A review of uncertainty visualization errors: Working memory as an explanatory theory

Lace Padilla[a,*], Spencer C. Castro[b], and Helia Hosseinpour[a]
[a]Cognitive and Information Sciences Department, University of California Merced, Merced, CA, United States
[b]Management of Complex Systems Department, University of California Merced, Merced, CA, United States
[*]Corresponding author: e-mail address: lace.padilla@ucmerced.edu

Contents

Abstract

Uncertainty communicators often use visualizations to express the unknowns in data, statistical analyses, and forecasts. Well-designed visualizations can clearly and effectively convey uncertainty, which is vital for ensuring transparency, accuracy, and scientific credibility. However, poorly designed uncertainty visualizations can lead to misunderstandings of the underlying data and result in poor decision-making. In this chapter, we present a discussion of errors in uncertainty visualization research and current approaches to evaluation. Researchers consistently find that uncertainty visualizations requiring mental operations, rather than judgments guided by the visual system, lead to more errors. To summarize this work, we propose that increased working memory

Psychology of Learning and Motivation
ISSN 0079-7421
https://doi.org/10.1016/bs.plm.2021.03.001

demand may account for many observed uncertainty visualization errors. In particular, the most common uncertainty visualization in scientific communication (e.g., variants of confidence intervals) produces systematic errors that may be attributable to the application of working memory or lack thereof. To create a more effective uncertainty visualization, we recommend that data communicators seek a sweet spot in the working memory required by various tasks and visualization users. Further, we also recommend that more work be done to evaluate the working memory demand of uncertainty visualizations and visualizations more broadly.

1. Introduction

From simple analyses, such as those used in introductory statistics textbooks, to the complex forecasts of pandemic projection models, uncertainty presents a difficult challenge for those seeking to represent and interpret it. Uncertainties that can arise throughout a modeling and analysis pipeline (Pang, Wittenbrink, & Lodha, 1997) are of interest to many fields. To constrain the complex category of uncertainty to its component parts, scholars commonly distinguish between several types of uncertainty: ontological (uncertainty created by the accuracy of the subjectively described reality depicted in the model), epistemic (limited knowledge producing uncertainty), and aleatoric (inherent irreducible randomness of a process; Spiegelhalter, 2017). Additionally, quantified forms of aleatoric and epistemic uncertainty are referred to as *risk* in decision-making domains (Knight, 2012). In this chapter, we define uncertainty to encompass quantifiable and visualizable uncertainty, such as a probability distribution.

Many people have difficulty reasoning with even simple forms of uncertainty (Gal, 2002). One study found that 16–20% of 463 college-educated participants could not correctly answer the question, "Which represents the larger risk: 1%, 5%, or 10%?" (Lipkus, Samsa, & Rimer, 2001). Other work finds that even experts with training in statistics commonly misunderstand how to interpret statistical significance from frequentist 95% confidence intervals (Belia, Fidler, Williams, & Cumming, 2005). These findings—that even simple forms of uncertainty are challenging for college graduates and statisticians to understand—should concern both the scientific community and society. We should be concerned because we all make both small- and large-scale decisions with uncertainty throughout our lives, such as picking stocks to invest in or evaluating our pandemic risk.

In the context of textual expressions of uncertainty, researchers propose that people have difficulty understanding probabilities when expressed as a percent (e.g., 10% chance of rain), because this framing is not how we experience probabilities in our daily lives (Gigerenzer & Hoffrage, 1995). A substantial body of research demonstrates that if we express uncertainty in the form of frequency (e.g., it will rain 1 of 10 times), the representation becomes more intuitive (e.g., Gigerenzer, 1996, 2008; Gigerenzer & Gaissmaier, 2011; Gigerenzer & Hoffrage, 1995; Gigerenzer, Todd, & ABC Research Group, 2000; Hoffrage & Gigerenzer, 1998). This line of inquiry takes the perspective that humans can effectively reason with uncertainty if, and only if, the information is presented in an intuitive way.

In addition to research on textural expressions of uncertainty, a large body of evidence demonstrates that communicating uncertainty visually can help people make more effective judgments about risk (for reviews see, Kinkeldey, MacEachren, Riveiro, & Schiewe, 2017; Kinkeldey, MacEachren, & Schiewe, 2014; Maceachren et al., 2005; Padilla, Kay, & Hullman, 2021). Researchers propose that visualizations leverage the substantial processing power of the visual system (Zacks & Franconeri, 2020), recruiting roughly half of the brain (Van Essen, Anderson, & Felleman, 1992). Visualizations allow a viewer's visual system to complete some complex processing efficiently, such as pattern recognition and data comparisons (Szafir, Haroz, Gleicher, & Franconeri, 2016), which would be more challenging to do mathematically. The power and efficiency of the visual system creates an advantage for visualizations over textual expressions of uncertainty. For example, consider how long it takes to read about the following two treatments and how challenging it is to decide which is riskier.

Treatment A: 3 of 10 patients have side effects.

Treatment B: 6 of 45 patients have side effects.

Now consider the same comparison of treatments but visualized using the icon array in Fig. 1.

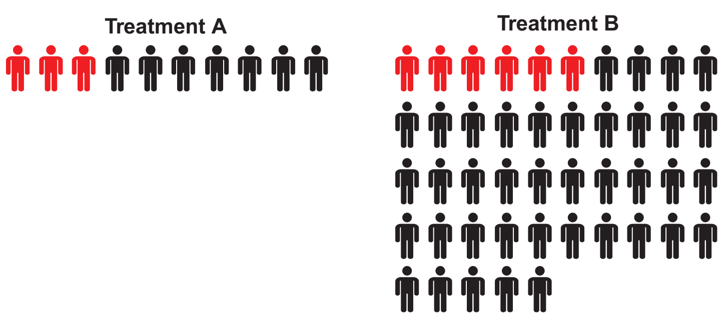


Fig. 1 Icon arrays showing the proportion of patients with side effects in red after receiving hypothetical treatments A or B.

The red icons in Treatment A represent a larger portion of side effects than the red icons in Treatment B. Icon arrays afford visual comparisons that are relatively quick and easy for the visual system to compute, using Gestalt grouping principles, which we will discuss in the *Early-Stage Processing Errors* section. The visual comparison process above does not necessarily require any mathematical calculation. A viewer can arrive at the correct answer, that treatment A is riskier than B, by visually comparing the proportion of side effects for each treatment and determining that A is larger. The viewer does not need to calculate the exact proportions to accomplish this task. Researchers have extensively studied icon arrays in the context of health care communication; they find that icon arrays consistently help people understand probabilities of risk and can be easier to understand than textual representations of probabilities (for reviews, see Fagerlin, Zikmund-Fisher, & Ubel, 2011; Garcia-Retamero & Cokely, 2017; Waters, Fagerlin, & Zikmund-Fisher, 2016).

Many researchers have demonstrated that visualizations of uncertainty can lead to better judgments than textual descriptions of the same information (Fagerlin, Wang, & Ubel, 2005; Feldman-Stewart, Brundage, & Zotov, 2007; Fernandes, Walls, Munson, Hullman, & Kay, 2018; Garcia-Retamero & Galesic, 2009a, 2009b; Garcia-Retamero, Galesic, & Gigerenzer, 2010; Garcia-Retamero, Okan, & Cokely, 2012; Hawley et al., 2008; Tait, Voepel-Lewis, Zikmund-Fisher, & Fagerlin, 2010; Waters et al., 2016; Waters, Weinstein, Colditz, & Emmons, 2006). For example, one study presented a mixed group of older adults and students with probabilities via text (e.g., "aspirin can reduce the risk of having a stroke or heart attack by 13%"). A second group was shown textual and icon arrays of this information (Galesic, Garcia-Retamero, & Gigerenzer, 2009). Participants were asked to estimate the number of people out of 1000 who had a stroke if they did and did not take aspirin. Participants who were provided with the icon arrays in addition to the textual information made significantly more accurate judgments. Researchers have also documented improvements compared to text for more complex visualizations (e.g., Fernandes et al., 2018) (for a review of effective uncertainty visualization technique, see Padilla et al., 2021).

In addition to mounting evidence illustrating the utility of uncertainty visualizations, a number of studies have also documented reasoning errors (e.g., Belia et al., 2005; Correll & Gleicher, 2014; Joslyn & LeClerc, 2013; Padilla, Creem-Regehr, & Thompson, 2020; Padilla, Ruginski, & Creem-Regehr, 2017; Ruginski et al., 2016). Errors due to interpreting uncertainty visualizations can exacerbate the difficulty people have in reasoning with uncertainty.

An uncertainty visualization can unintentionally mislead its viewers, which results in poorer decision-making (e.g., Broad, Leiserowitz, Weinkle, & Steketee, 2007). For example, the Cone of Uncertainty—produced by the National Hurricane Center—has become one of the most notorious uncertainty visualizations (see, Fig. 2). The Cone of Uncertainty is intended to show the forecasted path of a storm with the centerline representing the mean prediction and the edge of the cone denoting a 66.6% confidence interval around the mean. When people are not provided with additional information about what the cone is intended to represent, they believe that the cone shows the size of the hurricane growing over time (Padilla et al., 2017). Instead, the cone is intended to show that the uncertainty in the storm's path increases with time from the initial forecast. The concept of uncertainty increasing over time can be intuitive. For example,

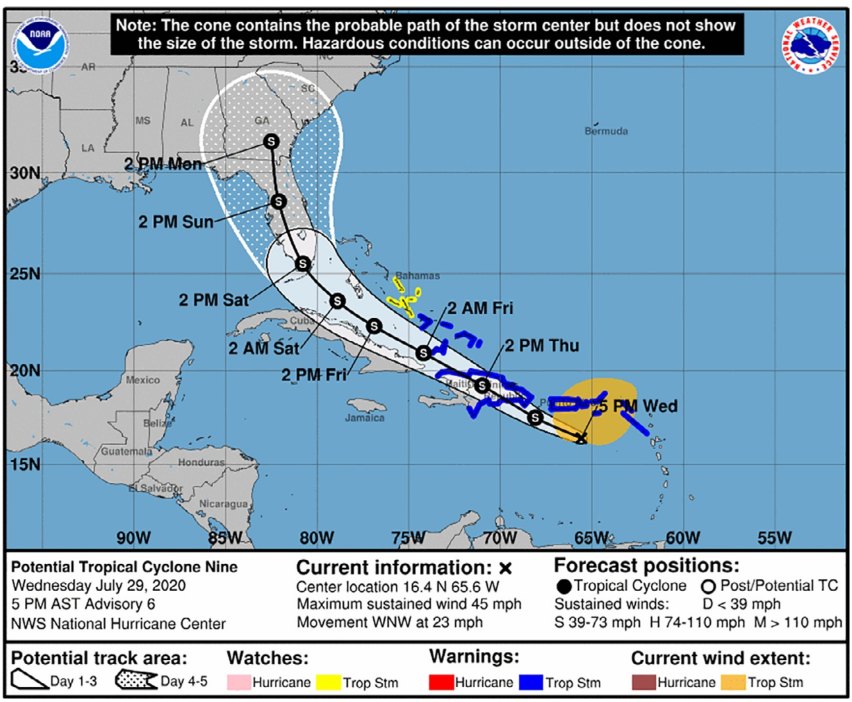


Fig. 2 Example hurricane track forecast cone produced by National Hurricane Center (https://www.nhc.noaa.gov/aboutcone.shtml).

it is easier to predict the temperature for tomorrow than the temperature for 2 weeks from now. However, when uncertainty in the storm's path is represented visually with a cone-like visualization, it requires effort to understand it as anything other than the size of the storm.

