This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3456381

Causal Priors and Their Influence on Judgements of Causality in Visualized Data

Arran Zeyu Wang, David Borland, Tabitha C. Peck, Wenyuan Wang, and David Gotz

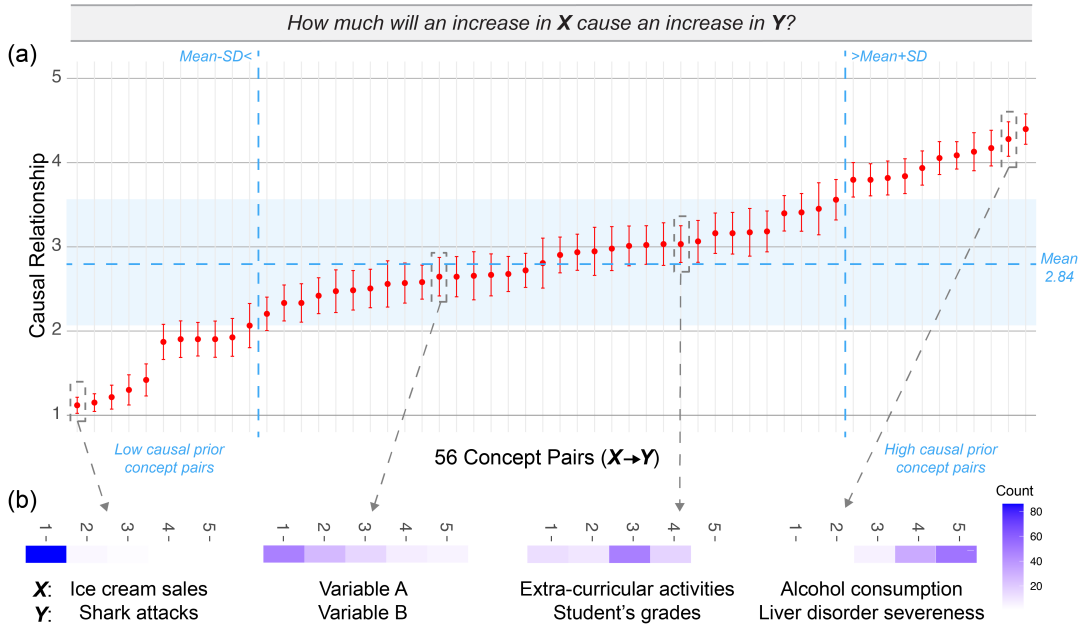

Fig. 1: Results from the first study in this paper of participant-rated causal relationships for 56 concept pairs curated from open-source datasets. Participants were asked questions in the form "How much will an increase in X cause an increase in Y?" for each X→Y concept pair. In (a), the Y-axis represents the participant-reported scores for concept causal relations (1 to 5: none to high). Each **concept pair** is placed in order by mean causal relation along the X-axis, showing 95% confidence intervals. The light blue horizontal band represents the mean score across all concept pairs ± one standard deviation (SD). The **vertical dashed lines** delineate concept pairs that we refer to as having either low causal priors (<mean-SD) or high causal priors (>mean+SD). Part (b) shows four example concept pairs from different parts of the causal prior spectrum. The heat maps show the number of participants in our study reporting each score on the 1-5 causal scale. As these results show, causal priors can vary widely across different concept pairs. In the second study in this paper (Section 3.4), we examine the impact of these causal priors on visualization interpretation.

Abstract— "*Correlation does not imply causation*" is a famous mantra in statistical and visual analysis. However, consumers of visualizations often draw causal conclusions when only correlations between variables are shown. In this paper, we investigate factors that contribute to causal relationships users perceive in visualizations. We collected a corpus of concept pairs from variables in widely used datasets and created visualizations that depict varying correlative associations using three typical statistical chart types. We conducted two MTurk studies on (1) preconceived notions on causal relations without charts, and (2) perceived causal relations with charts, for each concept pair. Our results indicate that people make assumptions about causal relationships between pairs of concepts even without seeing any visualized data. Moreover, our results suggest that these assumptions constitute causal priors that, in combination with visualized association, impact how data visualizations are interpreted. The results also suggest that causal priors may lead to over- or under-estimation in perceived causal relations in different circumstances, and that those priors can also impact users' confidence in their causal assessments. In addition, our results align with prior work, indicating that chart type may also affect causal inference. Using data from the studies, we develop a model to capture the interaction between causal priors and visualized associations as they combine to impact a user's perceived causal relations. In addition to reporting the study results and analyses, we provide an open dataset of causal priors for 56 specific concept pairs that can serve as a potential benchmark for future studies. We also suggest remaining challenges and heuristic-based guidelines to help designers improve visualization design choices to better support visual causal inference.

Index Terms—Causal inference, Perception and cognition, Causal prior, Association, Causality, Visualization

- *Arran Zeyu Wang, Wenyuan Wang, and David Gotz are with the University of North Carolina at Chapel Hill (UNC). E-mail: zeyuwang@cs.unc.edu, vaapad@live.unc.edu, gotz@unc.edu*
- *David Borland is with RENCI at UNC. E-mail: borland@renci.org*
- *Tabitha C. Peck is with Davidson College. E-mail: tapeck@davidson.edu*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

1 Introduction

Supporting causal inference for users is a foundational pursuit for visualization and visual analytics [5, 25, 30, 51, 65]. However, it is a challenging task and the assumption of causal relationships where there may be none can lead to significant misinterpretations, affecting decision-making processes across various domains [35].

Well-known analytical guidelines caution against conflating correlation with causation. However, when viewing visualizations (which

typically display correlative associations rather than causal relations), users often make assumptions regarding causality. Effectively communicating causal relationships, or the lack thereof, remains challenging for data visualizations [5]. Understanding how humans interpret causal relationships from visualized data is therefore crucial for effective visualization design.

Acknowledging that understanding relationships—causal or otherwise—is essential, this paper aims to gain insights into how humans infer causal relationships when they are consuming visualizations, a process that we refer to as *visual causal inference*. Previous studies have scrutinized how users examine the data presented in a visualization as well as the manner of visual presentation to understand how these factors influence users' visual causal inferences [27, 54, 62]. Concurrently, research into human perception has shed light on how it shapes the interpretation of visual information of associations and correlations [8, 41]. However, humans' visual causal inferences do not occur in isolation; they are made within a broader context that includes specific choices of tasks and may be influenced by cognitive effects.

Beyond these existing contextual factors, we theorized that underlying semantic causal priors—a person's preconceived notions about concepts and the underlying causal relationships between them [43]—play a critical role in the causal inferences drawn from charts. We therefore set out to investigate whether and how these underlying causal priors might affect the causal inferences people make (i.e., whether the same chart can lead to different interpretations under different causal priors).

This paper introduces two related studies aimed at answering the questions of (1) whether underlying causal priors influence causal inferences from visualizations, and if yes, (2) how these priors interact with the actual statistical associations depicted in visualizations presented with different chart types.

We first collected a corpus of concept pairs by selecting variables from widely-used datasets and created visualizations representing varying statistical association levels with three typical statistical chart types: scatterplots, line charts, and bar charts, representing continuous, temporal, and categorical data respectively. We then conducted two crowdsourced studies to: (1) measure the underlying causal priors for the concept pairs in our corpus as rated by participants when not seeing any visualization of data at all; and (2) examine the effects of the causal prior, visualized association, and chart type on participants' perceptions of causal relations in visualized data. Our results help provide a deeper understanding of how preconceived notions about causality combine with visual evidence to influence human cognition of causal inference. We analyze these results in light of previous studies that looked at human perception of causal inference in visualizations, and suggest actionable heuristic-based design guidelines for visual causal inference. Finally, we highlight some remaining challenges such as the need for a deeper understanding of the impact of specific chart and data types.

More specifically, the contributions of this paper include:

- **Empirical evidence and reference measurements of causal priors associated with concept pairs.** We report data from a study showing that people infer causal relationships between pairs of concepts absent of any other information or charts. Data from this study is provided as an open dataset of priors for 56 concept pairs that can serve as a benchmark for future studies.
- **Empirical evidence demonstrating the effect of causal priors on users' visual causal inferences.** We report data from a study showing that a user's assessment of causality when viewing a chart is influenced not only by the type of chart and the data being visualized, but also by the causal prior.
- **A model capturing the interaction between causal priors and other factors.** Based on empirical data gathered in our studies, we define a model that expresses the interaction between causal priors, visualized association, and chart type in influencing the user's reported strength of the causal relationship between visualized concepts.
- **A discussion of implications and heuristic-based design guidelines.** A detailed analysis of the study results is provided along with a discussion of the implications of these results on our understanding of visual causal inference. This includes a set of heuristic-based guidelines to help designers improve visual design choices to better support visual causal inference.

2 Background and Related Work

The research presented in this paper builds upon prior work in two broad areas of research: i) human inferences about causal relationships based on viewing visualizations of data and ii) the role of a person's prior context or knowledge in human cognition.

