COL SEPARATOR RESIZING	ROW/COL SEPARATORS	TRANSPPOSITION	SINGLE ROW/COL SORTING	AUTO REORDERING OF SUBSETS	ROW/COL AUTO REORDERING	ROW/COL MANUAL REORDERING	FIDELITY TO BERTIN'S STYLE	GLUE ROWS/COLUMNS	EXPORT CAPABILITIES	TEXT ANNOTATIONS	FONTS CUSTOMIZABILITY	
LIMITED FEATURES AND COMMAND-BASED	TMC CARTAX MATRIX																															
LIMITED FEATURES LIMITED INTERACTIVITY	TGINF MATRIXEXPLORER																															
MOSTLY MANUAL REORDERING	STEVE RUBIN, IN D3 CARTES & DONNÉES THE REORD. MATRIX																															
RICH FEATURES WITH COMPLEX WIMP USER INTERFACE	GAP PERMUTMATRIX VOYAGER VISULAB T_ALK AMADO																															
RICH FEATURES AND SCRIPT-BASED	BERTIN FOR R CHART																															
THIS ARTICLE	BERTIFIER																															

Fig. 3.5: Computer applications implementing Bertin's matrix method. Each row is a computer application. Crosses indicates unavailable data.

3.1.3 Bertifier Example: Survey of Applications

Our next example (Figure 3.5) was also made with BERTIFIER, but shows a recent dataset – a survey of computer applications. The figure visually summarizes the results of a survey we did of computer adaptations of Bertin's method, as part of our work on BERTIFIER. We used BERTIFIER to produce the matrix in Figure 3.5, which we included as a figure in our paper on BERTIFIER [PDF14]. Computer applications are in rows and their characteristics are in columns, such as the year the application was developed, the type of user interface it used, and the features it supported from Bertin's method. The name of each application is indicated on the second column (TMC, Cartax, etc.).

Overall, some early attempts – especially CHART [BK77] in 1977 – were rich in terms of functionality, but poor in terms of interactivity. On the other end of the spectrum, a few recent implementations – Like Rubin's implementation in d3 [Rub] – provide consistent, direct and easily-discoverable interactions, but at the cost of only providing a few features. There is a marked trade-off between the level of functionality on one side, and the level of interactivity and discoverability on the other side. Nevertheless, it is possible for a computer adaption of Bertin's method to be both highly functional and highly interactive, as illustrated by the last row showing the features of BERTIFIER.

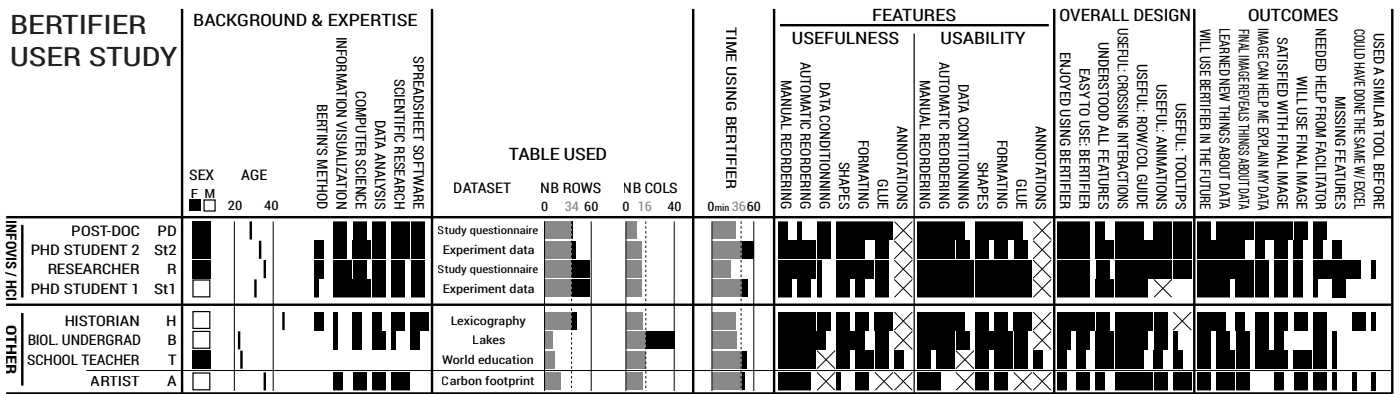


Fig. 3.6: Results of our usability study of BERTIFIER. Each row is a participant (N=8), with a visual summary of their demographics and of study data such as their responses to Likert items (the thicker a black rectangle the more positive the response on the scale). Crosses indicate unused features. Dashed gray lines indicate means.

3.1.4 Bertifier Example: User Study Data

Our next and last example (Figure 3.6) was also made with BERTIFIER, again with the goal of producing a figure to be included in our paper on BERTIFIER. This matrix summarizes the results of our usability study of BERTIFIER involving 8 participants (as rows). Columns are divided in six groups: 1) *Background and expertise*, indicating the participant’s basic demographics and level of expertise in different areas; 2) *Table used*, giving info about the numerical table used by them (they were each asked to bring a personal spreadsheet); 3) *Time spent using BERTIFIER*; 4) *Features*, i.e., responses to questions asking how useful and usable each BERTIFIER feature is; 5) *Overall design* – responses to more usability questions; and finally 6) *Outcomes* – responses to overall satisfaction and attitude questions. Those results are discussed in detail in [PDF14].

3.1.5 Conclusion

Bertin devised his matrix method and visual language to facilitate the communication of tabular data in research publications. Although visually encoding cell values was not a new idea at the time, he refined his method over many years. We presented examples of Bertin matrices, made with and without BERTIFIER. Although we did not cover the interactions aspects here, BERTIFIER implements novel interaction techniques for formatting and ordering tables that make it possible to easily and rapidly convert numerical tables into Bertin matrices (see video demo at <https://www.youtube.com/>

[watch?v=tJxAF_a_yBQ](#)). Our user study suggests BERTIFIER can help both researchers and a wider audience understand and communicate their data visually.

BERTIFIER can be especially useful to human-subject researchers who need to report responses to multiple Likert-type questions. Summary statistics and inferential statistics are commonly used to report such data, but they remove lots of potentially useful information. Bertin matrices allow researchers to report their raw data in a way that can be understood, and they can therefore usefully complement statistics, especially when inferential statistics are not applicable (e.g., for very small N). BERTIFIER has been indeed used in this manner in subsequent work [HJC14; RMM15; Thu+16; PBV16; Bot+16; Hur+16; Per+17; Bes+20]. BERTIFIER has been also used to summarize literature surveys [Hal+16; Raj+17; Koy+17; Fle18; ATD19]. However, most of these papers were co-authored by one of the authors of BERTIFIER or close collaborators, so we cannot claim that the tool has been widely adopted. Our goal was to design a tool that would finally popularize Bertin matrices, but for a tool to be widely adopted it needs to be 100% functional. Our tool was only 95% functional due to lacking features such as undo/redo and save, which likely prevented its broad adoption.

3.2 Communicating Results from Multiverse Analyses

In the previous section, I outlined a method researchers can use to communicate their data to their peers that is inspired by practices from 50 years ago. Here, I discuss another method researchers can use to communicate their data to their peers, but this time, it is inspired by a practice that has only recently emerged: that of reporting multiverse analyses. *Multiverse analysis* is a philosophy of statistical reporting where researchers report the outcomes of many different statistical analyses in their paper in order to show how fragile or robust their findings are [Ste+16]. In this section, I explore why it is interesting to combine this idea with the concept of *explorable explanations*, i.e., narratives that can be read as normal explanations but where the reader can also become active by dynamically changing some elements of the explanation [Vic11].

This section is based on a 2019 paper authored by myself, Yvonne Jansen, Abhraneel Sarma, Matthew Kay, and Fanny Chevalier [Dra+19]. Interactive examples are available at <https://explorablemultiverse.github.io/>.

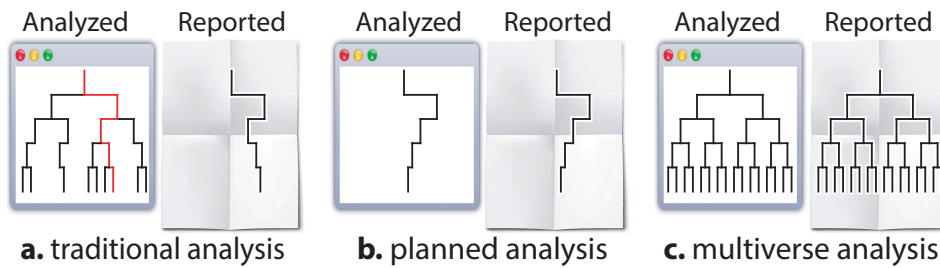


Fig. 3.7: Three reporting strategies, from the least transparent to the most transparent: a) traditional analysis with undisclosed flexibility; b) planned analysis; c) multiverse analysis. Each branching represents a choice between different analysis options [NT07].

3.2.1 Introduction

The recent replication crisis in psychology and other disciplines has greatly affected the credibility of human-subject research and prompted a movement of methodological reform [NSS18]. Much of this movement calls for more transparency in the way statistics are reported, so that findings become more trustworthy, more likely to be interpreted correctly, and easier to verify and replicate [Nos+15; Mun+17; Cum14]. Many solutions have been proposed to increase transparency. In this work, we seek to understand how we can augment research papers to support and promote transparent statistics reporting.

While there are many ways a statistical report can lack transparency, a common and damaging form of opacity is *undisclosed flexibility* (see Figure 3.7a), i.e., not reporting the different options that have been tried during the analysis [SNS11; Wic+16], or the options that would have been chosen had the data been different [GL13]. Undisclosed flexibility is damaging because it substantially increases the chances of reporting erroneous findings, while being invisible to the reader.

One response to the issue of undisclosed flexibility has been to encourage researchers to commit to a single statistical analysis that has been planned [Dra16; Ber12] and ideally registered [CGD18] ahead of time (Figure 3.7b). Although planning eliminates the problem of undisclosed flexibility, some statisticians and methodologists are starting to argue that more transparency can be achieved by letting researchers try many analyses and report all of them in their paper [SSN15; Ste+16] (Figure 3.7c). This is partly motivated by evidence that different researchers who analyze the same data will make different choices and will thus get slightly—and sometimes widely—different results [Sil+17].

In a *multiverse analysis*, researchers identify sets of defensible analysis choices (e.g., different ways of excluding outliers, different data transformations), implement them

all, and then report the outcomes of all analyses resulting from all possible choice combinations. This approach increases transparency because readers can appreciate the “fragility or robustness of a claimed effect” [SSN15] by checking whether the findings are dependent on arbitrary analysis choices.

Multiverse analysis promises an unprecedented level of transparency for research papers, but the idea is still very much in its infancy. Writing papers with a multiverse analysis is difficult and there is currently very little guidance. An important part of the difficulty lies in reporting the outcomes of potentially hundreds or thousands of analyses in a single research paper (past examples contain between 120 and 1728 analyses [Ste+16; SSN15]), causing challenges for authors, reviewers, and readers. There are currently two ways to approach this problem.

A first option consists of sharing multiverse analyses as supplementary material, letting readers look under the hood of a default analysis and try alternative analysis options in a different environment. This approach has long been promoted by the *reproducible research* movement, and has been the subject of a vast body of work [BD95; GT07; Nüs+17]. Although supplementary material is crucial for reproducibility and reuse, casual readers are very unlikely to engage with it. Thus proponents of multiverse analysis argue for acknowledging, reporting and discussing the multiverse in the research paper itself [SSN15; Ste+16].

A second option consists of summarizing the multiverse in the research paper itself using tables and custom plots. This is the approach put forward by the two papers introducing multiverse analysis [SSN15; Ste+16], which use as examples histograms of p -values [Ste+16], grids of p -values [Ste+16], and sorted dot-plots of point estimates where color encodes statistical significance [SSN15]. Although such visualizations are very useful as overviews, they force the paper author to collapse the results of each analysis into a single point estimate, p -value or statistical significance verdict. In doing so, a lot of the richness and nuance of well-crafted statistical reports is lost.

A third approach, which we propose here, consists of making the results section of the research paper interactive. More specifically, we introduce and explore the concept of *explorable multiverse analysis report* (EMAR). An EMAR is a multiverse analysis report that is presented in the form of an *explorable explanation* in Bret Victor’s sense [Vic11]. Explorable explanations are explanations that “enable and encourage truly active reading” and allow “the reader to play with the author’s assumptions and analyses, and see the consequences” [Vic11]. At the same time, they can be read like normal prose: “the reader

is not forced to interact in order to learn” [Vic11]. Consistent with this design philosophy, an EMAR looks like a regular results section, but readers are given the possibility to change analysis options and immediately see the results within the research paper itself.

We posit that EMARS can be a useful and compelling complement to existing multiverse reporting approaches. In this section, we will explore part of their design space through three examples of short interactive papers which we wrote to demonstrate the concept. More extensive discussions (e.g., positioning with respect to previous work on adjacent topics, a systematic description of the design space of EMARS, trade-offs between different EMAR reporting strategies, benefits and challenges raised by EMARS compared to alternative reporting strategies such as the use of supplemental material or static multiverse analysis reports) can be found in our EMAR paper [Dra+19].

We wrote five mini-papers, all of which are re-analyses of previously published studies for which data and R scripts were publicly available. They are self-contained and can be read online at <https://explorablemultiverse.github.io> or by clicking directly on a code name below. We present three of them here.

3.2.2 Example 1 – FREQUENTIST

The FREQUENTIST example [DJ18b] is a reanalysis of a CHI study evaluating physical visualizations [JDF13]. It is meant to illustrate a few basic multiverse analysis ideas for a typical frequentist analysis with confidence intervals (CIs). The results of the analysis are initially identical to the original paper, including the two figures reporting mean task completion time per technique and pairwise comparisons, with 95% CIs. Four aspects of the analysis can be changed by the reader, which has the effect of immediately updating the two plots and some text elements such as explanations and figure captions. Changes are made by clicking or dragging the elements of the text in blue as in Bret Victor’s explorable explanations [Vic11] (see Figure 3.8).

First, horizontally dragging the “95%” text has the effect of changing the confidence level (7 levels are provided from 50% to 99.9%) and updating the length of error bars in the two figures. This allows the reader to appreciate that the 95% level is arbitrary [McE18] and thus that CIs should not be interpreted in a strictly dichotomous manner [Cum14]. Meanwhile, readers who insist on interpreting effects as significant or non-significant have the option of changing the customary cutoff of $\alpha=.05$ (95% CIs), for example to the $\alpha=.005$ (99.5% CIs) criterion now advocated by some methodologists [Ben+18].

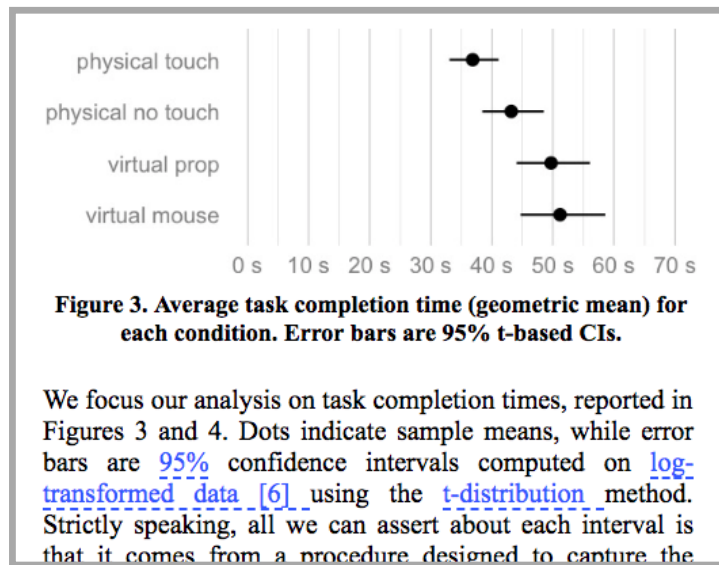


Fig. 3.8: Excerpt from the mini-paper [FREQUENTIST](#), showing widgets embedded in the text in Bret Victor’s [\[Vic11\]](#) style. Operating a widget changes one aspect of the analysis and immediately updates the figure.

Clicking the “transformed data” text toggles the text to “untransformed data” and updates the two figures with results from the corresponding analysis. Although some researchers recommend that completion times be log-transformed [\[SL10\]](#), other researchers may be suspicious of, or unfamiliar with data transformations—this option reassures them that the results hold for untransformed data. Similarly, clicking on “t-distribution” switches the text to “BCa bootstrap” and shows the results of the analysis using non-parametric bootstrap CIs, which tend to be liberal (i.e., too narrow) with small samples but do not require distributional assumptions [\[KG13\]](#).

Finally, the plot with the three planned pairwise comparisons (not shown in [Figure 3.8](#)) shows uncorrected CIs, but the reader can apply a Bonferroni correction by clicking on the text “not corrected for multiplicity”. Many recommend such a correction but it is not without drawbacks, and there is a complex literature on the topic [\[Dra16\]](#). To help the reader interpret the CIs correctly, the mini-paper contains a paragraph that gives the individual and the family-wise CI coverage and false positive rates, which are updated whenever Bonferroni correction is turned on or off, or whenever the confidence level is changed. More details can be found in the mini-paper itself [\[DJ18b\]](#).

The [FREQUENTIST](#) mini-paper covers a total of $7 \times 2 \times 2 \times 2 = 56$ unique analyses. The paper concludes that the findings from the original study (i.e., good evidence of a difference for

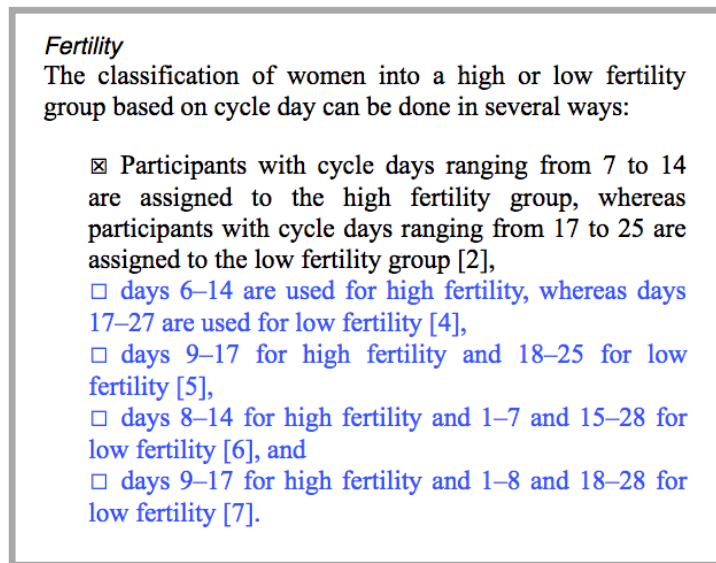


Fig. 3.9: Excerpt from the mini-paper [DATAVERSE](#), listing five different ways of dichotomizing a dependent variable. Elsewhere in the mini-paper, an interaction plot gets updated each time an option is chosen.

the first two comparisons, inconclusive results for the third one) are reasonably robust, as they hold across the sub-multiverse where the confidence level is at 95% or less.

All mini-papers offer the option of animating the multiverse by repeatedly drawing an analysis at random, which turns the paper’s plots into hypothetical-outcome plots [[HRA15](#); [Kal+18](#)]. This feature gives an overview of the multiverse and permits the reader to observe which parts of the article change substantially across the multiverse.

3.2.3 Example 2 – [DATAVERSE](#)

The [DATAVERSE](#) example [[KD18](#)] reproduces part of the multiverse analysis reported in Steegen et al. [[Ste+16](#)], which is itself a re-analysis of a famous and controversial study on the effect of ovulatory cycles on voting behavior [[DRG13](#)]. The [DATAVERSE](#) example is meant to illustrate alternative ways of processing experimental data (e.g., dichotomizing responses, excluding participants), and the use of interactive choice lists.

The “*Constructing the data multiverse*” section in Steegen et al. [[Ste+16](#)] goes through each data processing choice made in the original study [[DRG13](#)] and describes alternative choices that could have been reasonably made. The [DATAVERSE](#) mini-paper essentially

reproduces this section with the difference that the reader can select particular choices. The mini-paper first lists five ways of dichotomizing a particular dependent variable, and lets the reader choose one of them (Figure 3.9). Four other data processing operations are described afterwards, each with two to three options to choose from. The mini-paper ends with a figure showing the result of the selected analysis in the form of an interaction plot, which is updated each time a different option is chosen in the text.

The [DATAVERSE](#) mini-paper covers $5 \times 2 \times 3 \times 3 \times 2 = 180$ unique analyses. Steegen et al. [Ste+16] summarizes the multiverse by plotting the 180 corresponding p -values. While this summary provides an extremely useful overview clearly showing that the original findings are not robust, it does not allow the reader to examine detailed outcomes of specific analyses of interest. By making it possible to select any particular analysis and see the resulting effect sizes, the [DATAVERSE](#) mini-paper conveys more complete results than a simple summary of p -values. As in the [FREQUENTIST](#) mini-paper the multiverse can be animated, giving a striking demonstration of the variability of effect sizes across the multiverse that can usefully complement the p -value summary.

3.2.4 Example 3 – DANCE

The [DANCE](#) example [Dra18] is a reanalysis of a previous InfoVis study on the perception of correlations [Har+14]. It is meant to illustrate the use of simulated datasets to convey inferential information that can be missing from plots.

Here, the analysis procedure is the same across the multiverse but the raw dataset varies. More specifically, 100 alternative datasets were created from the original study's dataset using bootstrapping [Woo05]. A bootstrap dataset is created by sampling the original dataset with replacement. A remarkable property of bootstrapping is that the distribution of bootstrap samples tends to resemble the true sampling distribution, and thus bootstrap samples can be used to compute CIs [Efr92]. In the [DANCE](#) example, we use bootstrapping to derive a set of datasets that could have reasonably come up if the study was replicated with different participants. We then subject all datasets to the same statistical analysis and plotting procedure.

The mini-paper reproduces the analysis from the original study, with its four plots. It also lets readers replace the original dataset with any of the 100 bootstrap datasets. When the dataset changes, each of the 4 plots changes slightly. More interestingly, animating the multiverse yields hypothetical-outcome plots [HRA15; Kal+18] similar to Cumming's

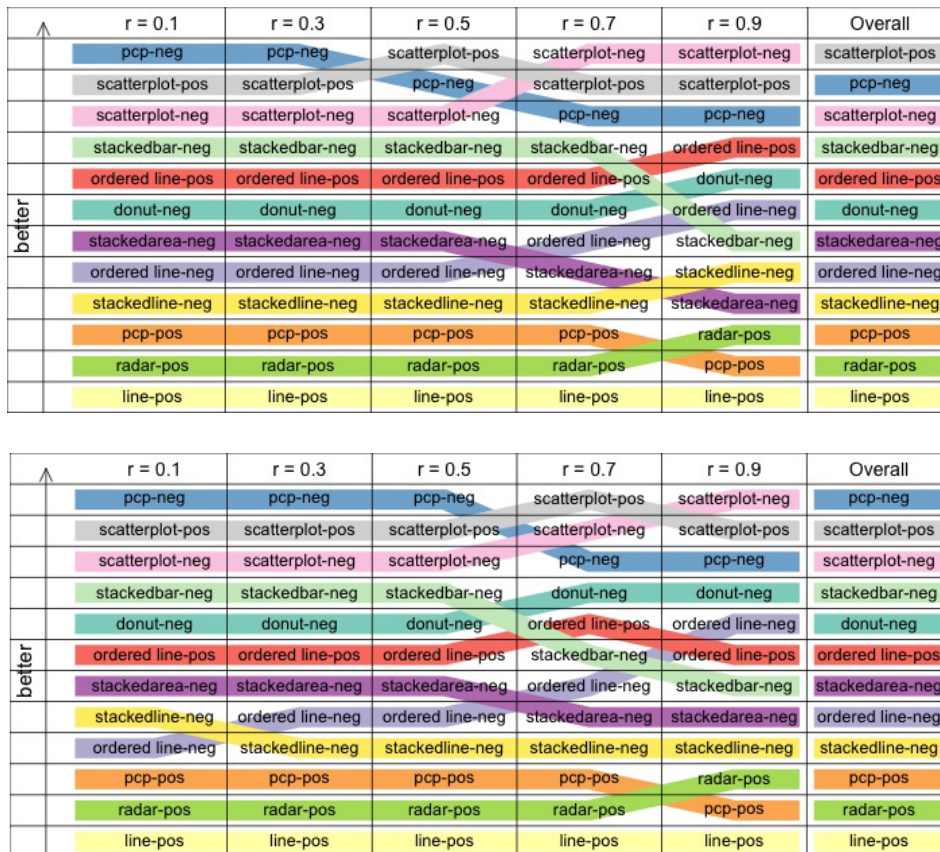


Fig. 3.10: *Top:* plot showing a ranking of visualizations in their ability to convey correlation [Har+14]. *Bottom:* an alternative plot that could have reasonably come up in an exact replication, created by bootstrapping the experimental dataset. Some results hold (e.g., the bottom of the ranking) while some do not (e.g., the top and middle of the ranking). The mini-paper DANCE allows to animate between 100 of those plots.

dance of p -values [Cum09] and other statistical dances [Dra17], with the difference that the sampling distribution is estimated from data rather than simulated.

