Heuristics for Supporting Cooperative Dashboard Design

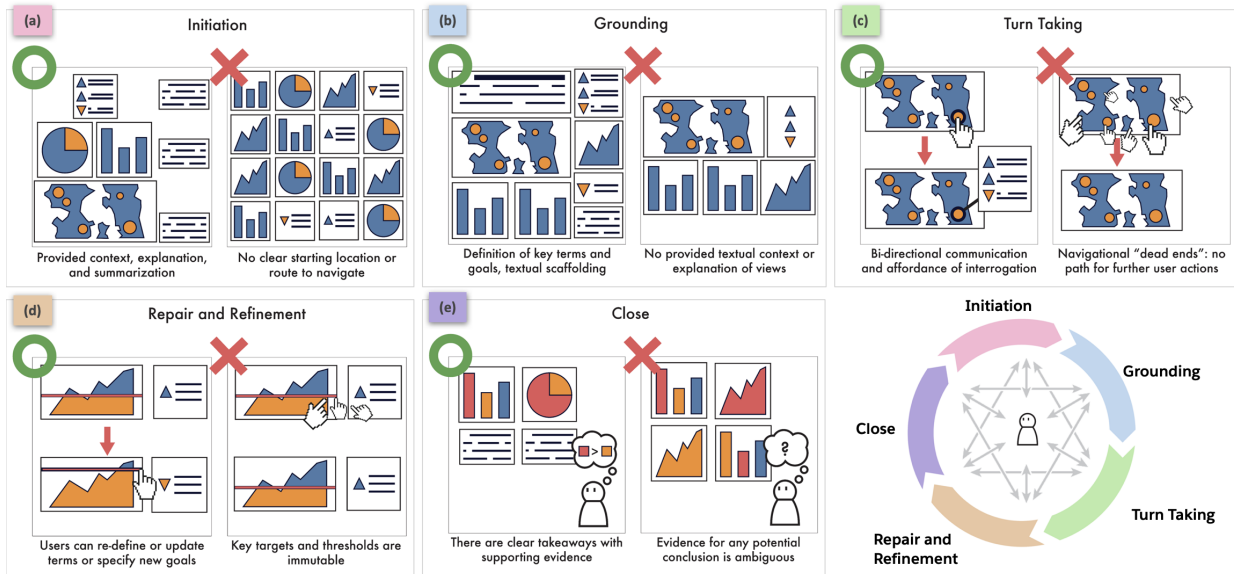Vidya Setlur (iD), Michael Correll (iD), Arvind Satyanarayan (iD), and Melanie Tory (iD)

Fig. 1: Successes and failures of *cooperative dashboard* design throughout the five analytic states of a conversation (a-e). Cooperative dashboards guide users through their data and, in contrast to static dashboards, provide bi-directional communication through interactivity to allow the user to change or refine their analytical goals, switch between topics of interest and levels of detail, correct or update the system if it provides irrelevant or incorrect information, and provide useful summaries of analytical actions. Note that these conversation states are not necessarily sequential and the analyst can move between these various states.

Abstract— Dashboards are no longer mere static displays of metrics; through functionality such as interaction and storytelling, they have evolved to support analytic and communicative goals like monitoring and reporting. Existing dashboard design guidelines, however, are often unable to account for this expanded scope as they largely focus on best practices for visual design. In contrast, we frame dashboard design as facilitating an *analytical conversation*: a cooperative, interactive experience where a user may interact with, reason about, or freely query the underlying data. By drawing on established principles of conversational flow and communication, we define the concept of a *cooperative dashboard* as one that enables a fruitful and productive analytical conversation, and derive a set of 39 dashboard design heuristics to support effective analytical conversations. To assess the utility of this framing, we asked 52 computer science and engineering graduate students to apply our heuristics to critique and design dashboards as part of an ungraded, opt-in homework assignment. Feedback from participants demonstrates that our heuristics surface new reasons dashboards may fail, and encourage a more fluid, supportive, and responsive style of dashboard design. Our approach suggests several compelling directions for future work, including dashboard authoring tools that better anticipate conversational turn-taking, repair, and refinement and extending cooperative principles to other analytical workflows.