2.1 Human Perception in Visual Causal Inference

The term *causal inference* is typically used to refer to techniques for identifying and characterizing relationships between variables that are indicative of cause and effect [35]. The mathematical foundations of these approaches have been developed within the statistics and machine learning communities, where the techniques have been widely applied [35, 37, 46].

Motivated in part by these advances, causal inference has also drawn attention as an emerging topic in visualization research [5, 25, 30]. For example, a number of visualization techniques have been proposed to support specific causal inference workflows, such as event sequence or time-series analysis [10, 26, 58], paradox and illusion detection [44, 53, 57], algorithm interpretation [14, 24], and general exploratory tasks [20, 28, 55, 64]. See Borland et al. for a more detailed discussion on recent advances in developing visualization techniques that can directly support visual causal inference [5].

A number of these techniques are designed to provide graphical representations and controls over an underlying *statistical causal inference* process. Others, however, examine or support the ways in which humans cognitively draw inferences about causal relationships when looking at data visualizations. We emphasize that this idea of a user's *visual causal inference* is closely related to but distinct from statistical inference approaches. Visual causal inferences are ones made by a user based in part on their perception and interpretation of the data they see during an analytical task.

Critically, previous research has shown that users make visual causal inferences even when viewing basic statistical charts that may focus on correlation or other non-causal statistical measures (e.g., [28]). For this reason, the understanding of how human perception relates to the ways that people draw visual causal inferences has become an active topic of research [5].

For example, Xiong et al. found that different visualization types do not equally support causal interpretation [62]. For instance, their results suggest that bar charts may be more suggestive of causal relationships than scatter plots. The same research also explored the impact of specific visual encodings (as variants of the same basic chart type) and found that increasing levels of data aggregation were associated with increasing levels of perceived causality. Similarly, the same study suggests that within a given type of chart, lines and dots were viewed as more suggestive of causality than bars.

In other work, Kale et al. [27] employed mathematical psychology [19] and a causal support model to study causal inference perception with visualizations. Their model design supports comparisons between that ground truth and users' perceptions compared to several other works [28, 62]. They found that users' causal inferences deviated from the ground truth, with either overestimation or underestimation of the underlying ground truth causal relationships. The results of this study highlight the difficulty in precisely assessing the level of evidential support that a particular dataset offers for a specific causal explanation.

Kaul et al. found that users tend to infer causal relations when filtering data using variable constraints, employing simple bar charts and line charts to show correlation [28]. Moreover, they found that using counterfactual visualizations to explore different data subsets can partially mitigate these effects when causal relationships are unsupported by the data. Building upon these findings, Wang et al. developed a causality comprehension model for counterfactual-based visualizations [54]. They also reported results from a study that showed that visualizing

counterfactuals within static charts provides benefits to users at various levels of causal comprehension.

Taken together, these studies provide significant insights into how people infer causal relationships based on their perceptions of various elements of how the data is visually represented. This focus is well-motivated and the effects of human perception are important to understand. However, in practice, perception only plays one part in the broader cognitive process of drawing visual causal inferences. In this paper, we aim to begin addressing this gap by exploring how certain cognitive factors—specifically those related to assumptions made by users based on semantics associated with relationships between concepts—impact how people make visual causal inferences.

2.2 Human Cognition and Underlying Causal Priors

Human knowledge and assumptions about the world can be complex and varied, informed by each person's lived experience and in response to various forces such as culture and training [31, 49]. This is particularly true in how people understand language, as reflected in the results from several studies that have explored how people assume relationships for and between concepts in human communication. For example, research has long shown that different English verbs communicate different levels of implied causality in short phrases, a result that has been replicated in several studies [43]. This same phenomenon was more recently replicated by Ferstl et al. in research that also produced an annotated corpus of 305 English verbs based on data gathered during the study [12].

Demonstrating that implied causality is not unique to English, Goikoetxea et al. studied the implied causality communicated by 100 interpersonal verbs in Spanish [15]. Like the previously mentioned studies of English verbs, this study found that different verbs implied different types and degrees of causal relationships. Moreover, they found that the results were similar for both adult and child participants.

These findings from the psychology and human behavior literature have led to related attempts within the NLP community to identify implied causal relationships within text corpora. For example, Riaz and Girju proposed an NLP model to recognize causality in verb-noun pairs by predicting words' semantics [42]. In another example, Salim et al. developed the *BeliefMiner* system to extract causal graph networks from unstructured text and used the approach to explore causal illusions about climate change [44].

These efforts demonstrate that people tend to implicitly associate causal relationships with short phrases or individual words. These implied causal properties exist even before those concepts might appear in a visualization, for example as part of an axis, legend, or some other chart annotation. We refer to this idea—that a causal relation can be implied by the use of words in and of themselves, even before those words are used within a visualization—as *causal prior*.

In this way, causal priors are similar to other factors describing prior experience or knowledge that have been shown within the visualization literature to influence how people interpret charts [31, 49]. For example, research has shown that both background knowledge and individual differences can influence visualization comprehension [39]. The efficiency of specific visualization designs depends in part on characteristics of the potential audience such as internal representations [32], graphical literacy [13], color understanding [45, 48], mental models [33], diverse backgrounds [21, 36, 56], experience levels [6, 18], and cross-domain knowledge gaps [34]. Most recently, Xiong et al. reported results that show that users' personal beliefs can bias their estimates of visualized correlations [63].

Yet despite this breadth of research exploring how users' priors influence their interpretations of visualizations, to our knowledge, there have been no previous studies that have examined the effect of causal priors on how people perform visual causal inference using visualizations. In this paper, we both: (1) develop a corpus of concept pairs annotated with causal priors; and (2) study how those priors influence users' visual causal inference behavior in different visualization contexts.

3 Methodology

To investigate the existence and influence of causal priors, we designed and conducted two distinct yet complementary empirical studies by recruiting users from Amazon's Mechanical Turk to assess human cognition of causal inference. Both studies were approved by the [Redacted] Institutional Review Board. The studies are designed to investigate potential visual and non-visual impact factors on human cognition when conducting causal inference

3.1 Terminology

Here we provide definitions for three key terms used throughout the paper:

- **Causal prior:** The causal relation level of a directed pair of concepts, as indicated by participants in Study 1 without any visualization stimulus. Values exist on a scale from 1 (no relation) to 5 (high causal relation).
- **Visualized association:** The statistical association level between a pair of concepts that is represented in a visualization. This association level is controlled by the visualization stimulus generation process used for Study 2 (see Section 3.4.2), comprising five levels from 0 to 1 with intervals of 0.25.
- **Perceived causal relationship:** The degree of causal relationship between concept pairs perceived by a user when viewing a visualization. In this paper, the values refer to our results from Study 2. As with Study 1, values exist on a scale from 1 (no relation) to 5 (high causal relation).

3.2 Hypotheses

In this paper we aim to explore the potential relations among causal priors, visualized associations, and perceived causal relationships. Based on this goal, we hypothesized that:

- **H1:** People have underlying causal priors reagarding the assumed strength of causal relationships for concept pairs.
- **H2:** Causal priors affect the perception of causal relationships from charts.
- **H3:** Visualized associations affect perceived causal relationships from charts.
- **H4:** The nature of the effects of H2 and H3 can, in part, be explained by the *disagreement* between causal priors and visualized associations.
- **H5:** The impact of causal priors and visualized associations on perceived causal relationships vary by chart type.

Through these hypotheses, our study aims to contribute to the broader discourse on graphical perception, specifically in the context of how human cognition interprets causality from visual data. The exploration of these hypotheses will not only enhance our understanding of cognitive processes, but also inform the design of more effective visualization-guided exploration systems for causal inference.

3.3 Study 1: Examining Causal Priors for Concept Pairs

Study 1 examines inherent beliefs about causality by measuring the prior causal relationship ascribed to a given directed concept pair. Participants in this study were asked to assess the strength of causal relationships between concept pairs on a 5-point Likert scale based solely on the concept names, without any visualizations. Instances of the questions are provided in Section 3.3.2. Participants also provided their confidence in their causal strength ratings on a 5-point Likert scale. Each participant saw each concept pair in a randomized order. Study 1 was designed to address **H1**, and provide the causal priors used to assess **H2-H5** in Study 2.

3.3.1 Participants

A cohort of 100 participants was recruited for Study 1, with an average engagement time of 12 minutes, leveraging the Amazon MTurk platform using CloudResearch with at least a 95% approval rating and IP addresses from the United States and Canada. These selection criteria aimed to ensure a certain level of homogeneity in cultural and educational background, which can influence cognitive processing. We excluded eight participants who failed at least two random attention checks (these excluded participants spent less than 3 minutes on average), and analyzed data from the remaining 92 participants, resulting in a 92% acceptance rate. The final study 1 participants included 65 men and 27 women, ranging from 24–58 years of age.

3.3.2 Dataset and Stimuli

Our empirical investigation employed a self-collected corpus comprising two related components: concept pairs, and causal questions.