Animating the multiverse of bootstrap datasets allows the reader to appreciate the reliability of the different quantities, trends and patterns depicted by each plot and to carry out “inference by eye” [CF05]: a pattern that is stable across the multiverse is a good indication that it is reliable. Compared to static representations of inferential information such as error bars, this technique has the advantage of being applicable to any plot. It is especially useful for revealing statistical uncertainty that is hidden in some plots, such as the ranking plot reproduced in Figure 3.10. More examples can be found in the mini-paper.

3.2.5 Limitations and Possible Objections to EMARS

There are several limits to our work. First of all, our examples are meant to capture elementary EMAR techniques, but many other more sophisticated techniques are possible to further enhance EMARS, such as coordinated views and linking between figures [Rob07], and reverse direct manipulation [Dra+08] to let readers manipulate plots and observe which analysis options lead to certain plots (e.g., “which options yield the largest effect?”). Our examples also cover relatively simple statistical analyses. More complex analyses would probably need to expose less parameters to preserve usability: but as pointed out by Steegen et al. [Ste+16], multiverses do not need to be large to be useful.

Importantly, our mini-papers are only proofs of concepts and for EMARS to be adopted, it is crucial that usable toolkits are developed. A future toolkit could build on modern interactive document authoring tools such as Idyll [CH18] and capitalize on initiatives to integrate academic writing with statistics environments such as the recently released Radix [Mis18], which combines the Distill framework with R Markdown. But even if tools are developed to support the authoring of EMARS, some objections to their widespread use will likely remain. We examine four potential objections:

Objection 1: writing EMARS will remain hard. With the proper tools, it is unclear whether writing EMARS will be harder than writing static multiverse analysis reports as in [Ste+16; SSN15]. It is clear however that writing EMARS will always be harder than writing single-universe analyses, and as with providing supplemental material, the extra effort may not come with tangible rewards [BVH]. The statistical analysis itself is more work-intensive and even if new libraries can be developed to facilitate it, it is unlikely that the job can be automated [Ste+16]. Similarly, it will never be possible to ensure that a multiverse analysis is complete or even well-chosen. But again, even a small multiverse analysis is superior to a single-universe analysis in terms of transparency [Ste+16]. So for researchers who want to signal or promote transparency, the option of writing EMARS can still be attractive, and may become more and more attractive as signals of research integrity and transparency get more and more rewarded [Gin12a; Gin12b].

Objection 2: reading EMARS will remain hard. In the full paper [Dra+19], we pointed out potential usability issues with EMARS, including difficulties with following changes that occur outside the viewport and predicting where changes will occur. It would be interesting in the future to investigate whether these issues can be mitigated by improvements to the paper reading UI. For example, the interface could present two views of the same article that can be scrolled independently, allowing readers to monitor

targets that are located far from the operated controls. Alternatively, changes outside the viewport could be highlighted [Che+10; BR03], or replayed back when brought into view [Bau+06; BDB06]. Regardless, engaging with EMARS will always require extra effort from the reader. Because of this, it is again crucial that EMARS can be read at two levels and understood without interacting [Vic11] by readers who quickly want to learn about the authors' conclusions and how they arrived at them.

Objection 3: reviewing EMARS will remain hard. At first sight, EMARS seem to pose a challenge to reviewing policies. Publishers typically impose strict cut-offs in terms of page length, and enabling interactive content implies that authors can provide a theoretically unbounded amount of content. EMARS also break the linear structure reviewers are used to. The problem is however not new, and already arose with the introduction of supplemental materials [Mar09]. Publishers and reviewers can simply consider all non-default analyses as supplemental material: in case reviewers are not required to review supplemental material (e.g., as in many venues), only the default analysis would need to be reviewed. Reviewers are still free to scrutinize other analyses and demand analysis options to be added [SSN15].

Objection 4: preregistration will remain preferable. Multiverse analyses are compatible with preregistration [Ste+16; SSN15]. One approach is to preregister the entire multiverse analysis [SSN15], which would require specifying the default analysis in the EMAR (if any), as choosing which analysis to emphasize is a major researcher's degree of freedom. A weaker but easier form of preregistration could include the default analysis only.

3.2.6 Conclusion

We presented *explorable multiverse analysis reports* (EMARS), a new approach to statistical reporting where readers of research papers can explore alternative analysis options by interacting with the paper itself. We hope our work will inspire more HCI research where the academic paper is treated as a user interface whose purpose is to convey scientific knowledge in an accurate and transparent manner. For now, this idea has not been implemented in a tool as far as we know, but it is relatively recent work so there is hope that it will inspire future tools. To further help researchers develop visualization tools and techniques to support multiverse analysis and its reporting, we (some of the same co-authors as well as Brian Hall and Yang Liu) subsequently published a survey of multiverse analysis visualizations that researchers use in their academic papers [Hal+22].

3.3 Do Trivial Charts Hinder Truthful Communication?

We have seen two (more or less) novel ways of conveying data to scientific audiences. Now I focus on communicating data to the general public. This topic spans a huge research space and I will only report two studies on very specific research questions.

In this section, I focus on whether charts can be counter-productive when the amount of data to be conveyed is extremely small. This question was prompted by a new study that garnered attention at the time [TW16], and according to which the answer is yes. I had strong reservations about what I thought was the paper's implied message – that charts hinder clear thinking and should be used sparingly in serious and honest communication. I found this advice misguided and the study's conclusions implausible, because I think visual encodings are always better (easier to process and more informative) than numerals, unless exact values are needed. A quick examination revealed serious flaws in the study. So with Yvonne Jansen, we decided to refute the study's conclusions with a new study.

This section is based on a 2017 paper authored by myself and Yvonne Jansen [DJ18a]. Supplementary material (stimuli, data, analyses and experimental software) is available at <https://aviz.fr/blinded>.

3.3.1 Introduction

In 2014, Tal and Wansink [TW16] published a study entitled “*Blinded with science: Trivial graphs and formulas increase ad persuasiveness and belief in product efficacy*” in the journal *Public Understanding of Science*. The study claims that adding a trivial chart to a description of a new drug increases people's belief in the drug's efficacy. Figure 3.11 shows the stimuli and question from their second experiment: when a chart was included, more people thought the drug was effective. The authors argued that the chart was redundant but persuasive merely because of its scientific aura. The article concludes by advising “*caution when encountering communications hinting at scientific credibility*”, and advising consumers to “*ignore spurious cues to a scientific basis*”. These findings were publicized by the media and cited in the scientific literature, including in information visualization [Pan+15; Sta+16; CG14].

The article summarized in this section is a reappraisal of this study, which from now on we refer to as the *BwS study* (as in “Blinded with Science”). Our goal was to examine an

A large pharmaceutical company has recently developed a new drug to boost peoples' immune function. It reports that trials it conducted demonstrated a drop of twenty percent (from eighty three to sixty three percent) in occurrence of the common cold. It intends to market the new drug as soon as next winter, following FDA approval.

Incidence of illness drops from 83% to 63% with the medication.

Please rate your agreement with the following statement:

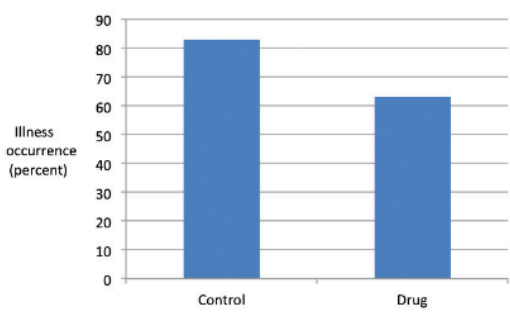
I believe the new drug is effective.

1 2 3 4 5 6 7 8 9

Strongly disagree Strongly agree

a

A large pharmaceutical company has recently developed a new drug to boost peoples' immune function. It reports that trials it conducted demonstrated a drop of twenty percent (from eighty three to sixty three percent) in occurrence of the common cold. It intends to market the new drug as soon as next winter, following FDA approval.



Group	Illness occurrence (percent)
Control	83
Drug	63

Please rate your agreement with the following statement:

I believe the new drug is effective.

1 2 3 4 5 6 7 8 9

Strongly disagree Strongly agree

b

Fig. 3.11: Stimuli used in our replication of experiment 2 from Tal and Wansink [TW16]. (a) *no-chart* condition; (b) *chart* condition.

alternative explanation for the results obtained in the BwS study, namely, that the charts allowed participants to better grasp the magnitude of the stated drug effect. Though the charts were informationally redundant, it is well known in information visualization that quantitative facts are often better understood if presented visually [KMS99; Fek+08; Mun14]. In the BwS study, the data only consisted of two quantities so the added value of the chart is unclear. Nevertheless, for the BwS study's conclusions to hold, it is important to ascertain that the chart did not give participants a better sense of the drug's efficacy.

Some chart experts advise against using charts for conveying very few numbers (as in Figure 3.11-b), but in reality, very little is known on whether such charts are actually helpful or harmful (for a review of the evidence, see our full paper [TW16]). In the BwS study, it could well be that participants found it hard to intuitively grasp the magnitude of the drug's effect when conveyed in textual form, and that the bar chart made it easier to see. The BwS study did attempt to establish that the chart had no positive effect on understanding, but the evidence is insufficient for several reasons: (1) the claim is based on accepting the null hypothesis with a single low-powered experiment, (2) the question did not test understanding but instead tested the retention of a specific number provided in the text (the percent reduction in illness), and (3) percent changes are known to be confusing, and are not the most intuitive measure of a drug's efficacy.

We conducted a series of four replications of the BwS study, which kept the stimuli and questions unchanged, and assessed people's understanding of the data provided about the drug. This allowed us to test whether the increased belief in drug's effectiveness could be explained by an improved understanding of its actual effectiveness.

The BwS study consisted of three experiments, all with positive results. Experiments 1 and 2 tested whether the addition of a simple bar chart affects people's belief in medication efficacy. Experiment 3 tested whether the addition of a chemical formula produced a similar effect. Since our focus is on charts, we did not consider experiment 3 and focused on replicating experiments 1 and 2. BwS' experiment 2 is an improved version of their experiment 1, and addresses a few issues with its design. Therefore, to save space here, we will not describe our replication of their experiment 1, and will move directly to our first replication of their experiment 2.

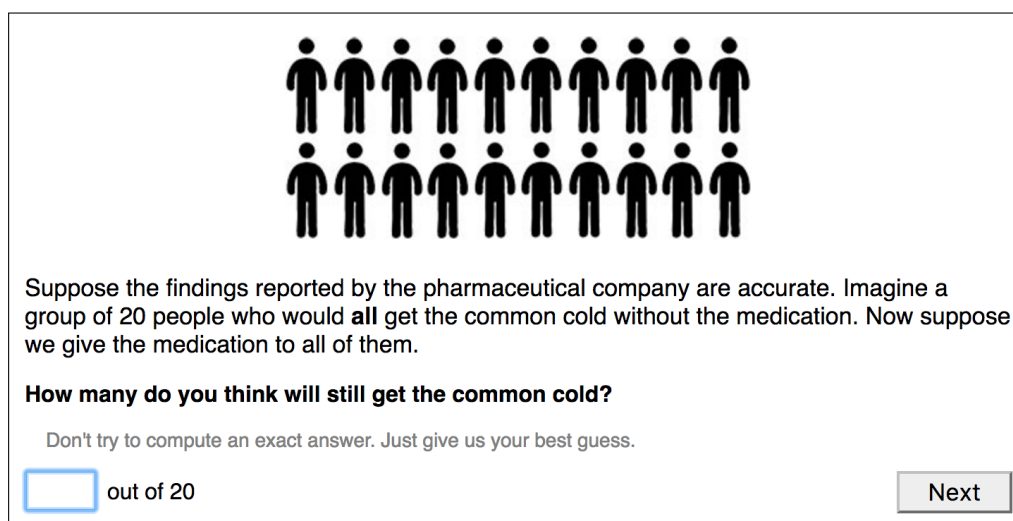


Fig. 3.12: Data comprehension test included in our first replication of BwS' second experiment.

3.3.2 First Replication of BwS' Second Experiment

The replicated stimuli and questions are shown in Figure 3.11. All participants were presented with a short text stating that a new drug was developed that was found to reduce the occurrence of the common cold by twenty percent, i.e., from eighty three percent without the drug to sixty three percent with the drug (the numbers were fully spelled out like this). For half of the participants, the text was followed by a restatement of the last two quantities using numerals (illness drops from 83% to 63%). Meanwhile, for the other half, the same two numbers were conveyed using a bar chart. All participants were then asked to report the extent to which they believed that the new drug was effective, on a 1–9 scale.

In the original experiment, a post-test was administered to rule out the possibility that the chart helped participants process the information – and thus to confirm that the chart was truly superfluous: about 30 minutes later, participants were asked to report “the percent by which the medication reduced illness” (the experiment was also part of a longer session and the question was asked at the very end). The correct answer was 20%. No statistical difference was found between the two conditions, leading the authors to conclude that “the effects of graphs [...] is not moderated by increased understanding or retention of information” [TW16]. This claim is however unsupported for a number of reasons, which are discussed in full in our paper [DJ18a].

To address these limitations, we designed our own comprehension test, so we can more reliably measure participants' intuitive understanding of the magnitude of the drug's effect in the fictional story. Our test, shown in [Figure 3.12](#), followed best practices in risk communication for medical decision making [[Gig+07](#); [MDF12](#)].

The best possible answer to the test can be derived by computing $20 \times 83/63 \approx 15.2$, which rounds to 15. The reasoning is as follows: let us first consider the entire population — not the people in the comprehension test. Out of 100 people there, 83 would have gotten sick (and 16 remained healthy) if no one took the drug (world *A*), while 63 would have gotten sick (and 27 remained healthy) if everyone took the drug (world *B*). It is fair to assume that the 17 people who were healthy in world *A* would also be healthy in world *B* (i.e., the drug does not cause some people to get sick). This means that there are $37 - 17 = 20$ people in world *B* whose cold was prevented by the drug. The drug thus prevents 20 colds out of 83, i.e., fails to prevent 63 colds out of 83. Therefore, among the 20 people in our comprehension test, $20 \times 63/83 \approx 15$ should be expected to get sick.

To be clear, this is not a trivial exercise and we did not expect that most participants would answer 15, especially since we instructed them not to calculate and to give us their best guess instead. Our assumption was that answers would be wrong overall, but if the chart did help understand the drug's true effect, answers would be on average less wrong with the chart. Accordingly, we used as dependent variable a continuous error metric (distance to the best answer, 15) instead of a binary measure of correctness.

We recruited 164 participants on a crowdsourcing platform, for a planned sample size of $N = 160$ (the data collection procedures and the analyses reported here were planned before data was collected).

Raw responses to the question “I believe the new drug is effective” are shown as histograms on the left side of [Figure 3.13](#). Although responses vary, most participants thought the drug was relatively effective, with 7 being the most common answer in both conditions. The mean response in each condition is reported in the middle of the figure: the two blue dots are the point estimates, and the two error bars are 95% confidence intervals. Thus there seems to be no evidence for a positive effect of the chart, as confirmed by the difference in means and its 95% CI, shown on the right side of [Figure 3.13](#). Point estimates from the original BwS experiment are shown in red for reference, with the 95% CI for the difference derived from the reported *p*-value [[AB11](#)].

Raw responses to the comprehension test are shown on [Figure 3.14](#). The accuracy of participants was rather poor: while the best answer was 15, responses were widely

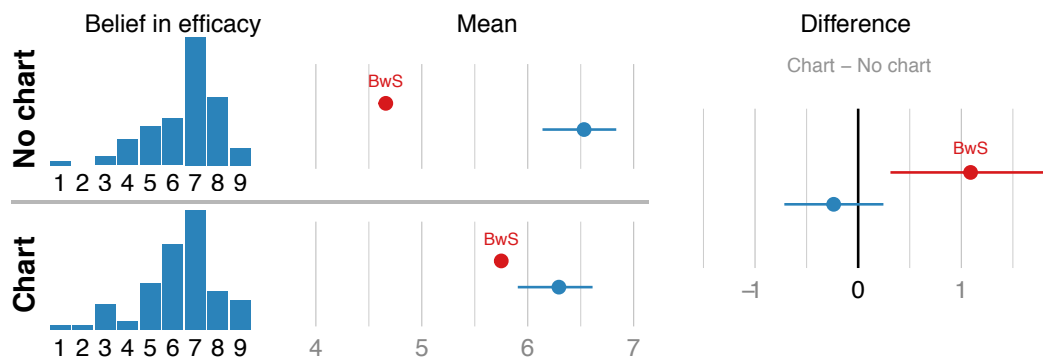


Fig. 3.13: Results from our replication of the second BwS experiment. *Left:* distribution of responses to the question “I believe the new drug is effective”, for the no-chart condition (top) and the chart condition (bottom). *Right:* means and difference in means (in blue). Red dots and error bars are the results from the original BwS experiment. Error bars are 95% CIs.

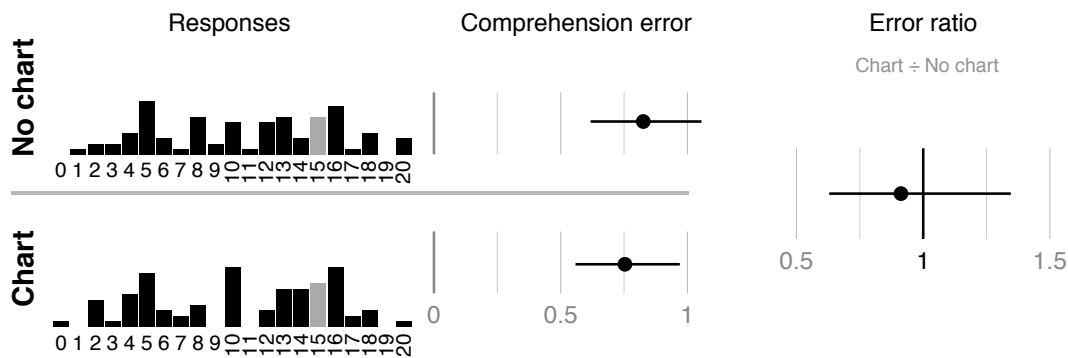


Fig. 3.14: Results for the comprehension test. *Left:* distribution of responses to the comprehension question, for the no-chart condition (top) and the chart condition (bottom); *Right:* mean response errors and their ratio. Error bars are 95% CIs.

distributed between 0 and 20. In addition, there is no evidence for a difference in mean comprehension error between the no-chart and the chart conditions.

Thus in this experiment, we were not able to replicate the original finding that the chart is better at **persuading** people that the drug is effective, despite our sample size being almost three times as large ($N = 164$ for us, vs. $N = 56$ for BwS). As we will see, this failure to replicate was consistent across our four experiments. In the experiment reported here, we were also unable to find that the chart increases participants’ **understanding** of the drug’s efficacy. As we will see, some of our other replications do provide weak evidence however, and combined together, there is some evidence for an effect.

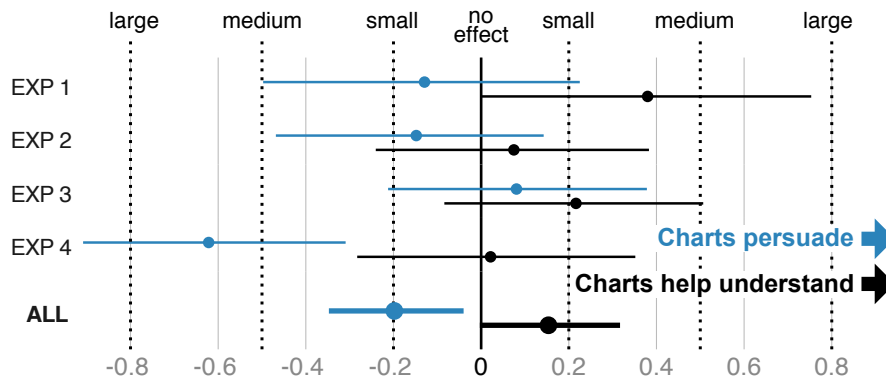


Fig. 3.15: Cohen's d for the chart's tendency to *increase* belief in efficacy (blue) and to *reduce* comprehension error (black). Error bars are 95% CIs.

3.3.3 Summary of All Experiments

We conducted four experiments in total, which were all similar except for a few details. For example, starting from our third experiment, we improved the design of our comprehension test to minimize inversion errors (i.e., participants erroneously reporting the number of people who are healthy rather than the number of people who are sick). In addition, our last experiment was conducted on Amazon MTurk with a US population (as in BwS), while our other experiments were conducted on CrowdFlower.

Figure 3.15 shows the standardized effect sizes we measured for each of our four experiments, and for all our experiments considered together (at the bottom). For blue estimates, the higher the value the larger the chart's contribution to *persuasion*. For black estimates, the larger the value the larger the chart's contribution to *understanding*. The results of the meta-analysis suggest that overall, the chart's contribution to understanding was most likely positive and its contribution to persuasion was very likely negative.

3.3.4 Conclusion

Considered together, our experiments suggest that as we expected, the chart used in the BwS study may have helped people to get a better sense of the data, but the effect is likely very small. Meanwhile, we were unable to replicate the original study's findings, as text with chart happened to be as persuasive or even less persuasive than text alone. This suggests that the effect may not be as robust as claimed and may need specific conditions to be reproduced. Regardless, within our experimental settings and considering our study

as a whole ($N = 623$), the chart's contribution to understanding was clearly larger than its contribution to persuasion. Therefore, the BwS study does not provide evidence that charts can persuade by their mere presence, just because they look scientific.

In order to establish that a chart biases judgment, it is necessary to also rigorously establish that it does not aid comprehension. Our study illustrates one way this can be done: if a chart's contribution to understanding is substantially lower than its contribution to persuasion in terms of standardized effect sizes (so the opposite result of Figure 3.15), it seems reasonable to assume that some persuasion has taken place through the peripheral route – i.e., based on extraneous cues rather than facts and logic [Pan+14]. One difficulty is this method's reliance on a particular comprehension test, which could be addressed by using a battery of comprehension tests.

Our focus was on closely replicating the BwS study, but there are many ways the study design can be improved if the goal is to investigate how charts in general can affect data comprehension or persuasion. In addition, the task may be too abstract and too artificial to capture persuasion in the real world. There seems to be more promise in testing more realistic tasks, e.g., asking people to assess real facts on meaningful social issues [Pan+14], exposing them to ads in their personal environment [Meh00], and measuring intentions and behavior instead of beliefs [NR15].

3.4 Communicating Humanitarian Data

The news are full of reports of human tragedies. While some people deplore a negativity bias, communicating about humanitarian issues is important if the audience can help alleviate them, e.g., by donating or by voting. In order to better ground stories in facts and better convey the magnitude of tragedies, journalists and public communicators have started to move beyond mere stories and visualize data. Academics have started studying this practice, but work so far has remained scattered, there is a lack of clear language for thinking about the space of problems and solutions, and most studies have yielded disappointing results. In this section, I summarize work that attempts to fill these gaps.

This section is based on two papers [Mor+20; Mor+21] authored by Luiz Morais, Yvonne Jansen, Nazareno Andrade, and myself, with elements from a pre-print by myself [Dra22b]. Supplementary material is at <https://luizaugustomm.github.io/pages/anthropographics-ds.html> and <https://osf.io/wfgmp/> (for the design space) and <https://anthropographics.github.io/> and <https://osf.io/epzub> (for the study).

Any Age Group ▾ Any Gender ▾ Any Location ▾ Any Date ▾ No Keywords ▾

Matched Deaths: **12,042** or more between Newtown and Dec. 31, 2013



Fig. 3.16: Newspaper visualization showing people who died by a gun shot between the Newtown attack in 2012, and December 2013. Each icon is a person. Sex and age are shown through the icon's shape. Additional information about each person can be obtained by clicking on their icon. Source: [Slate \[KK13\]](#).

3.4.1 Introduction

When communicating data about people, information designers and data journalists regularly create data visualizations meant to create an emotional connection with the persons whose data are represented. Figure 3.16 shows one example where the reader can see the story and personal information of each person who died in a public mass shooting in the USA. Meanwhile, the visualization in Figure 3.17 conveys the hardship of the life of refugees by narrating the story of S.W.G., a 26 years old refugee who left Pakistan and spent 651 days before arriving at his final destination. Both visualizations were crafted to bring readers closer to the persons whose data is visualized.