Index Terms—Gricean maxims, interactive visualization, conversation initiation, grounding, turn-taking, repair and refinement.

---◆---

1 INTRODUCTION

Dashboards have become ubiquitous for analyzing and communicating data because their expressive designs allow them to address a diverse range of purposes and contexts [66]. However, existing guidelines for dashboard design—whether in research or popular press [27, 95]—

- *Vidya Setlur is with Tableau Research. E-mail: vsetlur@tableau.com*
- *Michael Correll completed this work while at Tableau Research. E-mail: m.correll@northeastern.edu*
- *Arvind Satyanarayan is with MIT CSAIL. E-mail: arvindsatya@mit.edu*
- *Melanie Tory is with Northeastern University. E-mail: m.tory@northeastern.edu*

largely focus on issues of visual representation, perception, and graphic design. While important, this focus ignores the central role that interactivity and storytelling increasingly play in enabling users to explore, analyze, monitor, and track various data metrics [66]. What design guidance can we provide for these interactive capabilities?

Prior work has described interaction as *"engaging the data in dialogue"* [19, 85, 87]—an analogy to human-human conversation that we find productive for thinking about what it means for a dashboard's interaction to be designed effectively. Just as a human conversationalist can be circumlocutory, obscurant, or rude, so too can interactions in a dashboard be repetitive, unclear, or user-unfriendly. Moreover, gleaning insights from data is most productive and enjoyable when users can focus on answering the questions they have about their data rather than the mechanics of doing so [88]. But, what makes a dashboard an effective conversational partner?

We operationalize the conversation analogy by studying the *pragmatics* of language use, or how language shapes meaning [67]. We

define a dashboard as *cooperative* if it facilitates an interactive loop that follows the *Gricean Maxims* [34] — influential work in pragmatics that assesses the quality of a cooperative, communicative interaction based on the quantity, quality, relation, and manner of information communicated. Moreover, we draw on work by Beebe et al. [7] to model a cooperative analytical conversation as one where participants (in our case, the dashboard and the analyst) move between states of initiation, grounding, turn-taking, repair & refinement, and close.

Guided by these two frameworks, we enumerate a set of heuristics focused on interactive, cooperative communication between a dashboard and a user. Through an iterative process with 16 visualization practitioners, we distill down to a set of 39 design heuristics for promoting the design of cooperative dashboard conversations. To evaluate the utility of these heuristics in practice, we conduct two exercises with 52 computer science and engineering graduate students as part of optional, ungraded homework assignments. First, students were asked to use the heuristics to reflect on the efficacy of existing dashboard designs. Next, the students were asked to create a new dashboard or update the design of an existing dashboard based on the heuristics to better support cooperative conversational behavior with their target users.

Results of the classroom exercises indicate that our heuristics afford a new perspective for thinking about dashboard design. While dashboards tend to be effective at initiation and grounding a conversation, they are often weaker with respect to turn-taking, repair & refinement, and close. For instance, students noted how interactive results updating in place without any accompanying cues or messaging hinders turn-taking, as it can be difficult for a user to assess when an interaction is complete so they can resume their dialogue. When applying these heuristics to improve existing dashboard designs, students relied on textual annotation to provide contextual information and deliberately traded off visual aesthetics for clearer communication. Our results suggest opportunities for future work to study the impact of cooperative vs. uncooperative dashboard designs and to extend principles of cooperative conversation to analytical workflows beyond the dashboard.

2 RELATED WORK

Our work builds on three lines of research: understanding dashboard design and usage as representational media, conversational interactions with data, and design heuristics in HCI and visualization.

