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.2022.3232959

Reasoning Affordances with Tables and Bar Charts

Cindy Xiong, Elsie Lee-Robbins, Icy Zhang, Aimen Gaba, and Steven Franconeri

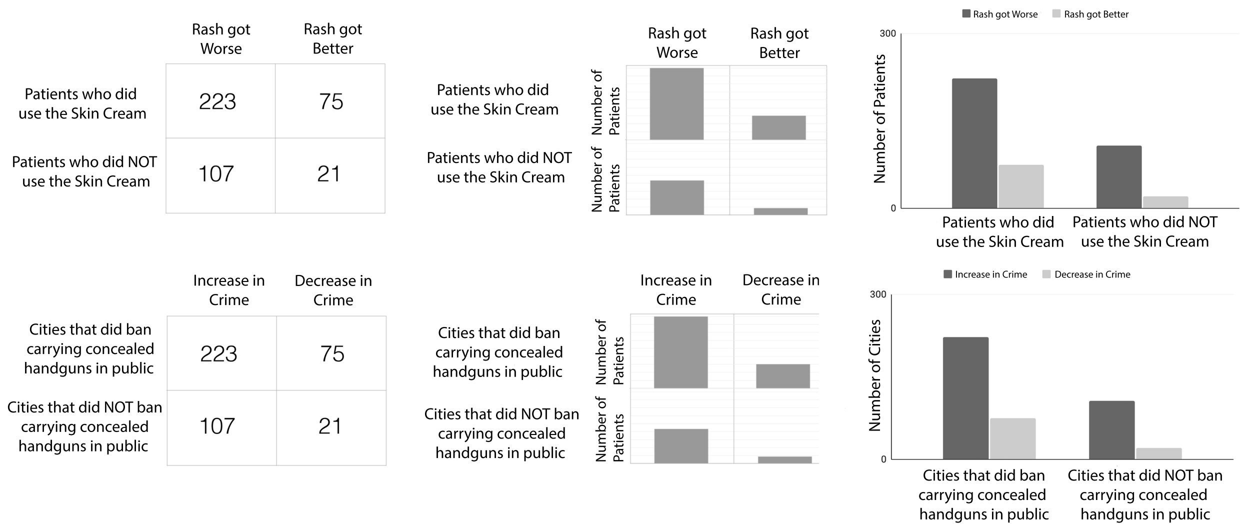


Fig. 1. What do the data above suggest about the effects of using the skin cream (top displays), or of banning concealed handguns in public (bottom displays)? We asked the same question for data shown in either a table, bar table, or a bar chart. These formats appear to afford different problem-solving strategies, with robustly higher accuracy for the tables (left) than for the visualizations (right).

Abstract—A viewer's existing beliefs can prevent accurate reasoning with data visualizations. In particular, confirmation bias can cause people to overweigh information that confirms their beliefs, and dismiss information that disconfirms them. We tested whether confirmation bias exists when people reason with visualized data and whether certain visualization designs can elicit less biased reasoning strategies. We asked crowdworkers to solve reasoning problems that had the potential to evoke both poor reasoning strategies and confirmation bias. We created two scenarios, one in which we primed people with a belief before asking them to make a decision, and another in which people held pre-existing beliefs. The data was presented as either a table, a bar table, or a bar chart. To correctly solve the problem, participants should use a complex reasoning strategy to compare two ratios, each between two pairs of values. But participants could also be tempted to use simpler, superficial heuristics, shortcuts, or biased strategies to reason about the problem. Presenting the data in a table format helped participants reason with the correct ratio strategy while showing the data as a bar table or a bar chart led participants towards incorrect heuristics. Confirmation bias was not significantly present when beliefs were primed, but it was present when beliefs were pre-existing. Additionally, the table presentation format was more likely to afford the ratio reasoning strategy, and the use of ratio strategy was more likely to lead to the correct answer. These findings suggest that data presentation formats can affect affordances for reasoning.

Index Terms—Data visualization, Tabular displays, Empirical evaluation, Reasoning

1 Introduction

People read, interpret, and understand trends and patterns in data to inform decisions. Data can be represented in various ways, such as text, tables, and bar charts. Each affords different ways of interacting with the data. When numbers are shown as verbal symbols as in tables or text, people must process the relationship between these numbers slowly over seconds or even minutes [22]. But when those same numbers are visualized, that task is offloaded to a powerful brain system that can process these numbers in a split second, across different visual channels like position, color, or size [8]. However, this rapid visual processing is limited to a superficial set of statistics, such as computing the mean position within a scatterplot or noticing the largest circle in a bubble chart [66]. More sophisticated comparisons and relationships among subsets of values require the viewer's visual system to isolate relevant values across time and pull back its throttle to be almost as slow as verbal symbol processing [20, 22, 51].

With experience, a faster set of automated associations increasingly allows us to quickly distinguish the relevant from the irrelevant, and combine multiple sources of information in parallel across a set of memorized routines. These routines allow physicians to quickly diagnose common diseases and pilots to fly airplanes in a more automated manner. However, fast decisions based on the typical conditions of past experiences can produce sub-optimal outcomes compared to a careful and slower reasoning process, especially for emotionally-charged or under-practiced scenarios. During an emergency, jumping to an intuitive move is extremely dangerous for pilots, because they can fail to consider unusual factors. To prevent pilots from making a fast and

- *Cindy Xiong is with UMass Amherst CICS. Email: cindy.xiong@cs.umass.edu.*
- *Elsie Lee-Robbins is with University of Michigan.*
- *Icy (Yunyi) Zhang is with UCLA.*
- *Aimen Gaba is with UMass Amherst.*
- *Steven Franconeri is with Northwestern University.*

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

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.2022.3232959

potentially destructive decision, an exhaustive checklist is used to force cautious and slow processing [28].

Visualizing data in different ways can change how people reason with data. For example, people tend to assume unwarranted causal connections between variables depicted with bar charts but are less likely to mistake correlation for causation when looking at scatterplots and line charts [84]. **Confirmation bias** can interfere with data reasoning. It occurs when people are influenced by their prior beliefs, leading them to focus on a conclusion from the data that is in agreement with their already-held beliefs [14, 16, 42]. When conducting visual analytics, existing beliefs can even bias how people extract objective statistics from data, such as correlations [86].

In this paper, we compare different ways of presenting information normally depicted in bar charts and simple tables to investigate reasoning affordances associated with visualization designs. We tested performance on a problem that requires analytic reasoning, adapted from literature on reasoning errors and biases [34]. Looking at the bottom left table in Figure 1 as an example, what do the data in the displays suggest about the effects of banning concealed handguns in public? In our experiments, we showed this data either with a table of numbers, or two types of visualizations, a bar table and a bar chart. The rows of the tables represent two groups: one consisting of cities that had recently enacted bans on concealed weapons and another that had cities with no such bans. In each row, the left column shows the number of cities that experienced increases in crime and the right column shows those that experienced decreases in crime.

Multiple strategies can be adopted to solve whether the handgun ban was effective [34]. Some strategies are less optimal and lead to inaccurate solutions. For example, a viewer might focus on the first top row only, seeing that more cities (223) had an increase in crime after banning concealed handguns, compared to fewer (75) that had a decrease. This strategy would lead them to the inaccurate conclusion that banning handguns was *ineffective* in decreasing the number of crimes in the city. But if this conclusion aligns with their existing belief that gun bans are not effective in decreasing crimes, they are likely less motivated to check their reasoning strategy and discover their solution to be inaccurate.

The correct way to interpret the data is to compare normalized percentages or ratios of the pairs of numbers in the top row to that of the bottom row, making use of all four cells. We see that approximately one in three cities that banned weapons became safer (75 vs 223), but approximately one in five cities that did not ban weapons became safer (21 vs 107). Comparing this normalized ratio between the two, we see that banning concealed weapons was associated with a decrease in crime. This ratio-comparison strategy requires an analytic reasoning process, requiring the solver to consider *two* comparisons across the columns, and then compare the results of those comparisons.

Contributions: We explore confirmation bias in visualization interpretation and whether confirmation bias can be mitigated by visual representation designs through the elicitation of different data reasoning strategies. We contribute two empirical experiments that explore the *reasoning affordance* of bar charts, bar tables, and tables, and show that tables most readily afford the complex ratio reasoning strategy. We adopt an established task from the reasoning literature in cognitive psychology [34]. Experiment 1 takes a controlled experimental approach where we primed people with beliefs using a neutral topic of skin cream effectiveness. Experiment 2 relies on a more emotionally-charged scenario of gun control laws, that has been used to explore how reasoning is associated with strong pre-existing beliefs.

We discuss the differing effects of primed beliefs versus pre-existing beliefs on how people make decisions with data, providing recommendations on leveraging visualization design to elicit the 'right' reasoning strategies for problem-solving.

2 Related Work

Our visual system can crunch vast arrays of numbers in a chart at a glance, offloading the work of extracting values and patterns needed to make decisions about data [43]. Visualizations provide immediate access to summary statistics, outliers, and trends [59,66]. But extracting other critical patterns in data requires a slower visual process: finding conjunctions of features or comparing particular values must be done slowly [3, 23, 27, 49, 52, 80], akin to reading through sentences in a paragraph [21,62]. This difficulty in comparing relations may influence an impatient viewer to focus their thinking on salient values drawn from a quick inspection of summary statistics.

A heuristic (usually relying on shortcuts) or intuition-based decision-making process often relies on an associative network based on previous experience, which can lead to quick and powerful processing in a complex problem [9]. However, this system can be prone to error and sub-optimal solutions, including availability bias (more heavily weighing options that more readily come to mind), 'search satisfaction' (stopping when any acceptable solution is found [54]), and loss aversion (weighing losses more heavily compared to equivalent gains [71]). Analytical reasoning can be more accurate but can take orders of magnitude more time to complete [26].

In addition, biases also influence the way we interpret visualizations and make decisions with visualizations [14,17,45,75,82]. Quick, initial judgments of visualizations might rely on faster heuristic processes [55], and these processes can be influenced by what grabs attention in a visualization design, such as unique colors or larger fonts [30, 79]. When asked to identify the most important parts of data visualizations, users tend to choose titles, labels, and max and min points in the data [7]. Another bias, the *attraction effect*, can also occur within both tables and visualizations. When deciding between two equal alternatives in a scatterplot, providing a third alternative that is only slightly worse than one alternative makes that alternative the preferred option [13]. When a story about a visualization highlights a subset of data values, participants predict that other people will find that subset more salient, demonstrating a *curse of knowledge* bias [88]. People also make belief-driven estimations of correlation. When they read scatterplots depicting a relationship congruent with their belief, they overestimate the correlation value [86].