The practice of visualizing data about people in a way that helps the audience relate has been studied in academia and called *anthropographics* [Boy+17; Ber17]. We subsequently defined anthropographics as “visualizations that represent data about people in a way that is intended to promote prosocial feelings or prosocial behavior” [Mor+20]. More recently, I introduced the term *humanitarian visualization* to refer to “data visualizations or infographics designed to promote human welfare” [Dra22a]. Although the two concepts are not identical, I noticed that the term anthropographics tends to confuse people, so

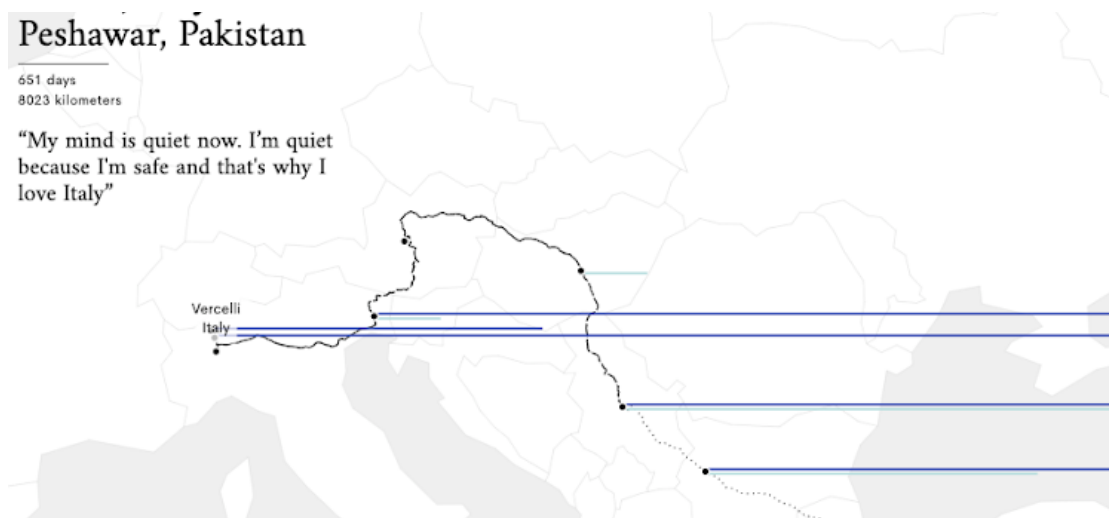


Fig. 3.17: Excerpt from an interactive web visualization showing stories of seven refugees who fled from their home country (here, a migrant who left Pakistan towards Italy). The stroke patterns in the path correspond to different means of transportation (e.g., by foot or by car); the dark blue lines represent the days of stay in a city while the light blue ones represent days of travel. Source: storiesbehindaline.com [FP19]

in this HDR thesis I will use the term humanitarian visualization instead. Although humanitarian visualization is very broad by definition, in this section I will be mostly focusing on conveying data about human tragedies.

Visualization designers have explored many strategies for creating humanitarian visualizations. A popular strategy is to use human-shaped symbols to represent individuals (e.g., in Figure 3.16), while other strategies include the use of text annotations to make each person appear unique (e.g., in Figure 3.17), or the use of visual metaphors (e.g., a red bar chart to symbolize blood or death). Although humanitarian visualizations have started to be studied in visualization research, the community still lacks a precise language for discussing them, as well as an overview of design strategies and opportunities for design. To address this, we propose a conceptual framework that is meant to help researchers and practitioners reason and communicate about humanitarian visualizations.

3.4.2 Quick Overview of the Design Space

Based on a previously-published design space [Boy+17] and the examination of a collection of 105 communicative data visualizations that convey data about people, we

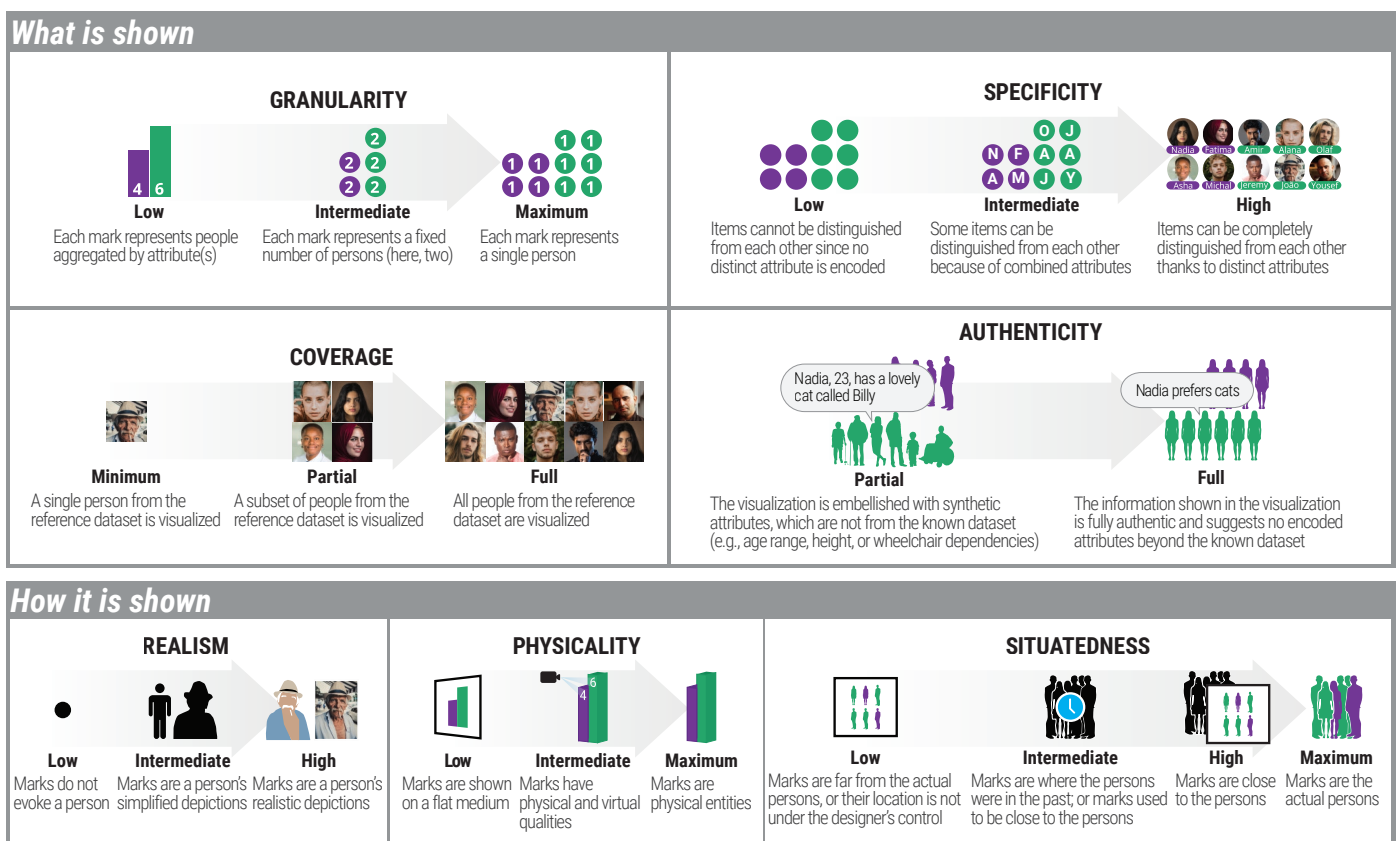


Fig. 3.18: The seven dimensions of our design space of humanitarian visualizations.

identified seven design dimensions that can be reasonably believed to have some effect on prosocial feelings or behavior, namely: *granularity*, *specificity*, *coverage*, *authenticity*, *realism*, *physicality*, and *situatedness*. Those seven design dimensions are grouped into two categories: *what is shown*, which are dimensions related to what information and how much information is represented in a visualization, and *how it is shown*, which is about how the information is represented.

We summarize these dimensions in Figure 3.18 using the minimalistic dataset of Table 3.1 as an example. The dimensions are as follows:

- **What is shown:**
 - **Granularity** refers to the extent to which the people in the dataset map to separate visual symbols. In a visualization with low granularity, each symbol stands for a large population, as is commonly the case in statistical charts. In a visualization with maximum granularity, each data item maps to a single person (e.g., Figure 3.16).

Name	Gender	Likes	Photograph
Yousef	Male	Dogs	(image)
Amir	Male	Dogs	(image)
Alana	Female	Dogs	(image)
Olaf	Male	Dogs	(image)
Jeremy	Male	Dogs	(image)
João	Male	Dogs	(image)
Fatima	Female	Cats	(image)
Nadia	Female	Cats	(image)
Asha	Female	Cats	(image)
Michal	Male	Cats	(image)

Tab. 3.1: Fictional dataset used to illustrate the dimensions in Figure 3.18.

- **Specificity** refers to how much information is shown about each person or group of people. In a visualization with low specificity, symbols convey little or no information and tend to be visually very similar. In a visualization with high specificity, symbols convey rich information and tend to be visually unique.
- **Coverage** refers to the extent to which the people visualized capture the population the visualization is about. In a minimum-coverage visualization, the visualization only features a single person from that population. A visualization like in Figure 3.17 provides partial coverage (six people) of the entire population of refugees the story is about. Full coverage means the full population is represented, aggregated or not.
- **Authenticity** refers to the proportion of genuine information in the visualization. The illustration of partial authenticity in Figure 3.18 shows a design where symbols are given different silhouettes to make them look unique: we can see, for example, a child and a person in a wheelchair, and personal information about people. However, none of this information is in the dataset (Table 3.1). Meanwhile, in a visualization with full authenticity, all visualized information is genuine.
- **How it is shown:**
 - **Realism** refers to how much the visualization’s symbols look like actual persons. A visualization with low realism represents people or groups of people using abstract shapes or symbols, e.g., dots, bars, or glyphs. In contrast, a visualization with high realism is made of shapes or images that closely resemble actual people.
 - **Physicality** refers to the degree to which a visualization’s marks are embodied in physical objects [Jan+15]. In a visualization with **low physicality**, the marks are shown on a flat medium, such as a computer screen or a sheet of paper. All visualizations designed for the web or for magazines fall in this category. In visualizations



Fig. 3.19: Art installation using grains of rice to show various statistics about people. Here, the piles of rice show victims of the 1906 earthquake, where each grain of rice stands for one victim. Image from <https://www.youtube.com/watch?v=zvv110w5lww>.

with **maximum physicality**, the visualization is made of physical objects or actual persons. For example, in [Figure 3.19](#), each grain of rice represents a person.¹

- **Situatedness** refers to how physically close a visualization is (or was) to the people it represents [[WJD17](#)]. In visualizations with low situatedness, which is the most common case, the visualization is either presented far from the people, or its physical location is not under the control of the designer. On the opposite end of the spectrum, a visualization has maximum situatedness if it is made of the people it represents. Examples include so-called human histograms, where people sort themselves by height or other attributes to form life-sized histograms (e.g., <http://dataphys.org/list/davenport-physical-distributions/>).

3.4.3 Overview of Previous Studies

In a subsequent paper [[Mor+21](#)], we used our design space to characterize 11 experiments (5 studies) that have been conducted between 2017 and 2020 in information visualization to test humanitarian visualization design strategies. [Figure 3.20](#) gives a visual summary of those studies. It would take too long to describe this diagram in detail

¹Using collections of physical objects to convey death tolls is a common theme in art installations and memorials. Other examples include nails (dataphys.org), artificial flowers (dataphys.org), toe tags (dataphys.org), shoes (cnn.com), tattoo dots (dataphys.org), and actual people (dataphys.org, classicfm.com).

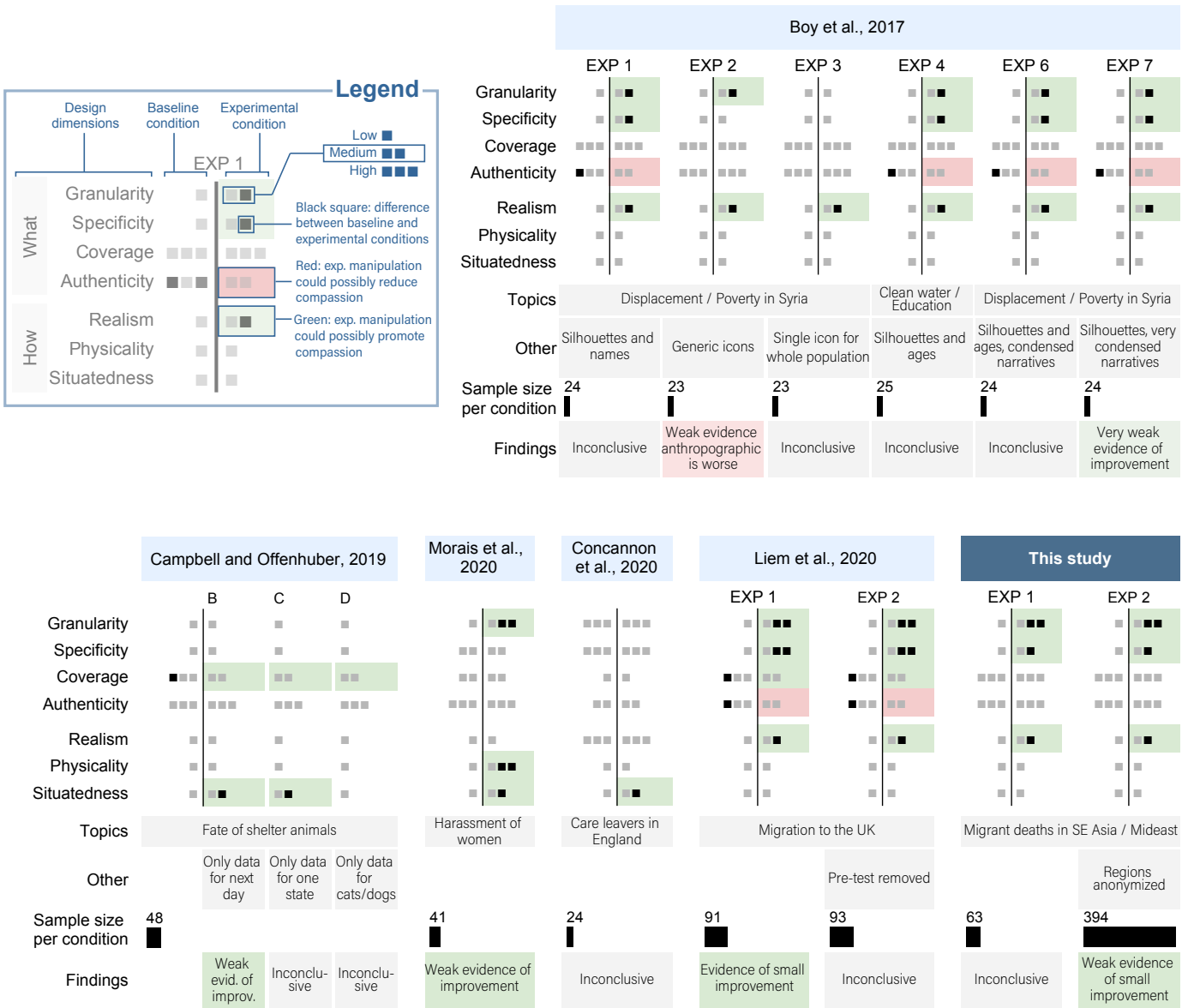


Fig. 3.20: Visual summary of five prior studies on humanitarian visualization (11 experiments in total), plus the new study reported later in this section (2 experiments).

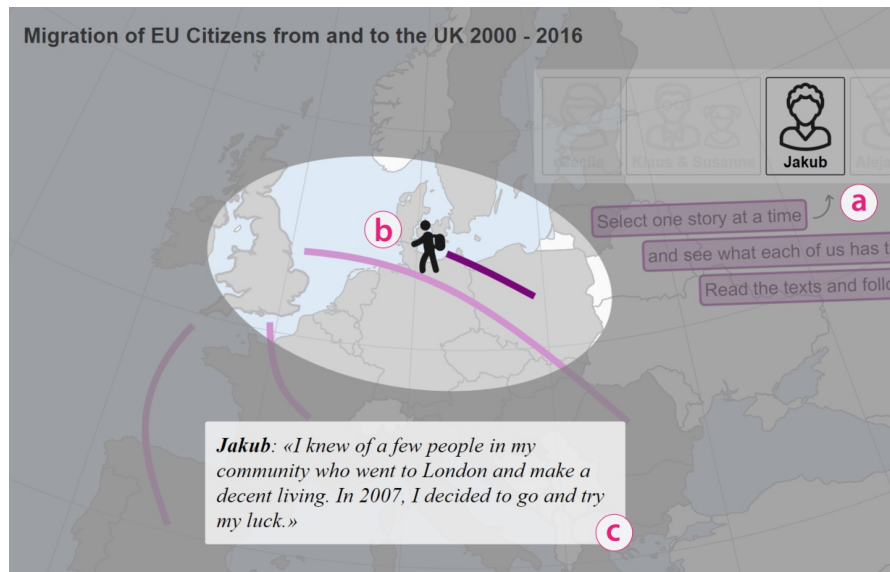


Fig. 3.21: Excerpt from an interactive visualization of UK migration data in the study by Liem et al. [LPW20]. Participants can go through the story of any of six fictional characters, represented as animated icons. The stories included the reasons why the person moved to the UK, what they gained from it, and what they did afterward.

(see [Mor+21] for more), so we will only skim over it. Each blue rectangle (and the stuff underneath it) is one study, and the dark blue rectangle on the bottom right is our own study. Each study reports one or more experiments (labeled EXP1, EXP2,...), and each experiment generally compares a control condition with an experimental condition. Conditions are visualization designs, which can be partly described using our design space: for example, a visualization design can have low or high granularity, low or high specificity, etc. Figure 3.20 summarizes those design dimensions for each experiment, for the control condition (to the left of the vertical black line) and for the experimental condition (to the right of the line). The findings are summarized at the very bottom.

Two key takeaways from Figure 3.20 are: 1) many experiments are inconclusive; and 2) many experiments compare statistical charts with designs that emphasize individuals, but where the information provided about individuals is not authentic (see Figure 3.21 as an example). More specifically, several past experiments use a control condition with low granularity, low specificity and low realism (i.e., the visualizations are made of symbols that stand for entire populations, do not convey detailed information, and do not look like people), and compare it with a design with high granularity, medium specificity and medium realism (i.e., visual symbols stand for individuals or small groups of individuals, they convey more detailed information, and they evoke people). These three design

strategies are expected to increase participants' concern about the population represented – which is highlighted in [Figure 3.20](#) with green rectangles. However, we think that visibly unauthentic information could simultaneously make participants distrust the visualization, and thus distance themselves from the people represented (which is highlighted with pink rectangles in the same figure). This may explain why those designs were less effective than the authors expected.

Based on this observation, we decided to run a new study that has analogous control and experimental conditions, but where the information provided about individuals is genuine. We summarize this study next.

3.4.4 Our Own Study

We chose to visualize data from the Missing Migrants Project (missingmigrants.iom.int), which “tracks deaths of migrants, including refugees and asylum-seekers, who have died or gone missing in the process of migration towards an international destination”. The data is collected from official records and media reports. Each entry in the dataset describes an incident, such as a drowning, a vehicle accident, or a shooting. Every incident contains the number of victims and survivors (split into men, women, children, and unknown), as well as where and when the event occurred.

We visualized subsets of this dataset in two ways: using an information-poor visualization design (*infoPoor*) and using an information-rich visualization design (*infoRich*). The *infoPoor* design ([Figure 3.22](#)-top) is a simple bar chart showing the number of migrants who died or survived an incident during their migration attempt in 2018. The *infoRich* design ([Figure 3.22](#)-bottom) shows one row for each incident that took place in 2018. A text label indicates the date, nature, and location of the incident. Underneath, each individual migrant involved in the incident is shown as a human silhouette conveying their gender according to the dataset (male, female, unknown), age group (child or adult), and whether they died or survived (red vs. light pink). Participants had to scroll to see the entire visualization. The inset on the right of [Figure 3.22](#) gives an indication of the full height of the visualization.

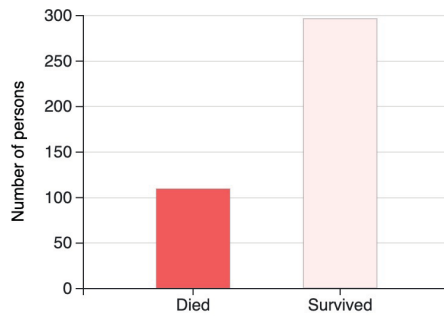
The two designs differ in three major respects: granularity, specificity and realism. Since the *infoRich* design shows real data, it has the same level of authenticity as the *infoPoor* design. The two visualization designs differ in other respects – importantly, *infoRich* requires a significant amount of scrolling, while *infoPoor* fits in the browser window.

Information-poor visualization (bar chart)

Migrant crisis in the Middle East

In the Middle East, irregular migrants are forced to leave their countries due to civil conflicts that have been occurring in the region for decades. Many migrants die in the process.

Here is a chart showing for 2018 how many migrants died due to accidents or attacks and how many survived (source: <https://missingmigrants.iom.int/>):



Next

Information-rich visualization (humanitarian design)

Migrant crisis in the Middle East

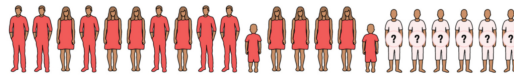
In the Middle East, irregular migrants are forced to leave their countries due to civil conflicts that have been occurring in the region for decades. Many migrants die in the process.

Here is a chart showing for 2018 how many migrants died due to accidents or attacks and how many survived (source: <https://missingmigrants.iom.int/>):

Legend: Each person in the chart represents a single migrant



January 19, 2018 - Hypothermia, frostbite (Masna'a, Lebanon, near the border with Syria)



January 23, 2018 - Unknown (Near Masna'a border crossing, Lebanon, near border with Syria)



Fig. 3.22: Two stimuli used in our study, showing data from the Missing Migrants Project (missingmigrants.iom.int). On the top is the information-poor control condition, and on the bottom is the information-rich experimental condition (participants had to scroll to see the entire visualization – see overview on the right).

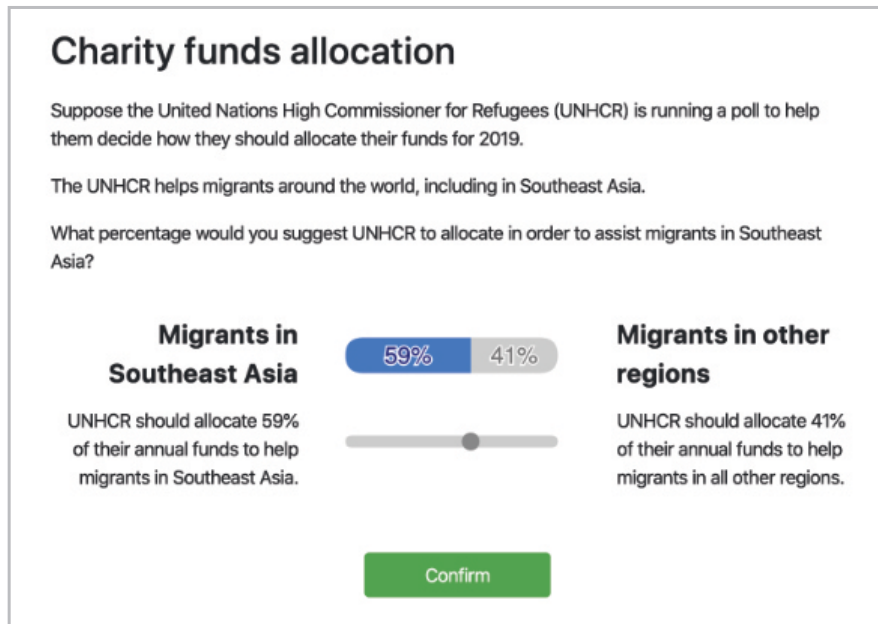


Fig. 3.23: Page from experiment 1 asking participants to allocate hypothetical funds.

Similarly, *infoRich* requires considerably more time to be fully examined than *infoPoor*. All these potential confounds will make it hard to identify the cause of an experimental effect should we find one. However, the simultaneous manipulation of multiple design dimensions was a deliberate feature of this initial experiment, whose goal was to confirm that we are indeed able to detect an effect if we try to amplify it as much as possible.

We showed those two visualizations designs to 128 participants on a crowdsourcing platform. Half of the participants saw a subset of the missing migrants dataset (migrants in Southeast Asia, year 2018) shown with the *infoRich* design, and then saw another subset (migrants in the Middle East, year 2018) shown with the *infoPoor* design. The other half of the participants saw the Southeast Asia dataset with the *infoPoor* design, and then the Middle East dataset with the *infoRich* design. After each visualization, we measured participants' concern for the migrant population by asking them how much hypothetical funds they think should be allocated to help the affected population vs. helping people elsewhere in the world (see [Figure 3.23](#)). At the end of the experiment, we also asked them to split hypothetical funds between the two populations.

Note that our questions were not designed to measure affect. While many essays and studies on anthropographics are concerned with promoting prosocial emotions such as empathy (e.g., [Boy+17; IDW18; Con+20]), empathy does not necessarily lead to helping

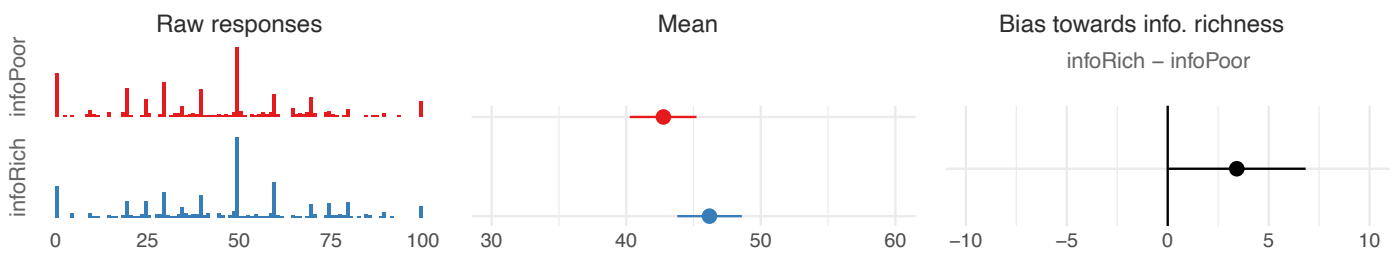


Fig. 3.24: Raw responses, mean responses, and difference in means for the funds allocation in experiment 2. Error bars are 95% CIs.

behavior [Blo17; WB10b; LD16]. Therefore, we chose to measure people’s judgment about the extent to which a certain population needs help, regardless of whether this judgment is the result of deliberate reasoning or of intuition and emotions.