Concept Pairs: A total of 56 concept pairs were curated for our study. We first collected 50 pairs of variable names from ten widely used open-source datasets, including the *Titanic*, *Census Income*, and *Heart Disease* datasets. These datasets were selected from the "Popular Datasets" list on the UCI data repository [1] and the "Trending Datasets" list on Kaggle in an attempt to cover a diverse range of concepts. Concept names for selected variables were determined by the authors, and then reviewed by five individuals with a maximum education level of a high-school diploma or equivalent to confirm that the names could be interpreted by the general public. We slightly revised the concept names from the original datasets based on their feedback, e.g., "computer speed" was used to simply replace "RAM size" which can be more complex to understand. These 50 pairs were supplemented with an additional 5 concept pairs, such as *divorce rate in Maine* and *per capita consumption of margarine*, that were selected from Spurious Correlations [52] to include pairs that we expected to be rated as having little or no causal relationship. Finally, we also included *Variable A* and *Variable B* to establish a baseline for a non-semantic concept pair.

For each concept pair, we separated them into one causal factor and one outcome: we employed the widely used factor-outcome relations of the 50 pairs from datasets (e.g., studying time can be a causal factor to impact students' grades), followed the original order of the 5 pairs from spurious correlations [52], and determined *Variable A* as the causal factor for the non-semantic concept pair. The causal factor is denoted as $\mathbf{X}$ (e.g., *Alcohol consumption* in Figure 1 (b)), and the outcome as $\mathbf{Y}$ (e.g., *Liver disorder severeness* in Figure 1 (b)). This assigned a causal direction to each concept pair, denoted $\mathbf{X}\rightarrow\mathbf{Y}$, and participants were only asked to consider the causal relationship for each concept pair in the specified direction.

Study Questions: Each concept pair was accompanied by a question designed to determine each participant's rating of the strength of the causal relationship between the two concepts, e.g., "How much will an increase in **computer speed** cause an increase in **laptop price**?". All questions were asked with an assumption of a positive causal relationship, i.e., how much does an increase in one cause an increase in the other. The questions were designed to provide sufficient background context to answer the questions, including phrases such as "in the Titanic sinking" or "in a cancer surgery." Further, we validated the questions with five individuals who hold high-school level or lower educational degrees to ensure the questions would be understandable for general users. Participants were asked to rate the strength on a 5-point Likert scale from 1 (no causal relationship) to 5 (high causal relationship). Participants were also asked to assess their confidence in each rating on a 5-point Likert scale from 1 (low confidence) to 5 (high confidence). In addition, three simple attention check questions were added, e.g., "Please choose the answer of 3+7.", displayed in random order and position within the first 1/3, second 1/3, and last 1/3 of all questions.

Details of concept pairs, used datasets, and study questions are all available in the supplemtns on OSF.

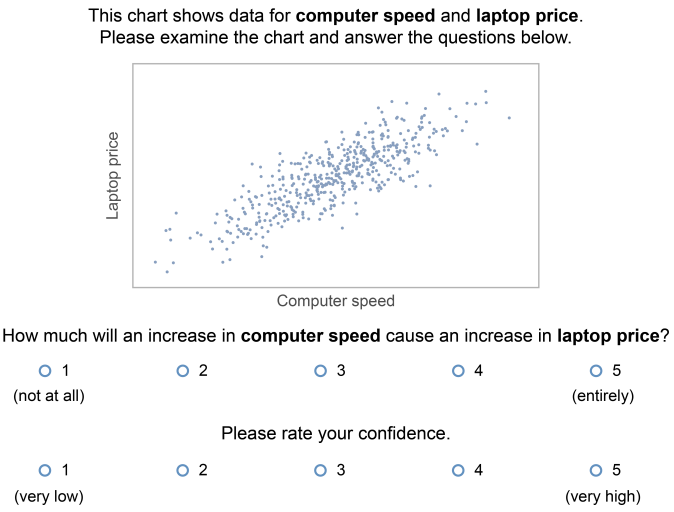

Fig. 2: An instance of a visualization stimulus combined with a causal question and a confidence question from Study 2.

3.4 Study 2: Exploring the Effects of Visualizations on Causality Perception

Study 2 aimed to investigate how the perception of causal inference is influenced by the causal priors from Study 1 as well as varying levels of visualized associations in charts. The tasks in Study 2 mirrored those of Study 1, however a visualization for the concept pairs was also provided. Participants reported perceived causal relationships based on these visualizations, which varied by three chart types (line, bar, scatterplot) and different levels of visualized correlative association (ranging from 0 to 1). Duplicate charts or questions were avoided, and neither the chart type nor the association level was repeated consecutively during the study to avoid potential biases and learning effects. A representative chart example is provided in Figure 2. Study 2 was designed to address **H2-H5**.

3.4.1 Participants

For Study 2, we recruited a new cohort of 250 participants again using Mechanical Turk and the same user restrictions as Study 1. Participants who completed Study 1 were ineligible to complete Study 2. 21 participants were excluded from the final analysis due to failing at least 2 out of 3 of the attention checks, resulting in a 91.6% acceptance rate. The final Study 2 participants included 138 men, 87 women, and 4 non-binary, ranging from 22–64 years of age. The experiment took approximately 15 minutes.

3.4.2 Visualization Stimuli

In Study 2 we employed the same set of concept pairs and study questions as Study 1. We classified the concept pairs into three categories based on their inherent data types: time series, categorical, and continuous. To balance study design simplicity with a reasonably representative sample of commonly-used visualizations, we chose three of the most widely applied chart types [38] for the three data types: line charts for time series, bar charts for categorical, and scatterplots for continuous. These three chart types have also been commonly included in various types of graphical perception studies such as visualization comprehension [39], counterfactual visualization [54], color perception [48], and causality visualization [62]. The breakdown of concept pairs per chart type is as follows. The line chart included 16 concept pairs from open-sourced datasets, the non-semantic concept pair, and the five spurious concept pairs [52] for a total of 22 line chart trials. The bar chart and scatterplot each included 17 concept pairs from open-sourced datasets, for a total of 17 trials each.

To explore the impact of causal priors on perceived causal relationships, we displayed data in each chart type using five distinct association levels (Figure 3):

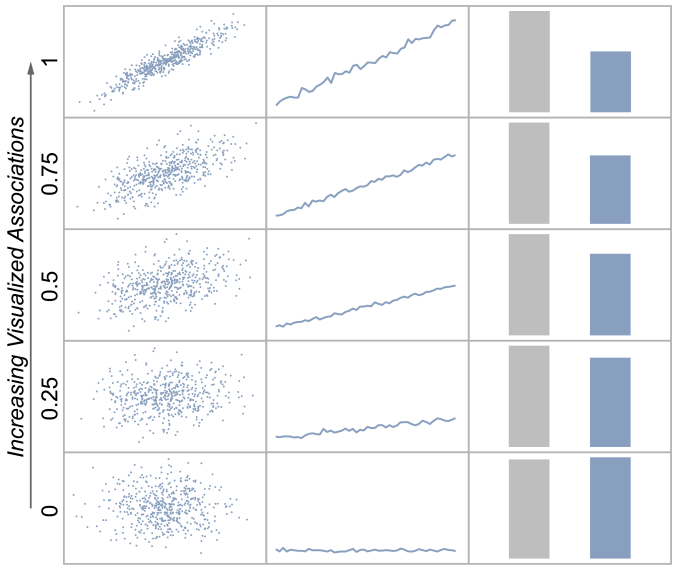

Fig. 3: This figure shows illustrations of the three chart types and five visualized association levels employed in Study 2. Left to right: scatterplots, line charts, and bar charts. Bottom to top: low association (around 0) to high association (around 1).

Scatterplots: We adjusted the multivariate covariance, with values spanning from 0 (weak association) to 1.0 (strong association), at intervals of 0.25. Following existing research on correlation judgments in scatterplots [9,40,41], we only tested positive correlations to simplify the study's complexity and avoid the impact of negative correlations.
Line charts: Previous studies on line charts [8, 23, 59, 60] show that the effectiveness of human perception is highest at around a 45° trend. We therefore generated 5 evenly spaced association levels from 0 to 1 with a linear mapping of *slope values* (i.e., the average y to x ratio of a line chart) from 0° (no association) to 45° (strong association).
Bar charts: We chose simple two-bar charts following the methodology from prior work [62] by manipulating the difference ratio between the bar heights to control the association difference. Differences for the 5 association levels ranged from no difference at 100% to 100%, to a maximum difference of 100% to 50%, with evenly spaced intervals in between.

Additionally, the visualized associations across all three chart types were generated by adding random noise with a variance in the open range (0, 0.1) to avoid duplicated charts and unrealistic cases, e.g. a scatterplot with a covariance of 1.0 that would form a straight line. The upper threshold of random noise (0.1) was derived from previous work, in which differences of around 10% were often chosen as a threshold for correlation or slope value comparisons for simple statistical charts [40, 50]. In this way, the difference between two neighboring association levels was close to 25%. In the following analysis, we refer to these charts using their base association values of 0, 0.25, 0.5, 0.75, and 1.