2.1 Cognitive Framework for Reasoning

Cognitive frameworks propose processing stages, such as attention, memory, and pattern recognition. A *dual-processing model* [55,58,91] suggests two types of processing when making a decision or response based on visualized data: a faster, impulsive system and a slower, deliberate system (though there is debate over whether these systems are truly separable [18]). The faster system relies on past experience and heuristics to make faster decisions [9]. On the other hand, a set of more slowly-operating processes requires more deliberate effort, but helps people reach more accurate conclusions, and to solve harder problems [35]. Decades of research in human cognition and decision-making have shown that this faster processing can also lead to biases, as people rely less on their analytic thinking and more on heuristics (shortcuts) and intuitions [35].

2.2 Belief, Confirmation Bias, and Priming

Confirmation bias is a classic demonstration of the shortcomings in decision-making caused by heuristic and intuition-based thinking. Confirmation bias has been studied through various psychology experiments. One classic task that demonstrates confirmation bias is the Wason Card selection task [76]. Four cards are laid on a table, each with a number or letter on the front: a vowel, a consonant, an even number, or an odd number (e.g., A, B, 4, and 7). The participant chooses which cards to flip over in order to determine the veracity of the statement: "If there is a vowel on one side of the card, then there is an even number on the other side." In this example, people readily chose to flip the card with an A on it, because if there was an even number on the other side, this card would confirm the previous statement. Fewer people chose to flip over the odd-numbered card, which could also falsify the statement if it had a vowel on the other side. This is an example of confirmation bias because people seek out information to confirm, but not to disprove, the original statement. In the real world, people are more likely to frame hypotheses in a way that would affirm if a person was suited for the job, rather than unsuited [63]. Medical

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.2022.3232959

professionals who have made a preliminary diagnosis will interpret new information through a biased lens that is more congruent with their belief and potentially reach the incorrect diagnosis [46]. On social media, people are more likely to follow people who share similar beliefs rather than opposing beliefs. This can create a filter bubble, in which users are only exposed to ideas that they already agree with [57]. Neuroscience studies also demonstrate that existing beliefs can alter the neural representation of the strength of presented information, which leads to confirmation bias and makes people less likely to change their minds even when presented later with incongruent information [36].

There are data analytics systems designed to mitigate biases. For example, cognitive biases in college admissions have shown to be mitigated by presenting data with visualizations, which encourages multiple interpretations of the data [65]. For analytical tasks that involve using many pieces of evidence, software tools, such as Sandbox, allow users to visually organize information to reduce cognitive biases and increase performance [81]. It encourages users to explore multiple hypotheses and record supporting or rejecting evidence when working with data. Another application is the U.S. Navy application JIGSAW (Joint Intelligence Graphical Situation Awareness Web, not to be confused with Jigsaw, a visual analytic system from [64]). The tool visually lays out the evidence on a horizontal axis, ranging from *strongly refutes* to *strongly supports*. U.S. Navy trainee analysts showed less biased decision-making when evidence was visually organized in this way, compared to only being labeled with equivalent information [11]. In an effort to mitigate the attraction bias in visual data decision-making, Dimara et al. found that highlighting the superior alternatives weakened the bias, but did not eliminate it [12].

Given that bias can be mitigated by adopting different visualization designs, we expand on this effort to further explore if visualization design can help mitigate biases. In addition, studies so far primarily examined the use of visualization to mitigate bias, while very few have investigated if confirmation bias exists when people are reasoning with data visualizations. Thus, inspired by the study that found confirmation bias for data shown in table format [34], we investigate whether a similar task would reveal confirmation bias in different visualizations of those data.

Models of intelligence analysis cycle were developed by [78] to mitigate confirmation bias in real-world situations, leveraging active search, information source collection, and quality assessment. This suggests that priming effects can strongly influence decision-making in visual data analysis. Priming effects tend to refer to situations where the presence of a stimulus changes one's subsequent reaction or interpretation to similar or related stimuli [35].

In cognitive psychology, priming effects can be observed in situations such as categorization tasks. For example, an ambiguous description of a person, which can be interpreted to be either positive or negative tends to be categorized in a manner consistent with the primed construct (e.g., positive prime leads to positive categorization). [32]. People are more likely to complete the word 'so?p' as 'soup' when primed with food and as 'soap' when primed with shower [35]. People's evaluation of brands can be impacted by priming, both through cognitive priming where participants are given certain product attributes, and affective priming, where participants are primed with certain emotional associations with the product [89]. Priming effects can also happen in the perceptual domain. Separability of previously seen scatterplots can influence people's judgment of the separability of subsequent scatterplots, such that people see scatterplot clusters as less separated when they saw a more separated plot prior [73].

In visual data analytics, priming effects can be elicited by prior beliefs to bias data interpretation. For example, existing belief can bias people to misestimate correlations in scatterplots, such that people who believe that a strong positive correlation should exist between two variables would see their correlation to be stronger in a scatterplot by an effect size of $r = 0.1$ [86]. Strong existing attitudes about politically polarized topics can make people less Bayesian when reasoning with data, updating their beliefs either overly conservative or aggressive when interacting with statistical data visualizations [37]. Participants were more likely to stop engaging with the information when their heuristic assessment of it gratified their political predispositions, even though the resulting inference that they drew about the result of the experiment was incorrect [38].

2.3 Affordances of Visualization Design

Perceived affordances are the relationships between the properties of an object that convey potential interactions and the capabilities of that object [50]. For example, a door with a handle can be pulled or pushed open, whereas a door with a metal plate surface can *only* be pushed open. Similar to physical objects, visual data representations also hold similar affordances, via common tasks done with those designs or conceptual associations driven by their underlying metaphors [72].

Even seemingly small design choices in a visualization can nudge viewers to see different patterns and produce different sentences [1, 6], with varying perceptual accuracies [10, 33], further highlighting the complexities of visual data interpretation. For example, bar graphs are discrete objects and elicit interpretations from comparing two distinct units (e.g., A is larger than B), while line graphs are single continuous objects and elicit interpretations of trends (e.g., As X increases, Y increases) [90]. Arrows afford conclusions of functionality: diagrams without arrows are more likely to be described as structural (e.g., The brake fluid is in the drum), but those with arrows are more likely to be described as functional (e.g., The brake fluid moves into the drum) [31]. When aggregating data from eight or sixteen marks to only two marks, viewers are more likely to infer causality [84], possibly because aggregated data can be more easily associated with experimental manipulations where such inferences are valid, or because the aggregated data seem less noisy and therefore more robust. When grouping bars into small multiples, showing them vertically versus horizontally [83], or varying the spatial proximity or color encoding [85] can elicit different comparisons across two groups. In icon arrays, depending on the internal arrangements of the icon grids, viewers can over or underestimate the percentage depicted by icon arrays [2, 82].

In risk communication, past work has shown that patient understanding of risks and uncertainty can depend on the format with which medical risks are communicated [19]. In a study testing people's understanding of risk data on 6 types of charts (bar charts, icon arrays, spark-plug, pie charts, clock graphs, and tables) through two main measures: whether people are able to read the graphs (e.g., How many people out of 100 would have side effects for drug A?) and whether people understand the essential message of the graph (e.g., Do more people get better with drug A or drug B?), more people answered the question about the essential message accurately for bar graphs (65%) than for tables (57%), but more people correctly extracted values for tables (67%) than for bar graphs (62%) [29].

A common framework used for decision-making with visualizations is Bayesian reasoning, which involves incorporating prior knowledge when calculating the probability of an event [41]. A common Bayesian reasoning error is *base-rate neglect*: failing to take into account the rate of an occurrence in the world before weighing evidence for it in an isolated case [35]. This error is similar to the one made in our present task, where participants failed to normalize values into a proportion of a total. Attempts to mitigate this error by replacing tables with visualizations have largely been unsuccessful across Euler diagrams, frequency grids, Sankey diagrams, flowcharts, tree diagrams, and hybrid diagrams [5, 38, 48, 53], though there is some evidence that interactive visualizations can improve performance [70]. We consider the more simple approach of a 2x2 table and bar table in this study. Both representations align the numbers both horizontally and vertically to afford ratio comparisons needed for Bayesian reasoning while preserving a general sense of familiarity to people, which can potentially boost the accuracy and confidence in reasoning tasks [39].

2.4 Tables versus Visualizations

The advantages and disadvantages of tables and visualizations have been studied across reasoning and judgment tasks. While results have been mixed, a review of 21 studies showed that there is little systematic difference between tables and bars for decision-making accuracy [61]. Some studies show that compared to tables, visualizations like line

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.2022.3232959

charts are faster, more accurate, and more preferred for finding trends such as correlation (e.g., Is there a correlation between movie budgets and the number of Oscars?) [47,60]. On the other hand, tables are faster, more accurate, and more preferred for tasks such as retrieving values (e.g., What is the duration of Frozen?) and deriving values (e.g., What is the total duration of all the movies?) [60]. Additionally, tables are significantly faster (but not more accurate) for comparing data points (e.g., In 2010, were there more action movies or horror movies?) [47].

Some studies suggest that visualizations can be more effective at changing attitudes than tables if there are no strong initial attitudes [56]. However, in some cases, visualizations can be more effective in changing attitudes even when those attitudes are strong, as in the surprising case of climate change attitudes [67], though other demonstrations of similar effects have failed to replicate [15]. More recently, tables have been shown to be effective tools for practitioners to directly manipulate their data to better their understanding and increase their confidence in the analytic process [4].

For estimating causality (e.g., Does having this virus make it more or less likely to contract this disease?), judgments are more accurate when using frequency trees that have numbers than a 2x2 table with countable objects or a text format [74]. The frequency tree showed the total number of people at the top, with a branch downwards stating how many people do or do not have the virus, with more branches from that stating how many do or do not have the disease. In the 2x2 tables, the rows represented people who had or did not have the virus, the columns represented people who had or did not have the disease, and in each cell, icon faces showed the number of people. Using frequency trees led to a better understanding of covariation and probabilistic scenarios compared to text or a table with countable icons.