2.1 Understanding Dashboard Design and Usage

Dashboards are pervasive. They operate as the primary portal to data for many people in work and daily life. Yet until recently, dashboards were given little attention by the visualization research community. A survey of dashboards in the wild [66] offered a classification of dashboards and highlighted their criticality as a means of circulating data within organizations. An extension by Bach et al. [5] identified six distinct dashboard genres and characterized content and composition design patterns. Dimara et al. discussed the role dashboards play in supporting data-driven decision making [20], Zhang et al. described the work practices and challenges of dashboard creators [99], Lee-Robins and Adar [49] characterized affective intents in visualizations and dashboards, and Tory et al. [86] discussed the work practices of dashboard users. Research into dashboard design and construction includes approaches to enable layout and view consistency [63] and semantic snapping [45]. Research into multiple coordinated views and composite visualizations is also relevant to dashboard design (for a survey see Roberts [64] or Deng et al. [17]). Dashboard design is typically a manual process that can be aided by design heuristics such as those introduced in this work. However, design heuristics may also be codified into systems that automatically generate dashboards (e.g. [41]) or provide mixed-initiative support for dashboard creation [12, 59, 97]. Our research focuses on better understanding the dialogue around dashboards by introducing a set of heuristics to support dashboard design and evaluation for analytical conversation.

2.2 Conversational Interaction with Data

The novelty in our heuristics stems from framing people's interaction with dashboards as a conversation. Designers have long recognized

the power of interacting with computers in ways that emulate our conversational interactions with people. A long history of research on chatbots and other conversational interfaces is summarized in several surveys [2, 3, 11, 52]. In recent years, this research theme has extended into interactions with data. A survey of natural language interfaces (NLIs) for data visualization was introduced by Shen et al. [75]. This body of work has led to an understanding of principles for cooperative communication design in conversational bots [11, 73], including behaviors such as communicability, conscientiousness, conciseness, manner, proactivity, and turn-taking (drawing on the Gricean Maxims [34]).

Recent work recognizes that these cooperative principles apply beyond the scope of interfaces that employ spoken or written language. Most relevant are papers that characterize interactions with data and/or dashboards as data *conversations* [28, 55, 86]. Muller et al. [55] described how data science workers engage in back-and-forth interactions with data, especially for data wrangling. Tory et al. [86] described how dashboards serve as a portal to data and a jumping-off point to further data activities. Their observation that dashboards alone are often ineffective in supporting these conversations, resulting in data being exported for use in spreadsheets, presentation tools, and reports, suggests a strong need for dashboards to evolve in ways that support more conversational forms of interaction. *BOLT* explores the use of NLIs for dashboard authoring, wherein NL utterances are mapped to prevalent dashboard objectives to generate dashboard recommendations [80].Our work contributes a set of heuristics that can support designers in creating such cooperative, conversational dashboards and the systems that generate them.

2.3 Heuristics in HCI and Visualization

Our dashboard heuristics build upon a long history of design heuristics for interfaces and visualizations. In interface design and evaluation, perhaps the most well-known are Nielsen's [56, 57] ten usability guidelines and Shneiderman et al.'s [77] eight golden rules. More specific heuristics have been developed for topics such as human-AI interaction [4], augmented reality [24, 31], and mobile computing [8], among many others. Researchers have proposed numerous heuristics specific to conversational interaction with chatbots and voice assistants [30, 36, 47, 53, 58, 82, 94].

Tory & Möller [87] explored usability heuristics as a way to evaluate visualizations. Subsequently, there have been numerous efforts to develop and evaluate visualization-specific heuristics [15, 21, 29, 83, 89, 100]. The numerous high-level books, guidelines, and principles around dashboard design are also relevant (e.g., [27, 95, 98]), as are frameworks of user goals or intents that may help to guide visualization design (e.g., [46, 49]), and design tools considered to support cognition [93]. More recently, Lin et al. [51] introduced a data-driven approach for identifying a set of dashboard design rules from dashboards mined from the web. The rules describe view-wise relationships in terms of data, encoding, layout, and interactions and subsequently develop a recommender for dashboard design.

However, heuristics and guidelines for dashboards tend to focus on layout, structure, data and its visual representation, and usability. Our work augments these guidelines based on principles of cooperative conversation. The conversational framing offers a different perspective that aligns with an evolution of dashboards away from autocratic information artifacts and towards cooperative conversational partners.