These values are shared across all chart types. However, we acknowledge that the chart types are each associated with specific data types. This results in different distributions across chart types and makes interpretations of the impact of chart type more difficult. Please see Section 5.4 and Section 7.1.4 for more discussions.

3.5 Procedure

Both Study 1 and 2 shared a similar procedure. The studies were administered using Qualtrics. Before beginning the study, participants completed an informed consent form, acknowledging their understanding of the study's purpose and their rights as participants. Participants completed 56 randomized trials, one for each concept pair, and three attention check questions, for a total of 59 trials. For each concept pair, participants were asked, "How much will an increase in **X** cause an increase in **Y**?" where **X** and **Y** were the concept pair. Participants answered on a 5-point Likert scale from 1 (not at all) to 5 (entirely). Participants were then asked, "Please rate your confidence" on a 5-point Likert scale from 1 (very low) to 5 (very high).

For Study 2, each concept pair question also included a visualization (Figure 2). To avoid learning and fatigue effects, each participant saw each concept pair once, with its associated chart type (see Section 3.4.2) showing a randomly chosen visualized association level. Visual association levels were evenly distributed for each participant such that each participant saw each association level 10-11 times. Consecutive trials prohibited repeated chart types and visual associations. E.g., a bar chart was never shown directly after another bar chat, and a visual association with level 1 was never shown directly after a visual association with level 1.

Across all participants, each association level for each concept pair visualization was shown a total of 50 times. After excluding participants who failed attention checks, the total number of completed trials for each concept pair and association level varied from 44 to 48 trials each. After completing the 59 concept pair and attention check trials, participants were asked to provide any feedback. In the end, we noticed users the data shown in our study were synthetic and did not depict any real-world relationships. On average, Study 1 took 12 minutes and Study 2 took 15 minutes.

3.6 Analysis Overview

In Section 4 and Section 5 we discuss significant results and statistical analysis for the two studies based on the independent factors considered in this paper with 95% bootstrapped confidence intervals (± 95% CI) for fair statistical communication [11]. The study data, analysis code, and the open dataset of priors are available at OSF.

4 Study 1 Analysis

To test **H1** we performed a repeated measures ANOVA on the 56 concept pairs. Our analysis reveals a significant effect of different concept pairs on causal rating ($F(55,5152) = 53.41, p < .0001, \eta^2 = .36$), indicating that people assume different causal relationship strengths for different concept pairs.

We then calculated the average value of users' reported causal relationships for each concept pair to serve as a key factor, causal priors, used in the analysis of Study 2. Figure 1 shows these average values with 95% confidence intervals for each concept pair.

The mean causal relationship strengths indicate an obvious difference in the score distribution across different concept pairs (see Figure 1 (a)). In addition, the relatively tight 95% confidence intervals indicate good agreement across users for the causal relationship ratings. The 5 spurious correlations also had the five lowest causal priors (e.g., see *ice cream sales* and *shark attacks* in Figure 1), further justifying the design. The results from Study 1 therefore support **H1**, indicating that people have underlying causal priors between specific concept pairs.

5 Study 2 Analysis

Given the experimental support for underlying causal priors shown in Study 1, we include the average causal relationship strength per concept pair from Study 1 as a variable, *causal prior*, to the analysis in Study 2.

A generalized multiple linear regression model was used to analyze the *perceived causal relationship* based on *causal prior*, *visualized association*, and *chart type*. This approach can be better balanced for the distribution differences between chart types as mentioned in Section 3.4.2. The regression of *causal prior*, *visualized association*, and *chart type* on *causal relationship* was statistically significant, ($F(11,14404) = 257$, $p < .0001$, $R^2 = .17$). Table 1 summarizes the significance results from the linear regressions for *perceived causal relationship*, which are presented in more detail in the following sections.

5.1 The Impact of Causal Priors on Perceived Causal Relationships

Our results support **H2**: Causal priors affect the perception of causal relationships from charts. *Causal prior* ($\beta = 0.56$, $p < .0001$) significantly affects *perceived causal relationship*. This supports that higher *causal priors* were associated with higher ratings for the strengths of *perceived causal relationships*.

Table 1: Linear regression results from Study 2 for *perceived causal relationship*. Significant effects are indicated in bold and the corresponding rows are highlighted in green. Note that line charts are the reference of the other two chart types in the model so it cannot be shown below.

	β	Std. Error	t-value	p-value
(Intercept)	0.91	0.07	11.50	**< .0001**
Causal Prior	0.56	0.05	10.14	**< .0001**
Visualized Association (VA)	1.74	0.12	13.39	**< .0001**
(Chart Type) Bar	0.36	0.19	1.94	**.05**
(Chart Type) Scatterplot	0.61	0.16	3.77	**.0001**
Causal Prior * VA	-0.14	0.04	-3.13	**.001**
Bar * Causal Prior	-0.07	0.06	-1.18	.23
Scatterplot * Causal Prior	-0.06	0.05	-1.21	.22
Bar * VA	-1.07	0.30	-3.46	**.0005**
Scatterplot * VA	-1.03	0.26	-3.91	**< .0001**
Bar * Causal Prior * VA	0.24	0.10	2.28	**.02**
Scatterplot * Causal Prior * VA	0.09	0.08	1.04	.29

5.2 The Impact of Visualized Associations on Perceived Causal Relationships

Our results support **H3**: Visualized association affects perceived causal relationships from charts. *Visualized association* ($\beta = 1.74$, $p < .0001$) significantly affects *perceived causal relationship*. The result indicates that higher *visualized association* is associated with higher ratings for the strengths of *perceived causal relationship*.

5.3 Interaction between Causal Priors and Visualized Associations

Our results support **H4**: The nature of the effects of **H2** and **H3** can, in part, be explained by the disagreement between causal priors and visualized associations.

A significant interaction was found between *causal priors* and *visualized associations* ($\beta = -0.14$, $p = .001$). Post-hoc analysis of simple slopes was performed at the *causal prior* mean $\pm$ standard deviation, in which $mean - SD = 2.01$, $mean = 2.84$, and $mean + SD = 3.66$. Note that the slope differences here are distinct from the *slope values* that were used to define associations in our line chart stimuli (Section 3.4.2). In this instance the slope differences from the simple slope analysis results indicate how the influence of *visualized association* changes as the *causal prior* changes.

Significant differences were found between all slope pairs. There was a significant difference between *mean-SD* (M=1.26, SE=0.04) and the *mean* (M=1.14, SE=0.03), ($t(14052) = 4.01$, $p = .0002$, $r = .82$), a significant difference between *mean* and *mean+SD* (M=1.01, SE=0.05), ($t(14052) = 4.01$, $p = .0002$, $r = .84$), and a significant difference between *mean+SD* and *mean-SD*, ($t(14052) = 4.01$, $p = .0002$, $r = .94$). See Figure 4 for the visualization of simple slopes.

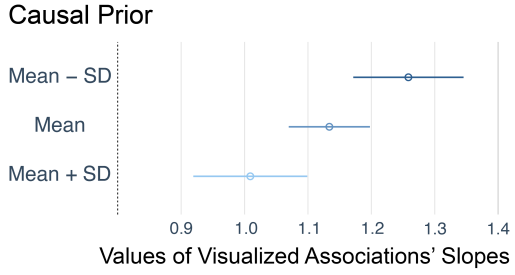

Fig. 4: Results of the simple slope analysis between causal priors and visualized associations. It shows that visualized associations would have a lower impact for higher causal priors. From top to bottom, the dark blue line shows the result for the mean-SD causal prior (2.01), blue shows the result for the mean causal prior (2.84), and light blue shows the result for the mean+SD causal prior (3.66).

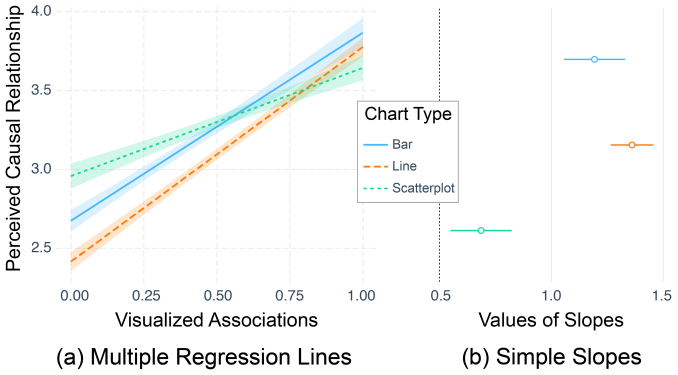

Fig. 5: Results of the simple slope analysis between chart types and visualized associations. The values hint scatterplots have a lower impact on perceived causal relationships compared to bar charts and line charts. On the left (a) are the simple slope regressions, and on the right (b) are slopes at each interval.

The results show that the *mean-SD* has a significantly steeper slope than both the *mean* and the *mean+SD*. This difference indicates that the *visualized association* has a greater influence on the *perceived causal relationship* at lower *causal prior*. As *causal prior* increases, the influence of *visualized association* decreases. Though, for all cases, the slope is positive indicating that *visualized association* positively impacts *perceived causal relationship*.