3 Study Overview and Design Motivation

In the current study, we compared the effectiveness of table, bar table, and bar charts in facilitating critical thinking with an established reasoning paradigm from cognitive psychology research [34]. The original paradigm was the same as we have presented in Section 1: the participants were shown a 2x2 table and were asked to determine if the skin cream (or gun bans) made the rashes worse or better (or cities safer or not). This task has been used to show a relationship between numeracy (which measures one's ability to use quantitative information) and motivated reasoning. We used this same task to explore reasoning affordances of visualizations for two reasons. First, this paradigm presents a difficult task that has been shown to elicit multiple forms of reasoning strategies from participants, which makes it an appropriate test bed for biases and reasoning. Second, while the original study only used the table representation, testing bar tables allows us to directly compare the effect of showing people exact numerical values versus not while controlling for everything else. The additional bar chart design preserves the visual information from the bar tables, but rearranges the chart elements to be familiar to an average viewer, as bar charts are amongst the most commonly used visualizations [68]. This allows us to also examine the effect of chart arrangement familiarity on reasoning strategies and confirmation bias.

We also explored the effect of participants' prior beliefs on their reasoning strategy with different visual representations of data, following the set-up from [34]. This allows us to compare the effect size across topics, as well as the nature (predisposed vs. primed) and strength (weak vs. strong) of beliefs to potentially draw more generalizable conclusions. Experiment 1 primes people with a belief on the topic of skin cream, while Experiment 2 measures people's existing belief on gun-related policies and depicts data on gun bans without any primes.

3.1 Hypotheses

Based on prior work discussed in Section 2, we hypothesize:

1. People will engage with the data using different reasoning strategies depending on the visual representations (table, bar table, and bar charts) and belief types (primed belief related to skin cream vs. existing belief related to gun bans).
2. People will exhibit confirmation bias when making decisions depending on the visual representations and belief types.
3. People will exhibit different levels of accuracy in reasoning depending on the visual representations and belief types.

4 Experiment 1 Primed Belief

In Experiment 1, we investigate how primed beliefs could influence how people reason with visualizations. We introduced a skin cream problem based on an existing paradigm [34] and primed participants to think that the skin cream is likely either effective or ineffective.

4.1 Participants

We recruited 326 participants from Amazon Mechanical Turk. In order to qualify for our study, participants needed to have a 95% acceptance rate and be located in the United States. We implemented a series of quality checks to ensure that workers were paying attention (e.g., 'How many years of experience in Computer Programming do you have? Please choose "10+" to show that you are paying attention.'). In addition to these attention check questions, we were able to track how long a person stayed on a page through Qualtrics. We used this to filter out workers who spent less than 10 seconds reading the prime, which is a 2 paragraph-long description of the skin cream company (our pilot study suggests the median time people spend reading these paragraphs is around 20 seconds). After excluding participants who did not complete the survey, failed our attention checks, and/or provided extremely-low quality responses (ranging from writing off-topic answers to unintelligible sentences), we ended up with 230 participants ($M_{age} = 36$, $SD_{age} = 10$, 77 women).

4.2 Method and Procedure

In a between-subject design experiment, we primed each participant with information about a skin cream production company. Participants were randomly primed to form a positive or negative impression of the company. The positive prime described a successful company with a history of high-quality products, while the negative prime described an unsuccessful company that had previously violated federal health and safety policies. We intended this prime to bias participants into believing that the skin cream will be either effective or ineffective, so we can investigate how prior belief influences problem-solving for different visualization designs in this task. In our analysis, we coded participant beliefs as 'congruent' if the correct conclusion from the data aligned with the prime (e.g., primed to be a good company and data shows effective skin cream), and 'incongruent' otherwise (e.g., primed to be a bad company but data shows effective skin cream).

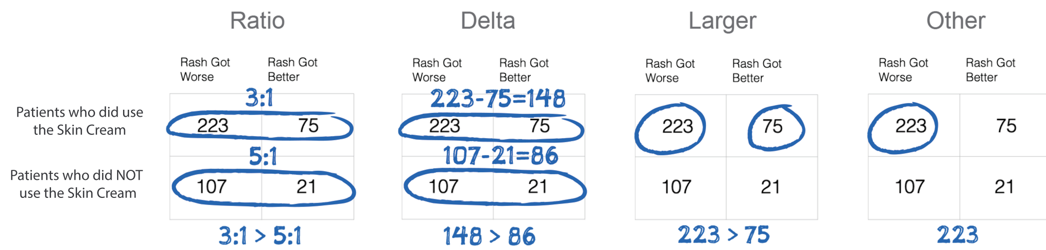
After reading the company description, participants viewed a skin cream data set as a tabular display (Figure 1 top left, adopted from [34, 77]), bar table display (Figure 1 top middle) or bar chart display (1 top right). Participants were asked whether using the skin cream made it more likely for rashes to get *worse* or *better*. They then typed their reasoning behind their decision into an open-response text box while referencing the data they saw, labeled with A, B, C, and D, as shown in Figure 3. Participants were instructed to explicitly refer to the cell labels when describing their strategies. Note that these letter labels were only shown for the open-ended responses and not for the initial task. These labels allowed us to more easily determine how participants used data in each cell to solve the problem.

All displays were counterbalanced by swapping the labels on the columns, such that for half of the participants the data suggested that the skin cream was effective. For the other half of the participants, the data suggested that the skin cream was ineffective. For the bar table and bar chart, we omitted numerical annotations (such as numeric value labels or numbers on the y-axis) to suppress verbal processing of the information, and to ensure a purely visual comparison. The experiment ended with participants self-reporting demographic information.

4.3 Qualitative Coding of Reasoning Strategies

Qualitative codes for possible strategies underwent several iterations before agreement on four categories: Ratio, Delta, Larger, and Other. The first four authors coded these responses, blind to the condition. We

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.2022.3232959



		Ratio	Delta	Larger	Other
Exp 1	Table	68%	4%	23%	5%
	Bar Table	43%	3%	4%	14%
	Bar Chart	25%	10%	33%	33%
Exp 2	Table	39%	12%	37%	13%
	Bar Table	17%	8%	54%	21%
	Bar Chart	26%	15%	46%	13%

Fig. 2. Qualitatively coded strategies for each experiment condition. Across each row (study), the color-coded table cells represent the proportion of participants in that condition who used the strategy. The visual annotations shown for each strategy are representative examples of the category, and are not exhaustive. Cell values are redundantly coded with color.

initially attempted to organize strategies by which cells (A, B, C, D) a participant mentioned in their open-ended response. For example, a common strategy was to compare cells A and C. However, we soon realized that it was more relevant to note the type of comparison that the participant was using. For example, if the participant was comparing whether B is larger than D, this is the same type of strategy as comparing whether A is larger than C. Both of these strategies use a heuristic of comparing two cells to each other, and it is not as important which pairs of cells they were comparing. Additionally, strategies could be complex, with people mentioning multiple comparisons, such as comparing *both* A $>$ B and A $>$ C. After many discussions between the authors, we decided to focus on the type of comparison instead of which cells were compared, and group responses into Ratio, Delta, Larger, and Other, which we explain below. See Figure 2 for a sketched example of each strategy and Table 1 for example quotes.

Ratio: Strategies were coded as Ratio if they mentioned "proportions," "percentages," or "ratio" and suggested a calculation of one of these values (e.g., for row 1) compared to another (e.g., compared to row 2). Some participants compared the rate of recovery of the people that used the skin cream (top row) to the rate of recovery of the people that did not use the skin cream (bottom row). Some participants used ratios not within the rows (as depicted in Figure 2), but instead within the columns. For example, they could compute that the ratio of the people who "got better" was about 2:1, but the ratio of the people who "got worse" was greater than 3:1. In all cases, the ratio strategy could involved using all four data points.

Delta: Strategies were coded as Delta if they mentioned the difference between cells. This calculation involved subtracting one cell from another cell (e.g., A-B). Participants commonly compared the deltas between two sets of cells (e.g., A-B is bigger than C-D). Note that this strategy is incorrect because subtraction is not a valid normalization operation for the present dataset, though it at least *attempts* to normalize the data, in spirit.

Larger: Strategies were coded as Larger if the response was a greater than or less than comparison. This was the most common heuristic. Participants compared cells A and B, or A and C, or B and D, or a combination of these.

Other: Any strategy that did not fit into the above categories but was still coherent was coded as Other. For example, one participant said "I wanted to choose that people got better by not using the cream as is conveyed in C. It wasn't a choice, however. Between the two choices, patient getting better by using the cream seem to be a better response." In this quote, the participant only mentions cell C and does not compare it with other cells or perform calculations. In another example, one participant mentioned "Without knowing how many people are in each group, it appears the percentage of those who got worse are in the group that used the new skin cream.", and it is unclear if they are making a comparison to anything. (Note that this participant appeared either to think that knowing the absolute value was critical or misunderstood that the bar represented a count.)

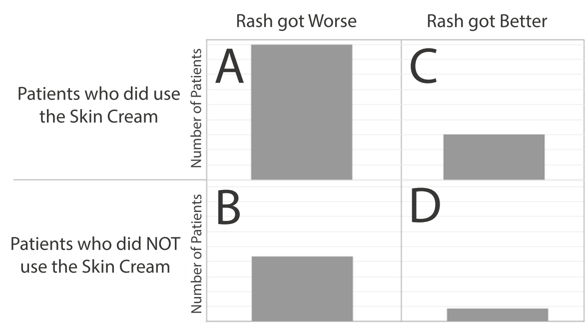


Fig. 3. Participants were asked to describe their strategies using the A, B, C, and D labels on the display. Letters were not present during the problem-solving stage. This figure is a recreation of the stimuli with larger labels, and the actual stimuli shown are in supplementary materials.

Discrepancies between coder ratings were discussed and resolved. Overall, our coders agreed 88.7% of the time in their ratings, with a high inter-rater reliability Kappa value of 0.87 ($z = 36.7$, $p < 0.001$). We further collapsed these strategies into a binary factor of whether it is the correct ratio strategy or a heuristic strategy (Delta, Larger, or Other) for later analysis.

4.4 Results Overview

Overall, the reasoning problem proved to be quite challenging for participants to solve. When the data was presented as a table, 54.7% of the participants got the question correct. The accuracy rate dropped to 38.1% for the bar table and to 35.9% for the bar chart.