3 ANALYTIC CONVERSATION STATES

The motivation for this work stems from exploring how cooperative conversation guidelines for human-computer interfaces could inform the design and evaluation of interactive dashboards. Conversation is highly structured and organized according to set principles. Sacks et al. [65] initiated the modern literature on conversational behavior by outlining a system of social interactions with specific properties. This interaction is characterized by a mechanism of exchange based on alternating dialogues of information.

Beebe et al. [7] break conversation down into five states (i.e., initiation, grounding, turn-taking, repair & refinement, and close) that we adapt here for our discussion around interactive dashboards. While

Gricean Maxims [34] provide guidelines for assessing the overall quality of a conversation, the conversation states specifically help define how an analytical conversation progresses through different interaction states; they also help organize the heuristics. We maintain, as per Tory et al. [86], that the users of dashboards are similarly engaged in "data conversations", so conversational structures (and pitfalls) can apply to dashboards and to considerations for their design. In this section, we introduce and apply these conversational states to dashboard interaction for supporting analytical conversation with the user (Figure 1).

3.1 Initiation

Initiation is the first stage of conversation and requires one to be open to interacting with the other conversational participant(s). Greetings such as, "Hello!" and "How nice to see you!" are common ways to set the tone to welcome further dialogue. Conversations can also be initiated without any preliminaries using utterances such as, "when will it stop raining?" or including vocative or attention-seeking utterances such as, "excuse me" or "hey!"

With respect to dashboard design, initiation can be thought of as both the state of the dashboard when the user first interacts with it, as well as any tutorials, explanations, or other tools for orienting the user to the dashboard's contents. Dhanoa et al. [18] suggest an "onboarding model" for new users of dashboards. A successful onboarding process, per this model, is mindful of the target user, the dashboard components that will need likely explanation, how these explanations will be serviced, and how this onboarding process connects to later patterns of usage. The means and goals of onboarding are then connected with an "onboarding narrative". For instance, a "depth-first narrative" might involve a serial explanation of every dashboard component (and their subcomponents) in detail. As in Figure 1, a successful initiation in dashboard design provides the user with **information and explanations** of components, but also clear options for where to begin to understand their data. A failure can occur either through the lack of appropriate onboarding (e.g., an insufficient quantity of onboarding for the user, insufficient relevance to their task, or missing context) or even by presenting a "data deluge" of too many unconnected or unstructured views without a clear reading order or spatial organization.

Other strategies for successful initiation are to provide users with **curated information and metadata**. For instance, as in Srinivasan et al. [79], dashboards can be augmented with "data facts" of potentially important relationships or patterns in the data. Or, as Gebru et al. [33], a "datasheet" or important context and metadata could be provided to a user prior to any analysis.

3.2 Grounding

Grounding refers to establishing the time, location, or actuality of a situation according to some reference point in the conversation [13]. Two people in a conversation need to coordinate not only the content of what they say but also how that message is delivered. For example, if Mary wants to get Clara to join her for lunch at a particular restaurant, she cannot simply email her with - "Let's meet at Sol at noon." After sending her invitation, Mary awaits evidence that Clara has received, understood, and committed to the lunch invitation. Meanwhile, Clara does not find a taxi as soon as she gets Mary's message but sends an email response. To be at common ground, if Mary and Clara need to further clarify or modify their plans, they may exchange additional emails before they consider their plan to meet at the restaurant.

With respect to dashboards, while onboarding (as discussed above) can assist in moving users through the grounding stage, there are other actions designers can take to build shared understanding of expectations and terms. The first might be the direct solicitation of priors and predictions from users. Hullman & Gelman [38] suggest that existing (ungrounded) "model free" visualizations are inherently limited for visual analytics, and point to examples where either asking the user to predict data [42] or, alternatively, showing users the predictions of others [43], can not only result in improved recall and retention of information, but also avoid drawing spurious conclusions. Shi et al. [76] similarly point to cases where data stories solicit information from users in order to ensure that the resulting information is **relevant,**

interesting, or contextualized for users, and Lin et al. [50] call for incorporating users' "data hunches" into charts.