The experiment yielded inconclusive results: we did not find clear evidence that the *infoRich* design increased people’s concern for the population visualized. Therefore, we ran a second experiment which was an improved and simplified version of the first one, with a higher statistical power. The new experiment only used one migrant dataset (and was thus fully between-subjects), adjusted the framing of the scenario and of the funds allocation question, and used a much larger participant sample ($N = 786$). The main result is reported in Figure 3.24. The left chart shows the distribution of responses to the funds allocation question for each of the two visualization conditions (e.g., a value of 25 indicates that the participant decided to allocate 25% of the available funds to the migrants and 75% to an alternative cause). The middle chart shows the mean responses, and the rightmost chart shows the difference of means. There is some evidence of an effect in favor of the information-rich design, which seems to prompt larger funds allocations on average, though the difference is small ($diff = 3.4\%$, 95% CI [0.02, 6.8]).

3.4.5 Conclusion

There is increasing interest in humanitarian visualization, and especially in ways to convey data about human tragedies in order to make readers compassionate with the victims. Humanitarian visualization is a rich and growing area, but the work so far has remained scattered. Despite preliminary empirical work and a few web essays written by practitioners, there is a lack of clear language for thinking about and communicating about such visualizations. We addressed this gap by introducing a conceptual framework and a design space for humanitarian visualizations that other researchers can use.

Empirical studies have started to test designers' recommendations and intuitions by examining whether various design strategies for humanitarian visualizations indeed promote prosocial feelings or behavior. Results have been mostly disappointing so far, in contradiction with the intuitions of many designers and researchers.

We applied our conceptual framework to provide a detailed overview of past experiments. We then conducted two new experiments that use large samples and a combination of design strategies to maximize the possibility of finding an effect. But our information-rich design had a small effect at best on funds allocation. Such a small effect may yield non-negligible outcomes in a large-scale donation campaigns, but it seems very difficult to study experimentally. Indeed, our second experiment was barely able to provide sufficient evidence, despite having involved about 800 participants for a cost of about 500 GBP (550€ or \$650). As this effect arises from three design strategies combined and other potentially influential factors (e.g., asking participants to scroll, different exposure times), disentangling these effects may require sample sizes that are impractical to achieve.²

It remains possible that there exist alternative design strategies that have a large effect on measures of judgment, decision making, or behavior. However, the overall inconclusive results from the range of experiments conducted by researchers so far call for some skepticism. There is presently no clear evidence that if designers employ current design strategies, this will have a clear and observable impact on people's decisions and behavior. Nevertheless, many areas of the vast design space of humanitarian visualization have not been tested, some of which appear promising (see, e.g., discussions in [Mor+20] and in the next chapter), and it is crucial to test them to advance knowledge.

It is also possible that different ways of measuring prosociality (e.g., observing donations in actual charity campaigns) would lead to different results, perhaps revealing a clear effect where crowdsourced studies could not. It would be surprising if an ecologically more valid measure of prosociality yielded a larger effect (since it is generally harder to measure an effect in the wild than in the lab), but such studies are nevertheless needed. Finally, other ways of using visualization to support humanitarian causes deserve to be considered; For example, it can be useful to study how visualizations can help existing donors better allocate their money to charities, which carries the potential of doing far more good than trying to maximize donations for any single charity [Ord19]. This avenue of research will be discussed in the next chapter.

²The relationship between sample size and effect size is not linear. For example, being able to detect an effect four times smaller than experiment 2 (Cohen's $d=0.05$) while maintaining a power of 0.8 would require $N=12,562$ participants.

Perspectives — Supporting Rational Decisions and Effective Communication on Humanitarian Issues

2023–2024 Update. This chapter is largely based on a pre-print I wrote back in 2022. In the light of recent revelations about serious problems in certain branches of the effective altruism (EA) movement, I no longer support EA as a whole. I do still support EA's general philosophy and its branch focusing on alleviating global poverty, which are the focus of this chapter. More info at <https://dragice.fr/ea.html>.

In this chapter, I explore possible future directions of research that combine the goals of the two main chapters of this thesis (supporting rational judgment and decisions, and supporting effective communication), and applies them to the area of humanitarian visualization (discussed in [section 3.4](#)), which is an area that I am particularly interested in developing further. Much of the discussions here are informed by the effective altruism movement. I have been aware of this movement for nearly a decade, but only after working on humanitarian visualization with Luiz Morais have I realized that it could inform my research. This chapter is based on an unpublished pre-print by myself [[Dra22a](#)]. In another pre-print, I elaborate on the opportunities offered by emerging user interface technologies [[Dra22b](#)].

4.1 Introduction

A lot has been written on how data visualizations can be useful for a variety of tasks [[Fek+08](#)], but only recently have researchers started to explicitly consider how they can be used to promote human welfare. In particular, as we have seen in the previous

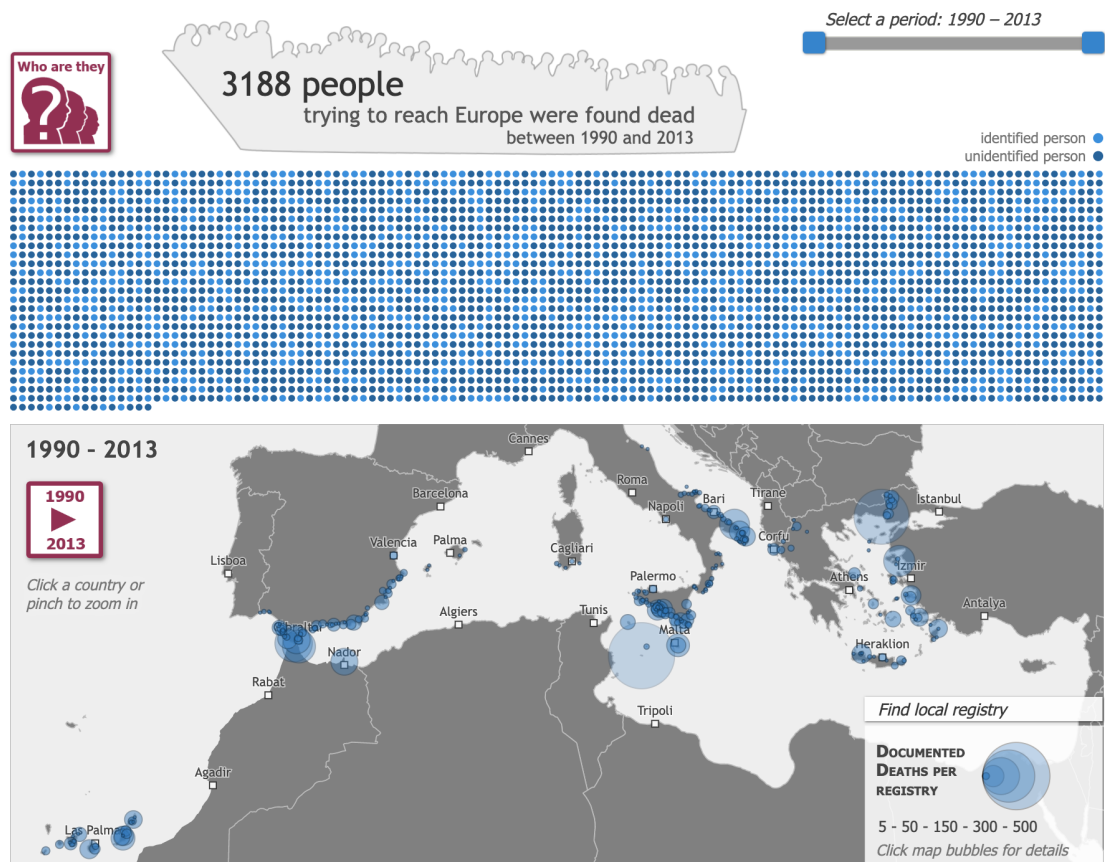


Fig. 4.1: Interactive web visualization showing deaths at the borders of Southern Europe between 1990 and 2013. Each small dot represents one dead migrant documented by local authorities. Source <http://www.borderdeaths.org/>.

chapter, visualization researchers have started to look at how data journalists use data visualizations to raise the public awareness about humanitarian issues (see Figure 4.1).

Despite the great potential of humanitarian visualization as a practice and a research area, its ability to make progress and have a positive impact is limited by the way problems are generally framed, and by the types of solutions and metrics of success typically considered. Suppose a newspaper publishes a striking infographic about the number of blind persons in the U.S. who need a guide dog, resulting in lots of people donating to a charity that provides guide dogs. Most practitioners and researchers would certainly consider this infographic a success, and a benefit to society. This is clearly the case if the counterfactual is a world where the same donors keep the money for themselves. However, the situation is less clear if the counterfactual is a world where the donors choose to give their money

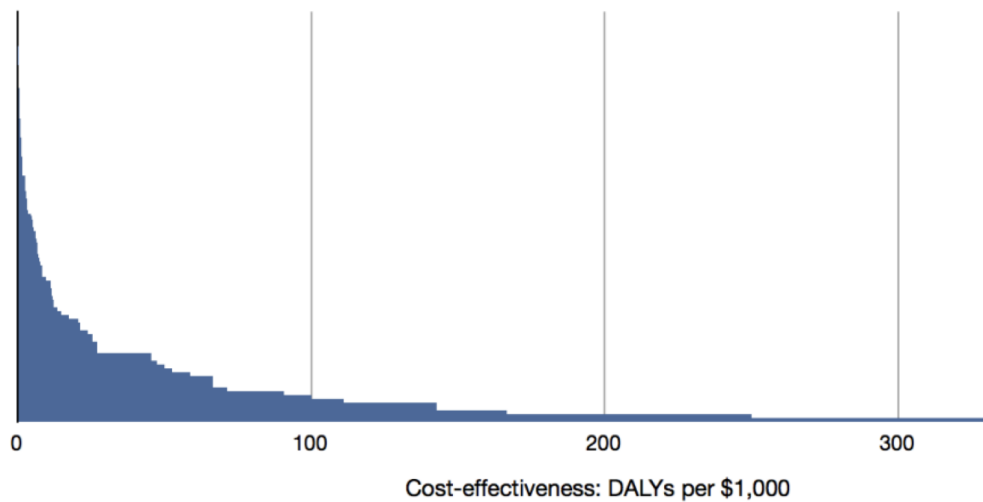


Fig. 4.2: Distribution of the cost-effectiveness of 100+ health interventions, expressed in disability-adjusted life years (DALYs) per \$1,000. Source [Ord19].

to a charity that provides even more benefits. In fact, training a guide dog in the U.S costs \$40,000, and the same amount of money can be used to cure more than 2,000 people in Africa of blindness by paying for surgeries to reverse the effects of trachoma [Ord19]. How money should be allocated in this case is a matter of personal judgment, but a lot can be gained by helping people – both the general public and decision makers – allocate their limited resources in an informed manner, with sufficient knowledge about the range of options they have at their disposal, and their expected impact.

Impact would not be a central concern if charity programs and interventions were roughly comparable in terms of their human impact, but this is very far from being the case. For example, a global health study has estimated the effectiveness of more than 100 health interventions in terms of disability-adjusted life years (DALYs) saved per \$1,000, finding an remarkably wide and skewed distribution (see Figure 4.2). Effectiveness ranges from 0.02 to 300 depending on the intervention (a factor of 15,000), with a median of 5, suggesting that the vast majority of health interventions are far less effective than the most effective ones. Thus, moving money from the many ineffective interventions to the most effective ones is likely to be helping people considerably more than donating even a lot of money to a random intervention.

Yet donating money to a random intervention is precisely the kind of task currently supported by most humanitarian visualizations, due to their focus on *case-by-case persuasion*: a cause is pre-identified, and the goal is to raise people’s concerns about that cause as

much as possible. Similarly, in research, the focus is on finding the design strategies that are the most effective independently from the cause.¹ However, strategies that help make a message persuasive are not necessarily the ones that promote the best decisions. As an example, a lot has been written on the power of storytelling in visualization [KM13]; But as Neil Halloran, the designer of the celebrated data-driven documentary *The Fallen of WWII* has pointed out, “you can tell a story about a crisis of any size, and tell a compelling story” [Hal17], implying that stories do not necessarily help people to think rationally about the extent of human suffering. Visualization researchers could greatly benefit from new thinking frameworks to help them move beyond case-by-case persuasion and reason about how to best use visualization to alleviate human suffering on a global scale. Effective altruism provides such a framework.

4.2 What is Effective Altruism?

The term *effective altruism* (often abbreviated as *EA*) was coined in 2011 at Oxford University, by a small group of academic philosophers and individuals involved in charity and philanthropy organizations [Mac19]. In 2016, the head of this group, William MacAskill, worked with many leaders involved in the EA community to write a definition that has been widely endorsed by the community:

Effective altruism is about using evidence and reason to figure out how to benefit others as much as possible, and taking action on that basis. [Mac19]

In 2018, using again input from many EA leaders, MacAskill proposed a more precise definition to be used in academic discussions:

Effective altruism is: (i) the use of evidence and careful reasoning to work out how to maximize the good with a given unit of resources, tentatively understanding ‘the good’ in impartial welfarist terms, and (ii) the use of the findings from (i) to try to improve the world. [Mac19]

This definition highlights the double aspect of EA as (i) an intellectual project (a research field) and (ii) a practical project (a social movement). The definition is non-normative: it

¹To be clear, these criticisms apply to my own research on humanitarian visualization which I have summarized in [section 3.4](#).

does not say how people should behave (e.g., that we should make personal sacrifices to help others). *Welfarist* means that views that assign intrinsic value to other things than well-being (e.g., biodiversity, art, or knowledge) are excluded, while *impartial* means that views that do not weigh people equally (e.g., prioritizing nationals over foreigners) are also excluded. *Tentative* means that the impartial welfarist view is a working assumption that can be debated and refined within EA. For example, while animal welfare is a central concern for many EA proponents, how much moral weight should they be given compared to humans remains unclear.

These broadly-accepted definitions are very helpful when discussing the merits and weaknesses of EA, because many criticisms of EA arise from people using their own interpretation of what it is. This leads to common misconceptions, such as: EA is just applied utilitarianism, it is only about fighting poverty, it is only about donations or earning to give, and it ignores systemic change (for discussions see [Mac19]).

4.3 How Can EA Inform Visualization

Because people have different moral intuitions, not all researchers working on – or considering working on – humanitarian visualization will find the EA philosophy compelling enough to embrace it. But for those who do, EA can provide a clear thinking framework in an area that has been lacking one. Indeed, many discussions so far have focused on how to design visualizations that elicit empathy, often ignoring that empathy does not necessarily promote helping behavior [Mor+20]. Even when a visualization does cause people to act, their actions can have a negligible, null, or possibly even negative impact on global human welfare. EA provides clear grounds to think about research goals and metrics of success.

The EA lens can also help researchers think out of the box and broaden the scope of humanitarian visualization research by identifying new types of solutions and approaches. In particular, some visualizations may not promote prosocial feelings or behavior – and thus might not be considered conventional humanitarian visualizations – but may still promote welfare. For example, a visualization that helps a charity director effectively allocate money across different health programs does not promote prosocial feelings or behavior (since all the money will be used to help people no matter what), but it can tremendously increase human welfare.

Intervention	Top charities implementing this intervention	Outcome	Cost of outcome (2021 average, rounded up to nearest \$500)	Output	Cost of output (2021 average, rounded to nearest dollar)
Malaria nets	Against Malaria Foundation	Lives saved	<u>\$5,500</u>	Number of nets delivered to households	<u>\$5</u>
Malaria medication	Malaria Consortium	Lives saved	<u>\$5,000</u>	Number of children treated with a full course of medicine	<u>\$7</u>
Incentives for immunizations	New Incentives	Lives saved	<u>\$5,000</u>	Number of infants vaccinated who would not have been vaccinated without the program	<u>\$160</u>
Vitamin A supplementation	Helen Keller International	Lives saved	<u>\$3,500</u>	Number of children receiving supplements	<u>\$1</u>
Deworming	SCI Foundation, Sightsavers, Evidence Action, The END Fund	N/A	<u>N/A</u>	Number of children dewormed	<u>\$1</u>
Cash transfers	Give Directly	Cash transferred	<u>83% of dollars donated</u>	N/A	<u>N/A</u>

Fig. 4.3: Table showing impact metrics for six charities identified as among the most effective by GiveWell in 2021. Source <https://www.givewell.org/cost-to-save-a-life>.

EA is a thinking framework but it is also a community. This community is full of people who are deeply knowledgeable about humanitarian issues or have been extensively involved in humanitarian actions, and thus visualization researchers could learn a lot by connecting with them. In addition, the EA community has unique needs that visualization could help address. For example, several EA organizations do research on the effectiveness of different charities and charity programs, in order to guide potential donors. GiveWell is a known example: it maintains a list of top effective charities, primarily based on the cost of life saved (see Figure 4.3). GiveWell shares a range of spreadsheets with data and calculations to explain how it arrived at its estimates. All such initiatives generate lots of useful data, but the amount of information can rapidly become overwhelming for potential donors. And yet, data is currently largely communicated through numbers and text, and very rarely through visualizations. Perhaps visualization could also be used by the EA communicators to better explain its general principles to naive audiences, and by EA researchers to help them analyze the effectiveness of different charity programs.

4.4 Using Visualization and Psychology to Support EA

People share many misconceptions and biases preventing them from helping effectively – for example, geographical and cultural proximity often greatly affect how much people feel like helping [CSG21]. Researchers studying humanitarian visualization can take inspiration from recent work on judgment and decision making with visualizations [Dim+18], and apply findings and methods from psychology to study how visualizations interact with cognitive biases, and whether visualizations can help alleviate those biases. Unfortunately, much like visualization research, psychology research has mostly focused on how to make people donate more, rather than more effectively. However, Lucius Caviola and colleagues [CSG21] have recently done a tremendous job at reframing past findings through an EA lens, leading them to identify major psychological obstacles to effectiveness, which fall in two categories:

1. *Motivational obstacles.* People think that whether and how to help is largely a matter of personal preference; they give based on how much they feel emotionally connected to the issue (e.g., they feel more strongly about diseases that are common in their country or have affected their loved ones); they dislike prioritizing some causes over others; they view people who try to donate rationally more negatively than those who donate based on empathy.

2. *Epistemic obstacles.* People think that charity overhead is wasteful, or find funding overhead unsatisfying; they think that effectiveness cannot be quantified; they do not think clearly about probabilities; they are not aware that charities differ greatly in their effectiveness; they don't know which charities are the most effective.

Caviola and colleagues also identified four types of strategies to increase effective giving: information, choice architectures and incentives, philosophical reasoning, and norm changes.

Information addresses epistemic obstacles through education. As I mentioned before, one of the areas where visualization can help is by conveying rich quantitative facts about charity effectiveness in a way that is easy to process. Visualization could also be used to argue for lesser-known EA causes such as wild animal suffering and global catastrophic risks, by conveying data about how serious, neglected and tractable these causes are. Finally, visualization could also help dispel misconceptions, for example by showing data about how charity overhead is employed, together with simulations illustrating how cutting overhead would likely yield less positive outcomes. Information is the type of

strategy where the possible benefits of visualizations are the most evident, and where a lot can be done in collaboration with the EA community.

Choice architectures and incentives address motivational obstacles by nudging (e.g., using effective charity programs as default options) and incentivization (e.g., using donation matching or tax deductions targeted to effective charity programs). Here, possible roles for visualization are less immediately evident, but this type of strategy can potentially lead to the most interesting innovations and contributions to knowledge. In particular, it could be interesting to study which nudging techniques can translate to visualizations. For example, a well-documented bias is the decrease of people's concern for individual victims as the number of victims increases: a tragedy that affects one million people typically does not generate 100 times more concern or donations than a tragedy that affects a thousand people [CSG21]. However, this effect is less pronounced when donors evaluate all options at the same time – which could mean seeing data about multiple tragedies visualized side-by-side – than if they evaluate the options sequentially.

The last two categories listed by Caviola and colleagues, *philosophical reasoning* (exposing people to philosophical arguments) and *norm changes* (pushing for a change of moral standards) are important but probably less directly relevant to visualization.

4.5 Combining Experiences with Quantitative Facts

Again, a major way in which visualization can support EA is by helping people compare charity programs. To take a trivial example, an EA website could include as an overview of its top programs a bar chart of the number of lives saved per unit of donation for each program.

However, this is only a minimalist example, and important visualization design challenges arise when a variety of outcomes need to be visualized and compared. For example, about 600 mosquito nets prevent the death of a child, but they also prevent 500 to 1,000 cases of malaria [Aga22]. This is an enormous benefit in and of itself, as malaria is a crippling disease with flu-like symptoms that can periodically return, can be highly disruptive for the life of households, and can leave children disabled [Ric12]. Similarly, GiveWell lists a charity that saves lives by giving vitamin A supplements to children, but even when it is not fatal, vitamin A deficiency causes a range of terrible problems such as repetitive infections and blindness [Wor22]. GiveWell sometimes go beyond lives saved

and considers charities expected to impact the recipient's lifetime earnings (treatments for parasitic worm infections) or their overall quality of life (cash transfers for extreme poverty). Another effective altruism website lists a charity that can use about \$20,000 to prevent a year of homelessness in the US or UK [Cla20], and another one that can use \$200–\$300 to prevent the equivalent of one year of severe major depressive disorder for a woman in Uganda [Hal19]. It is very hard to imagine how to visualize those widely different types of outcomes in a way that supports informed, effective-altruist decisions.

Ideally, a major donation or funds allocation decision should be based both on quantitative facts (e.g., the number of people affected, the cost of interventions) and a deep understanding of people's subjective experiences with and without the interventions, especially concerning the degree of physical and psychological suffering involved. However, it is hard for a person who has never contracted malaria or never had a vitamin A deficiency to have a reliable intuition of what those experiences entail. This is where stories – in the form of text, images, graphic novels, movies or video games – could play an important role by helping people understand subjective experiences on a visceral level. I have previously emphasized the limits of storytelling for EA purposes, but certain ways of combining stories with data may be very effective at supporting EA.

One potentially effective strategy could be to (i) use stories to give a qualitative understanding of the personal experiences involved in a human tragedy, and (ii) use data to give a quantitative understanding of the extent of the tragedy. It seems important that both elements are provided in order to support EA decisions. In particular, stories of personal tragedies provide a proof of existence but can give a distorted vision of reality in the presence of selection bias: news media, for example, often select atypical stories based on their shock value. But if personal stories are complemented with clear data about how representative they are, viewers will get a more accurate appreciation of the extent of the problems and of the magnitude of the human suffering involved.

It will likely be a major research challenge to find out more concretely how to effectively combine stories with quantitative data. There are at least two possible approaches: in a *data-then-story* approach, people would view statistical data about tragedies or social issues, and then zoom into individuals to see their personal stories, either real or hypothetical. The choice of individuals may be decided by the viewer following a detail-on-demand approach, or it may follow a random sampling scheme. Meanwhile, in a *story-then-data* approach, people would first see one or several typical stories (for example, the daily life of someone with disease A or disease B), and would then be able to explore statistical data (for example, the prevalence of those two diseases, and how they

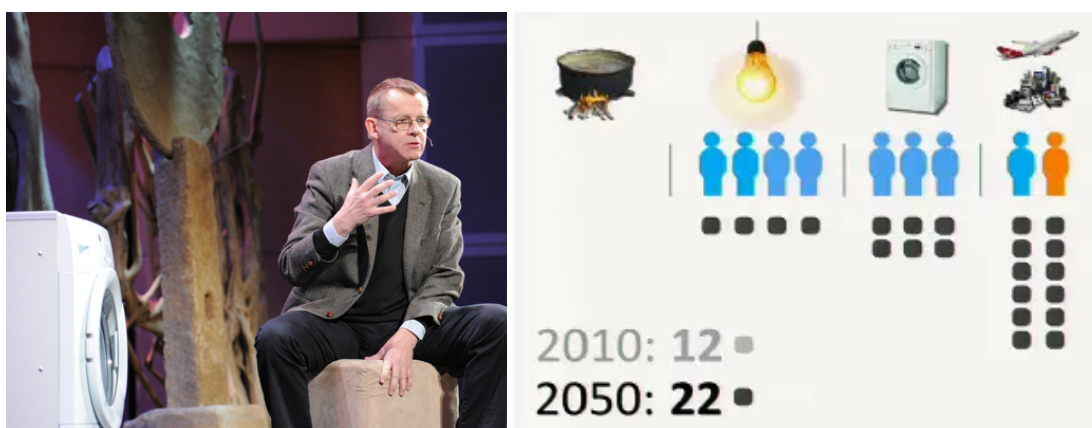


Fig. 4.4: Excerpts from Hans Rosling’s TED talk *The Magic Washing Machine*, which combines a personal story (left) with data (right). Source <https://www.ted.com/>.

could be reduced with different interventions). An example of story-then-data approach is Hans Rosling’s talk *The Magic Washing Machine* (Figure 4.4): he first tells a story that gives a powerful account of how life-changing washing machines are, and then goes through data about how many people in the world have access to them, and how this is likely to change with economic growth.

It is challenging to reconcile the world of numbers with the world of subjective experience, but not impossible – for example, if an effective altruist judges that having disease A is twice as bad as having disease B, they could conclude that preventing 10,000 cases of disease A is equally desirable as preventing 20,000 cases of disease B.