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.2022.3232959

Exp	Strategy	Design	Example Quote
1	Ratio	Table	Say the number of people who DID use the cream (A) is basically double that of those who DIDN'T (C), the number of people whose rash got worse (B) should be slightly more than doubled (D). Yet it's over 3x as big.
1		Bar Table	In panel A, the people who used the cream had the rash get substantially worse. It was almost double the number of people who did not use the cream (panel c). The number of people who got better was almost 1/3 the number of the people who got worse using the cream (panel b).
1		Bar Chart	I imagine that if the cream did not work, row B would be smaller and more even to the level of D. If I cut A in half, it would be almost equal to C. If I cut B in half, it would be more than D.
1	Delta	Table	The difference between A and C was higher than the difference between B and D. Also, A is higher than B. This showed me that there were more people who got worse while using the cream than got better while using the cream and that more people got worse while using the cream than people who got worse while not using the cream. I think the people who got worse because of the cream is a more important result than the people that got better while using the cream.
1		Bar Table	Both of the answers were actually true. If you look at B and D you will see that people who used the cream also were more likely to get worse. If you look at A and C people who used the cream were also more likely to get better. I chose my answer because the bigger difference in data looks to be A vs C.
1		Bar Chart	This is difficult because there aren't numbers to the left really. I just looked at the difference between A and B compared to C and D. The difference looks smaller with regards to A and B which makes me believe more patients got better taking the cream than not taking it.
1	Larger	Table	The count of people in table B is higher than the count of people in the table D.
1		Bar Table	The bar in A is a lot higher than the bar in graph C.
1		Bar Chart	Because the number was high for people who used the skin cream in A And C. I would not use it
2	Ratio	Table	Total cities that did ban by adding A + B = 298 while cities that did not ban (C + D) was 128. For those that did ban, about 1/4 saw crime rates decrease. For those that did not, only 1/6 saw a decrease. Other variables were surely in play but based on that alone, I feel the ban did help to decrease crime overall.
2		Bar Table	I felt that comparing A vs C and B vs D the ratios showed that A was about double the figure of C. Whereas B was about three times higher than D showing that banning handguns increased crime somewhat. It would seem like the obvious choice did not have the desired effect.
2		Bar Chart	It seems, the relative growth of C with regards to D, is higher than A to B. Which means C grows faster than A, for the same numbers.
2	Delta	Table	I attempted to calculate the net decrease in crime for cities that did and did not ban carrying concealed handguns in public. For cities that did choose to ban concealed guns, I got 148 (A - B). For cities that did not choose to ban concealed guns I got 86 (C - D). The net decrease was greater in cities that banned concealed guns.
2		Bar Table	I subtracted the right columns (B and D) from the left columns (A and C). The one with the smaller remaining number of cities was the option that worked best. This is because it means they had the lowest number of cities with a net increase in crime.
2		Bar Chart	I saw that B went up higher than D. It also appeared to me, but I am not certain, that the distance between AB was less than the distance between CD. D was much lower than B. I tried to only pay attention to the B and D.
2	Larger	Table	I looked at mostly A and C. From there is where I concluded. I feel like I got it right.
2		Bar Table	I used choice A because the number of crimes has jumped up too high and is almost to the top of the chart. Comparative to the other charts, there is some increase in crime, but the rate is exponentially more in choice A. The decrease in crime in choice B shows that it doesn't compare to the increase in crime in choice A.
2		Bar Chart	I only looked at the increase in crime and D was the smallest.

Table 1. Reasoning quotes from Experiments 1 and 2 highlighting the different strategies.

We conducted a general logistic model predicting accuracy with a binary factor of the presence of the ratio strategy. The results suggested that using the ratio strategy significantly increased the likelihood of getting the correct answer by more than 4-fold ($Est = 1.44, SE = 0.38, p < 0.001$, $OR = 4.22$). Figure 5 shows the percentage of participants who used these strategies (Ratio, Delta, Larger) in each condition. 71.3% of the participant that used the ratio strategy selected the correct answer, while 28.7% of the participant that used the ratio strategy still ended up selecting the wrong answer. Figure 5 provides a summary view of how strategy relates to accuracy.

In terms of congruency, participants selected the correct answer 47.2% of the time when the correct answer was congruent with their belief (e.g., primed to be a good company and the data suggest that the skin cream is effective), and they selected the correct answer 38.5% of the time when the answer was incongruent with their belief. Figure 4 provides a summary view of how belief congruency relates to accuracy.

4.5 Analysis of Response Accuracy and Strategy

To start, we investigated whether visualization design and belief congruency affected accuracy. We constructed a logistic regression model predicting answer accuracy with whether the problem was presented as a table, a bar table, or a bar chart, and whether the correct answer was congruent or incongruent with the primed belief, and their interactions. To account for the effect of counterbalancing the skin cream to be either effective or ineffective, we also added the counterbalanced condition 'effectiveness' into our model. We found a significant effect of visualization type (χ^2 = 7.86, p = 0.020), such that when the problem was presented as a bar table or a bar chart, participants did not differ in their likelihood of getting the question correct (p = 0.88), but on average 4.68 times more likely to get the answer correct when the problem was presented with a table compared to a bar table, and 4.34 times more so compared to bar charts. Surprisingly, there was no significant effect of belief congruency (χ^2 = 1.96, p = 0.16) nor a significant interaction (χ^2 = 2.19, p = 0.13), suggesting that people were not more likely to select the correct answer when the answer aligns with their belief. We did, however, find an effect of the effectiveness counterbalance, such that participants were on average 3.56 times more likely to answer the question accurately when the correct answer was that the skin cream was effective (χ^2 = 13.35, $p < 0.01$).

As previously discussed, we noticed that participants used different reasoning strategies when tackling the problem. To account for the differing strategies, we added the strategies the participants used into our regression analysis to see if they interacted with visualization design and belief congruency and whether the strategies additionally accounted for whether people selected the correct answer. We constructed a logistic regression model predicting answer accuracy with visualization design, belief, strategy, their interactions, and the counterbalancing factor of skin cream effectiveness. We found a significant effect of strategy, such that participants who used the ratio strategy were significantly more likely to answer the question correctly (χ^2 =

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.2022.3232959

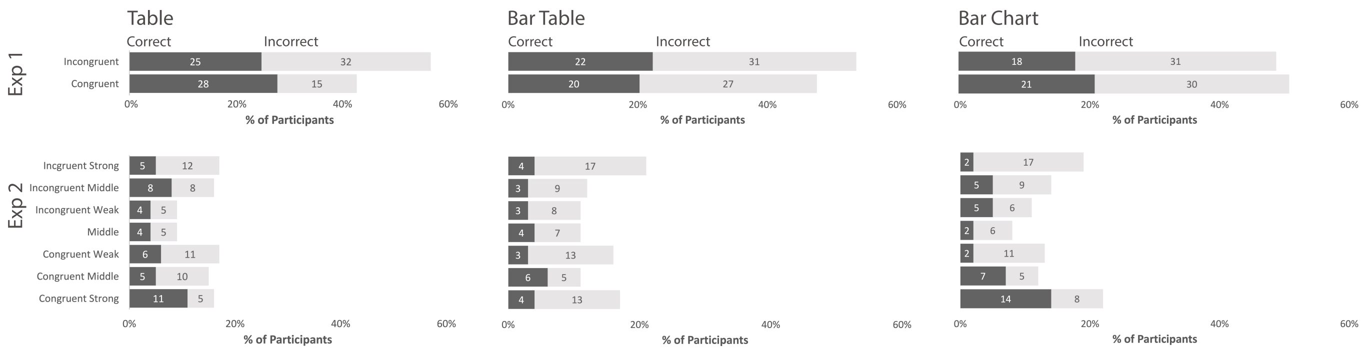


Fig. 4. Experiment 1 and 2 results showing how belief-data congruency relates to how likely the participants selected the correct answer.

59.34, $p < 0.01$). Interestingly, after the use of strategy was added to the model, the main effect of visualization went away (p = 0.34).

To explore why the effect of visualization design went away once we accounted for the reasoning strategies participants used, we performed mediation analysis using R Studio [69]. The outcome variable was response accuracy. The predictor variable was visualization type. The mediator variable was whether the participant used a ratio strategy or not. The indirect effect was tested using bootstrapping procedures for each of the 1000 bootstrapped samples. /icyThe average causal mediation effect (i.e. the indirect effect of visualization type on response accuracy) was found to be statistically significant (effect $= -0.12$, $p < 0.001$, $CI = [-0.20, -0.04]$). The significant indirect path suggests that the visualization design led participants to use different strategies, which impacted the likelihood of them selecting the correct answer. Follow-up regression analysis reveals that participants who viewed the problem presented as a table were significantly more likely to use the ratio strategy compared to those that viewed the bar table (OR = 2.83, p = 0.003) and bar chart (OR = 6.38, $p < 0.001$). Surprisingly, participants were also more likely to use the correct ratio strategy when they viewed the bar table, compared to the bar chart, by on average 2.25 folds (p = 0.021).

Given that not everyone who used the correct ratio strategy ended up selecting the correct answer, it is also possible that, in addition to visualization design, participants' beliefs can also impact whether participants used the ratio strategy or not and whether they subsequently selected the correct answer. To investigate whether the effect of congruency on response accuracy was mediated by the use of ratio strategy, the same mediation analysis was performed, but with congruency as the predictor variable. The average causal mediation effect (i.e. the indirect effect of congruency on response accuracy) was found to be statistically significant ($\text{Effect} = -0.05, p = 0.046, CI = [-0.12, 0.00]$). This means whether the correct answer was congruent with the participants' primed belief impacts whether they would use the ratio strategy or not, which then impacts the likelihood of them selecting the correct answer. Follow-up regression analysis reveals that a participant whose primed belief is congruent with the correct answer is on average 1.82 times more likely to use the ratio strategy compared to participants whose belief is incongruent with the correct answer.

4.6 Discussion

In Experiment 1, as shown in Figure 5, we found that presenting the problem as a table motivates people to use the ratio strategy to reason with data and increases the likelihood of them selecting the correct answer. Presenting the problem using bar tables and bar charts is far less likely to elicit the appropriate ratio strategy, and leads to lower accuracy. When the participants were primed with a belief that is congruent with the correct answer, they were more likely to reason with the ratio strategy to reach the correct answer. A comparison of effect size between experimental factors suggests that the reasoning strategies is the strongest predictor to task accuracy, with belief congruency and visual representation type being the weaker predictors.

There is a relationship between strategy and accuracy, but it is not a one-to-one relationship. The simple comparison strategy usually results in the wrong answer, but can sometimes lead to the correct answer by luck. For example, although heuristic users mostly compared A and C or A and B, a small subset of people compared B and D. This comparison that B is greater than D will lead to the correct answer, even if the participant's rationale is flawed.