3.3 Turn-taking

Turn-taking is a fundamental aspect of dialogue and occurs in a conversation when one person listens while the other person speaks [65]. As the conversation progresses, the listener and speaker roles are exchanged back and forth. Participants need to coordinate who is currently speaking and when the next person can start to speak. Humans are very good at this coordination and typically achieve fluent turn-taking with very small gaps and little overlap. A conversationalist who does not allow others a turn, or speaks over others, may be considered rude. An example of turn-taking in conversation is:

SPEAKER A: "Lovely weather this week."
SPEAKER B: "Isn't it? I hope it's nice on the weekend."
SPEAKER A: "Me too. I have plans to go for a hike."
SPEAKER B: "That's fun! Which trail are you going on?"

As dashboards move from static displays to more complex and interactive forms [66], there are an increasing number of examples of **bi-directional communication** between the user and a visualization system. Examples of this communication can be as simple as providing tooltips or annotations on a user's request, supporting filtering or aggregation options, to more complex forms such as soliciting personal information from the user [76] or even incorporating "analytical chatbots" [73] that respond to natural language queries. Failure to allow the user to perform follow-up actions (as in Figure 1) can result in frustrating analytical experiences where a user has a question or concern that the dashboard is not equipped to address. For example, [86] a sales dashboard that only allows a user to see a snapshot of the data at a single point in time can be frustrating if the user's next step is to try to understand the data in the context of the last month or year.

Dashboards systems can also take conversational initiative, and there are potential analytical benefits for such "proactive design." [88] An example is the Frontier system [48], where the user can select recommended views based on a set of analytical intents. Other forms of bidirectional interaction can be more subtle: for instance, the autocompletion metaphor in visual analytics [71] represents an attempt to match a user's utterance or intended action with the system's understanding of valid or popular alternatives. One consideration with turn-taking in dashboards is to allow bi-directional communication and useful division of labor between the person and the system, while respecting the user's agency and autonomy [35]. Systems that steal focus, override user choices, and lead to **dead-ends** in the communication flow, are "impolite" [96] and produce friction and user enmity.

3.4 Repair and Refinement

Conversational repair and refinement is the process conversation participants use to detect and resolve problems of speaking, hearing, and understanding [68]. If dialogue is to proceed smoothly, it is vital that there are opportunities for checking to understand and provide clarification when misunderstanding does occur. Everyday interaction is full of such checks and repairs, though these may be so automatic as to be almost seamless, rarely disturbing the flow of the interaction. In human conversation, there are continual implicit acknowledgments that communication is proceeding smoothly. The speaker monitors the participants in the conversation in different ways to see if they understand (e.g., using checking moves such as "Do you know what I mean?") and the other participants are often giving verbal acknowledgments to the speaker (e.g., "yes", "uh huh"). However, if the utterance is not understood, repair may be initiated. Through repair, participants display how they establish and maintain communication and mutual understanding during the turn-taking process.

Repair and refinement are both critical components of interactive dashboard design. NLIs for data provide a model for this sort of interaction, as natural language utterances (and the intents behind them) are often vague [72], under-specified [74], or misinterpreted by the natural language system. Some systems **afford follow-up conversations** for repairing or re-specifying intents. Perhaps more relevant to dashboard

design are systems like DataTone's [32] "ambiguity widgets" that explicitly afford the resolution of ambiguous queries. The inability to update a dashboard when information is stale, irrelevant, incorrect, or misaligned with the user's goal can lead to frustration, as in Figure 1.

Another way to support repair in analytic conversations with dashboards is to support fluid switching of tools and contexts if the existing dashboard is insufficient for a particular analytical task. Both Tory et al. [86] and Bartram et al. [6], in their interviews with "data workers": reveal a recurring need to move data between tools (for instance, into a spreadsheet tool for manual data cleaning or inspection, or into a presentation tool for curated storytelling), and frustration with existing dashboard software that makes this process difficult.

A last intriguing potential for repair in dashboard design is to **employ summaries or recommendations to prevent or ameliorate cognitive biases** on the part of the users [90, 92]. That is, a belief that the user is making a potential analytical error or oversight and intervening. For instance, Wall et al. [91] propose the incorporation of a user's interaction records to provide a summary report explaining whether they are interacting with biased subsamples of the whole dataset, or whether they have considered representative facets of the data.