Within traditional uncertainty visualization research, practitioners commonly recommend a set of best practices or general principles without positing cognitive theories as to *why* a visualization might produce errors. However, uncertainty visualization researchers are increasingly interested in cognitive perspectives (Fernandes et al., 2018; Hullman, Kay, Kim, & Shrestha, 2017; Kale, Kay, & Hullman, 2020; Kale, Nguyen, Kay, & Hullman, 2018; Kim, Walls, Krafft, & Hullman, 2019). Notably, Kim et al. (2019) propose a Bayesian cognitive modeling approach to incorporate prior beliefs and update evaluations of uncertainty visualizations. Also, Joslyn and Savelli (2020) detail the cognitive mechanisms associated with a specific type of reasoning error in uncertainty visualization. Although prior approaches have detailed the cognitive aspects of reasoning with uncertainty visualizations, they do not offer a unified theory that describes the sources of errors across visualization types. As a result, accurately predicting when a new type of uncertainty visualization will fall into the category of helpful or harmful is difficult.

The current chapter seeks to bridge this gap in knowledge by providing a unifying theory for why errors occur when making decisions with uncertainty visualizations. We begin this work by describing a cognitive framework for how decisions are made with visualizations (Padilla, Creem-Regehr, Hegarty, & Stefanucci, 2018), which we subsequently use as a tool to ground empirical work on errors in uncertainty visualization. Then, we review behavioral evidence of using uncertainty visualizations with a focus on when errors or misunderstandings occur, in order to find commonalities among these errors.

As a preview, researchers consistently observe errors when a visualization or task requires a viewer to perform a complex mental computation to accurately interpret the visual information. We propose that a unifying cognitive process that predicts these errors is increased working memory or cognitive effort. This chapter reviews research on working memory demand in the context of visualizations and how working memory as a mental process can potentially explain many of the errors observed in uncertainty-visualization use.

2. Visualization decision-making framework

Reasoning errors with uncertainty visualizations have the potential to arise at various stages in the decision-making process. If we have a clear understanding of where an error occurs in this process, we can more clearly develop interventions to help make more effective decisions with uncertainty visualizations.

For this chapter, we will utilize a cognitive model that describes decision-making with visualizations proposed by Padilla et al. (2018) (see Fig. 3). The Padilla et al. (2018) model integrates a dual-process theory of decision-making and a modern understanding of visualization comprehension and learning.

2.1 Visual array and attention

The Padilla et al. (2018) model begins with a *visual array*, which is the unprocessed neuronal firing in response to a stimulus. *Bottom-up* and *top-down attention* guide a viewer's gaze around the image. Bottom-up attention refers to how the visual system is guided to elements in the visualization based on visual salience. Errors that occur from bottom-up attention result from the visualization directing a viewer's attention to task-irrelevant information. Top-down attention is how the viewer controls his or her gaze around a visualization. Top-down attention is based on the viewer's goals, experiences, and other individual differences. For example, Kim et al. (2019) capture the influence of top-down attention on decision-making with

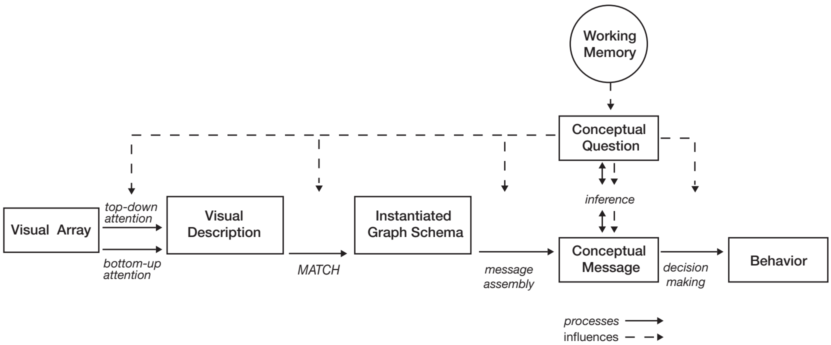


Fig. 3 Visualization decision-making model proposed by Padilla et al. (2018). *Reproduced per CC-BY license from Padilla, L., Creem-Regehr, S., Hegarty, M., & Stefanucci, J. (2018). Decision making with visualizations: A cognitive framework across disciplines.* Cognitive Research: Principles and Implications, 3, *29.*

Bayesian priors representing previous beliefs and experiences. Errors that arise due to top-down attention may come from users having an incomplete understanding of how to achieve their goals, having experiences that tell them to search through an image ineffectively, or other potential biases based on long-term knowledge.

2.2 Working memory

Working memory (located in the circle at the top of Fig. 3) is a cognitive process that can influence most visualization decision-making processes (Padilla et al., 2018). The debate about how to define the term working memory is ongoing, as it has differing characteristics in various fields (Cowan, 2017). For this review, we will use the definition of working memory defined by Cowan (2017), where working memory is a multi-component system. Working memory maintains a finite amount of information for a short time before that information is potentially stored in long-term memory. In the context of uncertainty visualization, the term *maintain* means that when viewers see a visualization, they store a mental representation of it in their mind to update or manipulate later. For example, viewers might see a scatter plot and want to find the data's central tendency. In their mind, they would mentally overlay a trend line onto their temporarily stored mental representation of the visualization. Within the traditional model of working memory, the *visual-spatial sketch pad* represents the mechanism that maintains information from a visualization; a separate mechanism maintains phonological information (Baddeley, 1992).

Working memory has a *central executive* that controls its multicomponent functions, and it works to control attention while suppressing automatic processes (Logie & Marchetti, 1991). For example, the process of explicitly directing top-down attention requires working memory (Shipstead, Harrison, & Engle, 2015), as in directing one's attention away from salient information in a visualization. An error occurs when bottom-up attention guides the visual system to visually salient but task-irrelevant elements in the visualization. The central executive exerts its control over the finite amount of working memory available to simultaneously suppress bottom-up attention shifting to task-irrelevant information and guide attention toward task-relevant stimuli.

Three types of errors may occur due to working memory relevant to uncertainty visualizations: capacity limitation errors, failure to utilize working memory, and temporal decay errors. Researchers have traditionally studied capacity limitation errors in the context of how many digits or items in sequence participants can remember (Miller, 1956). More recent work suggests that we tend to group information (e.g., chunk) rather than maintain the information separately and that we can remember between three to five chunks of information (Doumont, 2002). Errors may occur when viewing uncertainty visualizations if a visualization requires the viewer to maintain too much information in working memory, essentially surpassing the limited working memory capacity. As a simple example, imagine a visualization that maps elements of the data to color, opacity, texture, size, shape, and position. To interpret the visualization correctly, one must maintain in working memory how each variable relates to the data. Working memory capacity may be overloaded if people are asked to do a complex data analysis with such a working-memory demanding visualization. Capacity limitation errors include failing to integrate all of the relevant information in a visualization; not being able to perform a mental computation on a visualization; or failing to maintain, switch, or update task goals.

The second category of errors related to working memory encompasses viewers failing to use working memory when they should. By default, we tend to make fast and automated decisions that use as little working memory as possible (Type 1 processing) (Kahneman, 2011; Tversky & Kahneman, 1974). Type 1 processing is an adaptive strategy that we have developed to minimize effort because effort is metabolically costly. Researchers estimate that our brains account for 20–25% of our resting metabolism (Leonard & Robertson, 1994). Voluntary effort may not exclusively account for the mind's propensity toward Type 1 processing, but a combination of energy conservation and reserving limited capacity working memory validates the preference for fast and automated decisions (Kool & Botvinick, 2014). However, some visualizations require the use of working memory to be understood correctly (Type 2 processing). For example, when viewing the line chart in Fig. 4 that illustrates the impact of the Stand Your Ground law on gun deaths in Florida, a viewer might not notice that the Y-axis is inverted. Without using working memory, the viewer would assume that the Stand Your Ground law correlated with a drop in gun deaths in Florida. To interpret this visualization correctly, a viewer needs to activate working memory to recognize that the Y-axis is inverted and reimagine the data's appropriate relationships.

The third type of error related to working memory results from forgetting relevant information because working memory decays over time. For example, if asked to memorize the sequence 9,875,341,890, recalling the numbers

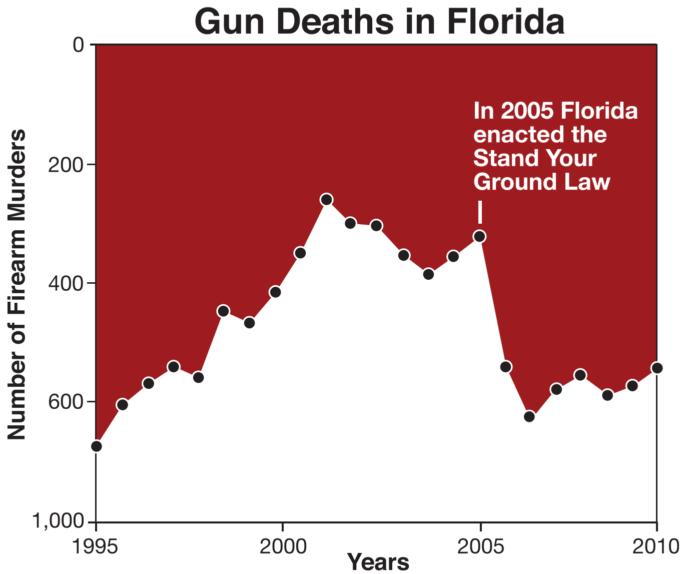


Fig. 4 Deceptive visualization showing the impact of the 2005 Stand Your Ground law in Florida and the number of murders from firearms with the Y-axis reversed. This example is based on a data visualization that was released to the public by Christine Chan at Reuters (Pandey, Rall, Satterthwaite, Nov, & Bertini, 2015). *Redrawn per CC-BY license from Padilla, L., Creem-Regehr, S., Hegarty, M., & Stefanucci, J. (2018). Decision making with visualizations: A cognitive framework across disciplines.* Cognitive Research: Principles and Implications, 3, *29*.

after holding them in working memory for 5 s is easier than after 5 min. To memorize such information and hold it in working memory for prolonged periods, people generally chunk information, such as (987) 534–1890, and then mentally rehearse the information. Without rehearsal, our ability to store information begins to decay after approximately 5–10 s (Cowan, 2017). The nature of the decay can vary due to the task, type of information, and individual capacities (Cowan, Saults, & Nugent, 1997). Longer sequential visualization tasks that require completion of longer-term goals may be error-prone due to the degradation of working memory over time.

2.3 Visual description

The *visual description* (second box in Fig. 3) is the resultant mental conception of the visualization's information after top-down and bottom-up processing have guided the extraction of information. Note that the visual description is not identical to the visualization; its generation is dependent on what the viewer focuses on and can be incomplete, biased, or skewed in its representation.

A visual description allows for enough understanding to mentally transform, interpret, and make decisions with the representation. These processes require the cognitive ability termed *mental imagery*, and again depend on working memory (for a review, see Kosslyn, 1995).

2.4 Graph schemas

Graph schemas (third box in Fig. 3) are templates, rules, graphic conventions, or strategies that people use to interpret a visualization. People might develop graph schemas during formal education if they were taught how to read different visualizations. For example, teachers commonly instruct students on how to read maps by explaining the purpose and use of a legend, scale, and compass rose. As another example, most educational institutions introduce the number line to students at a young age, cementing a *more is to the right* mapping (Winter & Matlock, 2013). These formal educational experiences likely establish many graph schemas. People may also develop a graph schema implicitly through practice viewing visualizations that utilize the same conventions. Errors may occur with graph schemas if a viewer has not developed the necessary schema to interpret a visualization. A viewer may lack graphic education or familiarity with a visualization, resulting in individual differences in graph literacy (Okan, Garcia-Retamero, Galesic, & Cokely, 2012). Alternatively, the visualization might be wholly novel and require the development of a new schema.