Some participants also mentioned multiple comparisons, such as mentioning B and D *and* mentioning A and C. In this case, making both of these comparisons could lead the participant to come to either the right or wrong answer. One participant noted "From the 2 options I felt that both were correct based on the graph shown. I noticed that A had a higher bar than C so I assessed that it meant more rashes got worse from using it than not using it. I also saw that B had a bar higher than D which led me to believe the other option was correct, and that more people who used the cream got better than those who did not. These reasons are why I was not completely sure that my answer was correct, as I didn't expect both to be true. I am still not sure if both are true or if I was wrong."

The ratio strategy can also lead to an incorrect answer if the participant made a calculation error. For example, one participant reasoned: "D is 5 times more likely to occur when patients did not use the skin screen in the group that wasn't treated vs B where patients were only 3 times more likely for the rash to worsen when being in the treatment group." This person should have made the calculation that D is 1/5 the size of C, instead, they calculated that D was 5 times bigger than C. Finally, in the bar condition, some participants noted that it was difficult to tell which ratio was larger: "Without looking at actual numbers, it's hard to see what the ratio is between A and B, and if it's better or worse than the ratio between C and D." This participant was not able to accurately determine the ratio of each pair of numbers because they were not able to precisely compare the bar lengths.

Furthermore, our counterbalancing factor of skin cream effectiveness appears to have an effect on overall task accuracy to be higher in conditions where the skin cream was more effective. In addition to demonstrating the importance of counterbalancing, this observation also suggests that participants more heavily weigh the effect of the absolute larger values in their decisions, both when that value was presented as a visually salient large bar, and when that value was presented in its numeric format. We discuss potential future directions to further investigate this effect in Section 6.

5 Experiment 2

To generalize our results beyond the existing case study of skin cream effectiveness, Experiment 2 presents a new problem on gun control laws that leveraged people's pre-existing opinions on gun control instead of using a prime. In Experiment 1, we primed participants to form an opinion about a skin cream company. Although we didn't observe a strong effect size in difference in accuracy driven by visualization type and belief congruency, we found strong differences in the *reasoning strategies* used across bars and tables. Experiment 2 presents a new

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.2022.3232959

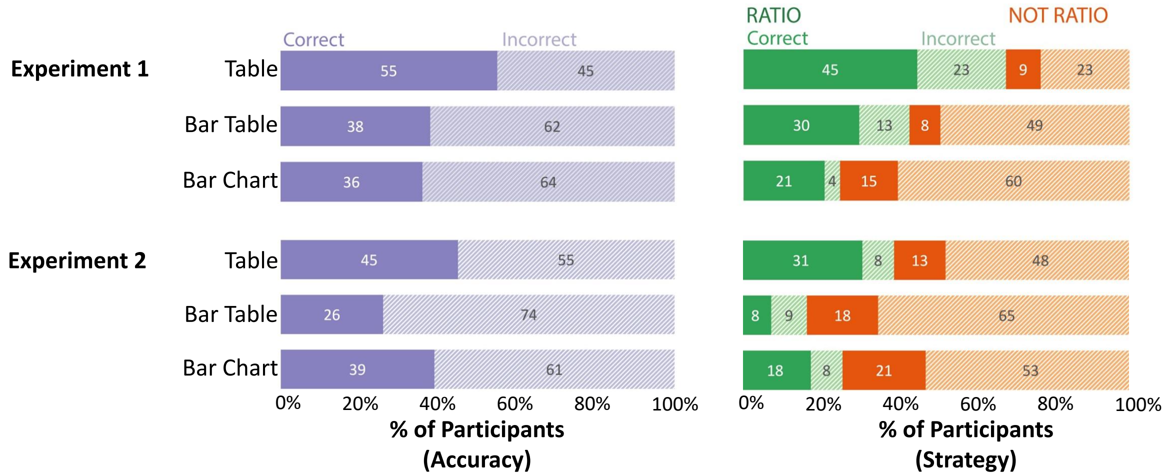


Fig. 5. Experiment 1 and Experiment 2 results. The left chart shows the percentage of participants who were accurate for each condition. The Table condition had higher rates of accuracy. The right graph shows the percentage of participants for each type of strategy. Diagonal line patterns on each color indicate incorrect trials for that strategy. There is a higher percentage of ratio strategies for the tables than the bar charts. Most of the ratio strategies resulted in accurate trials, while most of the 'not ratio' strategies resulted in inaccurate trials.

context to test the effect of data visualization, the effect of congruency between a prior belief and reasoning with a data set, and their interactions. We used a gun legislation problem inspired by previous work in reasoning [34], as shown in Figure 1.

5.1 Participants

We recruited 385 participants from Amazon's Mechanical Turk following the same exclusion criteria as before. After applying exclusion criteria, we were left with 359 participants (M_{age} = 38, $SD_{age} = 12$, 174 female).

5.2 Method and Procedure

We used a between-subject design and randomly assigned participants to solve a reasoning problem presented with either a table, a bar table, or a bar chart (as shown in Figure 1 (bottom row). Participants were asked whether a ban on carrying concealed weapons made it more likely for crime to *increase* or *decrease*. Their judgment on whether the ban on concealed carry was effective or not was coded as accurate if it was supported by the data they viewed. Similar to Experiment 1, they then explained the strategies behind their decisions.

Because people tend to already have prior beliefs on the effectiveness of gun control laws, we did not prime these participants. At the end of the session, we asked them to rate how much they agreed with the statement "greater gun restriction laws are necessary to reduce violence" on a scale from 1 (strongly disagree) and 7 (strongly agree). This self-reported information was used to identify people as being more likely to either believe that stricter gun control laws will reduce (anti-gun) or increase (pro-gun) crime. In our analysis, we coded participant beliefs as 'congruent' if the correct conclusion from the data aligned with their prior belief (e.g., a participant who believes that gun restrictions are necessary to reduce violence saw data showing that the ban on carrying concealed weapons decreased crime rates), and 'incongruent' otherwise (e.g., a participant who believes that gun restrictions are necessary to reduce violence saw data showing that the ban on carrying concealed weapons increased crime rates).

Each display was counterbalanced such that for half of the participants the data suggested that the ban on concealed carry was effective in reducing crime. For the other half of the participants, the data suggested that the ban on concealed carry was ineffective, by swapping the labels on the columns.

5.3 Qualitative Coding of Strategies

We coded the strategies each participant used to make their judgment. The coding structure was the same as in Experiment 1. Our coders agreed 85.4% of the time in their ratings, with a high inter-rater reliability Kappa value of 0.84 ($z = 50.7$, $p < 0.001$). Examples are listed in Table 1 and shown in Figure 2.

5.4 Results Overview

Overall, the reasoning problem we provided was once again proven quite challenging for participants. When the data was presented as a table, 44.5% of the participants got the question correct. The accuracy rate dropped to 26.27% for the bar table and to 38.5% for the bar chart.

We additionally looked at how accuracy might be influenced by the reasoning strategy participants used. Again, although the correct way to reason with this problem is to compare the ratio between two pairs of values, not everyone who used the ratio strategy ended up selecting the correct answer. 70.4% of the participants that used the ratio strategy selected the correct answer, while 29.6% of the participants that used the ratio strategy selected the wrong answer.

Because participants rated how much they agreed with the statement "greater gun restriction laws are necessary to reduce violence" on a scale from 1 to 7, the congruency factor in Experiment 2 is a seven-level variable, as opposed to a binary variable in Experiment 1. This allows us to observe how accuracy and strategy vary with the degree of congruency. We refer to selecting 1 and 7 on the scale as *strongly congruent* or *strongly incongruent*, depending on the data they saw. Selecting 2 and 6 on the scale is referred as *medium congruent* or *medium incongruent*. We referred to selecting 3 and 5 as *weakly congruent* or *weakly incongruent*. Finally, we referred to selecting 4 as *middle*. Most people held strong attitudes towards gun-restriction laws, with 38.71% of the participants falling into the *strongly congruent* and *strongly incongruent* categories. There were very few participants who held a neutral attitude toward gun-restriction laws (8.91%).

We found an overall effect of congruency on accuracy, such that the more congruent the participant's belief was with the ground truth the data depicted, the more likely they selected the correct answer (*strongly incongruent* *strongly congruent*). Specifically, with each unit increase in congruency, the odds of the viewer selecting the correct answer increased by 1.18 folds ($Est = 0.16$, $p = 0.0016$).

5.5 Analysis of Response Accuracy and Strategy

Following similar analysis procedures as that in Experiment 1, we first investigated whether visualization design and belief congruency affected accuracy. We performed a logistic regression model predicting answer accuracy with whether the problem was presented as a table, a bar table, or a bar chart, how strongly the correct answer aligns with

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.2022.3232959

the participants' existing beliefs, their interactions, and the counterbalancing factor of effectiveness (achieved through swapping the labels on the columns). We found a significant effect of visualization type ($\chi^2 = 7.42$, $p = 0.024$) and a significant effect of belief congruency ($\chi^2 = 15.72$, $p < 0.001$), but no significant interactions ($\chi^2 = 2.35$, $p = 0.31$). We also found an effect of the counterbalanced effectiveness ($\chi^2 = 14.01$, $p < 0.001$), such that accuracy was higher by on average 2.44 folds when the gun ban was not effective.

Post-hoc analysis suggests that participants were the most likely to select the correct answer when their belief aligned with the underlying data ($\chi^2 = 15.72$, $p < 0.001$) and when the problem was presented as a table, and the least likely to select the correct answer when the problem was presented as a bar table. Specifically, participants that viewed the table display selected the correct answer on average 1.87 times more often than participants that viewed the bar table display. Participants who viewed the bar table selected the correct answer 1.29 times more often than those that viewed the bar chart.

Next, to account for the different use of strategies, we again added the strategies the participants used into our regression analysis to see if they interacted with visualization design and belief congruency to additionally account for whether people selected the correct answer. We constructed a logistic regression model predicting answer accuracy with visualization design, belief, strategy, and their interactions. We again added the counterbalancing factor of effectiveness as a predictor. Similar to that in Experiment 1, there was a significant effect of strategy used on accuracy ($\chi^2 = 51.07$, $p < 0.001$), and visualization design no longer significantly impacted accuracy once we accounted for the strategy they used ($\chi^2 = 2.50$, $p = 0.29$). There was again a significant effect of effectiveness ($\chi^2 = 9.55$, $p < 0.01$), which suggests that counterbalancing the column labels was a critical design component.