3.5 Close

Close is the process by which two partners end a conversation by offering and accepting each other's final bids to close the conversation. Politeness strategies can avoid miscommunication when terminating the conversation. Coppock [14] proposed several strategies used to close the conversation: positive comment, excuse, and the imperative (e.g., "it looks like our time is up"). A positive comment implies that the conversation is pleasant, but the other does not want to continue. Excuse expresses an intent to end the conversation by providing an alternative motivation (e.g., "I better get back to work"). The imperative strategy explicitly employs an imperative tone to end the conversation (e.g., "It was nice talking to you").

While the end of a specific analytical *session* may be clear cut (say, navigating away from a website or closing a piece of software), a user's analytical *conversation* does not end when they stop looking at a dashboard; the notion of a final close is more fraught. As users of dashboards are commonly impacted by *reliance* on others [86] (either for data, stakeholder buy-in, or discussion of goals), there is often a step of sharing the insights gleaned from an analytical conversation with various levels of formality and practice [9]. **Providing useful summaries of information or insights in a dashboard**, and particularly summaries that can "travel" across different modalities, is, therefore, a critical (but often overlooked) component of dashboard design. Of particular interest to us is how summaries can concisely present not only the insights gained by the user over the course of an analytical conversation but also the supporting evidence for these insights (and the strength of this evidence).

Beyond post hoc summaries, we point to two potential examples of visualizations making good use of the *end* of analytical conversations. The first involves systems where past users can provide important context for future users, as with Feng et al.'s [26] Hindsight system where the interaction history of other users can be used to suggest potential starting places for new users, or in Kim et al. [43] where other viewers' predictions can help situate one's own expectations of the relationships between data values. The second example embraces the multiplicity of potential methods and the potential fragility of conclusions, as in Dragicevic et al.'s [22] multiverse analysis reports, where the goal is to produce a report (with included conclusion and discussion sections) that is robust across a variety of different analytical choices or even natural data variability.

4 ITERATIVE DEVELOPMENT OF CONVERSATION HEURISTICS

We apply the notion of cooperative conversation and its maxims by examining the conversational properties that are specifically relevant to interactive dashboards, drawing from the following sources:

- **Natural language interfaces for visual analysis**: We explore how language pragmatics in the context of natural language interfaces can help support analytical conversation. A review of