4.6 Emerging Technologies

In visualization research, there has been a lot of interest in conveying visualizations through other media than computer screens, like physical objects [DJV20] and mixed reality displays [Kra+21]. In a recent position paper [Dra22b], I discuss the interesting research opportunities offered by such media for the purpose of humanitarian visualization. I summarize them here.

Virtual reality. By providing a way for viewers (e.g., donors or charity managers) to immerse themselves more fully into personal stories, virtual reality (VR) may help enhance their visceral understanding of human issues and tragedies. VR documentaries already exist that cover topics such as war, migration, and diseases. Such immersive

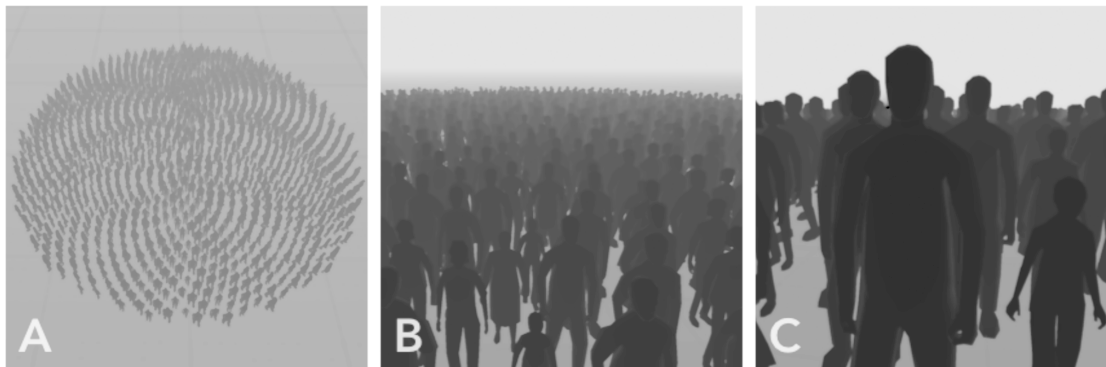


Fig. 4.5: VR visualization of mass shooting data in the US. Source [Iva+19].

stories could be combined with immersive data visualizations for EA purposes. This idea has started to be explored by Ivanov and colleagues [Iva+19], who designed a VR visualization of mass shooting casualties in the US (Figure 4.5). Each silhouette represents a person who died from a mass shooting in the US. Viewers can step back to get an overview of the dataset (A in the figure), or come closer to gather information about individual victims such as their age group or gender, which are encoded by the shape of the silhouette (B, C).

I see the concept from Ivanov et al. as only a starting point, as one could imagine conveying richer qualitative information about each victim like their physical appearance (as some memorials do by showing photo portraits) or elements of their personal stories, which viewers could choose to relive from a first-person perspective. Unfolding or hypothetical humanitarian issues could be conveyed in a similar manner using a combination of data visualizations and immersive video footage (such as already used in VR documentaries) or simulated scenes. VR could also be used to convey the positive outcomes of donations; For example, one could imagine an immersive version of GDLive (<https://live.givedirectly.org/>), a website that posts information and updates about recipients of cash transfers.

Augmented reality. Augmented reality (AR) can create illusions of objects and people around us, including objects and people that exist remotely. This opens up unprecedented possibilities for bringing the lives of distant suffering people closer to our own, and making humanitarian issues more salient or more memorable. For example, if a person walks in a refugee camp that has been temporarily relocated in their backyard, they may create a mental association and remember the refugees each time they see (or even think about) their backyard. Visualizing data about refugee camps in such a way could thus



Fig. 4.6: Physical data visualizations of 28 cases of sexual harassment – each plant conveys data reported by one person. Source [MAS22].

give a much more lasting impression. In contrast, VR can subjectively transport viewers in distant places, but once the viewers are back, the event is remote again. As with VR, AR could also be used to convey positive outcomes of charitable donations. In the context of a donor/recipient pairing program, future AR technology may even make it possible for a donor to meet a past recipient on the street and chat with them: a long-distance cash transfer may suddenly feel like helping out an acquaintance in a small village. Finally, in the future, effective altruists may be able to use wearable AR devices as commitment devices, e.g., to get regularly reminded of remote tragedies or ways they can redirect unnecessary personal expenses to humanitarian causes.

Data physicalization. Data physicalizations are physical entities whose shape or geometry encodes data [DJV20]. Public spaces already contain physical objects that convey past human tragedies, such as memorials, sculptures and cemeteries. However, few of them focus on current issues and few convey quantitative facts, both of which are important for EA purposes. Rare exceptions include data sculptures (artistic data physicalizations) with a focus on humanitarian data, and occasional explorations by visualization researchers like the *Harassment Plants* (Figure 4.6). Like AR visualizations, physical visualizations can be embedded in our everyday environment. But unlike AR visualizations, they are always present, they can be touched, and do not need special equipment to be seen. On the other hand, AR content can be created and displayed at will.

Ambient displays. Ambient displays are displays that “*present information within a space through subtle changes in light, sound, and movement, which can be processed in the background of awareness*” [Wis+98]. In particular, research projects have explored how ambient displays can support remote intimacy – for example, the color of a lamp may change according to the affective state of a remote intimate partner captured through a wearable biofeedback device. Similar devices could be used to convey quantitative information about the plight of large populations of distant and anonymous people, such as the number of hospitalizations during a pandemic or the number of war casualties. Such ambient displays could give a continuous impression of the severity of an ongoing humanitarian crisis without having to constantly poll news reports.

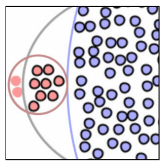
4.7 Conclusion

I discussed perspectives for future research that combine the goals of the two main chapters of this thesis (supporting rational judgment and decisions, and supporting effective communication), and applies them to the area of humanitarian visualization. I argued that such research would gain a lot from being informed by effective altruism (EA). EA offers both a new thinking framework and new questions and problems for visualization research. Yet, it appears that there has been virtually no collaboration so far between EA actors and visualization researchers, perhaps largely due to a lack of mutual awareness between the two communities. But this is changing, as EA is becoming mainstream and highly influential [Mat22]. There are many fascinating questions and problems at the intersection of the two areas and unique opportunities for collaboration, so it is time the two communities connect.

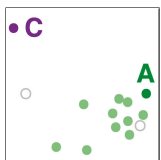
Conclusion

This habilitation thesis has presented research projects looking at how data visualizations can be used to support data-driven decisions and communicate quantitative facts. Here I list 17 key takeaways, noting that not all of them are supported by empirical evidence, and that any takeaway is tentative and can be challenged. Also, the research projects in this thesis are only dots in the vast space of research questions, and so the 17 takeaways listed here may appear to be a collection of statements without a clear overarching structure and without lots of coherence. I still hope they are useful.

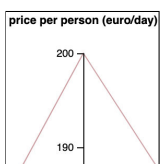
The first part of this thesis focused on supporting rational judgments and decisions with data visualizations. Some takeaways are:



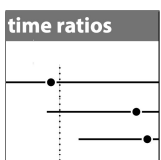
Takeaway 1. When giving people a nontrivial base-rate reasoning problem, adding a data visualization to the text does not help them much, except if all numbers are removed from the text in order to force people to perform visual estimation [MDF12].



Takeaway 2. The attraction effect generalizes to data visualizations: when people use scatterplots to visualize multiple alternatives to choose from, the number and position of irrelevant (dominated) alternatives can unduly influence their choice [DBD17c].



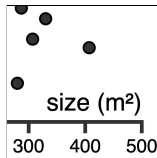
Takeaway 3. When evaluating visualizations for decision support using multi-attribute choice tasks, a ground truth can be approximated by asking people about their choice criteria and looking at the choices most consistent with those criteria [DBD17a].



Takeaway 4. When evaluating visualizations for decision support, time-on-task can be used as a tie-breaker when visualizations achieve similar decision accuracy [DBD17a].

country name	employment
Australia	72
Austria	72
Belgium	62
Canada	72
Czech Republic	65
Denmark	73
Estonia	61
Finland	68
France	84
Germany	71

Takeaway 5. Tabular visualizations seem to be a compelling choice for supporting multi-attribute choice tasks, despite them being less popular than other multidimensional data visualizations [DBD17a].



Takeaway 6. When evaluating data visualizations using crowdsourcing, it helps to add semantic context to the tasks (for example, explaining what axes mean in a plot), but elaborate narratives can make participants less attentive and less accurate [DBD17b].

Cognitive bias	Ref	Rd
Availability bias	[1873]	1
Conjunction fallacy	[1883]	1
Empathy gap	[1865]	1
Time-saving bias	[1868]	4
Anchoring effect	[1811]	2
Base rate fallacy	[1807a]	6
Dunning-Kruger effect	[1899]	3
Gambler's fallacy	[1875]	4
Hard-easy effect	[1877]	3
Hot-hand fallacy	[1876]	3
Inconsistency to sample size	[1875]	3
Regression bias	[1853]	4

Takeaway 7. A task-based taxonomy of cognitive biases can be a useful starting point for researchers interested in studying cognitive biases in a data visualization context [Dim+18].

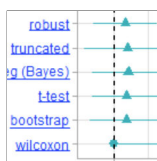


Takeaway 8. Studying cognitive biases in a data visualization context is subject to the same threats to validity as general cognitive bias research – a major threat being that it is often hard to identify a rational ground truth with certainty (section 2.4).

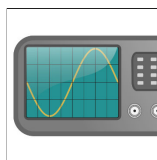
The second part of this thesis discussed how to support effective communication with data visualizations. Key takeaways are:



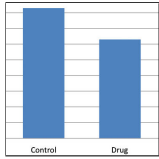
Takeaway 9. Bertin's matrices are a neglected family of tabular visualizations dating back to the 1970s, and that can greatly help both researchers and a wider audience understand and communicate their tabular data visually [PDF14; PFD19].



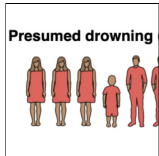
Takeaway 10. Multiverse analyses are a new and promising approach to transparent statistical reporting, and one interesting way of reporting them is through interactive research papers where readers can explore alternative analysis options by themselves [Dra+19].



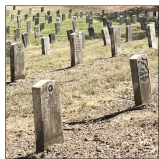
Takeaway 11. We still do not have conclusive evidence that charts can persuade by their mere presence, e.g., because people associate them with science [DJ18a].



Takeaway 12. There is no evidence that charts conveying only two values are trivial and unhelpful compared to reporting numerals [DJ18a].



Takeaway 13. When conveying data about humanitarian issues, common sense dictates that the audience is more likely to relate to the victims if the visualization emphasizes individuals, but studies have largely failed to confirm this intuition [Mor+21].



Takeaway 14. The design space of humanitarian visualizations is nevertheless very large, and most of it has remained unexplored [Mor+20].



Takeaway 15. Research on humanitarian visualization would gain from being informed by the effective altruism movement, which offers a useful thinking framework and a range of interesting research questions [Dra22a].

Two general takeaways from this HDR thesis are:

	A	B
variety	1	2
cleanliness	5	5

Takeaway 16. Although much of the data visualization literature focuses on large datasets, we often do not know how to best visualize even very small datasets.



Takeaway 17. Interesting data visualization research questions can arise from, and be tested with, small datasets.

Bibliography

- [Aga22] Against Malaria Foundation. *Why nets?* Web page <https://www.againstmalaria.com/WhyNets.aspx>. Last visited 2022-03-22. 2022 (cit. on p. 84).
- [Ajz77] Icek Ajzen. “Intuitive theories of events and the effects of base-rate information on prediction.” In: *Personality and Social Psychology* 35.5 (1977), pp. 303–314 (cit. on p. 13).
- [ATD19] Kevin Allain, Cagatay Turkay, and Jason Dykes. “Towards a WHAT-WHY-HOW Taxonomy of Trajectories in Visualization Research”. In: (2019) (cit. on p. 46).
- [AM85] Scott T Allison and David M Messick. “The group attribution error”. In: *Journal of Experimental Social Psychology* 21.6 (1985), pp. 563–579 (cit. on p. 36).
- [AB11] Douglas G Altman and J Martin Bland. “How to obtain the confidence interval from a P value”. In: *BMJ* 343 (2011), p. d2090 (cit. on p. 60).
- [AES05] Robert Amar, James Eagan, and John Stasko. “Low-level components of analytic activity in information visualization”. In: *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005*. IEEE. 2005, pp. 111–117 (cit. on p. 29).
- [AAJ03] Gennady Andrienko, Natalia Andrienko, and Piotr Jankowski. “Building spatial decision support tools for individuals and groups”. In: *Journal of Decision Systems* 12.2 (2003), pp. 193–208 (cit. on p. 24).
- [AA03] Natalia Andrienko and Gennady Andrienko. “Informed spatial decisions through coordinated views”. In: *Information Visualization* 2.4 (2003), pp. 270–285 (cit. on p. 24).
- [Ant+15] Nikos Antonopoulos, Andreas Veglis, Antonis Gardikiotis, Rigas Kotsakis, and George Kalliris. “Web Third-person effect in structural aspects of the information on media websites”. In: *Computers in Human Behavior* 44 (2015), pp. 48–58 (cit. on p. 36).
- [AP15] David Antons and Frank T Piller. “Opening the black box of “Not Invented Here”: attitudes, decision biases, and behavioral consequences”. In: *The Academy of Management Perspectives* 29.2 (2015), pp. 193–217 (cit. on p. 35).
- [ATS95] Toshiyuki Asahi, David Turo, and Ben Shneiderman. “Using treemaps to visualize the analytic hierarchy process”. In: *Information Systems Research* 6.4 (1995), pp. 357–375 (cit. on p. 24).

- [Ase+15] Bon Adriel Aseniero, Tiffany Wun, David Ledo, et al. “STRATOS: Using Visualization to Support Decisions in Strategic Software Release Planning”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM. 2015, pp. 1479–1488 (cit. on p. 28).
- [Att53] Fred Attneave. “Psychological probability as a function of experienced frequency.” In: *Journal of Experimental Psychology* 46.2 (1953), p. 81 (cit. on p. 35).
- [Bak10] Diane F Baker. “Enhancing group decision making: An exercise to reduce shared information bias”. In: *Journal of Management Education* 34.2 (2010), pp. 249–279 (cit. on p. 35).
- [Bar80] Maya Bar-Hillel. “The base-rate fallacy in probability judgments”. In: *Acta Psychologica* 44.3 (1980). Ed. by Roland WEditor Scholz, pp. 211–233 (cit. on p. 13).
- [BS07a] Aron K Barbey and Steven A Sloman. “Base-rate respect: From ecological rationality to dual processes”. In: *Behavioral and Brain Sciences* 30.03 (2007), pp. 241–254 (cit. on pp. 11, 35).
- [BS07b] Aron K Barbey and Steven A Sloman. “Base-rate respect: From statistical formats to cognitive structures”. In: *Behavioral and Brain Sciences* 30.3 (2007), pp. 287–297 (cit. on p. 11).
- [BBH88] Jonathan Baron, Jane Beattie, and John C Hershey. “Heuristics and biases in diagnostic reasoning: II. Congruence, information, and certainty”. In: *Organizational Behavior and Human Decision Processes* 42.1 (1988), pp. 88–110 (cit. on p. 35).
- [BGK93] Jonathan Baron, Rajeev Gowda, and Howard Kunreuther. “Attitudes toward managing hazardous waste: What should be cleaned up and who should pay for it?” In: *Risk Analysis* 13.2 (1993), pp. 183–192 (cit. on p. 35).
- [BH88] Jonathan Baron and John C Hershey. “Outcome bias in decision evaluation.” In: *Journal of personality and social psychology* 54.4 (1988), p. 569 (cit. on p. 35).
- [BHH02] Melissa Bateson, Susan D. Healy, and Hurly. “Irrational choices in hummingbird foraging behaviour”. In: *Animal Behaviour* 63.3 (Mar. 2002), pp. 587–596 (cit. on p. 18).
- [BR03] Patrick Baudisch and Ruth Rosenholtz. “Halo: a technique for visualizing off-screen objects”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. 2003, pp. 481–488 (cit. on p. 55).
- [Bau+06] Patrick Baudisch, Desney Tan, Maxime Collomb, et al. “Phosphor: explaining transitions in the user interface using afterglow effects”. In: *Proceedings of the 19th annual ACM symposium on User interface software and technology*. ACM. 2006, pp. 169–178 (cit. on p. 55).
- [BC08] Jeanette Bautista and Giuseppe Carenini. “An empirical evaluation of interactive visualizations for preferential choice”. In: *Proceedings of the working conference on Advanced visual interfaces*. ACM. 2008, pp. 207–214 (cit. on p. 28).

- [Ben+18] Daniel J Benjamin, James O Berger, Magnus Johannesson, et al. “Redefine statistical significance”. In: *Nature Human Behaviour* 2.1 (2018), p. 6 (cit. on p. 49).
- [BK77] William H. Benson and Bernard Kitous. “Interactive Analysis and Display of Tabular Data”. In: *SIGGRAPH Comput. Graph.* 11.2 (July 1977), pp. 48–53 (cit. on p. 44).
- [Ber12] Donald Berry. “Multiplicities in cancer research: ubiquitous and necessary evils”. In: *Journal of the National Cancer Institute* 104.15 (2012), pp. 1125–1133 (cit. on p. 47).
- [Ber07] Donald A Berry. “The difficult and ubiquitous problems of multiplicities”. In: *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry* 6.3 (2007), pp. 155–160 (cit. on p. 27).
- [Ber75] Jacques Bertin. *La graphique et le traitement graphique de l'information*. Nouvelle bibliothèque scientifique. Flammarion, 1975 (cit. on pp. 40, 42).
- [Ber67] Jacques Bertin. *Sémiologie graphique : Les diagrammes-Les réseaux-Les cartes*. Mouton/Gauthier-Villars, 1967 (cit. on p. 40).
- [BC94] Jacques Bertin and Jean-Hugues. Chauchat. “Logiciel AMADO (Analyse Graphique d'une Matrice de Données)”. In: *CISIA, France* (1994) (cit. on p. 41).
- [Ber81] Jaques Bertin. *Graphics and Graphic Information Processing*. Translation: William J. Berg, Paul Scott. De Gruyter, Berlin, 1981 (cit. on p. 40).
- [Ber17] Enrico Bertini. *Can Visualization Elicit Empathy? Our Experiments with "Anthropographics"*. [Online; accessed 13-November-2019]. 2017 (cit. on p. 64).
- [Bes+20] Lonni Besançon, Amir Semmo, David Biau, et al. “Reducing affective responses to surgical images and videos through stylization”. In: *Computer Graphics Forum*. Vol. 39. 1. Wiley Online Library. 2020, pp. 462–483 (cit. on p. 46).
- [BLP98] James R. Bettman, Mary F. Luce, and John W. Payne. “Constructive Consumer Choice Processes”. In: *J. of Consumer Research* 25.3 (Dec. 1998), pp. 187–217 (cit. on p. 22).
- [BDB06] Anastasia Bezerianos, Pierre Dragicevic, and Ravin Balakrishnan. “Mnemonic rendering: an image-based approach for exposing hidden changes in dynamic displays”. In: *Proceedings of the 19th annual ACM symposium on User interface software and technology*. ACM. 2006, pp. 159–168 (cit. on p. 55).
- [Blo17] Paul Bloom. “Empathy and its discontents”. In: *Trends in cognitive sciences* 21.1 (2017), pp. 24–31 (cit. on p. 74).
- [Bot+16] Fadi Botros, Charles Perin, Bon Adriel Aseniero, and Sheelagh Carpendale. “Go and grow: Mapping personal data to a living plant”. In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*. 2016, pp. 112–119 (cit. on p. 46).

- [Bou+12] Nadia Boukhelifa, Anastasia Bezerianos, Tobias Isenberg, and Jean-Daniel Fekete. “Evaluating Sketchiness as a Visual Variable for the Depiction of Qualitative Uncertainty”. In: *TVCG* 18.12 (2012), pp. 2769–2778 (cit. on p. 28).
- [Boy+14] J. Boy, R.A. Rensink, E. Bertini, and J.-D. Fekete. “A Principled Way of Assessing Visualization Literacy”. In: *Visualization and Computer Graphics, IEEE Transactions on* 20.12 (2014), pp. 1963–1972 (cit. on p. 28).
- [Boy+17] Jeremy Boy, Anshul Vikram Pandey, John Emerson, et al. “Showing People Behind Data: Does Anthropomorphizing Visualizations Elicit More Empathy for Human Rights Data?” In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM. 2017, pp. 5462–5474 (cit. on pp. 35, 36, 64, 65, 73).
- [Bra09] Gary L Brase. “Pictorial representations in statistical reasoning”. In: *Applied Cognitive Psychology* 23.3 (2009), pp. 369–381 (cit. on pp. 11, 13).
- [Bre81] Sharon S Brehm. “Psychological reactance and the attractiveness of unobtainable objects: Sex differences in children’s responses to an elimination of freedom”. In: *Sex Roles* 7.9 (1981), pp. 937–949 (cit. on p. 35).
- [BM13] Matthew Brehmer and Tamara Munzner. “A multi-level typology of abstract visualization tasks”. In: *IEEE transactions on visualization and computer graphics* 19.12 (2013), pp. 2376–2385 (cit. on p. 3).
- [Bre73] Malcolm Brenner. “The next-in-line effect”. In: *Journal of Verbal Learning and Verbal Behavior* 12.3 (1973), pp. 320–323 (cit. on p. 36).
- [BVH] Leen Breure, Hans Voorbij, and Maarten Hoogerwerf. “Rich Internet Publications: Show What You Tell”. In: *Journal of Digital Information* 12.1 () (cit. on p. 54).
- [Bro91] Alan S Brown. “A review of the tip-of-the-tongue experience.” In: *Psychological bulletin* 109.2 (1991), p. 204 (cit. on p. 36).
- [BM89] Alan S Brown and Dana R Murphy. “Cryptomnesia: Delineating inadvertent plagiarism.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15.3 (1989), p. 432 (cit. on p. 36).
- [BD95] Jonathan B Buckheit and David L Donoho. “Wavelab and reproducible research”. In: *Wavelets and statistics*. Springer, 1995, pp. 55–81 (cit. on p. 48).
- [BGP10] Roger Buehler, Dale Griffin, and Johanna Peetz. “The planning fallacy: cognitive, motivational, and social origins”. In: *Advances in experimental social psychology* 43 (2010), pp. 1–62 (cit. on p. 35).
- [CS99] W Keith Campbell and Constantine Sedikides. “Self-threat Magnifies the Self-serving Bias”. In: *Review of General Psychology* 3.1 (1999), pp. 23–43 (cit. on p. 36).
- [CL04] Giuseppe Carenini and John Loyd. “ValueCharts: analyzing linear models expressing preferences and evaluations”. In: *Proceedings of the working conference on Advanced visual interfaces*. ACM. 2004, pp. 150–157 (cit. on pp. 24, 27).

- [CR03] Melanie Cary and Lynne M Reder. “A dual-process account of the list-length and strength-based mirror effects in recognition”. In: *Journal of Memory and Language* 49.2 (2003), pp. 231–248 (cit. on p. 36).
- [CSG21] Lucius Caviola, Stefan Schubert, and Joshua D Greene. “The psychology of (in) effective altruism”. In: *Trends in Cognitive Sciences* 25.7 (2021), pp. 596–607 (cit. on pp. 83, 84).
- [CC69] Loren J Chapman and Jean P Chapman. “Illusory correlation as an obstacle to the use of valid psychodiagnostic signs.” In: *Journal of abnormal psychology* 74.3 (1969), p. 271 (cit. on p. 35).
- [Che+10] Fanny Chevalier, Pierre Dragicevic, Anastasia Bezerianos, and Jean-Daniel Fekete. “Using text animated transitions to support navigation in document histories”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2010, pp. 683–692 (cit. on p. 55).
- [CVG13] Fanny Chevalier, Romain Vuillemot, and Guia Gali. “Using concrete scales: A practical framework for effective visual depiction of complex measures”. In: *IEEE transactions on visualization and computer graphics* 19.12 (2013), pp. 2426–2435 (cit. on p. 4).
- [Chi93] Chicago Tribune. *A False HIV Test Caused 18 Months of Hell*. 1993 (cit. on p. 10).
- [Cho+18] Isaac Cho, Ryan Wesslen, Alireza Karduni, et al. “The Anchoring Effect in Decision-Making with Visual Analytics”. In: *IEEE Transactions on Visualization and Computer Graphics* (2018) (cit. on p. 35).
- [CHT13] Jongwoon Willie Choi, Gary W Hecht, and William B Tayler. “Strategy selection, surrogation, and strategic performance measurement systems”. In: *Journal of Accounting Research* 51.1 (2013), pp. 105–133 (cit. on p. 36).
- [CF14] H Choo and SL Franconeri. “Enumeration of small collections violates Weber’s law”. In: *Psychonomic bulletin & review* 21.1 (2014), pp. 93–99 (cit. on p. 35).
- [Cla20] Stephen Clare. *Homelessness in the US and UK Executive Summary*. Web article <https://www.founderspledge.com/stories/homelessness-in-the-us-and-uk-executive-summary>. 2020 (cit. on p. 85).
- [CGD18] Andy Cockburn, Carl Gutwin, and Alan Dix. “Hark no more: on the preregistration of CHI experiments”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM. 2018, p. 141 (cit. on p. 47).
- [Coh94] Jacob Cohen. “The Earth is round ($p < .05$)”. In: *American psychologist* 49.12 (1994), p. 997 (cit. on p. 27).
- [Col+17] John D Coley, Melanie Arenson, Yian Xu, and Kimberly D Tanner. “Intuitive biological thought: Developmental changes and effects of biology education in late adolescence”. In: *Cognitive psychology* 92 (2017), pp. 1–21 (cit. on p. 36).