5.4 The Impact of Chart Type

Our results may support **H5**: The impact of causal priors and visualized associations on perceived causal relationships vary by chart type. See Figure 5. Significant interactions between *chart type* and *visual association* were found. Post hoc analysis was performed with estimated marginal trends comparing each *chart type* pairwise, with Bonferroni corrections applied.

The Bar Chart (M=1.19, SE=0.07) and the Line Chart (M=1.36, SE=0.05) both had significantly steeper slopes compared to the Scatterplot (M=0.68, SE=0.07), ($t(14050) = 5.126$, $p < .0001$, $r = .96$) and ($t(14050) = 7.902$, $p < .0001$, $r = .98$) respectively. This result indicates that when visualizing bar charts and line charts, the *visualized association* had a greater influence on the *perceived causal relationship* than for scatterplots.

However, it is important to note that this observed impact may not be due only to the chart type. In our study, each chart type was associated with a particular data type (time series, categorical, or continuous), such that only one chart type was shown to users for each concept pair. Therefore any differences between chart types may be due, at least in part, to differences in data types. We therefore conclude that our results can only partly support **H5**. Section 7.1.4 provides further discussion on this limitation.

5.5 Confidence

Similar to the above analysis, we found the generalized linear regression on *confidence* was statistically significant, ($F(11, 14404) = 24.51$, $p < .0001$, $R^2 = .02$), with a significant interaction effect of *causal prior* and *visualized association* ($\beta = -0.27$, $p = .0002$). This suggests a larger disagreement between *causal prior* and *visualized association* may lead to lower subjective *confidence* and vice versa. Please see the OSF supplements for a more comprehensive reporting of these results.

6 Modeling Causality Perception

In addition to the significant results reported in Section 4 and Section 5, we report additional data patterns on causal priors observed in Study 1 in Section 6.1, differential data patterns observed in Study 2 in Section 6.2, and a model for perceived causal relationship in Section 6.3.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3456381

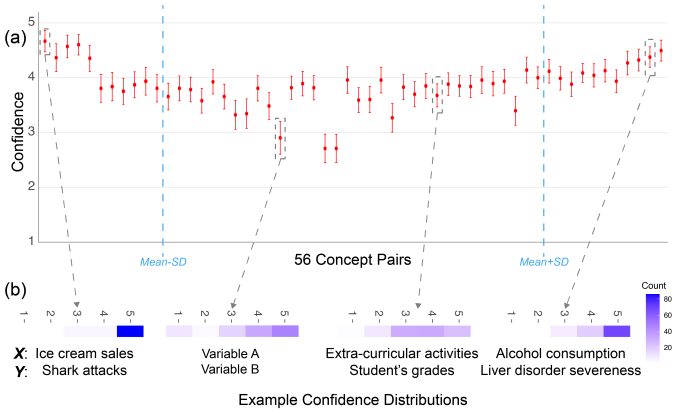

Fig. 6: Results of user-reported confidence from Study 1. Concept pairs are ordered with increasing causal prior. The cyan lines indicate mean ± SD cyan lines, as with Figure 1.

6.1 Causal Prior Data Patterns

Figure 1 (a) indicates that users' reported causal relationships for concept pairs are more tightly distributed, exhibiting smaller 95% confidence intervals, for more extreme causal priors with low or high values. For example, as shown in Figure 1 (b), when examining the causal relationship between *ice cream sales* and *shark attacks*, 86 users out of 92 chose a score of one, and for *alcohol consumption* and *liver disorder severeness*, 53 users chose a score of five and 32 a score of four.

Figure 1 (a) also shows that the causal relationship scores are more uniformly distributed, exhibiting larger 95% confidence intervals across the five rating levels, for concept pairs with intermediate causal priors, i.e., within the range of mean ± SD. For example, see the results between *variable A* and *variable B* in Figure 1 (b).

Interesting data patterns also appear in participant-reported confidence from Study 1. Figure 6 (a) indicates the confidence score distributions per increasing average causal relationship order of concept pairs, and Figure 6 (b) shows the confidence distribution of the same example cases.

These results indicate that for concept pairs with both very low or very high causal priors , the confidence is distributed in the high-value range. For example, in Figure 6 (b), more than 80 users scored five in confidence for *ice cream sales* and *shark attacks*, 64 users scored five and 18 scored four for *alcohol consumption* and *liver disorder severeness*.

But for concept pairs with intermediate causal priors, especially for those between mean ± SD, users' confidence becomes obviously lower. E.g., see *extra-curricular activities* and *students' grades* in Figure 6 (b)).

6.2 Exploratory Analysis for Differential Data Patterns

We then analyzed data patterns from Study 2 using two differential measures that calculate the difference between reported results from users who see visualizations in Study 2 with the causal priors. First, differential perceived causal relationship, Δ_{caus}, is defined as follows:

$$\Delta_{caus} = PCR - CausalPrior, \tag{1}$$

where *PCR* is the perceived causal relationship from Study 2.

A second measure, differential confidence (noted as Δ_{conf}), captures the corresponding difference in confidence as defined below:

$$\Delta_{conf} = PConf - ConfPrior \tag{2}$$

In this way, we can use these two measures to assess to what extent visualizations may lead to differences in users' responses compared to those reported without seeing visualizations. A positive Δ_{caus} value means that being exposed to the corresponding visualized data leads to an increased perceived causal relationship while a negative value refers to visualized data leading to a decreased perceived causal relationship. A positive or negative Δ_{conf} value can be interpreted similarly with respect to the difference in reported confidence.

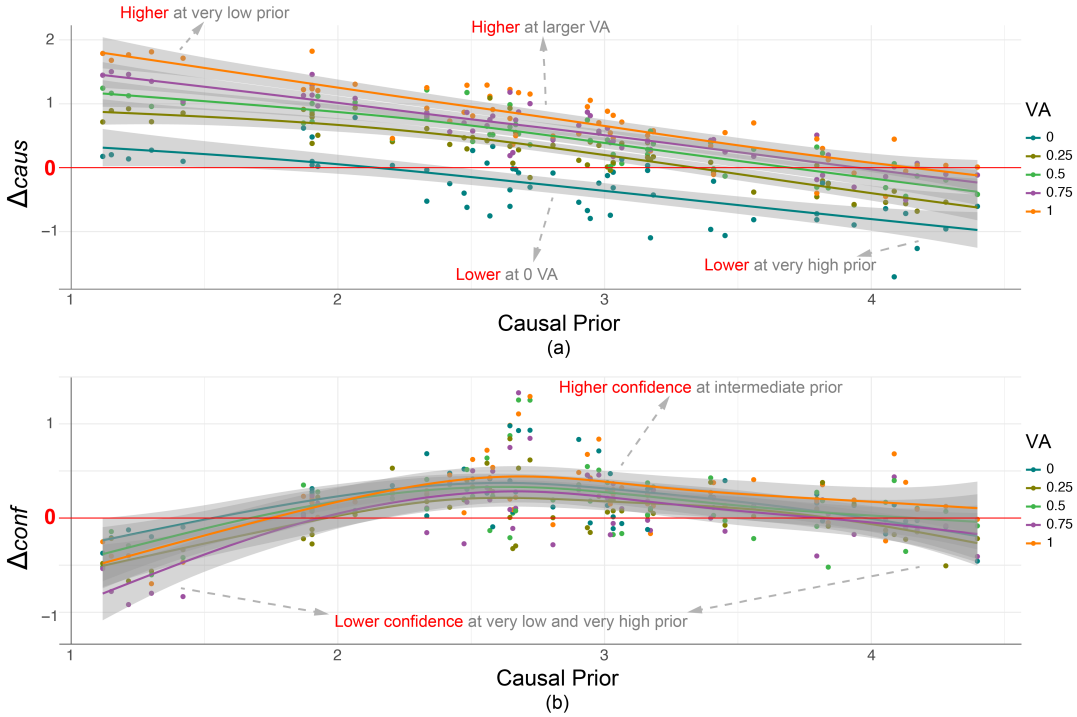

Fig. 7: Overall results of Δ_{caus} (a) and Δ_{conf} (b), as defined by Equation 1 and Equation 2. The x-axis is causal prior values of ordered concept pairs following Figure 1. Each point in this figure denotes an average value of Δ_{caus} or Δ_{conf}, which is the mean difference, for each concept pair. The trendlines are fitted per visualized association with generalized additive models [61] using Δ_{caus} or Δ_{conf} calculated via all user-reported results in all concept pairs. The colors denote different levels of visualized association (VA) from 0 to 1, at 0.25 intervals. (a) hints that people intend to judge higher causal relationships for concept pairs with very low causal priors, and vice versa. (b) shows people have lower confidence in concept pairs with both very high and low causal priors.

Figure 7 shows the overall results for Δ_{caus} and Δ_{conf} grouped by visualized association and plotted as a function of causal prior. Given this representation, two patterns emerge that can provide further insights into our findings between causal prior and visualized associations.