2.5 Matching process

The *matching* process between the visual description and the graph schema occurs when a viewer selects a graph schema to expedite the process of interpreting a visualization. The mechanism of how viewers select a particular schema remains unclear. Viewers may select a schema from the same broad category as a visualization. For example, a viewer may select a schema for a Cartesian coordinate plane when viewing a line graph. Viewers may also select schemas based on matching salient features of the visualization and the schema. When viewers see a bar chart, they may select a schema with similar rectangular objects. Viewers may select the schema that is easiest to recall, more recently stored, or one that has been primed. Errors may occur when viewers select the wrong schema for a visualization. For example, recent work shows that researchers could prime the type of schema participants used by telling them an interesting story about the data (Xiong, Van Weelden, & Franconeri, 2019). Different groups of people received different stories, and each story pointed out different features within the same visualization. When asked what other people might see as essential features in the data, participants were more likely to report that other people would see the information they were primed to think of as relevant (Xiong et al., 2019).

2.6 Instantiated graph schema

Instantiating the graph schema occurs when viewers update their mental representation of the visualization to include information from the graph schema. Errors can occur in this process if viewers select the wrong schema; they can also occur if the viewers have to perform a complex mental transformation to update their mental representation with the information from the schema. With a large mismatch between the schema and the viewer's mental representation of the visualization, a mental transformation may be required to combine the two (Vessey, 1991). For example, if the Y-axis is inversely ordered (low numbers at the top), the viewer may need to mentally transform the visualization to correctly order the Y-axis (according to their schema) before incorporating the visual description and schema. Increased errors and time to instantiate the graph schema will occur in cases in which a large mismatch and exorbitant mental transformations are required. Theses outcomes result from overloading the limited resources of working memory available as time decays the information and exceeds the capacity to hold chunks of information in memory (Doumont, 2002).

2.7 Message assembly

The *message assembly process* describes how viewers interpret their mental representation of the visualization after it has been updated by the graph schema. The resultant conceptualization of the meaning of the graphic is the *conceptual message* (fourth box in Fig. 3). Errors that may occur at this stage of the process result from taking the wrong meaning from the visualization's mental representation.

2.8 Conceptual question

The *conceptual question* (box below working memory in Fig. 3) refers to the question that the viewer asks of the visualization. A viewer may have specific goals, such as attempting a data analytics task, which could produce various direct conceptual questions, as in "Where are the outliers?" or "Which variables have meaningful relationships?." Conceptual questions can also be more general (e.g., What can this visualization tell me about my health?) or ill-defined (e.g., What am I looking at?). Many times, viewers may have a sequence of conceptual questions about the visualization, which may evolve.

In the Padilla, Creem-Regehr, Hegarty, & Stefanucci, 2018 framework, conceptual questions play a key role as they channel working memory. This framework suggests that the central executive (i.e., the resource allocation mechanism in working memory) applies working memory to answer the conceptual question during visualization reasoning. As a result, the conceptual question can:

1. Drive a viewer's top-down attention to relevant information
2. Guide which graph schemas are selected
3. Frame the conceptual message
4. Influence decisions

The viewer's specific question can influence all of the processes in this model except bottom-up attention. These processes can also form feedback loops or prime a specific graph schema (e.g., Xiong et al., 2019). Based on the conceptual message, a viewer may decide to update the question or goal and repeat some of the processes. Errors can occur as a result of the conceptual question if it is unclear to the viewer how to achieve his or her particular goals. The viewer might ask the wrong question to achieve their goals or use incorrect steps. Viewers might also have too many goals, which can be challenging to keep track of and require a significant amount of working memory to manage.

2.9 Decision-*making*

Once all the relevant conceptual questions have been answered for the viewer to feel comfortable making a decision, he or she completes the *decision* step. The majority of the widely documented decision-making biases and heuristics occurs in the decision-making step. This process involves taking the visual information stored in the mind and using Type 1 or Type 2 processing to reach a conclusion, usually in order to perform an action (Kahneman, 2011). Type 1 processing is relatively fast, unconscious, and intuitive. Type 2 processing involves working memory and is slower, more metabolically intensive, and more contemplative than Type 1 processing (Evans & Stanovich, 2013). Other models of decision-making characterize these processes differently. Here we note two processes in line with Evans and Stanovich (2013), one that requires the activation of working memory to make a decision and another process that does not require significant working memory. There exists a massive body of literature detailing numerous possible decision-making biases that can occur at this stage. Not all decision-making biases have been generalized to the context of decisions with visualizations, but many of these biases may influence reasoning with visualizations. However, more work is needed to examine if all previously documented decision-making biases generalize to the context of decision-making with visualizations.

2.10 Behavior

The final stage of the Padilla et al. (2018) model results in action or *behavior*. Errors, although not decision-making errors, might occur in this model's final stage when people cannot take the action that they have selected. For example, in hurricane forecasting, people might see a hurricane visualization, decide to evacuate, and then lack the necessary resources to evacuate or not know the appropriate evacuation route. These phenomena require exploration in the more applied social sciences and are beyond the scope of this chapter. However, failures to suppress heavily automated behaviors (e.g., in the case of addictions) due to reduced cognitive resources or poor executive control can also be observed during this stage.

3. Uncertainty visualization errors

Errors in understanding uncertainty visualizations can occur throughout the decision-making process. Here we will use the Padilla et al. (2018) cognitive model to organize and describe the widely documented errors as early-, middle-, and late-stage visualization processing errors (as seen in Fig. 5).

3.1 Early-stage processing errors

Early-stage visual processing errors are driven by the visual system and attentional processes. Occurring early in the decision-making process, these types of errors can be particularly hard to overcome as they influence all of the downstream processes. Researchers refer to early-stage errors driven by elements in the visualization as *visual-spatial biases* (Padilla et al., 2018). Researchers also speculate that visual-spatial biases are particularly hard to overcome because they may be due to bottom-up attention and Gestalt principles, both of which are difficult to cognitively control.

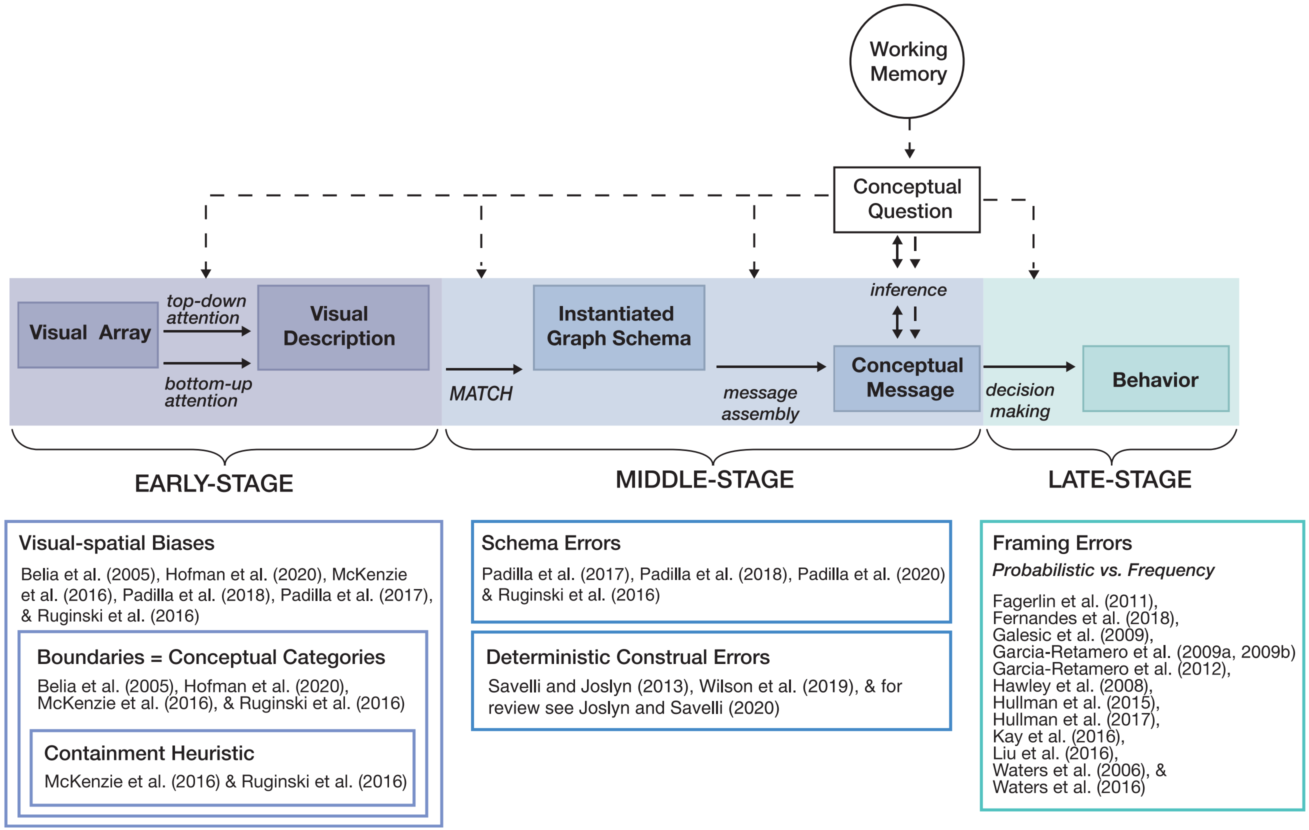


Fig. 5 Early-, middle-, and late-stage uncertainty visualization errors that are organized using the Padilla et al. (2018) visualization decision-making model.

3.1.1 Boundaries = conceptual categories

One visual-spatial bias that occurs often in uncertainty visualization arises when visualization creators delineate continuous data with isocontours, boundaries, intervals, bins, or other types of segmenting marks. Throughout our daily lives, we have learned to interpret the delineations in the world, such as fences, road lanes, and crosswalks, as indications of important information. Tversky (2011) writes, "Framing a picture is a way of saying what is inside the picture has a different status than outside of the picture" (p. 522). In our continuous world, physical delineations separate and categorize meaningful differences and space (Tversky, 2001). Delineations can also be metaphorical. We draw a proverbial line in the sand to indicate a boundary that should not be crossed (Lakoff & Johnson, 1980). As humans, we are adept at categorizing complex information, and we commonly do this by physically or mentally constructing boundaries.

The problem for uncertainty visualization emerges when the designer creates boundaries in probabilistic data, and the boundaries do not indicate categorically different information (Padilla et al., 2015, 2017). For example, 95% confidence intervals delineate probabilistic information to indicate that the true mean has a 95% chance of falling within the specified range. However, there is no categorical difference between the data inside and outside of the confidence interval. Said another way, 95% confidence is not unique, and scientists could have also chosen intervals at 96%, 94%, or 99%. Ninety-five percent confidence exists as a convention concerning the probability of error scientists consider acceptable to make certain inferences. Some fields have different conventions. The National Hurricane Center uses a 66.66% confidence interval to communicate the uncertainty in a hurricane forecast path.

When most viewers see an interval, they utilize the strategies they have developed throughout their lives and interpret it as a meaningful boundary that notes categorically different information (Tversky, Corter, Yu, Mason, & Nickerson, 2012; Zacks & Tversky, 2013). This error could also be considered a mismatch of the visual description and the instantiated graph schema. In a geospatial context, researchers have called this a containment strategy (McKenzie, Hegarty, Barrett, & Goodchild, 2016), where areas within a boundary are imbued with semantic homogeneity (Fabrikant & Skupin, 2005). For example, navigation applications show a user's location, but sometimes the location can have uncertainty (i.e., if the GPS signal is interrupted). One study examined different visualizations for representing the uncertainty in the viewer's location by comparing a gradient map to a 95% CI (see Fig. 6). When viewing the 95% confidence interval that looked

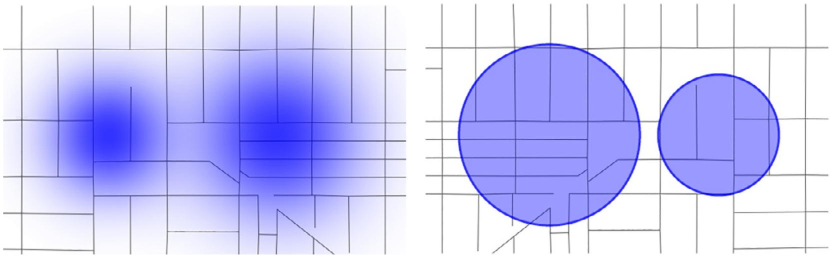


Fig. 6 Visualizations that show the uncertainty in two locations, using a gradient or a bounded circle right, used in McKenzie et al. (2016). *Reproduced per CC-BY license, from Padilla, L., Creem-Regehr, S., Hegarty, M., & Stefanucci, J. (2018). Decision making with visualizations: A cognitive framework across disciplines.* Cognitive Research: Principles and Implications, 3, *29.*

like a bounded circle, participants were more likely to take a containment strategy than when viewing the same positional uncertainty represented in a gradient (McKenzie et al., 2016).