We conducted a similar mediation analysis predicting response accuracy with visualization design, mediated by whether the participant used a ratio strategy. We tested the indirect effect using bootstrapping procedures. The indirect effect was computed for each of the 1000 bootstrapped samples. The average causal mediation effect (i.e. the indirect effect of visualization type on response accuracy) was found to be statistically significant (Effect $= 0.11$, $p < 0.002$, $CI = [0.03, 0.19]$). Consistent with our results from Experiment 1, the visualization design in which the problem was presented elicited differing strategies and subsequently impacted the likelihood of them selecting the correct answer. Follow-up regression analysis revealed that participants who viewed the problem presented as a table were on average 1.77 times more likely to use the ratio strategy compared to those that viewed the bar chart ($p = 0.04$), and 3.09 times more likely compared to those that viewed the bar table ($p < 0.001$). Although not statistically significant, participants were also on average 1.74 times more likely to use the correct ratio strategy when they viewed the bar chart compared to the bar table ($p = 0.08$).

We also explored the relationship between congruency and strategy using a mediation analysis, similar to our approach to analyzing results from Experiment 1. We predicted whether the participant selected the correct answer with belief congruency, using whether the participant used a ratio strategy as the mediator. Interestingly, we didn't find a significant causal mediation effect this time (Effect $= 0.01, p = 0.85, CI = [-0.10, 0.13]$). This means whether the correct answer was congruent with the participants' belief did not have an influence on whether they would use the ratio strategy or not. In other words, participants were not more likely to reason with the ratio strategy to reach the correct answer when their belief was congruent with the correct answer ($\chi^2 = 3.92$, $p = 0.14$).

5.6 Discussion

In summary, Experiment 2 replicated the finding from Experiment 1, where presenting the data as a table made it more likely for people to use the optimal ratio strategy. This in turn increased the likelihood of them selecting the correct answer. Showing the data using bar tables and bar charts led to overall lower accuracy, as they did not elicit the appropriate ratio strategy as strongly as the tables did. While Experiment 1, which relied on priming a belief, did not find congruency effects, Experiment 2, which relied on pre-existing beliefs, revealed an effect of belief congruency, such that participants were more likely to select the correct answer when the correct answer aligned with their belief. A comparison of effect size between experimental factors suggests that the reasoning strategies is the strongest predictor of task accuracy, followed by belief congruency, and visualization type being the weakest.

Participants in Experiment 2 also sometimes used ideological justifications for their decision as they attempted to solve the problem, instead of (or in addition to) referencing the data. For example, one participant mentioned "The data shows that cities that did ban handguns showed an increase in crime. This is likely due to rights being taken away." This shows that the context did have an effect on reasoning, with prior beliefs consciously (and potentially unconsciously) influencing what parts of the data people attended to and favored in their judgments. This also suggests that participants were driven to draw conclusions from data that were consistent with their beliefs.

Furthermore, we once again found an effect of the counterbalancing factor of effectiveness. But the direction of difference is opposite of that in Experiment 1. In Experiment 1, participants were more likely to get the answer correct when the underlying data shows the skin cream to be *effective*, whereas, in Experiment 2, participants were more likely to get the answer correct when the data shows that the gun policy was *not* effective. This suggests that this effect of the counterbalanced factor of effectiveness through label switching was more driven by the effect of the chart topic rather than a generalizable effect of labeling. We further discuss this in Section 6.

6 General Discussion and Future Directions

Contrary to our expectations, we did not find an effect of bar charts and bar tables in mitigating confirmation bias with our reasoning paradigm. However, we did find a strong effect of tables eliciting complex ratio reasoning strategies and observed confirmation bias in a real-world scenario, which supports **Hypothesis 1 and 2**, which we discuss in detail below.

6.1 Visualization Design and Reasoning Affordances

We found consistent results for the reasoning affordances of visual displays across both experiments. The two experiments reported here show that tables tended to elicit ratio strategies that helped participants reach the correct answer, while bar tables and bar charts did not. Exploration and coding of strategy reports revealed that for visual displays, participants tended to use quicker heuristic strategies, such as comparing the size of two bars, while the table format led participants toward the more complex strategy of considering relations among multiple data values.

Specifically, people who viewed tables were approximately *2 times more likely* to use the correct ratio strategy than people who viewed bar charts. These findings suggest that data presentation formats carry affordances for reasoning. When designers create visualizations to help people reason and make decisions, designers should think about the reasoning affordances of a visualization and pick the data displays that best afford the 'right' strategy to think about a particular problem. But why do these data displays afford different strategies?

Numeric Labels: One possibility is that the bar tables and bar charts do not contain numbers, while the absolute values presented in the table could provide additional context for the data. But, numbers are not required to determine whether the skin cream or the gun ban was effective – that answer always relied on a comparison between the *relative* sizes of values, or more specifically, the more complex comparison of those relative sizes.

Limitations and Future Directions: The current experimental design does not tease apart the specific effect of labels, but rather, focused on the strategies people used. It is possible that the mere presence of numerical labels triggers people to compute ratios. For example, people are more familiar with performing calculations on a series of numbers, and tables may be more associated with the type of strategic arithmetic practiced in school. Visual displays may be more typically used to perceive 'packaged' insights, and be less associated with the work

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.2022.3232959

needed to calculate insights. While existing work has demonstrated that numeric labels may increase the perceptual accuracy of visualized data [33], future work can further test the effect of the presence of numbers in a bar chart or bar table on people's reasoning strategies, which may lead to guidelines pointing to the potential benefits of labeling values in bar charts. For example, researchers can compare the accuracy in which people make ratio comparisons when looking at numeric tables versus bar charts or bar tables.

Effect of Arrangements: Perhaps the bar displays tempted participants with the salience of the largest sized bar, inviting a comparison of that cell to another [85, 87]. In contrast, for tables, all of the numbers were roughly the same physical size on the screen. Although the larger numerical values in the tables also attracted viewer attention and lead to higher weighting during the decision-making process in Experiment 1, this effect seems to be topic-dependent, as evident from the effect of the counterbalancing factor discussed in Experiment 1 turned out to be the opposite from that in Experiment 2. Participants were overall more consistently influenced by the larger value when they see it visually presented as bigger bars, and less consistently so when that larger value is presented numerically in a table. Here, the speed of graphical processing might actually lead participants to seek out one easily-extracted pattern, and that pattern's ease or salience could lock the viewer into a decision [24]. In contrast, the prevalence of the ratio strategy in the table condition may be higher because extracting data from text forces participants to slow down their thinking.

Limitations and Future Directions: Although we tested the effect of bar display type, we did not examine how data arrangement might affect the decisions people make. In the present study, larger values are always placed on the left. Placing the larger values on the right or alternating between having large values on the left and right across separate rows might lead to differing effects. Future researchers should further explore the effects of visual arrangement on viewer reasoning and decisions making.

Mitigating Bias with Visualizations: In the present study, although confirmation bias was present only when the participants held pre-existing beliefs relevant to the data they saw, participants reading both tables and bars/bar tables were more likely to select the answer that is congruent with their belief. Confirmation bias was less present in tables only to the degree that people were more likely to use a better overall reasoning strategy (ratio comparisons) when looking at tables.

Limitations and Future Directions: Although showing the data in a bar chart or a bar table did not mitigate confirmation bias, using a visual depiction may still hold the potential to mitigate biases. Future work could test whether other types of visualizations might mitigate this bias – in particular, icon arrays [25], which can improve statistical thinking about risk. Pie charts or stacked bars might more intuitively convey the data as parts of a whole, facilitating the ratio strategy for the same reason why showing the "proportion of people whose rashes got better" and normalizing each group to 100% might help. Finally, perhaps less-familiar visualizations could combine the advantage of visual processing power, but still encourage a slower reasoning process. Future work might also test whether other tabular or verbal representations (e.g., text explaining the numbers in a 'word problem') might similarly slow down reasoning, in a beneficial way.

Cognitive Framework for Reasoning As discussed in Section 2, the dual-processing model proposes a fast, impulsive system and a slower, deliberate system [55, 58, 91]. When shown the display on the left of Figure 1, a faster process might automatically determine that this is a table, that there are four numbers, and to isolate salient numbers, e.g. noticing that the top-left cell has the biggest number and the bottom right cell has the smallest number. Visualization displays could influence the processing mode for decision-making, potentially encouraging a fast or a slow process. On one hand, bar charts and bar tables can help offload cognitive tasks to the perceptual system, which might free up cognitive resources for people to more thoroughly process information and make more analytic decisions [21]. On the other hand, because visualizations tend to speed up information processing, it might lead the viewer to rely more on intuitions and heuristics, rather than effortful, slow, and analytical thinking.

Limitations and Future Directions: Future work can explore the underlying cognitive mechanism in data interpretation when people work with bar charts versus tables. For example, it is possible that bars trigger the intuitive, System 1 processing, and tables trigger the more analytic System 2 processing. Because system 2 processing is typically associated with slower processing speed and higher cognitive load, future researchers can measure or control for time spent on the decision task, or experiment with the amount of cognitive load on the participant while doing the reasoning task.

Data Complexity and Task Difficulty: In the present experiment, we only tested a simple dataset with four values in the current experiment following a commonly used paradigm from cognitive psychology [34]. Although this task has been carefully designed by psychologists to be difficult enough to afford multiple reasoning strategies, including less effortful but incorrect 'heuristics' (e.g., computing the deltas) and more effortful but correct strategies of comparing ratios, it is possible that replacing this dataset with other ones varying in difficulty and complexity might generate new insights into confirmation bias and mitigation strategies.

Limitations and Future Directions: Future work can test other datasets to cover a wider range of ratio values and combinations or expand the dataset in complexity to test for the generalizability of the effect. For example, analysts may interact with high-dimensional datasets in their workflow and the added complexity might create new pathways for confirmation bias and potential mitigation strategies.

7 Insights on Experimental Design with Primed versus Pre-Existing Beliefs

The present studies provide preliminary evidence for stronger congruency effects for pre-existing beliefs compared to primed beliefs. Although there was no interaction effect between visualization displays and belief congruency, primed beliefs were likely to push people toward the ratio strategy in Experiment 1. When belief was primed, participants were more likely to use the ratio strategy but were not more likely to select the answer that is congruent with their belief. One possible reason for this is that participants are willing to "work harder" to achieve the answer that they suspected was correct, but make sufficient calculation errors that this route doesn't always lead them to the correct answer. In contrast, pre-existing belief had no effect on the strategies used in Experiment 2, but it affected problem-solving accuracy. Participants were more likely to select the answer that was congruent with their belief but did not change their strategies depending on that belief.