First, Figure 7 (a) suggests that visualizations may increase users' perceived causal relationships for concept pairs with low causal priors (on the leftmost side), while also decreasing users' perceived causal relationships on concept pairs with high causal priors (on the rightmost side). We note, however, that this moderating effect may be related to a "regression to the mean" effect [2].

Second, Figure 7 (b) reveals that users' confidence may be impacted by visualized data in a way that is different from δ_{caus}. For concept pairs with both very low and very high causal priors, users seeing visualizations tended to report lower confidence compared to users who only saw concept pairs. However, for concept pairs with more intermediate causal prior, we found users who saw visualizations, no matter the visualized association levels, became more confident compared to those who only saw concept pairs. Overall, these results suggest that for concept pairs with middle causal priors, the impact of visualizations could be higher than those extreme cases.

6.3 Modeling Differential Perceived Causal Relationship

The results from Section 5.3 and Section 6.2 indicate that both causal priors and visualized associations impact perceived causal relationships. Here we present a model designed to help characterize this relationship.

First we calculate the difference between the visualized association (denoted as *VA* for the remainder of the paper, not to be confused with *VA* as an abbreviation of *visual analytics*) and causal prior, $\Delta_{VA,CP}$:

$$\Delta_{VA,CP} = VA - normalize(CausalPrior), \tag{3}$$

where the normalize function maps the causal prior from [1, 5] to [0, 1] to match the range of *VA*.

Figure 8 shows comparisons of the distributions of $\Delta_{VA,CP}$ to Δ_{caus} in (a) and Δ_{conf} in (b). The ranges of Δ_{caus} and Δ_{conf} were normalized to [0, 1] for easier comparison in this figure.

In Figure 8 (a), the densities clustered around the coordinate origin indicate that when $\Delta_{VA,CP}$ is near 0, Δ_{caus} also tends to be near 0. Furthermore, the trendline shows $\Delta_{VA,CP}$ exhibits a close to linear relationship with $Delta_{caus}$, suggesting a positive disagreement would lead to users' higher judgment of causality and vice versa, which aligns with our findings. Besides, we fit a linear regression model to the distribution between $\Delta_{VA,CP}$ and $Delta_{caus}$, with the result:

$$\Delta_{caus} \sim 0.37 * \Delta_{VA,CP}, \tag{4}$$

with an intercept at 0.06, $p < .0001, R^2 = .22$. This equation indicates that, on average, the difference between visualized association and causal prior will contribute 0.37 of the difference in the final perceived causal relationship. Further, beyond the overall trend that can be captured by this model, we noticed that there exist more complex data patterns. For example, we found some data points are clustered at around *(-1.0, -1.0)* and *(1.0, 1.0)* indicating an extreme disagreement may have a more severe impact than the general data pattern. We also found a densely clustered pattern near *x=0*, suggesting many users still keep their ratings near the original causal priors.

In Figure 8 (b), however, the densities cluster at different levels of Δ_{conf}, indicating that we cannot simply estimate how much users' confidences may change using the differences between visualized association and causal prior. This aligns with the differential patterns of confidence discussed in Section 6.2 and Figure 7 (b). Meanwhile, it's obvious that the densely clustered data points are mostly above zero and near horizontal lines, possibly indicating visualizations are more likely to increase users' confidence and the visualized associations may have a lower impact. These initial patterns indicate a need to better justify them in future research.

7 Discussion

Our study investigated the impact of underlying causal priors and visualized associations on causal strength ratings between concept pairs. We find that human cognition of causal inference is impacted by both preexisting causal priors and the displayed visualizations.

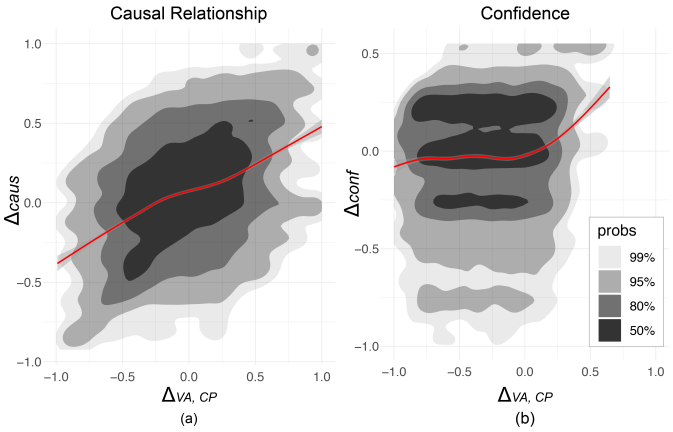

Fig. 8: Relationships between $\Delta_{VA,CP}$ and Δ_{caus} (a) and Δ_{conf} (b). The x-values denote the values of $\Delta_{VA,CP}$, while y-values show how much Δ_{caus} and Δ_{conf} change with respect to x. The heatmaps show the densities of all data points in each graph; both charts use the gray color scale shown on the bottom-right, indicating the proportion of the number of points shown within each color. The trendlines were fit with generalized additive models [61] on 95% confidence intervals.

7.1 Reflections on Previous Findings

Our results indicate many insights and relations related to underlying causal priors between concept pairs, associations conveyed via visualization, chart types, and human cognition of causal inference. The findings complement observations from past work focusing on related elements of visualization usage and visual causal inference.

7.1.1 Underlying Causal Prior in Mind

The first major finding indicates that people have preexisting causal priors (see Section 4). This insight aligns with the results from previous studies, such as Ferstl et al. [12] and Goikoetxea et al. [15], which found that there exist implicit causality norms from particular corpora of English and Spanish verbs respectively. Our results differ in that we gain insights between concepts represented by noun phrases, and because our public corpus may be more applicable for certain analytical tasks as most of the tested concepts are gathered from meaningful variable pairs from widely applied datasets.

7.1.2 Association in Visualizations

Furthermore, our results demonstrate that the correlations broadly represented in visualizations significantly impact human causal inference. The analysis provided further evidence built upon existing insights on how correlations in visualizations can impact the human perception of causal relations [8, 22, 29, 40, 41]. However, due to the significant impact of causal priors, we found that causal inference can be more complicated than just correlation estimation, thus the perceived causal relationship cannot be simply modeled with either a linear or log-linear model of visualized correlations such as those employed by many correlation estimation studies [22, 29, 41].

7.1.3 Estimation of Causal Relationships

We found that causal priors have a significant impact on the cognition of causal inference from visualizations. Although there exists limited previous research focus on priors in visual causal inference, our findings align with existing studies that focus on underlying relevant and background knowledge combined with correlation. For example, the impact of underlying causal priors shows a similar tendency to many existing studies which found that underlying background knowledge significantly impacts people's understandings of visualizations [32–34]. Additionally, this impact also aligns with Xiong et al.'s findings that people's beliefs impact their estimation of correlations [63].

Our results further support an important insight from Kale et al. [27]. Their study found that users' perception of causal inference tends to overestimate or underestimate the ground truth causal relations built upon the causal support model. As shown in Section 6.2, our findings align with their insight and provide evidence that for concept pairs, such cases more often appear with either very low or very high causal priors. This insight may also align with psychology phenomena such as the regression to the mean effect [2], which states that human predictions on extreme values, whether high or low, tend to be near the intermediate. However, our study is not designed to test psychological effects, further experiments will need to be conducted to validate these assumptions. For intermediate causal prior concept pairs, users' causality perception is largely impacted by visualizations. Additionally, our findings also show evidence that users' confidence varies based on concept pairs with different causal priors and visualized associations.

7.1.4 Chart Type

Finally, our results suggest that chart type may impact causal inference. We found that the perceived causal association for bar charts and line charts had a greater difference from the causal prior than for scatterplots, which aligns with prior studies [54, 62]. Xiong et al. [62] explained these differences could be due to the different aggregation levels of these charts. However, unlike their study, which tested different chart types for the same datasets, each chart type in our study was only used for a particular data type (e.g., scatterplots for continuous data). As a result, only one chart type was used for each concept pair. This limits our ability to disentangle the effects of chart type from the potential effects of data type. Future studies should assess how data type, chart type, and other potential factors

7.2 Implications for Design

In the light of existing perception experiments [27,28,54,62], our study confirms many insights and goes further in interpreting the cognition process of causal inference by connecting underlying causal priors and visualized associations. The results further lead to heuristic implications for the guidelines of future visual causal inference research and system design, resulting in the following key points.

Causal priors: Designers should recognize that users have pre-existing causal priors for many causal relationships. For specific analytical usage scenarios that may share a large number of similar variables, designers can collect analysts' underlying causal priors for those concept pairs in advance to better guide their interpretation of visualizations. This understanding could lead to the development of adaptive visualization systems that tailor the presentation of data based on the user's established causal prior knowledge base. However, there remain challenges to properly and precisely accessing the causal priors of target users, which may be hard and unrealistic to assume. Therefore, future design and research should also consider the causal priors of the users when presenting causal information, and provide ways to elicit, validate, and update them.