Work examining hurricane forecasts also finds that people use a containment strategy. In one study, researchers showed participants five visualizations of a hurricane's forecasted path (Fig. 7). The path visualizations were intended to show the forecasted direction of the storm and the uncertainty in the forecasted route. As the time increases from the initial forecast, it becomes increasingly more difficult to accurately predict the path of the storm, which is shown in the visualizations' spread increasing for B–E in Fig. 7. Researchers compared a version of the cone of uncertainty, which shows the main forecast path of the storm, along with a 66.6% confidence interval (C in Fig. 7), to a version with just the center line (A), a cone with no center line (B), a gradient mapping of the confidence interval (D), and a new visualization technique entitled an *ensemble visualization* (E). The ensemble visualization shows a subset of paths sampled from the hurricane's probabilistic forecast (Liu et al., 2016; Liu, Padilla, Creem-Regehr, & House, 2019). This research demonstrated that the visualizations that were cone-like (B–D in Fig. 7) elicited a containment strategy where participants rated areas inside of the cones to have more damage than areas outside of the cones. With the ensemble visualization, participants reported that areas near the center of the distribution would receive more damage and damage ratings decreased along with the distance to the center of the distribution (Ruginski et al., 2016). The response patterns observed for ensemble visualizations indicate that participants understand the distribution of uncertainty that the ensembles represent. In the context of hurricane forecasting, this experiment was the first to find an alternative

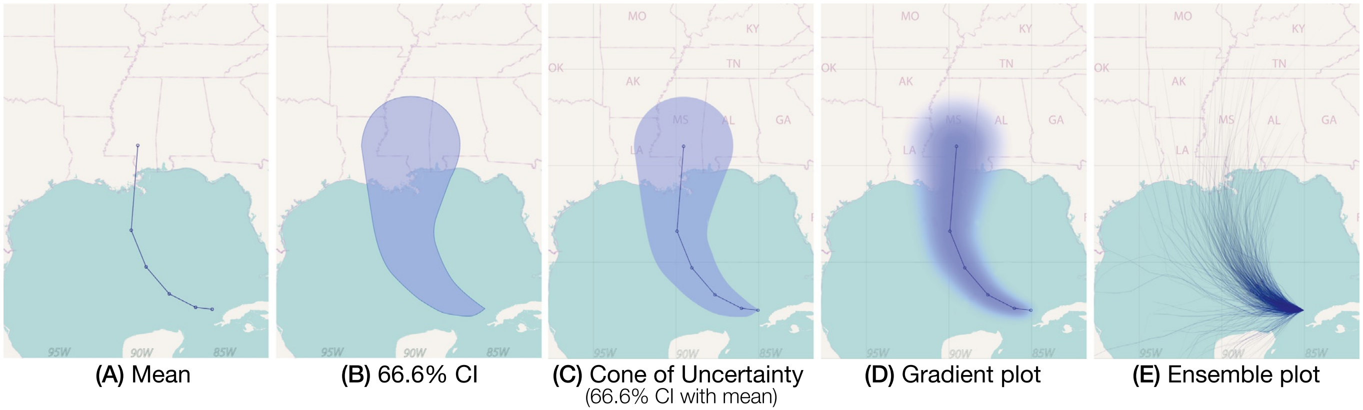


Fig. 7 Redrawn versions of hurricane forecast path of visualizations based on Ruginski et al. (2016).

to the cone of uncertainty that did not elicit the containment strategy. Researchers determined that the edge of the hard boundary elicits the highest visual salience and likely drives the containment strategy (Padilla et al., 2017).

The misunderstandings associated with delineations occur in one-dimensional data as well. Delineation errors can be understood as *boundaries creating conceptual categories* (Padilla et al., 2021). The boundaries creating conceptual categories error likely contributes to the numerous studies finding that people misunderstand how to interpret error bars and confidence intervals. Both well-trained experts in statistics and novices commonly misunderstand how to interpret statistical significance from frequentist 95% confidence intervals (e.g., Belia et al., 2005; Hofman, Goldstein, & Hullman, 2020). Researchers find that even trained experts incorrectly assume that no significant difference exists between two groups with overlapping intervals (Belia et al., 2005). When comparing two health treatments with visualized means and frequentist 95% confidence intervals, participants were more willing to overpay for treatment and to overestimate the effect size compared to when the same data were shown with predictive intervals (Hofman et al., 2020).

People tend to believe that error bars contain the distribution of values, resulting in the mismatch between the visual description and instantiated graph schema. If the two bars are far apart, the boundaries lead people to believe that these boundaries contain all the relevant values and therefore they incorrectly assume a statistically significant difference. A similar effect has also been found with bar charts. Researchers have demonstrated a "within the bar bias," where people believe that data points that fall within a bar are more likely to be part of a distribution than data points equal distance from the mean but outside of the bar (Newman & Scholl, 2012).

This *boundaries- create-conceptual-categories* error likely occurs early in the decision-making process. As demonstrated in Padilla et al. (2017), boundaries make up some of the most salient features in a visualization and can attract our bottom-up attention. As a result, we might spend more time looking at the boundaries in a visualization, which can produce an overweighting of the boundaries in our conceptualization of the data.

One of the reasons boundaries create conceptual categories is that they may reinforce Gestalt grouping principles, which are the visual system's propensity to group and categorize visual information based on similarities in properties such as shape, color, physical proximity, and other contextual information (Wertheimer, 1938). As an illustration, try to determine if patterns are depicted in Fig. 8. All the figure items may seem to be a part of one global grouping because they are all circular and loosely arranged in a circle.

Fig. 8 Ambiguous Gestalt grouping example.

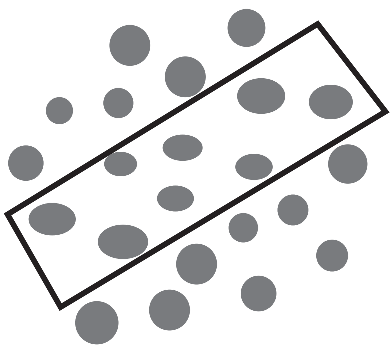


Fig. 9 Ambiguous Gestalt grouping example with boarder around the ovular items.

With effort, most viewers notice that some objects are larger or smaller and others circular or ovular. Identifying patterns becomes easier when boundaries are added, as in Fig. 9, which bounds the ovular items with a line.

When the boundaries are included, visually grouping the ovular objects and noticing they have an upward trend is much easier. The boundary works to precategorize some of the information for the visual system. Said another way, the boundaries offload cognition on the visualization by categorizing the objects before the visual system does. The categorization created by the boundaries occurs early in the decision-making process and reduces a visual system processing step. However, a problem arises when a viewer needs to group different information than what the boundary contains. When viewing Fig. 9, try to mentally group the smaller objects. Most people can successfully group the smaller objects and see their trend, but this process requires significant effort. Mentally grouping the small objects requires suppressing or ignoring the grouping formed by the boundary, requiring additional working memory.

The prominence of the boundaries on the patterns we can see in the data illustrates a visual-spatial bias, where the boundary can lead viewers to see different patterns within the same visualization or data. When visualization designers use boundaries, they define the types of patterns that viewers can see in their data. Their viewers will have difficulty seeing any other patterns within the data. The immutable effect of the boundary can be problematic when the boundaries are arbitrary (i.e., 95% CIs or 66% CIs in the Cone of Uncertainty), making viewers believe that categorical differences exist in the data when there are none, which hinders viewers from finding other important patterns.

Although the point at which the boundary enters the decision-making process occurs very early (e.g., the visual array), the impact of boundaries might be observed at multiple points throughout the decision-making process. Boundaries may be highly salient and direct viewers' bottom-up attention to information inside the boundaries. Viewers might form a strategy to assume that visualization designers are trying to communicate something meaningful with the boundaries and direct their top-down attention to the boundaries' information. Boundaries may evoke incorrect schemas and lead to misunderstandings about what the data represent. Boundaries could even evoke some traditional decision-making biases such as anchoring, where people are biased to make judgments in relationship to the boundaries.

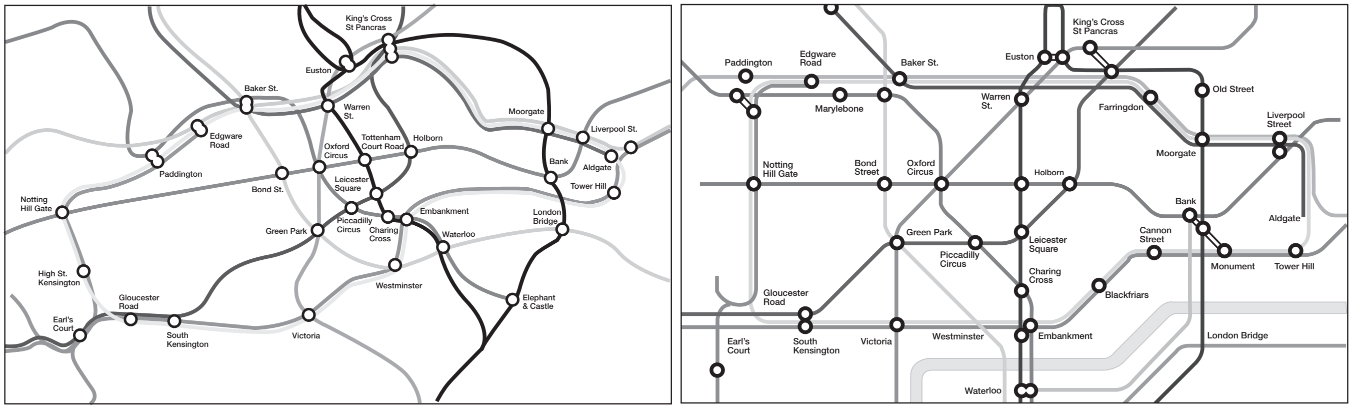
Such early-stage processing errors are some of the most consistent and widely documented, but little is known about *why* these errors occur. One theory that we propose here is that working memory is a crucial contributor to early-stage processing errors in uncertainty visualization. Early-stage processing errors represent a unique category because working memory cannot easily influence all these processes. In particular, bottom-up attention is difficult to control with effort. As noted throughout this section, many of the errors we reviewed might be fully explained by bottom-up attentional processes. For example, some work finds that boundaries in hurricane forecasts are highly salient and draw the viewers' attention (Padilla et al., 2017). All types of boundaries may draw viewers' attention, and therefore, they have an overstated impact on viewers' decision-making process compared to more task-relevant information in the visualization. Further, even when viewers consciously know not to focus on information, as with the Cone of Uncertainty boundary, they likely have difficulty suppressing saccadic movements toward such salient information. Thus, the boundary may increase the working memory required due to active inhibition.