- [Con+20] Shauna Concannon, Natasha Rajan, Parthiv Shah, et al. “Brooke Leave Home: Designing a Personalized Film to Support Public Engagement with Open Data”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu, HI, USA: Association for Computing Machinery, 2020, pp. 1–14 (cit. on p. 73).
- [CH18] Matt Conlen and Jeffrey Heer. “Idyll: A Markup Language for Authoring and Publishing Interactive Articles on the Web”. In: *ACM User Interface Software & Technology (UIST)*. 2018 (cit. on p. 54).
- [Cor+02] Joshua Correll, Bernadette Park, Charles M Judd, and Bernd Wittenbrink. “The police officer’s dilemma: using ethnicity to disambiguate potentially threatening individuals.” In: *J. Pers. Soc. Psychol.* 83.6 (2002), p. 1314 (cit. on p. 36).
- [CG14] Michael Correll and Michael Gleicher. “Bad for data, good for the brain: Knowledge-first axioms for visualization design”. In: *IEEE VIS 2014 (workshop DECISIVE)*. 2014 (cit. on p. 56).
- [CH17] Michael Correll and Jeffrey Heer. “Regression by Eye: Estimating Trends in Bivariate Visualizations”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM. 2017, pp. 1387–1396 (cit. on p. 35).
- [CT96] Leda Cosmides and John Tooby. “Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty”. In: *Cognition* 58.1 (1996) (cit. on pp. 10, 11, 13).
- [CT75] Fergus IM Craik and Endel Tulving. “Depth of processing and the retention of words in episodic memory.” In: *Journal of experimental Psychology: general* 104.3 (1975), p. 268 (cit. on p. 36).
- [CM60] Douglas P Crowne and David Marlowe. “A new scale of social desirability independent of psychopathology.” In: *Journal of consulting psychology* 24.4 (1960), p. 349 (cit. on p. 36).
- [CH89] Robert J Crutcher and Alice F Healy. “Cognitive operations and the generation effect.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15.4 (1989), p. 669 (cit. on p. 36).
- [Cum09] Geoff Cumming. *The dance of p-values (video)*. 2009 (cit. on p. 53).
- [Cum14] Geoff Cumming. “The new statistics: Why and how”. In: *Psychological science* 25.1 (2014), pp. 7–29 (cit. on pp. 47, 49).
- [CF05] Geoff Cumming and Sue Finch. “Inference by eye: confidence intervals and how to read pictures of data.” In: *American Psychologist* 60.2 (2005), p. 170 (cit. on p. 53).
- [Cus08] Fiery Cushman. “Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment”. In: *Cognition* 108.2 (2008), pp. 353–380 (cit. on p. 36).

- [DCM13] Mohammad Daradkeh, Clare Churcher, and Alan McKinnon. “Supporting informed decision-making under uncertainty and risk through interactive visualisation”. In: *Proceedings of the Fourteenth Australasian User Interface Conference-Volume 139*. Australian Computer Society, Inc. 2013, pp. 23–32 (cit. on p. 28).
- [Das+17] Aritra Dasgupta, Susannah Burrows, Kyungsik Han, and Philip J Rasch. “Empirical Analysis of the Subjective Impressions and Objective Measures of Domain Scientists’ Visual Analytic Judgments”. In: *ACM Conference on Human Factors in Computing Systems (CHI)*. 2017, pp. 1193–1204 (cit. on p. 35).
- [Das+15] Aritra Dasgupta, Jorge Poco, Yaxing Wei, et al. “Bridging theory with practice: An exploratory study of visualization use and design for climate model comparison”. In: *IEEE transactions on visualization and computer graphics* 21.9 (2015), pp. 996–1014 (cit. on p. 35).
- [DGG15] Andrea De Mauro, Marco Greco, and Michele Grimaldi. “What is big data? A consensual definition and a review of key research topics”. In: *AIP conference proceedings*. Vol. 1644. 1. American Institute of Physics. 2015, pp. 97–104 (cit. on p. 4).
- [Dim+19] Evanthia Dimara, Gilles Bailly, Anastasia Bezerianos, and Steven Franconeri. “Mitigating the Attraction Effect with Visualizations”. In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019) (cit. on p. 35).
- [DBD17a] Evanthia Dimara, Anastasia Bezerianos, and Pierre Dragicevic. “Conceptual and methodological issues in evaluating multidimensional visualizations for decision support”. In: *IEEE transactions on visualization and computer graphics* 24.1 (2017). [PDF], pp. 749–759 (cit. on pp. 6, 23, 26, 91, 92).
- [DBD17b] Evanthia Dimara, Anastasia Bezerianos, and Pierre Dragicevic. “Narratives in Crowd-sourced Evaluation of Visualizations: A Double-Edged Sword?” In: *ACM Conference on Human Factors in Computing Systems (CHI)*. [PDF]. Denver, United States, May 2017 (cit. on pp. 6, 28, 92).
- [DBD16] Evanthia Dimara, Anastasia Bezerianos, and Pierre Dragicevic. *Testing the Attraction Effect on Two Information Visualization Datasets*. Research Report RR-8895. <http://hal.inria.fr/hal-01295624>. Inria, Mar. 2016, p. 14 (cit. on pp. 21, 23).
- [DBD17c] Evanthia Dimara, Anastasia Bezerianos, and Pierre Dragicevic. “The attraction effect in information visualization”. In: *IEEE Transactions on Visualization and Computer Graphics* 23.1 (2017). [PDF], pp. 471–480 (cit. on pp. 6, 17, 24, 28, 35, 91).
- [DDB14] Evanthia Dimara, Pierre Dragicevic, and Anastasia Bezerianos. “Accounting for Availability Biases in Information Visualization”. In: *DECISIVE: Workshop on Dealing with Cognitive Biases in Visualizations. IEEE VIS*. 2014 (cit. on p. 35).

- [Dim+18] Evanthia Dimara, Steven Franconeri, Catherine Plaisant, Anastasia Bezerianos, and Pierre Dragicevic. “A task-based taxonomy of cognitive biases for information visualization”. In: *IEEE transactions on visualization and computer graphics* 26.2 (2018). [PDF], pp. 1413–1432 (cit. on pp. 6, 33, 37, 83, 92).
- [DS21] Evanthia Dimara and John Stasko. “A Critical Reflection on Visualization Research: Where Do Decision Making Tasks Hide?” In: *IEEE Transactions on Visualization and Computer Graphics* 28.1 (2021), pp. 1128–1138 (cit. on p. 3).
- [Dra18] Pierre Dragicevic. *Adding Inferential Information to Plots using Resampling and Animations*. Online. <https://explorablemultiverse.github.io/examples/dance/>. 2018 (cit. on p. 52).
- [Dra16] Pierre Dragicevic. “Fair Statistical Communication in HCI”. In: *Modern Statistical Methods for HCI*. [PDF]. Springer, 2016, pp. 291–330 (cit. on pp. 20, 47, 50).
- [Dra22a] Pierre Dragicevic. “Information Visualization for Effective Altruism”. In: *arXiv preprint arXiv:2209.00836* (2022). [PDF] (cit. on pp. 7, 64, 77, 93).
- [Dra17] Pierre Dragicevic. *Statistical Dances: Why no Statistical Analysis is Reliable and What to do About it (video)*. https://www.youtube.com/watch?v=UKX9iN0p5_A. Séminaire recherche reproductible, GRICAD, Grenoble.(2017). 2017 (cit. on p. 53).
- [Dra22b] Pierre Dragicevic. “Towards immersive humanitarian visualizations”. In: *arXiv preprint arXiv:2204.01313* (2022). [PDF] (cit. on pp. 63, 77, 86).
- [DJ18a] Pierre Dragicevic and Yvonne Jansen. “Blinded with Science or Informed by Charts? A Replication Study”. In: *IEEE transactions on visualization and computer graphics* 24.1 (2018). [PDF], pp. 781–790 (cit. on pp. 4, 7, 56, 59, 92, 93).
- [DJ18b] Pierre Dragicevic and Yvonne Jansen. *Re-Evaluating the Efficiency of Physical Visualizations: A Simple Multiverse Analysis*. Online. <https://explorablemultiverse.github.io/examples/frequentist/>. 2018 (cit. on pp. 49, 50).
- [DJ14] Pierre Dragicevic and Yvonne Jansen. “Visualization-Mediated Alleviation of the Planning Fallacy”. In: *DECISIVE: Workshop on Dealing with Cognitive Biases in Visualizations. IEEE VIS*. 2014 (cit. on p. 35).
- [Dra+19] Pierre Dragicevic, Yvonne Jansen, Abhraneel Sarma, Matthew Kay, and Fanny Chevalier. “Increasing the transparency of research papers with explorable multiverse analyses”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. [PDF]. 2019, pp. 1–15 (cit. on pp. 6, 46, 49, 54, 92).
- [DJV20] Pierre Dragicevic, Yvonne Jansen, and Andrew Vande Moere. “Data physicalization”. In: *Handbook of Human Computer Interaction* (2020). [PDF], pp. 1–51 (cit. on pp. 2, 86, 88).
- [Dra+08] Pierre Dragicevic, Gonzalo Ramos, Jacobo Bibliowicz, et al. “Video browsing by direct manipulation”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2008, pp. 237–246 (cit. on p. 54).

- [DM02] Karolina Duklan and Michael A Martin. *Communicating effectively with words, numbers and pictures: Drawing on experience*. School of Finance and Applied Statistics, Australian National University, 2002 (cit. on p. 4).
- [DMH89] David Dunning, Judith A Meyerowitz, and Amy D Holzberg. “Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability.” In: *Journal of personality and social psychology* 57.6 (1989), p. 1082 (cit. on p. 36).
- [DRG13] Kristina M Durante, Ashley Rae, and Vldas Griskevicius. “The fluctuating female vote: Politics, religion, and the ovulatory cycle”. In: *Psychological Science* 24.6 (2013), pp. 1007–1016 (cit. on p. 51).
- [Edd82] David M. Eddy. “Probabilistic reasoning in clinical medicine: problems and opportunities”. In: *Judgement under uncertainty: Heuristics and Biases*. Cambridge University Press, 1982, pp. 249–267 (cit. on pp. 10, 12–14).
- [Efr92] Bradley Efron. “Bootstrap methods: another look at the jackknife”. In: *Breakthroughs in statistics*. Springer, 1992, pp. 569–593 (cit. on p. 52).
- [ECD14] Serge Egelman, Ed H Chi, and Steven Dow. “Crowdsourcing in HCI Research”. In: *Ways of Knowing in HCI*. Springer, 2014, pp. 267–289 (cit. on pp. 28, 29, 32).
- [Ein+03] Gilles O Einstein, Mark A McDaniel, Carrie L Williford, Jason L Pagan, and R Dismukes. “Forgetting of intentions in demanding situations is rapid.” In: *Journal of Experimental Psychology: Applied* 9.3 (2003), p. 147 (cit. on p. 36).
- [ED15] Geoffrey Ellis and Alan Dix. “Decision Making Under Uncertainty in Visualisation?” In: *IEEE VIS2015*. 2015 (cit. on p. 35).
- [EBP83] J St BT Evans, Julie L Barston, and Paul Pollard. “On the conflict between logic and belief in syllogistic reasoning”. In: *Memory & cognition* 11.3 (1983), pp. 295–306 (cit. on p. 35).
- [Fek+08] Jean-Daniel Fekete, Jarke Van Wijk, John Stasko, and Chris North. “The value of information visualization”. In: *Information visualization* (2008). Ed. by Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and Chris North, pp. 1–18 (cit. on pp. 58, 77).
- [FB75] Baruch Fischhoff and Ruth Beyth. “I knew it would happen: Remembered probabilities of once—future things”. In: *Organizational Behavior and Human Performance* 13.1 (1975), pp. 1–16 (cit. on p. 36).
- [Fis80] Susan T Fiske. “Attention and weight in person perception: The impact of negative and extreme behavior.” In: *Journal of personality and Social Psychology* 38.6 (1980), p. 889 (cit. on p. 36).

- [Fle18] Tamara Flemisch. “Effects of Different Display Form Factors on InfoVis Applications: Exploring Selection Management and Brushing and Linking for Mobile Cross-Device Interaction”. PhD thesis. Masterarbeit, Dresden, Technische Universität Dresden, 2018, 2018 (cit. on p. 46).
- [FP19] Federica Fracapane and Alex Piacentini. *Stories Behind a Line*. <https://web.archive.org/web/20190719064902/http://www.storiesbehindaline.com/>. [Online; accessed 19-July-2019]. 2019 (cit. on p. 65).
- [FLB14] Shane Frederick, Leonard Lee, and Ernest Baskin. “The Limits of Attraction”. In: *Journal of Marketing Research* 51.4 (2014), pp. 487–507 (cit. on p. 22).
- [FK93] Barbara L Fredrickson and Daniel Kahneman. “Duration neglect in retrospective evaluations of affective episodes.” In: *Journal of personality and social psychology* 65.1 (1993), p. 45 (cit. on p. 36).
- [FB11] Adrian Furnham and Hua Chu Boo. “A literature review of the anchoring effect”. In: *The Journal of Socio-Economics* 40.1 (2011), pp. 35–42 (cit. on p. 35).
- [GST10] Stephen M Garcia, Hyunjin Song, and Abraham Tesser. “Tainted recommendations: The social comparison bias”. In: *Organizational Behavior and Human Decision Processes* 113.2 (2010), pp. 97–101 (cit. on p. 35).
- [GRD06] Andrew B Geier, Paul Rozin, and Gheorghe Doros. “Unit bias: A new heuristic that helps explain the effect of portion size on food intake”. In: *Psychological Science* 17.6 (2006), pp. 521–525 (cit. on p. 36).
- [GL13] Andrew Gelman and Eric Loken. “The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time”. In: *Department of Statistics, Columbia University* (2013) (cit. on p. 47).
- [GT07] Robert Gentleman and Duncan Temple Lang. “Statistical analyses and reproducible research”. In: *Journal of Computational and Graphical Statistics* 16.1 (2007), pp. 1–23 (cit. on p. 48).
- [GB05] Tim P German and H Clark Barrett. “Functional fixedness in a technologically sparse culture”. In: *Psychological Science* 16.1 (2005), pp. 1–5 (cit. on p. 35).
- [GVP08] Jeffrey A Gibbons, Andrew K Velkey, and Kathren T Partin. “Influence of recall procedures on the modality effect with numbers and enumerated stimuli”. In: *Journal of gen. psychology* 135.1 (2008), pp. 84–104 (cit. on p. 36).
- [Gig98] Gerd Gigerenzer. “Ecological intelligence: An adaptation for frequencies”. In: *Psychologische Beiträge*. Ed. by Colin Allen and Denise Editors Cummins Dellarosa. Vol. 39. S. Oxford University Press, 1998, pp. 9–29 (cit. on p. 10).
- [Gig18] Gerd Gigerenzer. “The bias bias in behavioral economics”. In: *Review of Behavioral Economics* 5.3-4 (2018), pp. 303–336 (cit. on p. 38).

- [Gig+07] Gerd Gigerenzer, Wolfgang Gaissmaier, Elke Kurz-Milcke, Lisa M Schwartz, and Steven Woloshin. “Helping doctors and patients make sense of health statistics”. In: *Psychological science in the public interest* 8.2 (2007), pp. 53–96 (cit. on pp. 10, 60).
- [GH95] Gerd Gigerenzer and Ulrich Hoffrage. “How to Improve Bayesian Reasoning Without Instruction: Frequency Formats.” In: *Psychological Review* 102.4 (1995), pp. 684–704 (cit. on pp. 11, 13).
- [Gil+00] Daniel T Gilbert, Ryan P Brown, Elizabeth C Pinel, and Timothy D Wilson. “The illusion of external agency.” In: *J. Pers. Soc. Psychol.* 79.5 (2000), p. 690 (cit. on p. 36).
- [GM95] Daniel T Gilbert and Patrick S Malone. “The correspondence bias.” In: *Psychological bulletin* 117.1 (1995), p. 21 (cit. on p. 36).
- [Gil+98] Douglas J Gillan, Christopher D Wickens, Justin G Hollands, and C Melody Carswell. “Guidelines for presenting quantitative data in HFES publications”. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 40.1 (1998), pp. 28–41 (cit. on p. 4).
- [GMS00] Thomas Gilovich, Victoria Husted Medvec, and Kenneth Savitsky. “The spotlight effect in social judgment: an egocentric bias in estimates of the salience of one’s own actions and appearance.” In: *J. Pers. Soc. Psychol.* 78.2 (2000), p. 211 (cit. on p. 35).
- [GVT85] Thomas Gilovich, Robert Vallone, and Amos Tversky. “The hot hand in basketball: On the misperception of random sequences”. In: *Cognitive psychology* 17.3 (1985), pp. 295–314 (cit. on p. 35).
- [Gin12a] Roger Giner-Sorolla. “Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science”. In: *Perspectives on Psychological Science* 7.6 (2012), pp. 562–571 (cit. on p. 54).
- [Gin12b] Roger Giner-Sorolla. “Will we march to utopia, or be dragged there? Past failures and future hopes for publishing our science”. In: *Psychological Inquiry* 23.3 (2012), pp. 263–266 (cit. on p. 54).
- [GG01] Vittorio Girotto and Michel Gonzalez. “Solving probabilistic and statistical problems: a matter of information structure and question form.” In: *Cognition* 78.3 (2001), pp. 247–276 (cit. on pp. 11, 15).
- [Goo95] Irving John Good. “When batterer turns murderer”. In: *Nature* 375.6532 (1995), p. 541 (cit. on p. 10).
- [GHW92] Sandra Graham, Cynthia Hudley, and Estella Williams. “Attributional and emotional determinants of aggression among African-American and Latino young adolescents.” In: *Developmental Psychology* 28.4 (1992), p. 731 (cit. on p. 36).

- [Gra+13] Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister, and Marc Streit. “Lineup: Visual analysis of multi-attribute rankings”. In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2277–2286 (cit. on pp. 24, 27).
- [Gre89] Robert L Greene. “Spacing effects in memory: Evidence for a two-process account.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15.3 (1989), p. 371 (cit. on p. 36).
- [HB05] Carolyn L Hafer and Laurent Bègue. “Experimental research on just-world theory: problems, developments, and future challenges.” In: *Psychological bulletin* 131.1 (2005), p. 128 (cit. on p. 36).
- [Hal+22] Brian D Hall, Yang Liu, Yvonne Jansen, et al. “A survey of tasks and visualizations in multiverse analysis reports”. In: *Computer Graphics Forum*. Vol. 41. 1. Wiley Online Library. 2022, pp. 402–426 (cit. on p. 55).
- [Hal+16] K Wm Hall, Charles Perin, Peter G Kusalik, Carl Gutwin, and Sheelagh Cappendale. “Formalizing emphasis in information visualization”. In: *Computer Graphics Forum*. Vol. 35. 3. Wiley Online Library. 2016, pp. 717–737 (cit. on p. 46).
- [Hal17] Niel Halloran. *Emotional Statistics*. Keynote presentation at Tapestry: The Data Storytelling Conference <https://youtu.be/TCqcpL8F99k>. 2017 (cit. on p. 80).
- [Hal19] John Halstead. *Founders Pledge – Mental Health Executive Summary*. Web article <https://founderspledge.com/stories/mental-health-report-summary>. 2019 (cit. on p. 85).
- [HJW12] Andrew S. Hanks, David Just, and Brian Wansink. *Trigger Foods: The Influence of “Irrelevant” Alternatives in School Lunchrooms*. Working Papers 05. Cornell Center for Behavioral Economics in Child Nutrition Programs, 2012 (cit. on p. 18).
- [Har+14] Lane Harrison, Fumeng Yang, Steven Franconeri, and Remco Chang. “Ranking visualizations of correlation using weber’s law”. In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), pp. 1943–1952 (cit. on pp. 35, 52, 53).
- [Has03] Martie G Haselton. “The sexual overperception bias: Evidence of a systematic bias in men from a survey of naturally occurring events”. In: *Journal of Research in Personality* 37.1 (2003), pp. 34–47 (cit. on p. 35).
- [HGT77] Lynn Hasher, David Goldstein, and Thomas Toppino. “Frequency and the conference of referential validity”. In: *Journal of verbal learning and verbal behavior* 16.1 (1977), pp. 107–112 (cit. on p. 35).
- [HC54] Albert H Hastorf and Hadley Cantril. “They saw a game; a case study.” In: *The Journal of Abnormal and Social Psychology* 49.1 (1954), p. 129 (cit. on p. 36).
- [Hea99] Chip Heath. “On the social psychology of agency relationships: Lay theories of motivation overemphasize extrinsic incentives”. In: *Organizational behavior and human decision processes* 78.1 (1999), pp. 25–62 (cit. on p. 35).

- [Hed00] James Hedlund. “Risky business: safety regulations, risk compensation, and individual behavior”. In: *Injury prevention* 6.2 (2000), pp. 82–89 (cit. on p. 36).
- [HB10] Jeffrey Heer and Michael Bostock. “Crowdsourcing graphical perception: using mechanical turk to assess visualization design”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2010, pp. 203–212 (cit. on p. 28).
- [HC04] Linda A Henkel and Kimberly J Coffman. “Memory distortions in coerced false confessions: A source monitoring framework analysis”. In: *Applied Cognitive Psychology* 18.5 (2004), pp. 567–588 (cit. on p. 36).
- [Hig96] Scott Highhouse. “Context-Dependent Selection: The Effects of Decoy and Phantom Job Candidates”. In: *Organizational Behavior and Human Decision Processes* 65.1 (1996), pp. 68–76 (cit. on p. 18).
- [HG04] Ulrich Hoffrage and Gerd Gigerenzer. “How to Improve the Diagnostic Inferences of Medical Experts”. In: *Experts in Science and Society* (2004). Ed. by Elke Kurz-Milcke and GerdEditors Gigerenzer, pp. 249–268 (cit. on p. 11).
- [Hof+00] Ulrich Hoffrage, Samuel Lindsey, Ralph Hertwig, and Gerd Gigerenzer. “Medicine: Communicating statistical information”. In: *Science* 290.5500 (2000), pp. 2261–2262 (cit. on pp. 10, 11).
- [HWS21] Matt-Heun Hong, Jessica K Witt, and Danielle Albers Szafir. “The Weighted Average Illusion: Biases in Perceived Mean Position in Scatterplots”. In: *IEEE Transactions on Visualization and Computer Graphics* 28.1 (2021), pp. 987–997 (cit. on p. 37).
- [Hse98] Christopher K Hsee. “Less Is Better: When Low-value Options Are Valued More Highly than High-value Options”. In: *Journal of Behavioral Decision Making* 11 (1998), pp. 107–121 (cit. on p. 35).
- [HZ04] Christopher K Hsee and Jiao Zhang. “Distinction bias: misprediction and mischoice due to joint evaluation.” In: *Journal of personality and social psychology* 86.5 (2004), p. 680 (cit. on p. 35).
- [HPP82] Joel Huber, John W Payne, and Christopher Puto. “Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis”. In: *Journal of Consumer Research* 9.1 (1982), pp. 90–98 (cit. on pp. 17, 18, 22, 35).
- [HPP14] Joel Huber, John W. Payne, and Christopher P. Puto. “Let’s Be Honest About the Attraction Effect”. In: *Journal of Marketing Research* 51.4 (2014), pp. 520–525. eprint: <http://dx.doi.org/10.1509/jmr.14.0208> (cit. on p. 22).
- [HRA15] Jessica Hullman, Paul Resnick, and Eytan Adar. “Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering”. In: *PLoS one* 10.11 (2015), e0142444 (cit. on pp. 51, 52).