Heuristic	Heuristic
H1: The dashboard supports specific analytical questions or tasks.	**H21:** There is sufficient contextual information that describes what the dashboard is about.
H2: The user should be able to explore the data using the dashboard.	**H22:** The charts within the dashboard make sense as an overall composition to help situate or contextualize the user.
H3: It is clear to the user where they need to start interacting with the dashboard.	**H23:** The text in the dashboard should emphasize the most salient points of what the visuals in the dashboard convey.
H4: The charts in the dashboard support tasks that represent the overall intended goal to the user.	**H24:** The dashboard communicates a certain style or mood to the user.
H5: There is a clear reading order within the dashboard and is it logical (e.g., top-down, bottom-up).	**H25:** The dashboard is interactive and supports the user in completing a new analytical task or starting a new line of inquiry.
H6: The text in the dashboard is legible, easy to read, and useful. The different parts of the chart (e.g. titles, captions, or narration) are well-described.	**H26:** Tasks in the dashboard should be achieved with minimal friction.
H7: The charts in the dashboard clearly identify, compare, or summarize a particular aspect of the data.	**H27:** When a user interacts with the dashboard, the dashboard should guide the user with their next step.
H8: The layout, placement of charts, and the flow in the visualization should be easy to follow.	**H28:** The dashboard should update its view based on what is selected, highlighted or filtered by the user.
H9: The visual systems in place (e.g., a consistent font and color brand) provide guidance in understanding the analysis.	**H29:** There are text and visual elements to frame or guide salient information.
H10: There is a clear and consistent visual style for elements such as titles, axes, and colors.	**H30:** There is a clear path and breadcrumbs for user actions within the dashboard.
H11: The conventions used in the dashboard such as icons are easy to understand for the user.	**H31:** There are clear strategies employed in the dashboard to mark charts or marks more prominently than others to encourage a user to interact with them, as well as de-emphasize items not relevant to the conversation.
H12: There are consistent rules for the formatting of elements such as gridlines, borders, and supporting ink elements.	**H32: Functional dead end:** The dashboard should not freeze, crash, display errors, or otherwise unexpectedly interrupt the user.
H13: The dashboard should disclose any biases. These can include the author's personal biases, design biases, and/or biases in the data.	**H33: Navigational dead end:** The dashboard supports interactivity (e.g., filters, search) to help a user navigate between different snapshots of the data being presented.
H14: The dashboard should communicate where the data came from, and what steps were taken to prepare the data.	**H34:** The dashboard should support input modalities (e.g., natural language interaction, filter widgets, and / or clicks) to answer any analytical question or task.
H15: Charts within the dashboard are at the appropriate level of detail to convey the intended message.	**H35:** The user can set or adjust their own definition of concepts or metrics, or otherwise include their own knowledge into the dashboard. (e.g., if a user is interested only in "tall" people, can they set their own definition of "tall")
H16: There is a clear presentation of context, theme, and detail as part of the analytical exploration with the dashboard.	**H36:** There should be a clear takeaway from the dashboard.
H17: The dashboard employs visual symbols and iconography to add meaning to the data being presented and help support the intended message.	**H37:** The dashboard should not contain any information that distracts from the key takeaway.
H18: If there are icons and symbols in the dashboard, the semantics of their placement and appearance are helpful to communicate patterns in the data. (e.g., changing a plane icon's location and direction for arrival vs. departure)	**H38:** The conclusions match what the charts in the dashboard show.
H19: The iconography supports or potentially replaces repetitive text directives.	**H39:** There is adequate evidence that the dashboard is truthful. The dashboard should be able to convince the key takeaway through credibility and trustworthiness
H20: Concepts or metrics (e.g., date formats , time, currency, measurement) are either easily understandable or clearly defined in the dashboard.	

Table 1: Through an iterative process, we distilled 39 heuristics for analytical conversation organized into the five conversation states: initiation, grounding, turn-taking, repair & refinement, and close. A more detailed version of this table with the labeled categories and assigned Gricean Maxims is provided in supplementary material.

previous academic prior art and software systems that implement techniques for supporting analytical conversation in the context of NLIs for visual analysis [32, 37, 39, 62, 69, 73, 81, 84] provided guidelines for informing the various heuristics for supporting the various conversational states when interacting with data.

- **Cooperative conversation behaviors in human-computer interfaces**: The design of such interfaces often draws inspiration from human-to-human conversation and mechanisms that facilitate the exchange of information between speaker and listener. There exists an expectation that the information shared is relevant and that intentions are clearly conveyed to support a cooperative conversation that is truthful, relevant, concise, and clear. A review of the various applications of Gricean Maxims and cooperative conversation guidelines in interactive interfaces and experiences between humans and computers, ranging from human-bot interaction, chatbots, smart assistants, and embodied agents [10, 23, 60] helped define the various heuristics that satisfy the maxims.

- **Practioner examples of dashboard design**: Interactive design guidelines for authoring functionally useful interactive dashboards as described in practitioner literature [44, 70], provided examples to help inform the creation of the initial set of heuristics.

However, as indicated in Section 2, many of the guidelines from the

visualization literature tend to focus on recommendations and best practices for layout, visual composition, data encodings, and chart types, as well as for natural language interfaces and systems. We instead focus on Grice's Cooperative Principle and its associated maxims as a way to identify themes to support analytical conversations in interactive dashboards. In particular, we apply the notion of *conversational implicature* as a way to systematize the properties of interactive dashboards. Conversational implicature, as introduced by Grice, is an indirect or implicit act within a conversation, determined by the conversational context that supports the primary dialogue [16, 34]. Implicature serves a variety of conversation goals towards effective communication, supporting pragmatics, maintaining good social relations, and overall efficiency in conveying the intended message. To come up with an initial set of heuristics, the co-authors adapted guidelines and heuristics developed for natural language interfaces to interactive dashboards (e.g., "Does the dashboard freeze, crash, display errors, or otherwise unexpectedly interrupt the user?") and drew inspiration from example dashboards authored by visualization experts (e.g., "Is there a clear reading order and is it logical (e.g., top-down, bottom-up)?".