The most notable characteristic of early-stage processing errors is that they are challenging to overcome. For example, in one study, participants were provided with extensive instructions on interpreting the Cone of Uncertainty (Boone, Gunalp, & Hegarty, 2018). Researchers instructed participants that the cone does not show the storm's size growing over time, but participants still made decisions as if the storm's size was increasing. Notably, at the end of the experiment, participants could accurately answer questions about interpreting the cone correctly (Boone et al., 2018). This work provides some evidence that even when viewers are aware that they should cognitively override the visual array's impacts, they find it challenging to do so.

Participants' inability to utilize working memory to make more effective decisions in the previous examples may be because working memory has difficulty impacting early processing errors. Working memory's problem in affecting early processing errors could be due to earlier errors biasing all the downstream processes. It could also be the case that early processing errors are primarily due to bottom-up attention and working memory may have little ability to impact bottom-up attention. However, no work has examined the exact nature of early-stage processing errors in visualization reasoning. More work is needed to understand the cause, prevalence, and unwavering nature of such errors.

3.2 Middle-stage processing errors

In the Padilla et al. (2018) framework, middle-stage processing errors occur after the visual system has created a mental representation of the visualization. At this stage, viewers apply a schema that they have stored in long-term memory to their mental representation of the visualization. For example, when viewing Fig. 10 (left), most people would categorize the picture as a map. Consciously or unconsciously, they would retrieve the schema for maps and make assumptions about the information, including that North is at the top and that a consistent relationship likely exists between the physical size of the areas shown. They would have made assumptions based on map schemas even though we excluded the map's compass rose and legend. As in this example, many of the assumptions we make about visualizations based on schemas create an advantage over their absence. Schemas help us interpret information correctly, efficiently, and quickly when a visualization adheres to known graphic conventions.



Geographically Accurate London Tube Map

Diagrammatic London Tube Map

Fig. 10 Left geographically accurate transit map and right diagrammatic map of the London Tube. *Redrawn from Guo, Z. (2011). Mind the map! The impact of transit maps on path choice in public transit.* Transportation Research Part A: Policy and Practice, *45(7), 625–639.*

The Padilla et al. (2018) schema instantiation process has three steps, as illustrated in Fig. 11. A viewer must first correctly classify the visualization type. With standard visualizations (e.g., line or bar charts), accurate classification occurs relatively easily. However, errors can arise when ambiguity exists in classifying the category or type for a visualization.

A famous example of classification involves the London Underground map by Harry Beck (see an example based on Beck's innovation in Fig. 10, right). Beck helped define a new cartographic convention that departed from the historical approach of superimposing subway lines on a geographically accurate map (Guo, 2011). In Beck's redesign, he opted to arrange the layout in a diagrammatic fashion that focused on improving the legibility of routes, transfers, and stops, inspired by electrical circuits. Initially, transit officials scoffed at the design, but it was ultimately adopted in 1933. Some of the apprehension about Beck's map began because officials thought that riders might see it as a standard map, fail to realize that the distances between stops were not based on physical distance, become confused, and miss their stops. Researchers continue to discuss whether Beck's design should be classified as a map or as a diagram (Cartwright, 2012).

When new innovations change visualization design, viewers might become confused about how to classify a new type of visualization, which can affect how they determine and implement an appropriate schema. Today, Beck's approach has been utilized worldwide for close to a century, and most transit riders have developed a specific schema for diagrammatic subway maps. Beck's success is likely due in part to the design being different enough from standard approaches that the design prompted riders to recognize that a standard map-based schema would not work. Additionally, the design reduced directional information to three axes, reducing the memory required to match viewers' destination goals with their visual description.

In the next step of the graph instantiation process, viewers retrieve the relevant schema based on how they classified the visualization. Errors can occur in this process when viewers have not learned an appropriate schema. When no schema is available for a graph type, the viewers might utilize a schema from a different visualization type or context. For example, see the new coordinate system in Fig. 12 and try to determine the values for B.

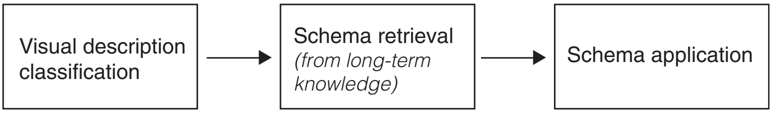


Fig. 11 Three-step schema instantiation process.

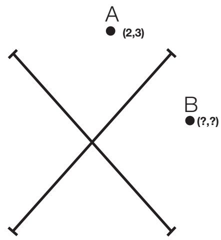


Fig. 12 Hypothetical new coordinate plane.

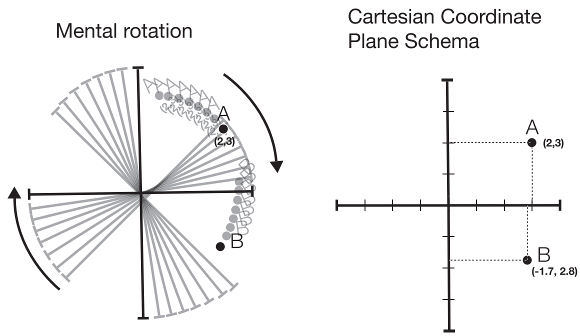


Fig. 13 Example of a mental rotation needed to apply at the schema for Cartesian coordinate plane to a hypothetical new coordinate plain and derived B.

One strategy is to notice that A and B both have two values and a coordinate plane. Dot plots use similar Cartesian coordinate planes but have different axes than in the example. One could apply the Cartesian coordinate schema to interpret the new hypothetical coordinate plane and then derive B's values, as illustrated in Fig. 13.

The problem with applying the schema for a Cartesian coordinate plane to the new coordinate plane is that the planes do not adhere to the same graphic conventions. The angles of the axes in Fig. 14 are not 90°. Applying a schema for a Cartesian coordinate to the new coordinate plane incorrectly is easy, as they share similar properties. When the appropriate schema is unknown, viewers commonly retrieve a different visualization schema to interpret the new information, which can work out well in some cases or can lead them to systematic misinterpretations. Graph schemas that viewers can easily remember and those frequently used are more likely to be applied to an ambiguous visualization type.

In the final stage of the schema instantiation process, viewers must apply the schema that they have retrieved to the visualization in order to answer

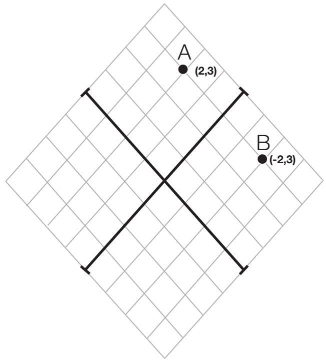


Fig. 14 Illustration of how the axes for the new coordinate plane are not 90° angles.

the conceptual question. When a mismatch between the schema and the visualization occurs, as illustrated in the prior example, a transformation is required to make the two align. Cognitive Fit Theory describes how errors occur when a mismatch between the schema and the visualization requires exorbitant mental computations (Vessey, 1991). A large mismatch between the schema and visualization requires significant working memory to make the two align, which results in increased errors and time to complete the task (Padilla et al., 2018). Note that the Padilla et al. (2018) model suggests that the schema matching process and all other processes (other than bottom-up attention) are in service of the conceptual question. Even if viewers do not think they are trying to answer a specific question, they always have a goal, which could be as simple as understanding what they see.

3.2.1 Schema errors in hurricane visualizations

Uncertainty visualizations of hurricane forecasts represent one of the most highly studied types of schema errors (Padilla et al., 2017; Padilla, Creem-Regehr, et al., 2020; Ruginski et al., 2016). As previously discussed, viewers assume that the National Hurricane Center's Cone of Uncertainty represents the storm's size growing over time, even though it does not communicate storm size information (Padilla et al., 2017). Researchers have also observed the misunderstanding that the cone's area represents the size of the storm when blurry or fuzzy boundaries border the cone.

One key source of these errors involves the schema that people utilize when seeing hurricane forecast maps. Viewers looking at a hurricane forecast map reasonably use the schema that they have learned for maps, which dictates that physical distance on a map should correspond to physical distance in the world. However, cone-like hurricane forecasts violate cartographic expectations by using physical distance to represent uncertainty in the storm's path. To interpret the forecast correctly, viewers must maintain the base-map schema but then suppress the map schema when looking at the cone. Viewers must keep schemas for both maps and uncertainty in working memory and apply each where appropriate in the visualization. Flexibly switching between schemas is highly demanding on working memory. Such a high working memory demand required by cone-like hurricane forecasts may overtax viewers' limited working memory capacity.

When the working memory demand of a visualization exceeds a viewer's working memory capacity, the viewer may drop one schema (e.g., use only the map schema). Viewers who utilize only a map schema commonly report that the cone-like visualizations represent a danger zone, where areas inside the cone are at risk and areas outside the cone are relatively safe (Ruginski et al., 2016). When forced to drop a schema, we argue that people will likely maintain the schema with which they have the strongest associations. As most people have seen and used maps for large portions of their lives, the map schema will take prominence over the uncertainty schema, which they may have less training or experience using.

We were initially surprised to find that viewers of a blurry cone also see a similar danger zone, as researchers have suggested that blur/fuzziness/transparency may be a more intuitive way to communicate uncertainty (MacEachren et al., 2012). Researchers continue to test blurry cones as an alternative approach to the Cone of Uncertainty and see no benefits of blur (Millet et al., 2020). The interest in testing alternative metaphorical expressions of uncertainty (e.g., blur, fuzziness, transparency, fogginess, and sketchiness), including our own, occurred mainly due to a misattribution of *why* the Cone of Uncertainty leads to misunderstandings. We argue that the principal error inherent in cone-like visualizations is that they force viewers to hold multiple schemas in working memory, which is the case for cones with both rigid and blurry boundaries. Blur, fuzziness, transparency, fogginess, and sketchiness express uncertainty *explicitly* as an additional attribute of the visualization that requires a second schema. More modern uncertainty visualization techniques *implicitly* communicate the uncertainty in animations (Hullman, Resnick, & Adar, 2015) or color (Correll, Moritz, & Heer, 2018) and may prove to be more effective because they do not require the viewer to hold multiple schemas in their mind.

Blur or distributional visualizations can be highly successful if a second schema is not required to understand the visualization. For example, researchers have found that gradient plots of 1D data can outperform interval plots of the same information (Correll & Gleicher, 2014). Gradient plots of 1D data require only a single schema for mapping opacity to probability.

Researchers have also documented how ensemble visualizations, which are the most effective hurricane forecast visualization technique (Ruginski et al., 2016), can also suffer from schema errors (Padilla et al., 2017; Padilla, Creem-Regehr, et al., 2020). The approach of this work was to identify the schema participants use when viewing ensemble visualizations. Ensemble visualizations have been developed as a technique relatively recently (Liu et al., 2016), and we can reasonably assume that people have not developed a specific schema for ensembles.

After reviewing all commonly available visualization techniques, researchers noted that the ensemble visualization shared many similar properties to map-based navigation applications (Padilla et al., 2017). Both map-based travel applications and ensemble hurricane forecasts have a base map that adheres to standard cartographic principles and overlays of lines. Researchers speculated that when viewing an ensemble visualization, people utilize the schema that they have developed for understanding travel applications (Padilla et al., 2017). An essential benefit to using a travel application schema is that participants would not have to hold multiple schemas in their minds (e.g., one for maps and one for uncertainty). The use of a single schema could be one reason why ensemble visualizations outperform cone-like hurricane forecasts (Padilla et al., 2017; Ruginski et al., 2016).