- [Hul11] Jessica R Hullman. “Not All HITs Are Created Equal: Controlling for Reasoning and Learning Processes in MTurk”. In: *Proceedings of the ACM CHI Workshop on Crowdsourcing and Human Computation*. Available from <http://crowdresearch.org/chi2011-workshop/papers/hullman.pdf>. Citeseer. 2011 (cit. on p. 28).
- [HJC14] Samuel Huron, Yvonne Jansen, and Sheelagh Carpendale. “Constructing visual representations: Investigating the use of tangible tokens”. In: *IEEE transactions on visualization and computer graphics* 20.12 (2014), pp. 2102–2111 (cit. on p. 46).
- [Hur+16] Samuel Huron, Alice Thudt, Bon Adriel Aseniero, Anthony Tang, and Sheelagh Carpendale. “Constructive visualization: A new paradigm to empower people to author visualization”. In: *SURFNET/Designing Digital Surface Applications*, Frank Maurer (Ed.). NSERC Surfnet, University of Calgary (2016), pp. 169–191 (cit. on p. 46).
- [Iva+19] Alexander Ivanov, Kurtis Danyluk, Christian Jacob, and Wesley Willett. “A walk among the data”. In: *IEEE Computer Graphics and Applications* 39.3 (2019), pp. 19–28 (cit. on p. 87).
- [IDW18] Alexander Ivanov, Kurtis Thorvald Danyluk, and Wesley Willett. “Exploration & Anthropomorphism in Immersive Unit Visualizations”. In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM. Montreal QC, Canada: Association for Computing Machinery, 2018, LBW008 (cit. on p. 73).
- [JJ82] W Burke Jackson and James V Jucker. “An empirical study of travel time variability and travel choice behavior”. In: *Transportation Science* 16.4 (1982), pp. 460–475 (cit. on p. 35).
- [Jan+10] Amy C Janes, Diego A Pizzagalli, Sarah Richardt, et al. “Neural substrates of attentional bias for smoking-related cues: an fMRI study”. In: *Neuropsychopharmacology* 35.12 (2010), pp. 2339–2345 (cit. on p. 36).
- [JP96] Ashok Jansari and Alan J Parkin. “Things that go bump in your life: Explaining the reminiscence bump in autobiographical memory.” In: *Psychology and Aging* 11.1 (1996), p. 85 (cit. on p. 36).
- [Jan14] Yvonne Jansen. 2014 – 888,246 Ceramic Poppies to Commemorate Fallen Soldiers in WW1. <http://dataphys.org/list/tower-poppies-888246-ceramic-poppies-to-commemorate-fallen-soldiers-in-ww1/>. 2014 (cit. on p. 5).
- [JDF13] Yvonne Jansen, Pierre Dragicevic, and Jean-Daniel Fekete. “Evaluating the efficiency of physical visualizations”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2013, pp. 2593–2602 (cit. on p. 49).
- [Jan+15] Yvonne Jansen, Pierre Dragicevic, Petra Isenberg, et al. “Opportunities and challenges for data physicalization”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2015, pp. 3227–3236 (cit. on pp. 2, 67).

- [JCM06] Steve MJ Janssen, Antonio G Chessa, and Jaap MJ Murre. “Memory for time: How people date events”. In: *Memory & cognition* 34.1 (2006), pp. 138–147 (cit. on p. 36).
- [JL69] Jon Jecker and David Landy. “Liking a person as a function of doing him a favour”. In: *Human Relations* 22.4 (1969), pp. 371–378 (cit. on p. 36).
- [JL97] Karen Jenni and George Loewenstein. “Explaining the identifiable victim effect”. In: *Journal of Risk and Uncertainty* 14.3 (1997), pp. 235–257 (cit. on p. 35).
- [JS94] Hollyn M Johnson and Colleen M Seifert. “Sources of the continued influence effect: When misinformation in memory affects later inferences.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20.6 (1994), p. 1420 (cit. on p. 36).
- [JPC02] John T Jost, Brett W Pelham, and Mauricio R Carvalho. “Non-conscious forms of system justification: Implicit and behavioral preferences for higher status groups”. In: *Journal of Experimental Social Psychology* 38.6 (2002), pp. 586–602 (cit. on p. 36).
- [KT72] Daniel Kahneman and Amos Tversky. “Subjective probability: A judgment of representativeness”. In: *Cognitive psychology* 3.3 (1972), pp. 430–454 (cit. on p. 35).
- [Kal+18] Alex Kale, Francis Nguyen, Matthew Kay, and Jessica Hullman. “Hypothetical Outcome Plots Help Untrained Observers Judge Trends in Ambiguous Data”. In: *IEEE transactions on visualization and computer graphics* (2018) (cit. on pp. 51, 52).
- [Kap+16] Oral Kaplan, Goshiro Yamamoto, Yasuhide Yoshitake, et al. “In-situ visualization of pedaling forces on cycling training videos”. In: *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*. IEEE. 2016, pp. 000994–000999 (cit. on p. 35).
- [KMS99] Stuart T Kard, Jock D Mackinlay, and Ben Scheiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999 (cit. on pp. 2, 58).
- [KLS09] Niklas Karlsson, George Loewenstein, and Duane Seppi. “The ostrich effect: Selective attention to information”. In: *Journal of Risk and uncertainty* 38.2 (2009), pp. 95–115 (cit. on p. 36).
- [KD18] Matthew Kay and Pierre Dragicevic. *A Multiverse Reanalysis of Likert-Type Responses*. Online. <https://explorablemultiverse.github.io/examples/likert/>. 2018 (cit. on p. 51).
- [Kei+08] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, et al. “Visual analytics: Definition, process, and challenges”. In: *Information visualization*. Springer, 2008, pp. 154–175 (cit. on p. 2).
- [KK96] Daniel A Keim and H-P Kriegel. “Visualization techniques for mining large databases: A comparison”. In: *IEEE Transactions on knowledge and data engineering* 8.6 (1996), pp. 923–938 (cit. on p. 23).

- [KJW05] Dave Kelly, Jaap Jasperse, and Ian Westbrooke. “Designing science graphs for data analysis and presentation”. In: *Department of Conservation Technical Series 32* (2005) (cit. on p. 4).
- [KH02] Boaz Keysar and Anne S Henly. “Speakers’ overestimation of their effectiveness”. In: *Psychological Science* 13.3 (2002), pp. 207–212 (cit. on p. 35).
- [Kha+15] Azam Khan, Simon Breslav, Michael Glueck, and Kasper Hornbæk. “Benefits of visualization in the Mammography Problem”. In: *International Journal of Human-Computer Studies* 83 (2015), pp. 94–113 (cit. on pp. 35, 37).
- [KBH18] Azam Khan, Simon Breslav, and Kasper Hornbæk. “Interactive instruction in bayesian inference”. In: *Human-Computer Interaction* 33.3 (2018), pp. 207–233 (cit. on p. 37).
- [KG13] Kris N Kirby and Daniel Gerlanc. “BootES: An R package for bootstrap confidence intervals on effect sizes”. In: *Behavior research methods* 45.4 (2013), pp. 905–927 (cit. on p. 50).
- [KK13] Chris Kirk and Dan Kois. *How Many People Have Been Killed by Guns Since Newtown?* https://web.archive.org/web/20200305065626/http://www.slate.com/articles/news_and_politics/crime/2012/12/gun_death_tally_every_american_gun_death_since_newtown_sandy_hook_shooting.html. [Online; accessed 17-July-2019]. 2013 (cit. on p. 64).
- [KCS08] Aniket Kittur, Ed H Chi, and Bongwon Suh. “Crowdsourcing user studies with Mechanical Turk”. In: *Proc. of the SIGCHI conference on human factors in computing systems*. ACM. 2008, pp. 453–456 (cit. on p. 28).
- [Koe97] Jonathan J Koehler. “One in Millions, Billions and Trillions: Lessons from People V. Collins (1968) for People V. Simpson (1995)”. In: *Legal Education* 47 (1997), pp. 214–223 (cit. on p. 10).
- [KGP00] Asher Koriat, Morris Goldsmith, and Ainat Pansky. “Toward a psychology of memory accuracy”. In: *Annual review of psychology* 51.1 (2000), pp. 481–537 (cit. on p. 36).
- [Kos12] Robert Kosara. *All Responses to Gelman and Unwin in One Convenient Posting*. Blog post <https://eagereyes.org/blog/2012/responses-gelman-unwin-convenient-posting>. 2012 (cit. on p. 2).
- [KM13] Robert Kosara and Jock Mackinlay. “Storytelling: The next step for visualization”. In: *Computer* 46.5 (2013), pp. 44–50 (cit. on p. 80).
- [Koy+17] Philipp Koytek, Charles Perin, Jo Vermeulen, Elisabeth André, and Sheelagh Carpendale. “Mybrush: Brushing and linking with personal agency”. In: *IEEE transactions on visualization and computer graphics* 24.1 (2017), pp. 605–615 (cit. on p. 46).
- [Kra+21] Matthias Kraus, Karsten Klein, Johannes Fuchs, et al. “The value of immersive visualization”. In: *IEEE computer graphics and applications* 41.4 (2021), pp. 125–132 (cit. on p. 86).

- [Kru99] Justin Kruger. "Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments." In: *J. Pers. Soc. Psychol.* 77.2 (1999), p. 221 (cit. on p. 35).
- [KD99] Justin Kruger and David Dunning. "Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments." In: *J. Pers. Soc. Psychol.* 77.6 (1999), p. 1121 (cit. on p. 35).
- [KG99] Justin Kruger and Thomas Gilovich. "'Naive cynicism" in everyday theories of responsibility assessment: On biased assumptions of bias." In: *J. Pers. Soc. Psychol.* 76.5 (1999), p. 743 (cit. on p. 35).
- [Kua+12] X Kuang, H Zhang, S Zhao, and Michael J McGuffin. "Tracing tuples across dimensions: A comparison of scatterplots and parallel coordinate plots". In: *Computer Graphics Forum*. Vol. 31. 3pt4. Wiley Online Library. 2012, pp. 1365–1374 (cit. on p. 24).
- [LB11] Tanya Latty and Madeleine Beekman. "Irrational decision-making in an amoeboid organism: transitivity and context-dependent preferences". In: *Proceedings of the Royal Society B: Biological Sciences* 278.1703 (2011), p. 307 (cit. on p. 18).
- [Law+21] Po-Ming Law, Leo Yu-Ho Lo, Alex Endert, John Stasko, and Huamin Qu. "Causal perception in question-answering systems". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, pp. 1–15 (cit. on p. 37).
- [LMV10] Jing Li, Jean-Bernard Martens, and Jarke J Van Wijk. "Judging correlation from scatterplots and parallel coordinate plots". In: *Information Visualization* 9.1 (2010), pp. 13–30 (cit. on p. 24).
- [LF77] Sarah Lichtenstein and Baruch Fischhoff. "Do those who know more also know more about how much they know?" In: *Organizational behavior and human performance* 20.2 (1977), pp. 159–183 (cit. on p. 35).
- [LPW20] Johannes Liem, Charles Perin, and Jo Wood. "Structure and Empathy in Visual Data Storytelling: Evaluating their Influence on Attitude". In: *Computer Graphics Forum*. Vol. 39. 3. 2020, pp. 277–289 (cit. on p. 70).
- [LD16] Daniel Lim and David DeSteno. "Suffering and compassion: The links among adverse life experiences, empathy, compassion, and prosocial behavior." In: *Emotion* 16.2 (2016), p. 175 (cit. on p. 74).
- [LJ89] D Stephen Lindsay and Marcia K Johnson. "The eyewitness suggestibility effect and memory for source". In: *Memory & Cognition* 17.3 (1989), pp. 349–358 (cit. on p. 36).
- [Loe05] George Loewenstein. "Hot-cold empathy gaps and medical decision making." In: *Health Psychology* 24.4S (2005), S49 (cit. on p. 35).

- [LP74] Elizabeth F Loftus and John C Palmer. “Reconstruction of automobile destruction: An example of the interaction between language and memory”. In: *Journal of verbal learning and verbal behavior* 13.5 (1974), pp. 585–589 (cit. on p. 36).
- [Mac19] William MacAskill. “The definition of effective altruism”. In: *Effective altruism: Philosophical issues*. Ed. by Hilary Greaves and Theron Pummer. Oxford University Press, 2019. Chap. 10 (cit. on pp. 80, 81).
- [Mah77] Michael J Mahoney. “Publication prejudices: An experimental study of confirmatory bias in the peer review system”. In: *Cognitive therapy and research* 1.2 (1977), pp. 161–175 (cit. on p. 35).
- [MHH13] Selin A. Malkoc, William Hedgcock, and Steve Hoeffler. “Between a rock and a hard place: The failure of the attraction effect among unattractive alternatives”. In: *Journal of Consumer Psychology* 23.3 (2013), pp. 317–329 (cit. on pp. 18, 20).
- [MRO12] Dietrich Manzey, Juliane Reichenbach, and Linda Onnasch. “Human performance consequences of automated decision aids: The impact of degree of automation and system experience”. In: *Journal of Cognitive Engineering and Decision Making* 6.1 (2012), pp. 57–87 (cit. on p. 35).
- [Mar+93] Frank M Marchak, William S Cleveland, Bernice E Rogowitz, and Christopher D Wickens. “The psychology of visualization”. In: *IEEE Visualization: Proceedings of the 4 th conference on Visualization’93: San Jose, California*. Vol. 25. 29. 1993, pp. 351–354 (cit. on p. 2).
- [Mar09] Emilie Marcus. “Taming supplemental material”. In: *Immunity* 31.5 (2009), p. 691 (cit. on p. 55).
- [MW10] Winter Mason and Duncan J Watts. “Financial incentives and the performance of crowds”. In: *ACM SigKDD Explorations Newsletter* 11.2 (2010), pp. 100–108 (cit. on p. 28).
- [MC05] Mara Mather and Laura L Carstensen. “Aging and motivated cognition: The positivity effect in attention and memory”. In: *Trends in cognitive sciences* 9.10 (2005), pp. 496–502 (cit. on p. 36).
- [MJ00] Mara Mather and Marcia K Johnson. “Choice-supportive source monitoring: Do our decisions seem better to us as we age?” In: *Psychology and aging* 15.4 (2000), p. 596 (cit. on p. 36).
- [Mat22] Dylan Matthews. “How effective altruism went from a niche movement to a billion-dollar force”. In: *Vox* (2022) (cit. on p. 89).
- [MMS95] John D Mayer, Laura J McCormick, and Sara E Strong. “Mood-congruent memory and natural mood: New evidence”. In: *Personality and Social Psychology Bulletin* 21.7 (1995), pp. 736–746 (cit. on p. 36).

- [MD02] Dawn M McBride and Barbara Anne Doshier. “A comparison of conscious and automatic memory processes for picture and word stimuli: A process dissociation analysis”. In: *Consciousness and cognition* 11.3 (2002), pp. 423–460 (cit. on p. 36).
- [McD86] Gilles O. McDaniel Mark A.; Einstein. “Bizarre imagery as an effective memory aid: The importance of distinctiveness.” In: *Journal of Exp. Psychology: Learning, Memory, and Cognition* 12 (1 1986) (cit. on p. 36).
- [McD+07] Mark A McDaniel, Janis L Anderson, Mary H Derbish, and Nova Morrisette. “Testing the testing effect in the classroom”. In: *European Journal of Cognitive Psychology* 19.4-5 (2007), pp. 494–513 (cit. on p. 36).
- [McE18] Richard McElreath. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press, 2018 (cit. on p. 49).
- [MT00] Matthew S. McGlone and Jessica Tofighbakhsh. “Birds of a Feather Flock Conjointly (?): Rhyme as Reason in Aphorisms”. In: *Psychological Science* 11.5 (2000), pp. 424–428 (cit. on p. 35).
- [Meh00] Abhilasha Mehta. “Advertising attitudes and advertising effectiveness”. In: *Journal of advertising research* 40.3 (2000), pp. 67–72 (cit. on p. 63).
- [MDF12] Luana Micallef, Pierre Dragicevic, and Jean-Daniel Fekete. “Assessing the effect of visualizations on bayesian reasoning through crowdsourcing”. In: *IEEE TVCG* 18.12 (2012). [PDF], pp. 2536–2545 (cit. on pp. 6, 9, 11, 28, 35, 37, 60, 91).
- [Mil63] Stanley Milgram. “Behavioral Study of obedience.” In: *The Journal of abnormal and social psychology* 67.4 (1963), p. 371 (cit. on p. 35).
- [MN75] Dale T Miller and Stephen A Norman. “Actor-observer differences in perceptions of effective control.” In: *J. Pers. Soc. Psychol.* 31.3 (1975), p. 503 (cit. on p. 36).
- [Mis18] Misc autors. *Radix for R Markdown*. <https://rstudio.github.io/radix/>. Accessed: 2018-09-16. 2018 (cit. on p. 54).
- [Mit+97] Terence R Mitchell, Leigh Thompson, Erika Peterson, and Randy Cronk. “Temporal adjustments in the evaluation of events: The “rosy view””. In: *Journal of Experimental Social Psychology* 33.4 (1997), pp. 421–448 (cit. on p. 36).
- [MM01] Benoit Monin and Dale T Miller. “Moral credentials and the expression of prejudice.” In: *Journal of personality and social psychology* 81.1 (2001), p. 33 (cit. on p. 36).
- [MAS22] Luiz Morais, Nazareno Andrade, and Dandara Sousa. “Exploring How Visualization Design and Situatedness Evoke Compassion in the Wild”. In: *Eurographics Conference on Visualization (EuroVis) 2022*. Rome, Italy, June 2022 (cit. on p. 88).
- [Mor+21] Luiz Morais, Yvonne Jansen, Nazareno Andrade, and Pierre Dragicevic. “Can anthropographics promote prosociality? a review and large-sample study”. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. [PDF]. 2021, pp. 1–18 (cit. on pp. 7, 63, 68, 70, 93).

- [Mor+20] Luiz Morais, Yvonne Jansen, Nazareno Andrade, and Pierre Dragicevic. “Showing data about people: A design space of anthropographics”. In: *IEEE Transactions on Visualization and Computer Graphics* 28.3 (2020). [PDF], pp. 1661–1679 (cit. on pp. 7, 63, 64, 75, 81, 93).
- [MG15] Carey K Morewedge and Colleen E Giblin. “Explanations of the endowment effect: an integrative review”. In: *Trends in cognitive sciences* 19.6 (2015), pp. 339–348 (cit. on p. 35).
- [Mor+13] Manfred Mortell, Hanan H Balkhy, Elias B Tannous, and Mei Thiee Jong. “Physician ‘defiance’ towards hand hygiene compliance: Is there a theory–practice–ethics gap?” In: *Journal of the Saudi Heart Association* 25.3 (2013), pp. 203–208 (cit. on p. 35).
- [MCP71] John Morton, Robert G Crowder, and Harvey A Prussin. “Experiments with the stimulus suffix effect.” In: (1971) (cit. on p. 36).
- [Mox+03] Annette Moxey, Dianne O’Connell, Patricia McGettigan, and David Henry. “Describing treatment effects to patients: How they are expressed makes a difference”. In: *Journal of General Internal Medicine* 18.11 (2003), pp. 948–959 (cit. on p. 38).
- [Mun+17] Marcus R Munafò, Brian A Nosek, Dorothy VM Bishop, et al. “A manifesto for reproducible science”. In: *Nature Human Behaviour* 1.1 (2017), p. 0021 (cit. on p. 47).
- [Mun08] Tamara Munzner. “Process and pitfalls in writing information visualization research papers”. In: *Information visualization*. Springer, 2008, pp. 134–153 (cit. on p. 2).
- [Mun14] Tamara Munzner. *Visualization analysis and design*. CRC press, 2014 (cit. on p. 58).
- [Mur62] Bennet B Murdock Jr. “The serial position effect of free recall.” In: *Journal of experimental psychology* 64.5 (1962), p. 482 (cit. on p. 36).
- [NCG93] Richard Nadeau, Edouard Cloutier, and J-H Guay. “New evidence about the existence of a bandwagon effect in the opinion formation process”. In: *International Political Science Review* 14.2 (1993), pp. 203–213 (cit. on p. 36).
- [NSS18] Leif D Nelson, Joseph Simmons, and Uri Simonsohn. “Psychology’s renaissance”. In: *Annual review of psychology* 69 (2018) (cit. on p. 47).
- [NW77] Richard E Nisbett and Timothy D Wilson. “The halo effect: Evidence for unconscious alteration of judgments.” In: *Journal of personality and social psychology* 35.4 (1977), p. 250 (cit. on p. 36).
- [NT07] Deborah Nolan and Duncan Temple Lang. “Dynamic, interactive documents for teaching statistical practice”. In: *International Statistical Review* 75.3 (2007), pp. 295–321 (cit. on p. 47).
- [NHP09] Loran F Nordgren, Frenk van Harreveld, and Joop van der Pligt. “The restraint bias: How the illusion of self-restraint promotes impulsive behavior”. In: *Psychological Science* 20.12 (2009), pp. 1523–1528 (cit. on p. 35).

- [NMA12] Michael I Norton, Daniel Mochon, and Dan Ariely. “The IKEA effect: When labor leads to love”. In: *Journal of Consumer Psychology* 22.3 (2012), pp. 453–460 (cit. on p. 35).
- [Nos+15] Brian A Nosek, George Alter, George C Banks, et al. “Promoting an open research culture”. In: *Science* 348.6242 (2015), pp. 1422–1425 (cit. on p. 47).
- [Nüs+17] Daniel Nüst, Markus Konkol, Marc Schutzeichel, et al. “Opening the publication process with executable research compendia”. In: *D-Lib Magazine* 23.1/2 (2017) (cit. on p. 48).
- [NR15] Brendan Nyhan and Jason Reifler. “Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information”. In: *Vaccine* 33.3 (2015), pp. 459–464 (cit. on p. 63).
- [NR10] Brendan Nyhan and Jason Reifler. “When corrections fail: The persistence of political misperceptions”. In: *Political Behavior* 32.2 (2010), pp. 303–330 (cit. on p. 36).
- [OM85] Edward J O’Brien and Jerome L Myers. “When comprehension difficulty improves memory for text.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 11.1 (1985), p. 12 (cit. on p. 36).
- [OP95] Yigang Pan Sue O’Curry and Robert Pitts. “The attraction effect and political choice in two elections”. In: *Journal of Consumer Psychology* 4.1 (1995), pp. 85–101 (cit. on pp. 17, 18).
- [OPR16] Mehrnoosh Oghbaie, Michael J Pennock, and William B Rouse. “Understanding the efficacy of interactive visualization for decision making for complex systems”. In: *Systems Conference (SysCon), 2016 Annual IEEE*. IEEE. 2016, pp. 1–6 (cit. on p. 35).
- [OL03] Maria Cristina Ferreira de Oliveira and Haim Levkowitz. “From visual data exploration to visual data mining: a survey”. In: *IEEE Transactions on Visualization and Computer Graphics* 9.3 (2003), pp. 378–394 (cit. on p. 23).
- [OMD09] Daniel M Oppenheimer, Tom Meyvis, and Nicolas Davidenko. “Instructional manipulation checks: Detecting satisficing to increase statistical power”. In: *JESP* 45.4 (2009), pp. 867–872 (cit. on p. 31).
- [Ord19] Toby Ord. “The moral imperative toward cost-effectiveness in global health”. In: *Effective Altruism: Philosophical Issues*. Oxford: Oxford University Press, 2019, pp. 29–36 (cit. on pp. 75, 79).
- [Ott+15] Alvitta Ottley, Evan M Peck, Lane T Harrison, et al. “Improving Bayesian reasoning: The effects of phrasing, visualization, and spatial ability”. In: *IEEE transactions on visualization and computer graphics* 22.1 (2015), pp. 529–538 (cit. on p. 37).
- [Paj+17] Stephan Pajer, Marc Streit, Thomas Torsney-Weir, et al. “WeightLifter: Visual Weight Space Exploration for Multi-Criteria Decision Making.” In: *IEEE Transactions on Visualization and Computer Graphics* 23.1 (2017), p. 611 (cit. on pp. 24, 27).

- [Pal03] Gilles Palsky. *L'esprit des cartes: approches historiques, sémiologiques et sociologiques en cartographie: diplôme d'habilitation à diriger des recherches*. Université de Paris XII-Val-de-Marne, 2003 (cit. on p. 40).
- [Pan+14] Anshul Vikram Pandey, Anjali Manivannan, Oded Nov, Margaret Satterthwaite, and Enrico Bertini. “The persuasive power of data visualization”. In: *IEEE TVCG* 20.12 (2014), pp. 2211–2220 (cit. on p. 63).
- [Pan+15] Anshul Vikram Pandey, Katharina Rall, Margaret L Satterthwaite, Oded Nov, and Enrico Bertini. “How deceptive are deceptive visualizations?: An empirical analysis of common distortion techniques”. In: *Proceedings of CHI*. ACM. 2015, pp. 1469–1478 (cit. on p. 56).
- [PR82] Bernadette Park and Myron Rothbart. “Perception of out-group homogeneity and levels of social categorization: Memory for the subordinate attributes of in-group and out-group members.” In: *J. Pers. Soc. Psychol.* 42.6 (1982), p. 1051 (cit. on p. 35).
- [PWA98] Amanda Parker, Edward Wilding, and Colin Akerman. “The von Restorff effect in visual object recognition memory in humans and monkeys: The role of frontal/perirhinal interaction”. In: *Journal of Cognitive Neuroscience* 10.6 (1998), pp. 691–703 (cit. on p. 36).
- [PKF09] Helen M Paterson, Richard I Kemp, and Joseph P Forgas. “Co-witnesses, confederates, and conformity: Effects of discussion and delay on eyewitness memory”. In: *Psychiatry, Psychology and Law* 16.sup1 (2009), S112–S124 (cit. on p. 36).
- [PBV16] Charles Perin, Jeremy Boy, and Frédéric Vernier. “Using gap charts to visualize the temporal evolution of ranks and scores”. In: *IEEE Computer Graphics and Applications* 36.5 (2016), pp. 38–49 (cit. on p. 46).
- [PDF14] Charles Perin, Pierre Dragicevic, and Jean-Daniel Fekete. “Revisiting bertin matrices: New interactions for crafting tabular visualizations”. In: *IEEE transactions on visualization and computer graphics* 20.12 (2014). [PDF], pp. 2082–2091 (cit. on pp. 6, 23, 27, 39, 41, 44, 45, 92).
- [PFD19] Charles Perin, Jean-Daniel Fekete, and Pierre Dragicevic. “Jacques Bertin’s legacy in information visualization and the reorderable matrix”. In: *Cartography and Geographic Information Science* 46.2 (2019). [PDF], pp. 176–181 (cit. on pp. 6, 39, 41, 92).
- [Per+17] Charles Perin, Tiffany Wun, Richard Pusch, and Sheelagh Carpendale. “Assessing the graphical perception of time and speed on 2d+ time trajectories”. In: *IEEE transactions on visualization and computer graphics* 24.1 (2017), pp. 698–708 (cit. on p. 46).
- [PW00] Jonathan C Pettibone and Douglas H Wedell. “Examining Models of Nondominated Decoy Effects across Judgment and Choice”. In: *Organizational Behavior and Human Decision Processes* 81.2 (2000), pp. 300–328 (cit. on p. 35).