Together, these results suggest that people rely more on their pre-existing beliefs than a primed belief in information processing and decision-making. When a task requires the viewer to reason with data, if the task is strongly associated with a pre-existing belief, the viewer might rely so strongly on their belief that they fail to fully engage with the actual data. But if the task is only weakly associated with a recently primed belief, the congruent primed belief makes people more likely to choose the complex ratio strategy to reason with the data, rather than jumping to a solution that simply aligns with that belief. Visualization researchers should consider eliciting the type of qualitative user input analyzed here in similar future studies. Because a pre-existing belief can drive heuristic strategies, participants might rely on previous beliefs instead of only processing the provided data.

Limitations in Current Study and Future Directions: One future approach to further tease apart the role of belief in reasoning strategies might be looking at alternative ways to capture participants' reasoning strategies, to capture the trade-off between the strength of one's belief and their effort to find belief-congruent evidence in data. In the present experiment, we coded open-ended responses to infer the strategies of participants. While we are confident that these codes provide a reliable way to differentiate strategy affordances, future work might explore using more detailed strategy reports such as sketches (as in Figure 2), formal arithmetic expressions, or talk-aloud interviews that unpack strategies in more detail, including why they were chosen, and what alternatives might exist.

Along with accuracy and reasoning, future work could also record other dependent measures to see if they can give us further insight as to what features of the display encourage multiple types of reasoning. In future experiments, we could ask people to explicitly describe their calculations of the ratios, in order to dissociate calculation errors from classifications of intended strategies.

Additionally, in the present experiment, people's belief was captured at the end of the study via a Likert scale. Although existing work has demonstrated that belief measured on Likert scales before and after an experiment (that is not aimed at changing beliefs) are extremely similar [86], future work should consider belief measurements before *and* after the experiment to control for any potential influence of the experiment itself. Furthermore, future work can investigate other approaches to measuring belief, such as drawing or verbalizing beliefs (for more possible metrics, see [40, 44, 45]).

8 Potential Design Guidelines

In this study, we found that tables are far more likely to elicit the correct ratio strategy, which leads to the correct answer more often, compared to showing the same data in bar charts which elicit a heuristic strategy that tend to lead to the incorrect answer. In the qualitative responses, we found that many people did mathematical calculations using the numbers in the table. This feature was not included in our bar displays because we wanted to restrict the bar tables and the bar charts to a purely visual calculation. Given the low task accuracy in the graphical displays, one clear prescription would be to explicitly encode the critical numerical information if the design requires the visualization reader to compute and compare ratios or conduct similar high-numeracy tasks, corroborating existing work [33]. In our work, participants would likely perform with much higher accuracy if the ratios between neighboring numbers were explicitly shown as percentages or if the differences were visualized, instead of requiring the reader to calculate them.

It's important to note that these results are not suggesting that visualization designers and practitioners stop using bar charts and bar tables to prevent falling victim to confirmation bias. While tables are associated with less confirmation bias, this relationship is mediated by the reasoning strategy rather than a causal connection. In other words, with simple datasets, tables tend to afford a complex ratio comparison reasoning strategy. Based on this, we encourage designers and practitioners to be more intentional in their design and potentially mitigate confirmation bias by helping people make complex comparisons and supporting the 'right' reasoning strategies.

For example, imagine a scenario where the dataset is large and multi-dimensional. People looking at the data in a table format will likely *not* use the ratio comparison reasoning strategy, but rather be confused and overwhelmed by the many rows and columns. In this case, designers should think about how to help people find patterns and think more analytically with visualizations, which have the power to offload effortful relation comparisons and ratio computations to the perceptual system.

9 Conclusions

We compared the effectiveness of table, bar table, and bar charts in facilitating critical thinking with an established reasoning paradigm. We qualitatively examined the strategies that viewers employed when viewing these displays and investigated the affordances of tabular and bar displays. Overall, we found that participants were far more likely to use a more difficult – but more accurate – reasoning strategy when viewing a table compared to bar tables and bar charts, supporting Hypothesis 1 and 2. We additionally explored the effect of participants' prior beliefs on the strategy they used to reason with data and explored whether beliefs impacted how participants interacted with different types of visualizations. In Experiment 1, we imparted beliefs with a story prime to create a controlled environment that minimized any effects of participants' existing knowledge and beliefs. In Experiment 2, we intentionally chose a political topic typically associated with strong prior beliefs to increase the ecological validity and generalizability of our findings. We found that primed beliefs impacted how participants reasoned, and pre-existing beliefs influenced the final decisions that people made.

Acknowledgments

The authors wish to thank the members of the Visual Thinking Lab and Akira Wada for their feedback and helpful comments. This work was supported by NSF award CHS-1901485.

References

[1] K. Ajani, E. Lee, C. Xiong, C. N. Knaflic, W. Kemper, and S. Franconeri. Declutter and focus: Empirically evaluating design guidelines for effective data communication. *IEEE Transactions on Visualization & Computer Graphics*, (01):1–1, 2021.

[2] E. B. Andrade. Excessive confidence in visually-based estimates. *Organizational Behavior and Human Decision Processes*, 116(2):252–261, 2011.

[3] A. D. Baddeley and D. Longman. The influence of length and frequency of training session on the rate of learning to type. *Ergonomics*, 21(8):627–635, 1978.

[4] L. Bartram, M. Correll, and M. Tory. Untidy data: The unreasonable effectiveness of tables. *arXiv preprint arXiv:2106.15005*, 2021.

[5] K. Binder, S. Krauss, and G. Bruckmaier. Effects of visualizing statistical information–an empirical study on tree diagrams and 2× 2 tables. *Frontiers in psychology*, 6:1186, 2015.

[6] A. Burns, C. Xiong, S. Franconeri, A. Cairo, and N. Mahyar. How to evaluate data visualizations across different levels of understanding. In *2020 IEEE Workshop on Evaluation and Beyond-Methodological Approaches to Visualization (BELIV)*, pages 19–28. IEEE, 2020.

[7] Z. Bylinskii, N. W. Kim, P. O'Donovan, S. Alsheikh, S. Madan, H. Pfister, F. Durand, B. Russell, and A. Hertzmann. Learning visual importance for graphic designs and data visualizations. In *Proceedings of the 30th Annual ACM symposium on user interface software and technology*, pages 57–69, 2017.

[8] M. Card. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[9] S. Chaiken. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of personality and social psychology*, 39(5):752, 1980.

[10] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387):531–554, 1984.

[11] M. B. Cook and H. S. Smallman. Human factors of the confirmation bias in intelligence analysis: Decision support from graphical evidence landscapes. *Human Factors*, 50(5):745–754, 2008.

[12] E. Dimara, G. Bailly, A. Bezerianos, and S. Franconeri. Mitigating the attraction effect with visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):850–860, 2018.

[13] E. Dimara, A. Bezerianos, and P. Dragicevic. The attraction effect in information visualization. *IEEE transactions on visualization and computer graphics*, 23(1):471–480, 2016.

[14] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, and P. Dragicevic. A task-based taxonomy of cognitive biases for information visualization. *IEEE transactions on visualization and computer graphics*, 2018.

[15] P. Dragicevic and Y. Jansen. Blinded with science or informed by charts? a replication study. *IEEE transactions on visualization and computer graphics*, 24(1):781–790, 2017.

[16] G. Ellis. *Cognitive Biases in Visualizations*. Springer, 2018.

[17] G. Ellis and A. Dix. Decision making under uncertainty in visualisation? In *IEEE VIS2015*, 2015.

[18] J. S. B. Evans and K. E. Stanovich. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, 8(3):223–241, 2013.

[19] D. Feldman-Stewart, N. Kocovski, B. A. McConnell, M. D. Brundage, and W. J. Mackillop. Perception of quantitative information for treatment decisions. *Medical Decision Making*, 20(2):228–238, 2000.

[20] S. L. Franconeri. The nature and status of visual resources. 2013.

[21] S. L. Franconeri. Three perceptual tools for seeing and understanding visualized data. *Current Directions in Psychological Science*, in press.

[22] 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.

[23] S. L. Franconeri, J. M. Scimeca, J. C. Roth, S. A. Helseth, and L. E. Kahn. Flexible visual processing of spatial relationships. *Cognition*, 122(2):210–227, 2012.

[24] A. Gaba, V. Setlur, A. Srinivasan, J. Hoffswell, and C. Xiong. Comparison conundrum and the chamber of visualizations: An exploration of how language influences visual design. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[25] R. Garcia-Retamero and M. Galesic. Communicating treatment risk reduction to people with low numeracy skills: a cross-cultural comparison. *American journal of public health*, 99(12):2196–2202, 2009.

[26] G. Gigerenzer and U. Hoffrage. How to improve bayesian reasoning without instruction: frequency formats. *Psychological review*, 102(4):684, 1995.

[27] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011.

[28] L. Guohua. Pilot-related factors in aircraft crashes: a review of epidemiologic studies. *Aviation, Space, and Environmental Medicine*, 65:944–952, 1994.

[29] S. T. Hawley, B. Zikmund-Fisher, P. Ubel, A. Jancovic, T. Lucas, and A. Fagerlin. The impact of the format of graphical presentation on health-related knowledge and treatment choices. *Patient education and counseling*, 73(3):448–455, 2008.

[30] M. Hegarty, M. S. Canham, and S. I. Fabrikant. Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1):37, 2010.

[31] J. Heiser and B. Tversky. Arrows in comprehending and producing mechanical diagrams. *Cognitive science*, 30(3):581–592, 2006.

[32] E. T. Higgins, J. A. Bargh, and W. J. Lombardi. Nature of priming effects on categorization. *Journal of experimental psychology: Learning, Memory, and Cognition*, 11(1):59, 1985.

[33] J. K. Hink, M. S. Wogalter, and J. K. Eustace. Display of quantitative information: are grables better than plain graphs or tables? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 40, pages 1155–1159. SAGE Publications Sage CA: Los Angeles, CA, 1996.

[34] D. M. Kahan, E. Peters, E. C. Dawson, and P. Slovic. Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, 1(1):54–86, 2017.

[35] D. Kahneman. *Thinking, fast and slow*. Macmillan, 2011.