Disagreement impact: Disagreement between a user's causal prior and the visualized association can lead to cognitive dissonance on overestimation or underestimation of causal relationships. Since the visualized association can be pre-calculated by visual analytics systems, we would recommend adding functionalities of disagreement checks using our suggested model in Section 6.3, if the causal prior has been collected in advance. It's also important for future research to find a balance for such impact to avoid undervaluing human knowledge and intuition and overvaluing the visualized data.

Visualization types: Previous work has shown that choices of chart type play a pivotal role in the interpretation of causal relationships, and our results seem to agree. Developers should be aware the overall cognitive process of causal inference that chart types may evoke should be an important consideration of effective visualization design. Future design efforts should focus on identifying which types of visual representations are most conducive to accurate causal inference.

7.3 Limitations and Future Directions

Data corpus: Although we collected concept names from variables from widely applied datasets, to achieve an efficient MTurk study design our corpus only contains 56 concept pairs. This introduces some limitations on the distribution of causal priors. For example, we didn't find any concept pair that has an extremely high average causal prior, e.g., larger than 4.6, whereas the highest causal prior we found was around 4.5. Future research should focus on enlarging the concept pair corpus to better cover all possible distributions of underlying causal priors. In addition, we tested one direction in the causal questions, by predetermining causal factors and outcome concepts (Section 3.3.2). Further experiments could consider the impact of different directions between concept names in causal questions.

Visualization stimuli design: Additionally, to simplify the study design, we only considered three typical chart types, following previous work [39, 54, 62], and did not test more complex visual encodings, such as aggregation levels within each chart type. Therefore, although we confirmed many existing insights regarding these charts, we cannot directly address insights related to other visual encodings, such as how text tables were found to perform similarly to bar charts [27], and increasing aggregation levels in specific chart types were found to improve the ability to convey causality [62]. Moreover, our chart types were associated with data types, so we are unable to decouple the potential impacts of data types and chart types, as discussed in Section 7.1.4. We also did not test visualizations showing negative associations, which could have a different impact on human perception in certain cases [4]. In addition, we removed chart axes since determining appropriate axis value complicated the study design. However, showing such values may also impact users' perceptions. We plan to incorporate more design factors in visualizations such as additional chart types, varying visual encodings, and negative associations in future work.

Expertise and visualization literacy: Our study did not aim to assess the impact of visualization expertise, and was conducted on the general population of MTurk users. Our results can therefore not provide insights related to existing perception experiments specifically designed for novices [6,7,18,36] or experienced users [3,16,17,47]. Future work could focus on examining how our findings may be affected by novice or expert populations.

8 Conclusion

This paper aims to provide insights on the human cognition process of causal inference from visualizations, and provides new perspectives on guidelines from graphical perception for visual causal inference. Our study has illuminated the significant role of causal priors, i.e., the preconceived notions about causality between concept pairs, in shaping users' interpretations of visualized data. The reported empirical evidence underscores the influence of causal priors, which, when combined with evidence of associations in visualized data, can significantly impact cognitive processes related to causal inference. The results additionally suggest that chart type may also have an impact. The introduction of a model capturing the differential patterns in perceived causal relationships caused by causal priors and visualized associations offers a novel perspective on how these elements converge to affect users' causality judgments. Furthermore, our research contributes to the field by providing a comprehensive dataset of causal priors associated with concept pairs and visualizations. This dataset not only serves as a benchmark for future studies but also aids in the development of heuristic-based design guidelines aimed at enhancing visual design choices to better support visual causal inference based on specific variables in data. In light of our findings, we advocate for heightened awareness among visualization designers regarding the potential for causal priors to lead to the impact of perceived causal relationships in the visualized data. By integrating these heuristic-based guidelines into visualization design, we can facilitate more accurate interpretations of visualized data, thereby improving decision-making processes across various domains.

Acknowledgments

We thank the reviewers for their insightful comments. This material is based upon work supported by the National Science Foundation under Grant No. 2211845.

References

[1] A. Asuncion and D. Newman. Uci machine learning repository, 2007. 4

[2] A. G. Barnett, J. C. Van Der Pols, and A. J. Dobson. Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology*, 34(1):215–220, 2005. doi: 10.1093/ije/dyh299 8, 9

[3] L. Battle and J. Heer. Characterizing exploratory visual analysis: A literature review and evaluation of analytic provenance in tableau. *Computer Graphics Forum*, 38(3):145–159, 2019. doi: 10.1111/cgf.13678 9

[4] S. S. Beheshitha, M. Hatala, D. Gašević, and S. Joksimović. The role of achievement goal orientations when studying effect of learning analytics visualizations. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, pp. 54–63. ACM, New York, 2016. doi: 10.1145/2883851.2883904 9

[5] D. Borland, A. Z. Wang, and D. Gotz. Using counterfactuals to improve causal inferences from visualizations. *IEEE Computer Graphics and Applications*, 44(1):95–104, 2024. doi: 10.1109/MCG.2023.3338788 1, 2

[6] A. Burns, C. Lee, R. Chawla, E. Peck, and N. Mahyar. Who do we mean when we talk about visualization novices? In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–16. ACM, New York, 2023. doi: 10.1145/3544548.3581524 3, 9

[7] A. Burns, C. Xiong, S. Franconeri, A. Cairo, and N. Mahyar. How to evaluate data visualizations across different levels of understanding. In *IEEE Workshop on Evaluation and Beyond-Methodological Approaches to Visualization (BELIV)*, pp. 19–28. IEEE, Washington DC, 2020. doi: 10.1109/BELIV51497.2020.00010 9

[8] W. S. Cleveland, M. E. McGill, and R. McGill. The shape parameter of a two-variable graph. *Journal of the American Statistical Association*, 83(402):289–300, 1988. doi: 10.1080/01621459.1988.10478598 2, 5, 8

[9] W. S. Cleveland and R. McGill. An experiment in graphical perception. *International Journal of Man-Machine Studies*, 25(5):491–500, 1986. doi: 10.1016/S0020-7373(86)80019-0 5

[10] Z. Deng, D. Weng, X. Xie, J. Bao, Y. Zheng, M. Xu, W. Chen, and Y. Wu. Compass: Towards better causal analysis of urban time series. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1051–1061, 2021. doi: 10.1109/tvcg.2021.3114875 2

[11] P. Dragicevic. Fair statistical communication in hci. *Modern Statistical Methods for HCI*, pp. 291–330, 2016. doi: 10.1007/978-3-319-26633-6_13 5

[12] E. C. Ferstl, A. Garnham, and C. Manouilidou. Implicit causality bias in english: A corpus of 300 verbs. *Behavior Research Methods*, 43:124–135, 2011. doi: 10.3758/s13428-010-0023-2 3, 8

[13] S. L. Franconeri, L. M. Padilla, P. Shah, J. M. Zacks, and J. Hullman. The science of visual data communication: What works. *Psychological Science in the Public Interest*, 22(3):110–161, 2021. doi: 10.1177/15291006211051956 3

[14] B. Ghai and K. Mueller. D-bias: A causality-based human-in-the-loop system for tackling algorithmic bias. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):473–482, 2022. doi: 10.1109/TVCG.2022.3209484 2

[15] E. Goikoetxea, G. Pascual, and J. Acha. Normative study of the implicit causality of 100 interpersonal verbs in spanish. *Behavior Research Methods*, 40(3):760–772, 2008. doi: 10.3758/BRM.40.3.760 3, 8

[16] D. Gotz and Z. Wen. Behavior-driven visualization recommendation. In *Proceedings of the 14th International Conference on Intelligent User Interfaces*, pp. 315–324. ACM, New York, 2009. doi: 10.1145/1502650.1502695 9

[17] D. Gotz and M. X. Zhou. Characterizing users' visual analytic activity for insight provenance. *Information Visualization*, 8(1):42–55, 2009. doi: 10.1057/ivs.2008.31 9

[18] L. Grammel, M. Tory, and M.-A. Storey. How information visualization novices construct visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):943–952, 2010. doi: 10.1109/TVCG.2010.164 3, 9

[19] T. L. Griffiths and J. B. Tenenbaum. Structure and strength in causal induction. *Cognitive Psychology*, 51(4):334–384, 2005. doi: 10.1016/j.cogpsych.2005.05.004 2

[20] G. Guo, E. Karavani, A. Endert, and B. C. Kwon. Causalvis: Visualizations for causal inference. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–20. ACM, New York, 2023. doi: 10.1145/3544548.3581236 2

[21] K. W. Hall, A. Kouroupis, A. Bezerianos, D. A. Szafir, and C. Collins. Professional differences: A comparative study of visualization task performance and spatial ability across disciplines. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):654–664, 2021. doi: 10.1109/TVCG.2021.3114805 3

[22] L. Harrison, F. Yang, S. Franconeri, and R. Chang. Ranking visualizations of correlation using weber's law. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1943–1952, 2014. doi: 10.1109/TVCG.2014.2346979 8

[23] J. Heer and M. Agrawala. Multi-scale banking to 45 degrees. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):701–708, 2006. doi: 10.1109/TVCG.2006.163 5

[24] M. N. Hoque and K. Mueller. Outcome-explorer: A causality guided interactive visual interface for interpretable algorithmic decision making. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4728–4740, 2021. doi: 10.1109/TVCG.2021.3102051 2