- [Pet79] Thomas F Pettigrew. “The ultimate attribution error: Extending Allport’s cognitive analysis of prejudice”. In: *Personality and social psychology bulletin* 5.4 (1979), pp. 461–476 (cit. on p. 36).
- [PE66] Lawrence D Phillips and Ward Edwards. “Conservatism in a simple probability inference task.” In: *Journal of experimental psychology* 72.3 (1966), p. 346 (cit. on p. 35).
- [PNM10] Linda M Phillips, Stephen P Norris, and John S Macnab. “A history of visualization in psychology and science”. In: *Visualization in mathematics, reading and science education*. Springer, 2010, pp. 9–18 (cit. on p. 2).
- [Plo21] Kim Plowright. 1968 – Grace Hopper’s Nanoseconds. <http://dataphys.org/list/grace-hopper-nanoseconds/>. 2021 (cit. on p. 5).
- [Pop+08] J Poppenk, G Walia, AR McIntosh, et al. “Why is the meaning of a sentence better remembered than its form? An fMRI study on the role of novelty-encoding processes”. In: *Hippocampus* 18.9 (2008), pp. 909–918 (cit. on p. 36).
- [Pre06] Norma Presmeg. “Research on visualization in learning and teaching mathematics: Emergence from psychology”. In: *Handbook of research on the psychology of mathematics education*. Brill, 2006, pp. 205–235 (cit. on p. 2).
- [Pro+01] Emily Pronin, Justin Kruger, Kenneth Savtisky, and Lee Ross. “You don’t know me, but I know you: The illusion of asymmetric insight.” In: *J. Pers. Soc. Psychol.* 81.4 (2001), p. 639 (cit. on p. 36).
- [PK07] Emily Pronin and Matthew B Kugler. “Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot”. In: *Journal of Experimental Social Psychology* 43.4 (2007), pp. 565–578 (cit. on p. 36).
- [PR06] Emily Pronin and Lee Ross. “Temporal differences in trait self-ascription: when the self is seen as an other.” In: *J. Pers. Soc. Psychol.* 90.2 (2006), p. 197 (cit. on p. 36).
- [PF00] Pearl Pu and Boi Faltings. “Enriching buyers’ experiences: the SmartClient approach”. In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM. 2000, pp. 289–296 (cit. on p. 24).
- [RS09] Priya Raghurib and Joydeep Srivastava. “The denomination effect”. In: *Journal of Consumer Research* 36.4 (2009), pp. 701–713 (cit. on p. 35).
- [Raj+17] Fateme Rajabiyazdi, Charles Perin, Jo Vermeulen, et al. “Differences that matter: In-clinic communication challenges”. In: *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 2017, pp. 251–260 (cit. on p. 46).
- [RMM15] Benjamin Renoust, Guy Melançon, and Tamara Munzner. “Detangler: Visual analytics for multiplex networks”. In: *Computer graphics forum*. Vol. 34. 3. Wiley Online Library. 2015, pp. 321–330 (cit. on p. 46).

- [Ric12] Francesco Ricci. “Social implications of malaria and their relationships with poverty”. In: *Med. j. of hematology and infectious diseases* 4.1 (2012) (cit. on p. 84).
- [RB90] Ilana Ritov and Jonathan Baron. “Reluctance to vaccinate: Omission bias and ambiguity”. In: *Journal of Behavioral Decision Making* 3.4 (1990), pp. 263–277 (cit. on p. 35).
- [Rob07] Jonathan C Roberts. “State of the art: Coordinated & multiple views in exploratory visualization”. In: *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV’07. Fifth International Conference on*. IEEE. 2007, pp. 61–71 (cit. on p. 54).
- [RKK77] Timothy B Rogers, Nicholas A Kuiper, and William S Kirker. “Self-reference and the encoding of personal information.” In: *Journal of personality and social psychology* 35.9 (1977), p. 677 (cit. on p. 36).
- [RR97] Ralph Rosnow and Robert Rosenthal. *People studying people: Artifacts and ethics in behavioral research*. WH Freeman, 1997 (cit. on p. 35).
- [RGH77] Lee Ross, David Greene, and Pamela House. “The “false consensus effect”: An egocentric bias in social perception and attribution processes”. In: *Journal of experimental social psychology* 13.3 (1977), pp. 279–301 (cit. on p. 35).
- [RS91] Lee Ross and Constance Stillinger. “Barriers to conflict resolution”. In: *Negotiation Journal* 7.4 (1991), pp. 389–404 (cit. on p. 35).
- [RS+79] Michael Ross, Fiore Sicoly, et al. “Egocentric biases in availability and attribution”. In: *Journal of personality and social psychology* 37.3 (1979), pp. 322–336 (cit. on p. 36).
- [RM90] Steven F Roth and Joe Mattis. “Data characterization for intelligent graphics presentation”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 1990, pp. 193–200 (cit. on p. 29).
- [RBW15] Joanna Różycka-Tran, Paweł Boski, and Bogdan Wojciszke. “Belief in a zero-sum game as a social axiom: A 37-nation study”. In: *Journal of Cross-Cultural Psychology* 46.4 (2015), pp. 525–548 (cit. on p. 36).
- [Rub] S. Rubin. Web app previously available at www.eecs.berkeley.edu/~srubin/toys/reorderable/. Seems to have gone offline. (cit. on p. 44).
- [RG04] Laurie A Rudman and Stephanie A Goodwin. “Gender differences in automatic in-group bias: Why do women like women more than men like men?” In: *Journal of personality and social psychology* 87.4 (2004), p. 494 (cit. on p. 36).
- [Sac+16] Dominik Sacha, Hansi Senaratne, Bum Chul Kwon, Geoffrey Ellis, and Daniel A Keim. “The role of uncertainty, awareness, and trust in visual analytics”. In: *IEEE transactions on visualization and computer graphics* 22.1 (2016), pp. 240–249 (cit. on p. 35).

- [SZ88] William Samuelson and Richard Zeckhauser. “Status quo bias in decision making”. In: *Journal of risk and uncertainty* 1.1 (1988), pp. 7–59 (cit. on p. 35).
- [SS04] Lawrence J Sanna and Norbert Schwarz. “Integrating temporal biases: The interplay of focal thoughts and accessibility experiences”. In: *Psychological science* 15.7 (2004), pp. 474–481 (cit. on p. 35).
- [SL10] Jeff Sauro and James R Lewis. “Average task times in usability tests: what to report?” In: *Proceedings of CHI*. ACM. 2010, pp. 2347–2350 (cit. on p. 50).
- [SG03] Kenneth Savitsky and Thomas Gilovich. “The illusion of transparency and the alleviation of speech anxiety”. In: *Journal of experimental social psychology* 39.6 (2003), pp. 618–625 (cit. on p. 35).
- [Saw96] Günther Sawitzki. “Extensible Statistical Software: On a Voyage to Oberon”. In: *Journal of Computational and Graphical Statistics* 5.3 (1996), pp. 263–283 (cit. on p. 41).
- [SH19] Simon Scarr and Marco Hernandez. *Drowning in plastic: Visualising the world’s addiction to plastic bottles*. Reuters. <https://graphics.reuters.com/ENVIRONMENT-PLASTIC/0100B275155/index.html>. 2019 (cit. on p. 5).
- [SK98] David A Schkade and Daniel Kahneman. “Does living in California make people happy? A focusing illusion in judgments of life satisfaction”. In: *Psychological Science* 9.5 (1998), pp. 340–346 (cit. on p. 36).
- [Sch19] Water Science School. *How Much Water is There on Earth?* <https://www.usgs.gov/special-topics/water-science-school/science/how-much-water-there-earth>. 2019 (cit. on p. 5).
- [SG01] P Sedlmeier and G Gigerenzer. “Teaching Bayesian reasoning in less than two hours”. In: *Experimental Psychology General* 130.3 (2001), pp. 380–400 (cit. on p. 11).
- [SDT97] Eldar Shafir, Peter Diamond, and Amos Tversky. “Money illusion”. In: *The Quarterly Journal of Economics* 112.2 (1997), pp. 341–374 (cit. on p. 35).
- [SWS02] Sharoni Shafir, Tom A. Waite, and Brian H. Smith. “Context-Dependent Violations of Rational Choice in Honeybees (*Apis mellifera*) and Gray Jays (*Perisoreus canadensis*)”. In: *Behavioral Ecology and Sociobiology* 51.2 (2002), pp. 180–187 (cit. on p. 18).
- [Sha+07] Tali Sharot, Alison M Riccardi, Candace M Raio, and Elizabeth A Phelps. “Neural mechanisms mediating optimism bias”. In: *Nature* 450.7166 (2007), pp. 102–105 (cit. on p. 35).
- [Sha70] Kelly G Shaver. “Defensive attribution: Effects of severity and relevance on the responsibility assigned for an accident.” In: *J. Pers. Soc. Psychol.* 14.2 (1970), p. 101 (cit. on p. 36).

- [Sii99] Harri Siirtola. “Interaction with the Reorderable Matrix”. In: *Proc. IV '99. IEEE*, 1999, pp. 272–277 (cit. on p. 41).
- [Sil+17] Raphael Silberzahn, Eric Luis Uhlmann, Dan Martin, et al. “Many analysts, one dataset: Making transparent how variations in analytical choices affect results”. In: (2017) (cit. on p. 47).
- [SNS11] Joseph P Simmons, Leif D Nelson, and Uri Simonsohn. “False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant”. In: *Psychological science* 22.11 (2011), pp. 1359–1366 (cit. on p. 47).
- [SSN15] Uri Simonsohn, Joseph P Simmons, and Leif D Nelson. “Specification Curve: Descriptive and Inferential Statistics on All Reasonable Specifications”. In: (2015) (cit. on pp. 47, 48, 54, 55).
- [Sim89] Itamar Simonson. “Choice Based on Reasons: The Case of Attraction and Compromise Effects”. In: *Journal of Consumer Research* 16.2 (1989), pp. 158–174 (cit. on p. 35).
- [Sim14] Itamar Simonson. “Vices and Virtues of Misguided Replications: The Case of Asymmetric Dominance”. In: *J. of Marketing Research* 51.4 (2014), pp. 514–519 (cit. on p. 22).
- [Sla68] Norman J Slamecka. “An examination of trace storage in free recall.” In: *Journal of experimental psychology* 76.4p1 (1968), p. 504 (cit. on p. 36).
- [Slo+03] Steven A. Sloman, David Over, Lila Slovak, and Jeffrey M. Stibel. “Frequency illusions and other fallacies”. In: *Organizational Behavior and Human Decision Processes* 91.2 (2003), pp. 296–309 (cit. on pp. 11, 13).
- [SLW11] Betsy Sparrow, Jenny Liu, and Daniel M Wegner. “Google effects on memory: Cognitive consequences of having information at our fingertips”. In: *science* 333.6043 (2011), pp. 776–778 (cit. on p. 36).
- [SMB91] Mark Spranca, Elisa Minsk, and Jonathan Baron. “Omission and commission in judgment and choice”. In: *Journal of experimental social psychology* 27.1 (1991), pp. 76–105 (cit. on p. 36).
- [Sta+16] Julian Stahnke, Marian Dörk, Boris Müller, and Andreas Thom. “Probing projections: Interaction techniques for interpreting arrangements and errors of dimensionality reductions”. In: *IEEE TVCG* 22.1 (2016), pp. 629–638 (cit. on p. 56).
- [Sta76] Barry M Staw. “Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action”. In: *Organizational behavior and human performance* 16.1 (1976), pp. 27–44 (cit. on p. 35).
- [Ste+16] Sara Steegen, Francis Tuerlinckx, Andrew Gelman, and Wolf Vanpaemel. “Increasing transparency through a multiverse analysis”. In: *Perspectives on Psychological Science* 11.5 (2016), pp. 702–712 (cit. on pp. 46–48, 51, 52, 54, 55).

- [Sti96] Gerald J Stine. *Acquired immune deficiency syndrome: biological, medical, social, and legal issues*. Prentice Hall, 1996 (cit. on p. 10).
- [SB17] Sergej Stoppel and Stefan Bruckner. “Vol 2 velle: Printable Interactive Volume Visualization”. In: *IEEE transactions on visualization and computer graphics* 23.1 (2017), pp. 861–870 (cit. on p. 35).
- [SLH10] Hannah Summerfelt, Louis Lippman, and Ira E Hyman Jr. “The effect of humor on memory: Constrained by the pun”. In: *The Journal of General Psychology* 137.4 (2010), pp. 376–394 (cit. on p. 36).
- [Sun02] Cass R Sunstein. “Probability neglect: Emotions, worst cases, and law”. In: *The Yale Law Journal* 112.1 (2002), pp. 61–107 (cit. on p. 35).
- [Sve08] Ola Svenson. “Decisions among time saving options: When intuition is strong and wrong”. In: *Acta Psychologica* 127.2 (2008), pp. 501–509 (cit. on p. 35).
- [TW16] Aner Tal and Brian Wansink. “Blinded with science: Trivial graphs and formulas increase ad persuasiveness and belief in product efficacy”. In: *Public Understanding of Science* 25.1 (2016), pp. 117–125 (cit. on pp. 56–59).
- [TKB04] James W Tanaka, Markus Kiefer, and Cindy M Bukach. “A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study”. In: *Cognition* 93.1 (2004), B1–B9 (cit. on p. 36).
- [TD81] Donald M Taylor and Janet R Doria. “Self-serving and group-serving bias in attribution”. In: *The Journal of Social Psychology* 113.2 (1981), pp. 201–211 (cit. on p. 36).
- [Tha81] Richard Thaler. “Some empirical evidence on dynamic inconsistency”. In: *Economics letters* 8.3 (1981), pp. 201–207 (cit. on p. 35).
- [Tha+97] Richard H Thaler, Amos Tversky, Daniel Kahneman, and Alan Schwartz. “The effect of myopia and loss aversion on risk taking: An experimental test”. In: *The Quarterly Journal of Economics* 112.2 (1997), pp. 647–661 (cit. on p. 35).
- [Tho99] Suzanne C Thompson. “Illusions of control: How we overestimate our personal influence”. In: *Current Directions in Psychological Science* 8.6 (1999), pp. 187–190 (cit. on p. 35).
- [Thu+16] Alice Thudt, Jagoda Walny, Charles Perin, et al. “Assessing the readability of stacked graphs”. In: *Proceedings of Graphics Interface Conference (GI)*. 2016 (cit. on p. 46).
- [Tob+88] Jerome Tobacyk, Gary Milford, Thomas Springer, and Zofia Tobacyk. “Paranormal beliefs and the Barnum effect”. In: *Journal of Personality Assessment* 52.4 (1988), pp. 737–739 (cit. on p. 35).
- [Tri13] Evangelos Triantaphyllou. *Multi-criteria decision making methods: a comparative study*. Vol. 44. Springer Science & Business Media, 2013 (cit. on p. 26).

- [Tru12] JenniferS. Trueblood. “Multialternative context effects obtained using an inference task”. English. In: *Psychonomic Bulletin and Review* 19.5 (2012), pp. 962–968 (cit. on p. 18).
- [Tuf01] Edward Tufte. “The visual display of quantitative information”. In: *Graphics Press, Cheshire, USA* 4.5 (2001), p. 6 (cit. on p. 4).
- [Tul74] Endel Tulving. “Cue-dependent forgetting: When we forget something we once knew, it does not necessarily mean that the memory trace has been lost; it may only be inaccessible”. In: *American Scientist* 62.1 (1974), pp. 74–82 (cit. on p. 36).
- [TK73] Amos Tversky and Daniel Kahneman. “Availability: A heuristic for judging frequency and probability”. In: *Cognitive psychology* 5.2 (1973), pp. 207–232 (cit. on p. 35).
- [TK83] Amos Tversky and Daniel Kahneman. “Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment.” In: *Psychological review* 90.4 (1983), p. 293 (cit. on p. 35).
- [TK75] Amos Tversky and Daniel Kahneman. “Judgment under uncertainty: Heuristics and biases”. In: *Utility, probability, and human decision making*. Springer, 1975, pp. 141–162 (cit. on p. 35).
- [TK86] Amos Tversky and Daniel Kahneman. “Rational choice and the framing of decisions”. In: *Journal of business* (1986), S251–S278 (cit. on p. 35).
- [TK81] Amos Tversky and Daniel Kahneman. “The framing of decisions and the psychology of choice.” In: *Science* (1981) (cit. on p. 35).
- [TK94] Amos Tversky and Derek J Koehler. “Support theory: A nonextensional representation of subjective probability.” In: *Psychological review* 101.4 (1994), p. 547 (cit. on p. 35).
- [UN93] JoNell A Usher and Ulric Neisser. “Childhood amnesia and the beginnings of memory for four early life events.” In: *Journal of Experimental Psychology: General* 122.2 (1993), p. 155 (cit. on p. 36).
- [VZS17a] André Caldero Valdez, Martina Ziefle, and Michael Sedlmair. “A Framework for Studying Biases in Visualization Research”. In: *DECISIVE: Workshop on Dealing with Cognitive Biases in Visualizations. IEEE VIS*. 2017 (cit. on p. 35).
- [VZS17b] Andre Calero Valdez, Martina Ziefle, and Michael Sedlmair. “Priming and Anchoring Effects in Visualizations”. In: *IEEE Transactions on Visualization and Computer Graphics* (2017) (cit. on p. 35).
- [VLB67] Françoise Vergneault, Roland Lamontagne, and Jacques Bertin. “Traitement graphique d’une information: les marines royales de France et de Grande-Bretagne (1697-1747)”. In: *Annales. Économies, Sociétés, Civilisations* 22.5 (1967), pp. 991–1004 (cit. on pp. 42, 43).

- [Vic11] Bret Victor. *Explorable Explanations*. Online. <http://worrydream.com/ExplorableExplanations/>. 2011 (cit. on pp. 46, 48–50, 55).
- [VFP11] Joel L Voss, Kara D Federmeier, and Ken A Paller. “The potato chip really does look like Elvis! Neural hallmarks of conceptual processing associated with finding novel shapes subjectively meaningful”. In: *Cerebral Cortex* 22.10 (2011), pp. 2354–2364 (cit. on p. 35).
- [WK85] Willem A Wagenaar and Gideon B Keren. “Calibration of probability assessments by professional blackjack dealers, statistical experts, and lay people”. In: *Organizational Behavior and Human Decision Processes* 36.3 (1985), pp. 406–416 (cit. on p. 35).
- [WV14] Drew Walker and Edward Vul. “Hierarchical encoding makes individuals in a group seem more attractive”. In: *Psychological Science* 25.1 (2014), pp. 230–235 (cit. on p. 35).
- [WS09] W Richard Walker and John J Skowronski. “The fading affect bias: But what the hell is it for?” In: *Applied Cognitive Psychology* 23.8 (2009), pp. 1122–1136 (cit. on p. 36).
- [Wal+19] Emily Wall, Leslie Blaha, Celeste Paul, and Alex Endert. “A formative study of interactive bias metrics in visual analytics using anchoring bias”. In: *IFIP Conference on Human-Computer Interaction*. Springer. 2019, pp. 555–575 (cit. on p. 37).
- [Wal+17] Emily Wall, Leslie M Blaha, Lyndsey Franklin, and Alex Endert. “Warning, bias may occur: A proposed approach to detecting cognitive bias in interactive visual analytics”. In: *IEEE Conference on Visual Analytics Science and Technology (VAST)*. 2017 (cit. on p. 36).
- [Was60] Peter C Wason. “On the failure to eliminate hypotheses in a conceptual task”. In: *Quarterly journal of experimental psychology* 12.3 (1960), pp. 129–140 (cit. on p. 35).
- [WCE10] Adam Waytz, John Cacioppo, and Nicholas Epley. “Who sees human? The stability and importance of individual differences in anthropomorphism”. In: *Perspectives on Psychological Science* 5.3 (2010), pp. 219–232 (cit. on p. 36).
- [WC98] Martin Weber and Colin F Camerer. “The disposition effect in securities trading: An experimental analysis”. In: *Journal of Economic Behavior & Organization* 33.2 (1998), pp. 167–184 (cit. on p. 35).
- [Wed91] Douglas H Wedell. “Distinguishing among models of contextually induced preference reversals.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17.4 (1991), p. 767 (cit. on pp. 18, 21).

- [WL90] Stephen Wehrend and Clayton Lewis. “A problem-oriented classification of visualization techniques”. In: *Proc. of Visualization’90*. IEEE Computer Society Press. 1990, pp. 139–143 (cit. on p. 29).
- [Wei80] Neil D Weinstein. “Unrealistic optimism about future life events.” In: *Journal of personality and social psychology* 39.5 (1980), p. 806 (cit. on p. 35).
- [WB10a] H Gilbert Welch and William C Black. “Overdiagnosis in Cancer”. In: *J National Cancer Institute* 102.9 (2010), pp. 605–613 (cit. on p. 10).
- [Wic+16] Jelte M Wicherts, Coosje LS Veldkamp, Hilde EM Augusteijn, et al. “Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking”. In: *Frontiers in Psychology* 7 (2016), p. 1832 (cit. on p. 47).
- [WB10b] Mark Ottoni Wilhelm and René Bekkers. “Helping behavior, dispositional empathic concern, and the principle of care”. In: *Social Psychology Quarterly* 73.1 (2010), pp. 11–32 (cit. on p. 74).
- [WJD17] Wesley Willett, Yvonne Jansen, and Pierre Dragicevic. “Embedded data representations”. In: *IEEE transactions on visualization and computer graphics* 23.1 (2017), pp. 461–470 (cit. on p. 68).
- [WS92] Christopher Williamson and Ben Shneiderman. “The Dynamic HomeFinder: Evaluating Dynamic Queries in a Real-estate Information Exploration System”. In: *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR ’92. Copenhagen, Denmark: ACM, 1992, pp. 338–346 (cit. on pp. 18, 24, 28).
- [Wis+98] Craig Wisneski, Hiroshi Ishii, Andrew Dahley, et al. “Ambient displays: Turning architectural space into an interface between people and digital information”. In: *International Workshop on Cooperative Buildings*. Springer. 1998, pp. 22–32 (cit. on p. 89).
- [Woo+11] Jo Wood, Donia Badawood, Jason Dykes, and Aidan Slingsby. “BallotMaps: Detecting name bias in alphabetically ordered ballot papers”. In: *Visualization and Computer Graphics, IEEE Transactions on* 17.12 (2011), pp. 2384–2391 (cit. on p. 35).
- [Woo05] Michael Wood. “Bootstrapped confidence intervals as an approach to statistical inference”. In: *Organizational Research Methods* 8.4 (2005), pp. 454–470 (cit. on p. 52).
- [Wor22] World Health Organization. *Vitamin A deficiency*. Web article <https://www.who.int/data/nutrition/nlis/info/vitamin-a-deficiency>. Last visited 2022-03-22. 2022 (cit. on p. 84).

- [Xio+22] Cindy Xiong, Elsie Lee-Robbins, Icy Zhang, Aimen Gaba, and Steven Franconeri. “Reasoning Affordances with Tables and Bar Charts”. In: *IEEE Transactions on Visualization & Computer Graphics* 01 (2022), pp. 1–13 (cit. on p. 37).
- [XWF17] Cindy Xiong, Lisanne van Weelden†, and Steven Franconeri. “The Curse of Knowledge in Visual Data Communication”. In: *DECISIVE: Workshop on Dealing with Cognitive Biases in Visualizations. IEEE VIS*. 2017 (cit. on p. 35).
- [YL14] Sybil Yang and Michael Lynn. “More Evidence Challenging the Robustness and Usefulness of the Attraction Effect”. In: *Journal of Marketing Research* 51.4 (2014), pp. 508–513. eprint: <http://dx.doi.org/10.1509/jmr.14.0020> (cit. on p. 22).
- [Yi+05] Ji Soo Yi, Rachel Melton, John Stasko, and Julie A Jacko. “Dust & magnet: multivariate information visualization using a magnet metaphor”. In: *Information Visualization* 4.4 (2005), pp. 239–256 (cit. on pp. 24, 28).
- [ZM04] Per-Henrik Zahl and Jan Maehlen. “Overdiagnosis in mammography screening”. In: *Tidsskr Norske Laege* 124.17 (2004), pp. 2238–2239 (cit. on p. 10).
- [Zaj01] Robert B Zajonc. “Mere exposure: A gateway to the subliminal”. In: *Current directions in psychological science* 10.6 (2001), pp. 224–228 (cit. on p. 35).
- [ZBK15] Yunfeng Zhang, Rachel KE Bellamy, and Wendy A Kellogg. “Designing information for remediating cognitive biases in decision-making”. In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM. 2015, pp. 2211–2220 (cit. on p. 35).
- [ZC07] Torre Zuk and Sheelagh Carpendale. “Visualization of uncertainty and reasoning”. In: *International Symposium on Smart Graphics*. Springer. 2007, pp. 164–177 (cit. on pp. 35, 36).