[36] A. Kappes, A. H. Harvey, T. Lohrenz, P. R. Montague, and T. Sharot. Confirmation bias in the utilization of others' opinion strength. *Nature neuroscience*, 23(1):130–137, 2020.

[37] A. Karduni, D. Markant, R. Wesslen, and W. Dou. A bayesian cognition approach for belief updating of correlation judgement through uncertainty visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):978–988, 2020.

[38] A. Khan, S. Breslav, M. Glueck, and K. Hornbæk. Benefits of visualization in the mammography problem. *International Journal of Human-Computer Studies*, 83:94–113, 2015.

[39] C. P. Killen, J. Geraldi, and A. Kock. The role of decision makers' use of visualizations in project portfolio decision making. *International Journal of Project Management*, 38(5):267–277, 2020.

[40] Y.-S. Kim, K. Reinecke, and J. Hullman. Explaining the gap: Visualizing one's predictions improves recall and comprehension of data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1375–1386, 2017.

[41] Y.-S. Kim, L. A. Walls, P. Krafft, and J. Hullman. A bayesian cognition approach to improve data visualization. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pages 1–14, 2019.

[42] J. Klayman. Varieties of confirmation bias. *Psychology of learning and motivation*, 32:385–418, 1995.

[43] J. H. Larkin and H. A. Simon. Why a diagram is (sometimes) worth ten thousand words. *Cognitive science*, 11(1):65–100, 1987.

[44] S. Mahajan, B. Chen, A. Karduni, Y.-S. Kim, and E. Wall. Vibe: A design space for visual belief elicitation in data journalism. 2022.

[45] P. Mantri, H. Subramonyam, A. L. Michal, and C. Xiong. How do viewers synthesize conflicting information from data visualizations? *IEEE transactions on visualization and computer graphics*, 2022.

[46] R. Mendel, E. Traut-Mattausch, E. Jonas, S. Leucht, J. M. Kane, K. Maino, W. Kissling, and J. Hamann. Confirmation bias: why psychiatrists stick to wrong preliminary diagnoses. *Psychological Medicine*, 41(12):2651–2659, 2011.

[47] J. Meyer, M. K. Shamo, and D. Gopher. Information structure and the relative efficacy of tables and graphs. *Human Factors*, 41(4):570–587, 1999.

[48] L. Micallef, P. Dragicevic, and J.-D. Fekete. Assessing the effect of visualizations on bayesian reasoning through crowdsourcing. *IEEE transactions on visualization and computer graphics*, 18(12):2536–2545, 2012.

[49] A. L. Michal and S. L. Franconeri. Visual routines are associated with specific graph interpretations. *Cognitive Research: Principles and Implications*, 2(1):1–10, 2017.

[50] D. Norman. *The design of everyday things: Revised and expanded edition*. Basic books, 2013.

[51] C. Nothelfer and S. Franconeri. Visual processing of spatial relations within and between objects. *Journal of Vision*, 18(10):321–321, 2018.

[52] C. Nothelfer and S. Franconeri. Measures of the benefit of direct encoding of data deltas for data pair relation perception. *IEEE transactions on visualization and computer graphics*, 26(1):311–320, 2019.

[53] A. Ottley, E. M. Peck, L. T. Harrison, D. Afergan, C. Ziemkiewicz, H. A. Taylor, P. K. Han, and R. Chang. Improving bayesian reasoning: The effects of phrasing, visualization, and spatial ability. *IEEE transactions on visualization and computer graphics*, 22(1):529–538, 2015.

[54] E. O'Sullivan and S. Schofield. Cognitive bias in clinical medicine. *JR Coll Physicians Edinb*, 48(3):225–32, 2018.

[55] 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):29, 2018.

[56] A. V. Pandey, A. Manivannan, O. Nov, M. Satterthwaite, and E. Bertini. The persuasive power of data visualization. *IEEE transactions on visualization and computer graphics*, 20(12):2211–2220, 2014.

[57] E. Pariser. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin, 2011.

[58] R. E. Patterson, L. M. Blaha, G. G. Grinstein, K. K. Liggett, D. E. Kaveney, K. C. Sheldon, P. R. Havig, and J. A. Moore. A human cognition framework for information visualization. *Computers & Graphics*, 42:42–58, 2014.

[59] E. Picon and D. Odic. Finding maximal and minimal elements in a set is capacity-unlimited and massively-parallel. *Journal of Vision*, 17(10):1284–1284, 2017.

[60] B. Saket, A. Endert, and C. Demiralp. Task-based effectiveness of basic visualizations. *IEEE transactions on visualization and computer graphics*, 2018.

[61] J. Schaubroeck and K. Muralidhar. A meta-analysis of the relative effects of tabular and graphic display formats on decision-making performance. *Human Performance*, 4(2):127–145, 1991.

[62] P. Shah and E. G. Freedman. Bar and line graph comprehension: An interaction of top-down and bottom-up processes. *Topics in cognitive science*, 3(3):560–578, 2011.

[63] M. Snyder and N. Cantor. Testing hypotheses about other people: The use of historical knowledge. *Journal of Experimental Social Psychology*, 15(4):330–342, 1979.

[64] J. Stasko, C. Görg, and Z. Liu. Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization*, 7(2):118–132, 2008.

[65] P. T. Sukumar and R. Metoyer. A visualization approach to addressing reviewer bias in holistic college admissions. In *Cognitive Biases in Visualizations*, pages 161–175. Springer, 2018.

[66] D. A. Szafir, S. Haroz, M. Gleicher, and S. Franconeri. Four types of ensemble coding in data visualizations. *Journal of vision*, 16(5):11–11, 2016.

[67] A. Tal and B. Wansink. Blinded with science: Trivial graphs and formulas increase ad persuasiveness and belief in product efficacy. *Public Understanding of Science*, 25(1):117–125, 2016.

[68] J. Talbot, V. Setlur, and A. Anand. Four experiments on the perception of bar charts. *IEEE transactions on visualization and computer graphics*, 20(12):2152–2160, 2014.

[69] D. Tingley, T. Yamamoto, K. Hirose, L. Keele, and K. Imai. Mediation: R package for causal mediation analysis. 2014.

[70] J. Tsai, S. Miller, and A. Kirlik. Interactive visualizations to improve bayesian reasoning. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 55, pages 385–389. SAGE Publications Sage CA: Los Angeles, CA, 2011.

[71] A. Tversky and D. Kahneman. Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics*, 106(4):1039–1061, 1991.

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.2022.3232959

[72] B. Tversky. Visualizing thought. In *Handbook of human centric visualization*, pages 3–40. Springer, 2014.

[73] A. C. Valdez, M. Ziefle, and M. Sedlmair. Priming and anchoring effects in visualization. *IEEE transactions on visualization and computer graphics*, 24(1):584–594, 2017.

[74] F. Vallée-Tourangeau, T. Payton, and R. A. Murphy. The impact of presentation format on causal inferences. *European Journal of Cognitive Psychology*, 20(1):177–194, 2008.

[75] E. Wall, J. Stasko, and A. Endert. Toward a design space for mitigating cognitive bias in vis. In *2019 IEEE Visualization Conference (VIS)*, pages 111–115. IEEE, 2019.

[76] P. C. Wason. Reasoning about a rule. *Quarterly journal of experimental psychology*, 20(3):273–281, 1968.

[77] E. A. Wasserman, W. Dorner, and S. Kao. Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(3):509, 1990.

[78] L.-C. Winter. Mitigation and prediction of the confirmation bias in intelligence analysis. 2017.

[79] J. M. Wolfe. Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, 1(2):202–238, 1994.

[80] J. M. Wolfe. What can 1 million trials tell us about visual search? *Psychological Science*, 9(1):33–39, 1998.

[81] W. Wright, D. Schroh, P. Proulx, A. Skaburskis, and B. Cort. The sandbox for analysis: concepts and methods. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 801–810. ACM, 2006.

[82] C. Xiong, A. Sarvghad, D. G. Goldstein, J. M. Hofman, and Ç. Demiralp. Investigating perceptual biases in icon arrays. In *CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2022.

[83] C. Xiong, V. Setlur, B. Bach, E. Koh, K. Lin, and S. Franconeri. Visual arrangements of bar charts influence comparisons in viewer takeaways. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):955–965, 2021.

[84] 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.

[85] C. Xiong, C. Stokes, and S. Franconeri. Visual salience and grouping cues guide relation perception in visual data displays. *Journal of Vision*, 21(9):2095–2095, 2021.

[86] C. Xiong, C. Stokes, Y.-S. Kim, and S. Franconeri. Seeing what you believe or believing what you see? belief biases correlation estimation. 2022.

[87] C. Xiong, C. Stokes, A. Lovett, and S. Franconeri. What does the chart say? grouping cues guide viewer comparisons and conclusions in visualizations. 2022.

[88] C. Xiong, L. van Weelden, and S. Franconeri. The curse of knowledge in visual data communication. *IEEE transactions on visualization and computer graphics*, 2019.

[89] Y. Yi. Cognitive and affective priming effects of the context for print advertisements. *Journal of advertising*, 19(2):40–48, 1990.

[90] J. Zacks and B. Tversky. Bars and lines: A study of graphic communication. *Memory & Cognition*, 27(6):1073–1079, 1999.

[91] J. M. Zacks and S. L. Franconeri. Designing graphs for decision-makers. *Policy Insights from the Behavioral and Brain Sciences*, 7(1):52–63, 2020.

Cindy Xiong is an Assistant Professor in the College of Information and Computer Sciences at UMass Amherst. By investigating how humans perceive and interpret visualized data, she aims to improve visualization design, data storytelling, and data-driven decision-making.

Elsie Lee-Robbins is a PhD candidate in the School of Information at the University of Michigan, Ann Arbor. Her research interests are data visualization, human-computer interaction, and decision-making.

Icy Zhang is a PhD student in Developmental Psychology at UCLA. She is interested in the role of embodied learning in supporting students' understanding of difficult content such as statistics and applying these theory-based interventions to real classrooms.

Aimen Gaba is a PhD student in the College of Information and Computer Sciences at UMass Amherst. Her research interests lie at the intersection of Human-Computer Interaction and Data Visualization. She investigates how to make visualization recommendation systems that use natural language queries better.

Steven Franconeri is a Professor of Psychology at Northwestern University and Director of the Northwestern Cognitive Science Program. He studies visual thinking and visual communication, across psychology, education, and information visualization.