[25] J. Hullman and A. Gelman. Designing for interactive exploratory data analysis requires theories of graphical inference. *Harvard Data Science Review*, 3(3):10–1162, 2021. doi: 10.1162/99608F92.3AB8A587 1, 2

[26] Z. Jin, S. Guo, N. Chen, D. Weiskopf, D. Gotz, and N. Cao. Visual causality analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1343–1352, 2020. doi: 10.1109/TVCG.2020.3030465 2

[27] A. Kale, Y. Wu, and J. Hullman. Causal support: modeling causal inferences with visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1150–1160, 2021. doi: 10.1109/tvcg.2021.3114824 2, 9

[28] S. Kaul, D. Borland, N. Cao, and D. Gotz. Improving visualization interpretation using counterfactuals. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):998–1008, 2021. doi: 10.1109/TVCG.2021.3114779 2, 9

[29] M. Kay and J. Heer. Beyond weber's law: A second look at ranking visualizations of correlation. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):469–478, 2015. doi: 10.1109/TVCG.2015.2467671 8

[30] Y.-S. Kim, P. Kayongo, M. Grunde-McLaughlin, and J. Hullman. Bayesian-assisted inference from visualized data. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):989–999, 2020. doi: 10.1109/TVCG.2020.3028984 1, 2

[31] K. Koffka. *Principles of Gestalt psychology*. routledge, 2013. 3

[32] J. H. Larkin and H. A. Simon. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11(1):65–100, 1987. doi: 10.1111/j.1551-6708.1987.tb00863.x 3, 8

[33] Z. Liu and J. Stasko. Mental models, visual reasoning and interaction in information visualization: A top-down perspective. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):999–1008, 2010. doi: 10.1109/TVCG.2010.177 3, 8

[34] L. M. Padilla, S. H. Creem-Regehr, M. Hegarty, and J. K. Stefanucci. Decision making with visualizations: a cognitive framework across disciplines. *Cognitive Research: Principles and Implications*, 3(1):1–25, 2018. doi: 10.1186/s41235-018-0120-9 3, 8

[35] J. Pearl. *Causality*. Cambridge University Press, Cambridge, 2009. doi: 10.1017/CBO9780511803161 1, 2

[36] E. M. Peck, S. E. Ayuso, and O. El-Etr. Data is personal: Attitudes and perceptions of data visualization in rural pennsylvania. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12. ACM, New York, 2019. doi: 10.1145/3290605.3300474 3, 9

[37] M. Prosperi, Y. Guo, M. Sperrin, J. S. Koopman, J. S. Min, X. He, S. Rich, M. Wang, I. E. Buchan, and J. Bian. Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence*, 2(7):369–375, 2020. doi: 10.1038/s42256-020-0197-y 2

[38] G. J. Quadri and P. Rosen. A survey of perception-based visualization studies by task. *IEEE transactions on visualization and computer graphics*, 28(12):5026–5048, 2021. doi: 10.1109/TVCG.2021.3098240 4

[39] G. J. Quadri, A. Z. Wang, Z. Wang, J. Adorno, P. Rosen, and D. A. Szafir. Do you see what i see? a qualitative study eliciting high-level visualization comprehension. In *ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 1–26. ACM, New York, 2024. doi: 10.1145/3613904.3642813 3, 4, 9

[40] R. A. Rensink. The nature of correlation perception in scatterplots. *Psychonomic Bulletin & Review*, 24:776–797, 2017. doi: 10.3758/s13423-016-1174-7 5, 8

[41] R. A. Rensink and G. Baldridge. The perception of correlation in scatterplots. vol. 29, pp. 1203–1210, 2010. doi: 10.1111/j.1467-8659.2009.01694.x 2, 5, 8

[42] M. Riaz and R. Girju. Recognizing causality in verb-noun pairs via noun and verb semantics. In *Proceedings of the EACL 2014 Workshop on Computational Approaches to Causality in Language (CAtoCL)*, pp. 48–57. ACL, Gothenburg, 2014. doi: 10.3115/v1/W14-0707 3

[43] U. Rudolph and F. Försterling. The psychological causality implicit in verbs: A review. *Psychological bulletin*, 121(2):192, 1997. doi: 10.1037/0033-2909.121.2.192 2, 3

[44] S. Salim, M. N. Hoque, and K. Mueller. Belief miner: A methodology for discovering causal beliefs and causal illusions from general populations. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1):1–37, 2024. doi: 10.1145/3637298 2, 3

[45] K. B. Schloss, C. C. Gramazio, A. T. Silverman, M. L. Parker, and A. S. Wang. Mapping color to meaning in colormap data visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):810–819, 2018. doi: 10.1109/TVCG.2018.2865147 3

[46] P. Spirtes, C. N. Glymour, and R. Scheines. *Causation, prediction, and search*. MIT Press, Cambridge, 2000. doi: 10.1007/978-1-4612-2748-9 2

[47] C. D. Stolper, A. Perer, and D. Gotz. Progressive visual analytics: User-driven visual exploration of in-progress analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1653–1662, 2014. doi: 10.1109/TVCG.2014.2346574 9

[48] D. A. Szafir. Modeling color difference for visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):392–401, 2018. doi: 10.1109/TVCG.2017.2744359 3, 4

[49] D. A. Szafir, R. Borgo, M. Chen, D. J. Edwards, B. Fisher, and L. Padilla. *Visualization Psychology*. Springer Nature, New York, 2023. doi: 10.1007/978-3-031-34738-2 3

[50] J. Talbot, J. Gerth, and P. Hanrahan. An empirical model of slope ratio comparisons. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2613–2620, 2012. doi: 10.1109/TVCG.2012.196 5

[51] J. W. Tukey et al. *Exploratory Data Analysis*, vol. 2. Addison-Wesley, Reading, 1977. doi: 10.1007/978-3-031-20719-8_2 1

[52] T. Vigen. *Spurious correlations*. Hachette UK, 2015. 4

[53] A. Z. Wang, D. Borland, and D. Gotz. Countering simpsons paradox with counterfactuals. In *IEEE VIS Posters*, 2023. 2

[54] A. Z. Wang, D. Borland, and D. Gotz. An empirical study of counterfactual visualization to support visual causal inference. *Information Visualization*, 23(2):197–214, 2024. doi: 10.1177/14738716241229437 2, 4, 9

[55] A. Z. Wang, D. Borland, and D. Gotz. A framework to improve causal inferences from visualizations using counterfactual operators. *Information Visualization*, 2024. doi: 10.1177/14738716241265120 2

[56] J. Wang, X. Li, C. Li, D. Peng, A. Z. Wang, Y. Gu, X. Lai, H. Zhang, X. Xu, J. Zhou, X. Liu, and C. Wei. AVA: An automated and ai-driven intelligent visual analytics framework. *Visual Informatics*, 2024. doi: 10.1016/j.visinf.2024.06.002 3

[57] J. Wang and K. Mueller. The visual causality analyst: An interactive interface for causal reasoning. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):230–239, 2015. doi: 10.1109/TVCG.2015.2467931 2

[58] J. Wang and K. Mueller. Domino: Visual causal reasoning with time-dependent phenomena. *IEEE Transactions on Visualization and Computer Graphics*, 29(12):5342–5356, 2022. doi: 10.1109/TVCG.2022.3207929 2

[59] Y. Wang, Z. Wang, C.-W. Fu, H. Schmauder, O. Deussen, and D. Weiskopf. Image-based aspect ratio selection. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):840–849, 2018. doi: 10.1109/TVCG.2018.2865266 5

[60] Y. Wang, Z. Wang, L. Zhu, J. Zhang, C.-W. Fu, Z. Cheng, C. Tu, and B. Chen. Is there a robust technique for selecting aspect ratios in line charts? *IEEE Transactions on Visualization and Computer Graphics*, 24(12):3096–3110, 2017. doi: 10.1109/TVCG.2017.2787113 5

[61] S. N. Wood. *Generalized additive models: an introduction with R*. Chapman and Hall/CRC, London, 2017. doi: 10.1201/9781420010404 7, 8

[62] C. Xiong, J. Shapiro, J. Hullman, and S. Franconeri. Illusion of causality in visualized data. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):853–862, 2019. doi: 10.1109/TVCG.2019.2934399 2, 4, 5, 9

[63] C. Xiong, C. Stokes, Y.-S. Kim, and S. Franconeri. Seeing what you believe or believing what you see? belief biases correlation estimation. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):493–503, 2022. doi: 10.1109/TVCG.2022.3209405 3, 8

[64] C.-H. E. Yen, A. Parameswaran, and W.-T. Fu. An exploratory user study of visual causality analysis. vol. 38, pp. 173–184, 2019. doi: 10.1111/cgf.13680 2

[65] Z. Zhou, X. Wen, Y. Wang, and D. Gotz. Modeling and leveraging analytic focus during exploratory visual analysis. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–15. ACM, New York, 2021. doi: 10.1145/3411764.3445674 1