However, the problem with using a travel application schema for hurricane ensembles is that the schema could lead to errors in specific cases. Researchers tested an additional hypothesis that people see each line of the hurricane forecast ensemble as a specific path the hurricane could take (Padilla et al., 2017). The schema for geospatial travel visualizations dictates that the application shows a finite list of possible discrete routes and not a distribution of routes. Whereas for the ensemble visualization, each line depicts a subset of a distribution. In other words, the ensemble lines show the spread of uncertainty in the path of the storm. They do not show an exhaustive list of every possible path the storm could take. If people use a schema for geospatial travel applications and one of the ensemble members intersects a location of interest, they may incorrectly think the likelihood is higher that the storm will hit that location (Padilla et al., 2017).[a]

[a] Note that researchers provided participants little information about how to interpret the ensembles, which simulates the conditions in which they would see hurricane forecast in the news (i.e., on average, hurricane forecasts are shown on TV for 1.52 min; Padilla, Creem-Regehr, et al., 2020; Padilla, Powell, Kay, & Hullman, 2020).

Researchers tested the hypothesis that viewers use a geospatial travel application schema to interpret ensemble displays by showing participants ensemble hurricane forecasts with two indicated locations (See Fig. 15) (Padilla et al., 2017). On each trial, an ensemble member intersected one of the locations. Researchers found that when the ensemble member intersected a location, the participants believed that the location would receive more damage than the location that was not intersected by an ensemble member. This overreaction due to the colocation with the ensemble member persisted regardless of the damage probability (Padilla et al., 2017).

Follow-up research provided converging evidence that ensemble visualizations evoke a geospatial travel application schema by replicating the overreaction when an ensemble member intersects a point of interest and demonstrating that the number of lines shown moderates this effect (Padilla, Creem-Regehr, et al., 2020). Researchers reduced the overreaction bias by increasing the number of lines shown. As an illustration, when shown an ensemble visualization with 5 or 10 paths and one path intersects a location, people commonly report a 20% chance the storm will hit the location with 5 paths and 10% with 10 paths. Researchers found that increasing the lines from 9 to 14 to 33 meaningfully reduced the overreaction bias (Padilla, Creem-Regehr, et al., 2020).

However, researchers were never able to entirely eliminate the overreaction by changing the number of lines (Padilla, Creem-Regehr, et al., 2020). In a final attempt to reduce the overreaction bias, researchers

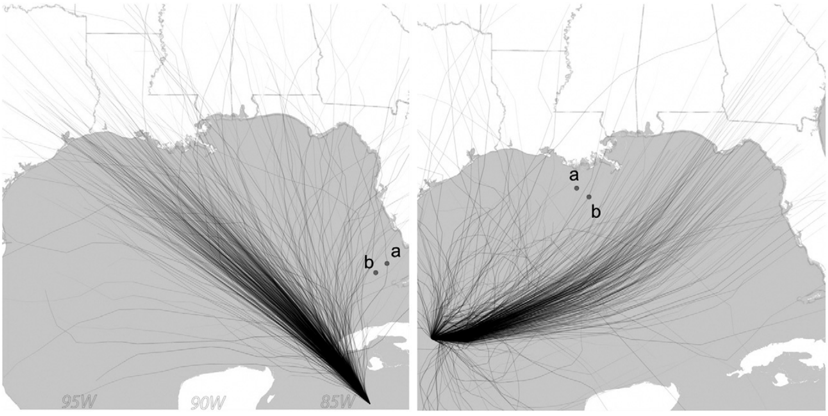


Fig. 15 Example ensemble hurricane forecast visualizations with two locations from Padilla et al. (2017). In each visualization, one location is intercepted by an ensemble member. *Reproduced per CC-BY license, from Padilla, L., Creem-Regehr, S., Hegarty, M., & Stefanucci, J. (2018). Decision making with visualizations: A cognitive framework across disciplines.* Cognitive Research: Principles and Implications, 3, *29.*

tested if participants could override the graph schema using working memory to control their overreaction cognitively. Researchers provided participants with extensive instructions on interpreting an ensemble visualization and how to perform the task correctly. Participants with extensive instructions were able to reduce their bias but not entirely remove it. At the end of the study, participants who received extensive instructions could report the correct strategy, but these participants still overreacted in their behavioral judgments, albeit to a lesser degree (Padilla, Creem-Regehr, et al., 2020).

In summary, ongoing research on hurricane forecast visualizations demonstrates multiple schema-related errors. Errors are highly likely when working memory demand from a visualization is increased, by maintaining two schemas or attempting to cognitively override one schema. The majority of geospatial uncertainty visualizations will likely encounter similar errors because superimposing the uncertainty visualization on the base map will likely evoke the viewer's map schema.

Future visualization designers interested in communicating geospatial uncertainty that does not evoke a traditional cartographic schema could utilize the approach pioneered by Harry Beck in the London Underground map. One possible reason that the London Underground map does not produce large schema-based errors is that its differences sufficiently separate the visualization from a traditional map, which makes people aware that a conventional map schema is not appropriate. If the visualization alerts the viewer to its novelty, it could trigger the viewer to develop a new schema.

3.2.2 Deterministic construal errors

Many schema-based errors may also be explained by viewers ignoring the uncertainty and instead interpreting uncertainty visualizations as communicating deterministic data, called *deterministic construal errors* (Joslyn & Savelli, 2020). Researchers first identified deterministic construal errors in 1D temperature forecasts, when they presented participants with uncertainty in mean temperature forecasts with confidence intervals visualized as bars (Savelli & Joslyn, 2013). The researchers found that 36% of participants believed that the confidence intervals represented high- and low-temperature forecasts rather than uncertainty around the mean (Savelli & Joslyn, 2013). Savelli and Joslyn then tested alternative visualization techniques, including dotted lines and blurry boundaries, and found that the participants still assumed that the intervals around the means were high- and low-temperature forecasts. The researchers went further by creating an obvious key that instructed viewers on how to interpret the confidence intervals accurately, which did not reduce the deterministic construal error (Savelli & Joslyn, 2013). Other researchers have demonstrated similar effects with color encoding, where viewers interpret the probability of rain fall as cumulative rainfall (Wilson, Heinselman, Skinner, Choate, & Klockow-McClain, 2019).

In a recent review of deterministic construal errors, Joslyn and Savelli (2020) propose that the psychological cause closely relates to *attribute substitution*. Attribute substitution is where people opt to use an easy and often incorrect mental process rather than doing a challenging mental computation (Kahneman & Frederick, 2002). Indeed, the schema associated with most uncertainty visualizations places high demand on working memory, and viewers may opt for a more easily interpreted schema, essentially reducing their working memory demand. In the context of textual information, scholars in a wide range of fields have documented an aversion to working memory demand associated with uncertainty, termed ambiguity aversion (e.g., Bach, Hulme, Penny, & Dolan, 2011; Curley, Yates, & Abrams, 1986; Einhorn & Hogarth, 1985; Ellsberg, 1961; Highhouse, 1994; Huettel, Stowe, Gordon, Warner, & Platt, 2006).

We propose that a schema hierarchy may be an additional contributor to deterministic construal errors, presenting a unique challenge to uncertainty visualizations where viewers unconsciously use only the dominant schema. The effectiveness of emerging visualization techniques supports this assertion, such as hypothetical outcome plots (Hullman et al., 2015) that force users to utilize a schema that includes uncertainty. Hypothetical outcome plots consist of animated visualizations that sample from a distribution. Each frame of the animation shows one sample from a probabilistic distribution. Hypothetical outcome plots force viewers to utilize a schema that incorporates uncertainty and have been found to outperform other modern visualization techniques (Hullman et al., 2015; Kale et al., 2018).

3.3 Late-stage errors

After viewers have gathered relevant information from a visualization, used graph schemas to interpret the visualization, and attempted to answer their conceptual question, the final stage of the process consists of making a decision with all of that information and acting. Late-stage processing errors in visualization decision making can occur when viewers apply universal decision-making heuristics. Universal decision-making heuristics are not specific to reasoning with visual information and are studied significantly by researchers in psychology and economics (e.g., Gigerenzer, 2008; Kahneman & Tversky, 1982; Montibeller & Von Winterfeldt, 2015). In this section we will review studies that have generalized decision-making biases to reasoning with uncertainty visualizations.

3.3.1 Framing errors: Probabilistic vs frequency

In the context of uncertainty communication, researchers find that textual information that uses a frequency framing (1 out of 10) is more intuitive than probabilistic framing (10%) (Gigerenzer & Hoffrage, 1995) and requires less working memory (Yin et al., 2020). The general theory suggests that people have difficulty reasoning with probabilities because they rarely experience risk in the form of probabilistic expressions (Gigerenzer, 1996, 2008; Gigerenzer & Gaissmaier, 2011; Gigerenzer et al., 2000). Emerging work in uncertainty visualization demonstrates that visualization techniques that utilize frequency framing can be highly successful, including icon arrays (Galesic et al., 2009), quantile dot plots (Fernandes et al., 2018), hypothetical outcome plots (Hullman et al., 2015), and ensemble plots (Liu et al., 2016, 2019).

Visualizations that utilize frequency framing allow the viewers' visual system to interpret probabilities rather than requiring them to consider numeric expressions. When communicated in text, researchers found that individuals with low working memory capacity can more easily interpret statements expressed as frequencies rather than probabilities (Yin et al., 2020). It is possible that visualizations that display frequency information rather than probability information will also require less working memory.

For example, the bottom of Fig. 16 shows a quantile dot plot of nighttime low temperatures, and the top shows the cumulative distribution function used to generate the dot plot. In this example, each dot represents a 5% probability. If the researcher asks a viewer to determine the probability that the nighttime temperature will be 32° or below using Fig. 16, the viewer could simply count the number of dots. In this example, each dot represents a 5% probability. If the same information is visualized with a density plot (e.g., a bell curve), the viewer would have to mentally compute the integral under the curve, which is a highly challenging judgment, resulting in viewers likely substituting this process for an easier one. Scholars suggest that frequency-framing uncertainty visualizations afford a simple and effective heuristic that requires minimal working memory compared to probabilistic techniques (Padilla et al., 2021). However, no work as directly tested the working memory demand of uncertainty visualizations.

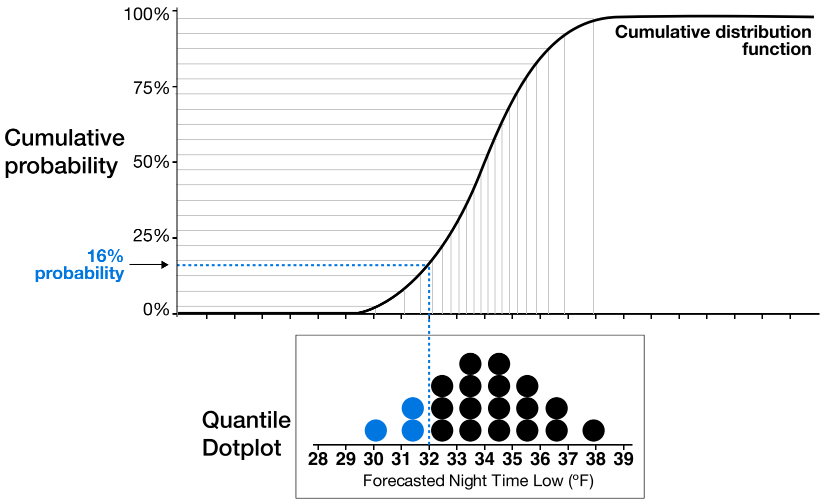


Fig. 16 The bottom illustrates a quantile dot plot that shows a forecasted nighttime low temperature, and the top shows the cumulative distribution function that was used to create the quantile dot plot. *Redrawn from, Padilla, L. M., Powell, M., Kay, M., & Hullman, J. (2020). Uncertain about uncertainty: How qualitative expressions of forecaster confidence impact decision-making with uncertainty visualizations.* Frontiers in Psychology*, 11, 3747.*

A slight conflict in the field of uncertainty visualization exists between visualizations that provide more expressive information about the distribution of data (e.g., distributional visualizations) and visualizations that are more simple (e.g., visualizations that show summary statistics). Distributional visualizations can represent essential features about the distribution, including the shape, skewing, or outliers (Padilla et al., 2021). A rule of thumb in visualization design is that more expressive visualizations are preferred because they give a more fine-grained and thorough representation of the data (Mackinlay, 1986; Munzner, 2014). However, more expressive visualizations depict more information, which might have the unintended consequence of increasing working memory demand. For example, distributional visualizations such as quantile dot plots convey more attributes of the data than simpler visualizations such as error bars or means. Interestingly, frequency framing distributional visualizations seem to hit the sweet spot by conveying the distributional data simplistically. Mounting evidence suggests that quantile dot plots improve accuracy and memory compared to density plots (Hullman et al., 2017; Kay, Kola, Hullman, & Munson, 2016) and outperform summary plots, density plots, and text descriptions of uncertainty for decisions with risk (Fernandes et al., 2018).

Icon arrays, which are nondistributional frequency-framing visualizations, have also consistently shown decision-making advantages over text communications of health risk (for review, see Fagerlin et al., 2011; Waters et al., 2016). Icon arrays can also reduce common decision-making biases, such as individuals focusing on the numerator and neglecting the denominator (for a review, see Garcia-Retamero et al., 2012), anecdote bias (e.g., prioritizing anecdote information over data; for a review, see Fagerlin et al., 2005) and *side effect aversion* (Waters, Weinstein, Colditz, & Emmons, 2007). Side effect aversion is a common bias where patients over-weight negative effects of treatment when making health decisions (Waters et al., 2007). Using a large sample ($n=4248$), researchers found that including icon arrays describing the likelihood of developing cancer with and without a hypothetical preventative drug decreased side effect aversion (Waters et al., 2007). Scholars propose that side effect aversion is closely related to risk aversion (Kahneman, Knetsch, & Thaler, 1990) and may have similar cognitive mechanisms (Waters et al., 2007). Icon arrays' use of frequency framing also helps people with low numeracy to interpret probabilities correctly (e.g., Galesic et al., 2009; Garcia-Retamero & Galesic, 2009a, 2009b; Hawley et al., 2008). While not directly tested, icon arrays may guide the viewer's attention to task-relevant information, which may naturally counteract the classical decision-making biases that tend to occur later in the decision process.

In summary, the majority of work in uncertainty visualization that examines universal decision-making biases finds that frequency-framing visualizations consistently outperform probabilistic depictions of the same data. Some studies also find that frequency-framing visualizations can reduce common decision-making biases. One explanation for the superiority of frequency-framing visualizations is that they evoke an effective heuristic for interpreting uncertainty, using less working memory (Kahneman & Frederick, 2002). However, more work is needed to directly test working memory in the context of decision-making biases with uncertainty visualizations.

4. Conclusions

In this chapter, we reviewed research on biases in uncertainty visualization. Using the Padilla et al. (2018) framework, we discussed biases at early-, middle-, and late-stage decision-making processes. We proposed a unifying theory that increased working memory demand or lack thereof contributes to many of the biases reviewed in this paper. If visualization designers can understand the source of the biases, they may be less likely to repeat visualization mistakes. Further, visualization designers will find novel visualization solutions more quickly by focusing on the problem's source (optimizing working memory demand) rather than on the symptoms of the problem. The problem's source can occur in:

- *The Early stage*: Due to bottom-up attention and Gestalt principles, visual-spatial biases may result in a poorly perceived and poorly understood visualization. These biases include boundaries of continuous data. Boundaries may cause conceptual categories that do not exist, and distort categories that do exist. Some boundaries invoke a containment heuristic, resulting in viewers reducing continuous data to a binary understanding, as in the Cone of Uncertainty (Boone et al., 2018; Padilla et al., 2017).
- *The Middle stage*: Schema errors occur when the visual description does not match the instantiated schema (e.g., judging walking distance and direction from the diagrammatic London Tube map, Fig. 10). Other errors may occur at this stage due to the viewer's unfamiliarity with uncertainty, resulting in a deterministic interpretation (Savelli & Joslyn, 2013).
- *The Late stage*: When people make decisions and perform actions, framing the data in complex or unfamiliar domains (i.e., probability) leads to poor decision-making. However, this outcome can be circumvented by reframing the data in a more intuitive (i.e., frequency) framing, allowing early and middle stage processes to lead to fully informed decision-making.

A key takeaway from this work is that a sweet spot exists in working memory demand. If uncertainty visualizations require too much working memory demand, as in those that require multiple schemas, viewers will become overloaded and not be able to complete a task accurately. However, in some cases, if a viewer fails to use working memory, they may rely on an ineffective or misleading heuristic. More work is needed to identify this sweet spot in the working memory demand of uncertainty visualizations. A path forward would be to measure working memory more directly in visualization experiments. Methods such as pupillometry, EEG, and fNIRS can provide relatively accurate working memory demand proxies but are rarely used in visualization research (for exceptions see, Padilla, Castro, Quinan, Ruginski, & Creem-Regehr, 2019; Peck, Yuksel, Ottley, Jacob, & Chang, 2013). More work is needed to determine which visualization techniques and tasks are working-memory demanding to predict when errors will occur more accurately.

We argue that visualizations that utilize frequency framing are more likely to find the sweet spot in working memory demand, including icon arrays (Zikmund-Fisher et al., 2014), quantile dot plots (Fernandes et al., 2018), hypothetical outcome plots (Hullman et al., 2015), and some ensemble plots (Liu et al., 2019). Frequency-framing visualizations convey probabilistic data in a way that is more intuitive to understand, requiring less working memory then other techniques. Further, frequency-framing visualizations capitalizes on the visual system's substantial processing power to interpret the probabilistic data. However, such visualizations are not entirely free of errors. For example, ensemble hurricane visualizations can lead viewers to think that all the forecasted hurricane paths are shown, which is a misunderstanding of the probabilistic data (Padilla, Creem-Regehr, et al., 2020).

Of the visualization research reviewed in this chapter, visualizations that summarize probabilistic data using ranges, boundaries, or intervals produce systematic and consistent reasoning errors. Part of the reason that summary uncertainty visualizations consistently lead to poor performance is that they can produce errors at every stage of the decision-making process. Of note are the errors produced by bottom-up attention, which are challenging for working memory to overcome. Concerningly, summary uncertainty visualizations are the most common visualization type used in scientific journals (e.g., confidence intervals and means). Researchers interested in effectively communicating the uncertainty in their science should opt for distributional visualization techniques, particularly those that use frequency framing. For an in-depth review of effective uncertainty visualization techniques, see Padilla et al. (2021).

References

Bach, D. R., Hulme, O., Penny, W. D., & Dolan, R. J. (2011). The known unknowns: Neural representation of second-order uncertainty, and ambiguity. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(13), 4811–4820.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556.

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, *10*(4), 389.

Boone, A., Gunalp, P., & Hegarty, M. (2018). The influence of explaining graphical conventions on interpretation of hurricane forecast visualizations. *Journal of Experimental Psychology: Applied*, *24*(3), 275.

Broad, K., Leiserowitz, A., Weinkle, J., & Steketee, M. (2007). Misinterpretations of the "cone of uncertainty" in Florida during the 2004 hurricane season. *Bulletin of the American Meteorological Society*, *88*(5), 651–668.

Cartwright, W. (2012). *Beck's representation of London's underground system: Map or diagram? Paper presented at the GSR.*

Correll, M., & Gleicher, M. (2014). Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 2142–2151.

Correll, M., Moritz, D., & Heer, J. (2018). Value-suppressing uncertainty palettes. In *Paper presented at the proceedings of the 2018 CHI conference on human factors in computing systems*.

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, *24*(4), 1158–1170.

Cowan, N., Saults, J. S., & Nugent, L. D. (1997). The role of absolute and relative amounts of time in forgetting within immediate memory: The case of tone-pitch comparisons. *Psychonomic Bulletin & Review*, *4*(3), 393–397.

Curley, S. P., Yates, J. F., & Abrams, R. A. (1986). Psychological sources of ambiguity avoidance. *Organizational Behavior and Human Decision Processes*, *38*(2), 230–256.

Doumont, J.-L. (2002). Magical numbers: The seven-plus-or-minus-two myth. *IEEE Transactions on Professional Communication*, *45*(2), 123–127.

Einhorn, H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review*, *92*(4), 433.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, *75*, 643–669.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241.

Fabrikant, S. I., & Skupin, A. (2005). Cognitively plausible information visualization. In *Exploring geovisualization* (pp. 667–690). Elsevier.

Fagerlin, A., Wang, C., & Ubel, P. A. (2005). Reducing the influence of anecdotal reasoning on people's health care decisions: Is a picture worth a thousand statistics? *Medical Decision Making*, *25*(4), 398–405.

Fagerlin, A., Zikmund-Fisher, B. J., & Ubel, P. A. (2011). Helping patients decide: Ten steps to better risk communication. *Journal of the National Cancer Institute*, *103*(19), 1436–1443.

Feldman-Stewart, D., Brundage, M. D., & Zotov, V. (2007). Further insight into the perception of quantitative information: Judgments of gist in treatment decisions. *Medical Decision Making*, *27*(1), 34–43.

Fernandes, M., Walls, L., Munson, S., Hullman, J., & Kay, M. (2018). Uncertainty displays using quantile dotplots or cdfs improve transit decision-making. In *Paper presented at the proceedings of the 2018 CHI conference on human factors in computing systems*.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*(1), 1–25.

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: Overcoming low numeracy. *Health Psychology*, *28*(2), 210.

Garcia-Retamero, R., & Cokely, E. T. (2017). Designing visual aids that promote risk literacy: A systematic review of health research and evidence-based design heuristics. *Human Factors*, *59*(4), 582–627.

Garcia-Retamero, R., & Galesic, M. (2009a). Communicating treatment risk reduction to people with low numeracy skills: A cross-cultural comparison. *American Journal of Public Health*, *99*(12), 2196–2202.

Garcia-Retamero, R., & Galesic, M. (2009b). Trust in healthcare. In M. W. Kattan (Ed.), *Encyclopedia of medical decision making* (pp. 1153–1155). Sage.

Garcia-Retamero, R., Galesic, M., & Gigerenzer, G. (2010). Do icon arrays help reduce denominator neglect? *Medical Decision Making*, *30*(6), 672–684.

Garcia-Retamero, R., Okan, Y., & Cokely, E. T. (2012). Using visual aids to improve communication of risks about health: A review. *The Scientific World Journal*, *2012*, 1–10.

Gigerenzer, G. (1996). The psychology of good judgment: Frequency formats and simple algorithms. *Medical Decision Making*, *16*(3), 273–280.

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, *3*(1), 20–29.
Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482.
Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684.
Gigerenzer, G., Todd, P. M., & ABC Research Group. (2000). Simple heuristics that make us smart. *Evolution and cognition*. USA: Oxford University Press.
Guo, Z. (2011). Mind the map! The impact of transit maps on path choice in public transit. *Transportation Research Part A: Policy and Practice*, *45*(7), 625–639.
Hawley, S. T., Zikmund-Fisher, B., Ubel, P., Jancovic, A., Lucas, T., & Fagerlin, A. (2008). The impact of the format of graphical presentation on health-related knowledge and treatment choices. *Patient Education and Counseling*, *73*(3), 448–455.
Highhouse, S. (1994). A verbal protocol analysis of choice under ambiguity. *Journal of Economic Psychology*, *15*(4), 621–635.
Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine*, *73*(5), 538–540.
Hofman, J. M., Goldstein, D. G., & Hullman, J. (2020). How visualizing inferential uncertainty can mislead readers about treatment effects in scientific results. In *Paper presented at the proceedings of the 2020 CHI conference on human factors in computing systems*.
Huettel, S. A., Stowe, C. J., Gordon, E. M., Warner, B. T., & Platt, M. L. (2006). Neural signatures of economic preferences for risk and ambiguity. *Neuron*, *49*(5), 765–775.
Hullman, J., Kay, M., Kim, Y.-S., & Shrestha, S. (2017). Imagining replications: Graphical prediction & discrete visualizations improve recall & estimation of effect uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, *24*(1), 446–456.
Hullman, J., Resnick, P., & Adar, E. (2015). Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PLoS One*, *10*(11), e0142444.
Joslyn, S., & LeClerc, J. (2013). Decisions with uncertainty: The glass half full. *Current Directions in Psychological Science*, *22*(4), 308–315.
Joslyn, S., & Savelli, S. (2020). Visualizing uncertainty for non-expert end users: The challenge of the deterministic construal error. *Frontiers in Computer Science*, *2*, 58.
Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy*, *98*(6), 1325–1348. Chicago.
Kahneman, D., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases* (1st ed.). Cambridge; New York: Cambridge University Press.
Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge University Press.
Kale, A., Kay, M., & Hullman, J. (2020). *Visual reasoning strategies for effect size judgments and decisions*. arXiv e-prints, arXiv: 2007.14516.
Kale, A., Nguyen, F., Kay, M., & Hullman, J. (2018). Hypothetical outcome plots help untrained observers judge trends in ambiguous data. *IEEE Transactions on Visualization and Computer Graphics*, *25*(1), 892–902.
Kay, M., Kola, T., Hullman, J. R., & Munson, S. A. (2016). When (ish) is my bus? user-centered visualizations of uncertainty in everyday, mobile predictive systems. In Proceedings of the 2016 chi conference on human factors in computing systems (pp. 5092-5103).
Kim, Y.-S., Walls, L. A., Krafft, P., & Hullman, J. (2019). A Bayesian cognition approach to improve data visualization. In *Paper presented at the proceedings of the 2019 CHI conference on human factors in computing systems*.

Kinkeldey, C., MacEachren, A. M., Riveiro, M., & Schiewe, J. (2017). Evaluating the effect of visually represented geodata uncertainty on decision-making: Systematic review, lessons learned, and recommendations. *Cartography and Geographic Information Science*, *44*(1), 1–21. https://doi.org/10.1080/15230406.2015.1089792.

Kinkeldey, C., MacEachren, A. M., & Schiewe, J. (2014). How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, *51*(4), 372–386.

Kool, W., & Botvinick, M. (2014). A labor/leisure tradeoff in cognitive control. *Journal of Experimental Psychology: General*, *143*(1), 131.

Kosslyn, S. M. (1995). In *Vol. 2. Mental imagery*. *Visual cognition: An invitation to cognitive science* (pp. 267–296).

Knight, F. H. (2012). *Risk, uncertainty and profit*. Courier Corporation.

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science*, *4*(2), 195–208.

Leonard, W. R., & Robertson, M. L. (1994). Evolutionary perspectives on human nutrition: The influence of brain and body size on diet and metabolism. *American Journal of Human Biology*, *6*(1), 77–88.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*(1), 37–44.

Liu, L., Boone, A., Ruginski, I., Padilla, L., Hegarty, M., Creem-Regehr, S. H., et al. (2016). Uncertainty visualization by representative sampling from prediction ensembles. *IEEE Transactions on Visualization and Computer Graphics*, *23*(9), 2165–2178.

Liu, L., Padilla, L., Creem-Regehr, S. H., & House, D. (2019). Visualizing uncertain tropical cyclone predictions using representative samples from ensembles of forecast tracks. *IEEE Transactions on Visualization Computer Graphics*, *25*(1), 882–891.

Logie, R. H., & Marchetti, C. (1991). Visuo-spatial working memory: Visual, spatial or central executive? In R. H. Logie, & M. Denis (Eds.), *Vol. 80. Advances in psychology* (pp. 105–115). Elsevier.

Maceachren, A. M., Robinson, A., Gardner, S., Murray, R., Gahegan, M., & Hetzler, E. (2005). Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science*, *32*, 160.

MacEachren, A. M., Roth, R. E., O'Brien, J., Li, B., Swingley, D., & Gahegan, M. (2012). Visual semiotics amp;amp; uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2496–2505. https://doi.org/10.1109/TVCG.2012.279.

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, *5*(2), 110–141.

McKenzie, G., Hegarty, M., Barrett, T., & Goodchild, M. (2016). Assessing the effectiveness of different visualizations for judgments of positional uncertainty. *International Journal of Geographical Information Science*, *30*(2), 221–239.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.

Millet, B., Cairo, A., Majumdar, S. J., Diaz, C., Evans, S. D., & Broad, K. (2020). *Beautiful visualizations slain by ugly facts: Redesigning the National Hurricane Center's "Cone of Uncertainty" Map*. VisComm Workshop.

Montibeller, G., & Von Winterfeldt, D. (2015). Cognitive and motivational biases in decision and risk analysis. *Risk Analysis*, *35*(7), 1230–1251.

Munzner, T. (2014). *Visualization analysis and design*. CRC Press.

Newman, G. E., & Scholl, B. J. (2012). Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias. *Psychonomic Bulletin & Review*, *19*(4), 601–607. https://doi.org/10.3758/s13423-012-0247-5.

Okan, Y., Garcia-Retamero, R., Galesic, M., & Cokely, E. T. (2012). When higher bars are not larger quantities: On individual differences in the use of spatial information in graph comprehension. *Spatial Cognition & Computation*, *12*(2–3), 195–218.

Padilla, L. M., Hansen, G., Ruginski, I. T., Kramer, H. S., Thompson, W. B., & Creem-Regehr, S. H. (2015). The influence of different graphical displays on nonexpert decision making under uncertainty. *Journal of Experimental Psychology: Applied*, *21*(1), 37. Chicago.

Padilla, L., Castro, S., Quinan, P. S., Ruginski, I. T., & Creem-Regehr, S. H. (2019). Toward objective evaluation of working memory in visualizations: A case study using pupillometry and a dual-task paradigm. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 332–342.

Padilla, L., Creem-Regehr, S., Hegarty, M., & Stefanucci, J. (2018). Decision making with visualizations: A cognitive framework across disciplines. *Cognitive Research: Principles and Implications*, *3*(1), 1–25.

Padilla, L. M. K., Creem-Regehr, S. H., & Thompson, W. (2020). The powerful influence of marks: Visual and knowledge-driven processing in hurricane track displays. *Journal of Experimental Psychology: Applied*, *26*(1), 1–15. https://doi.org/10.1037/xap0000245.

Padilla, L., Kay, M., & Hullman, J. (2021). Uncertainty visualization. In N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri, & J. L. Teugels (Eds.), *Wiley StatsRef: Statistics Reference Online* Springer Science. In Press.

Padilla, L. M., Powell, M., Kay, M., & Hullman, J. (2020). Uncertain about uncertainty: How qualitative expressions of forecaster confidence impact decision-making with uncertainty visualizations. *Frontiers in Psychology*, *11*, 3747.

Padilla, L., Ruginski, I. T., & Creem-Regehr, S. H. (2017). Effects of ensemble and summary displays on interpretations of geospatial uncertainty data. *Cognitive Research: Principles and Implications*, *2*(1), 40.

Pandey, A. V., Rall, K., Satterthwaite, M. L., Nov, O., & Bertini, E. (2015). How deceptive are deceptive visualizations? An empirical analysis of common distortion techniques. In *Paper presented at the proceedings of the 33rd annual ACM conference on human factors in computing systems*.

Pang, A. T., Wittenbrink, C. M., & Lodha, S. K. (1997). Approaches to uncertainty visualization. *The Visual Computer*, *13*(8), 370–390. https://doi.org/10.1007/s003710050111.

Peck, E. M. M., Yuksel, B. F., Ottley, A., Jacob, R. J., & Chang, R. (2013). Using fNIRS brain sensing to evaluate information visualization interfaces. In *Paper presented at the proceedings of the SIGCHI conference on human factors in computing systems*.

Ruginski, I. T., Boone, A. P., Padilla, L., Liu, L., Heydari, N., Kramer, H. S., et al. (2016). Non-expert interpretations of hurricane forecast uncertainty visualizations. *Spatial Cognition & Computation*, *16*(2), 154–172.

Savelli, S., & Joslyn, S. (2013). The advantages of predictive interval forecasts for non-expert users and the impact of visualizations. *Applied Cognitive Psychology*, *27*(4), 527–541.

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2015). Working memory capacity and the scope and control of attention. *Attention, Perception, & Psychophysics*, 77(6), 1863–1880.

Spiegelhalter, D. (2017). Risk and uncertainty communication. *Annual Review of Statistics and Its Application*, *4*, 31–60.

Szafir, D. A., Haroz, S., Gleicher, M., & Franconeri, S. (2016). Four types of ensemble coding in data visualizations. *Journal of Vision*, *16*(5), 11. https://doi.org/10.1167/16.5.11.

Tait, A. R., Voepel-Lewis, T., Zikmund-Fisher, B. J., & Fagerlin, A. (2010). The effect of format on parents' understanding of the risks and benefits of clinical research: A comparison between text, tables, and graphics. *Journal of Health Communication*, *15*(5), 487–501.

Tversky, B. (2001). Spatial schemas in depictions. In *Paper presented at the spatial schemas and abstract thought*.

Tversky, B. (2011). Visualizing thought. *Topics in Cognitive Science*, *3*(3), 499–535.

Tversky, B., Corter, J. E., Yu, L., Mason, D. L., & Nickerson, J. V. (2012). *Representing category and continuum: Visualizing thought*.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.
Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: An integrated systems perspective. *Science*, *255*(5043), 419–423.
Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, *22*(2), 219–240.
Waters, E. A., Fagerlin, A., & Zikmund-Fisher, B. J. (2016). Overcoming the many pitfalls of communicating risk. In M. Diefenbach, S. Miller-Halegoua, & D. Bowen (Eds.), *Handbook of health decision science* (pp. 265–277). Springer.
Waters, E. A., Weinstein, N. D., Colditz, G. A., & Emmons, K. (2006). Formats for improving risk communication in medical tradeoff decisions. *Journal of Health Communication*, *11*(2), 167–182.
Waters, E. A., Weinstein, N. D., Colditz, G. A., & Emmons, K. M. (2007). Reducing aversion to side effects in preventive medical treatment decisions. *Journal of Experimental Psychology: Applied*, *13*(1), 11.
Wertheimer, M. (1938). *Laws of organization in perceptual forms*. Kegan Paul, Trench, Trubner & Company.
Wilson, K. A., Heinselman, P. L., Skinner, P. S., Choate, J. J., & Klockow-McClain, K. E. (2019). Meteorologists' interpretations of storm-scale ensemble-based forecast guidance. *Weather, Climate, and Society*, *11*(2), 337–354.
Winter, B., & Matlock, T. (2013). More is up… and right: Random number generation along two axes. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*(35).
Xiong, C., Van Weelden, L., & Franconeri, S. (2019). The curse of knowledge in visual data communication. *IEEE Transactions on Visualization and Computer Graphics*, *26*(10), 3051–3062.
Yin, L., Shi, Z., Liao, Z., Tang, T., Xie, Y., & Peng, S. (2020). The effects of working memory and probability format on Bayesian reasoning. *Frontiers in Psychology*, *11*, 863.
Zacks, J. M., & Franconeri, S. L. (2020). Designing graphs for decision-makers. *Policy Insights From the Behavioral and Brain Sciences*, 7(1), 52–63.
Zacks, J., & Tversky, B. (2013). Bars and lines: A study of graphic communication. *Memory & Cognition*, *27*(6), 1073–1079. https://doi.org/10.3758/BF03201236.
Zikmund-Fisher, B. J., Witteman, H. O., Dickson, M., Fuhrel-Forbis, A., Kahn, V. C., Exe, N. L., et al. (2014). Blocks, ovals, or people? Icon type affects risk perceptions and recall of pictographs. *Medical Decision Making*, *34*(4), 443–453.