All the co-authors iteratively developed a set of heuristics, organized into themes, that support conversational implicature through *both* the presentation and interaction of dashboards with a human. Each co-author picked one of three dashboard examples of their choice that they encountered recently - a Tableau Public dashboard showing the best states to retire in the US [25], a COVID-19 Dashboard [1], and a Tableau World Indicators Business Dashboard [78] and independently reviewed the current heuristics to assess if they were relevant (including whether they were supported or violated) to the corresponding dashboard example. Subsequently, the co-authors collectively discussed and compared insights on what it meant for a dashboard to be cooperative.

We initially collected 95 potential heuristics. Note that we chose the term 'heuristic' defined as "serving as an aid to learning, discovery, or problem-solving by experimental and especially trial-and-error methods" and "relating to exploratory problem-solving techniques that utilize self-educating techniques to improve performance" [54] as a means to help guide a dashboard author. Through our experience in using the heuristics, we clustered them into related themes and iteratively reworded and clarified them to minimize unfamiliar jargon or other vague terms. This process resulted in 56 heuristics.

4.1 Phase 1: Pilot Review

We tested the modified set of heuristics with two pilot participants. The instructions asked each participant to pick an interactive dashboard that they recently authored, run the heuristic checklist by the dashboard, and respond with detail about whether the dashboard supported the given heuristic or not. Lastly, they were asked to indicate if any of the heuristics were confusing to understand or apply. Based on feedback from this exercise, we updated the instructions to include an example along with a screenshot of a sample dashboard for a heuristic, refined and consolidated the heuristics further, resulting in 53 heuristics under 13 themes (analytical conversation support, multi-modal conversation support, use of semiotics, clarification of vague concepts, communication goal, summaries and takeaways, exposition, integrating text with visual information, composition and layout, visual scaffolding, level of detail, trust and transparency, register).

4.2 Phase 2: Expert Feedback

We then conducted a self-reflection exercise with 16 expert visualization researchers and practitioners. Based on self-reporting, experts comprised six business intelligence analysts, five data visualization consultants, and five Ph.D. visualization students with at least three years of experience authoring visualizations and dashboards. One participant did not complete the exercise, leading to a total of 15 completed exercises. The goals of Phase 2 were to 1) understand how the heuristics are applied when critiquing the design and interaction of a dashboard and 2) get feedback about the clarity and usefulness of the heuristics.

4.2.1 Expert Reflection Exercise

The user study was designed as a self-reflection exercise where participants were asked to evaluate each heuristic against a dashboard example that did not contain confidential or proprietary data. We asked them to include a link and a screenshot of the dashboard they picked with an explanation for their choice. We included a link to a spreadsheet of heuristics, and for each heuristic, the spreadsheet asked participants to first determine if the heuristic applied to their chosen dashboard and, if not, to explain the reason. They were also asked to rate the extent of the application or violation on a 5-point semantic differential scale ranging from "Strong violation" to "Strong application." The spreadsheet also requested participants to provide visual examples of applications and violations of the chosen dashboard for each heuristic wherever possible. To help the participants understand the expectations for the exercise, we provided an example response to one of the heuristics.

After the participants completed the heuristics spreadsheet, they were requested to answer a set of questions:

- Were the heuristics and/or themes useful? How? Which ones in particular? Explain in detail.

- Were any of the heuristics not helpful or confusing to you? If so, please elaborate.

- Did any of the heuristics make you think of dashboard design in a new way?

- Were there any heuristics that you thought were missing?

- What changes would you make to your dashboard based on this assessment? Please describe in detail.

- Do you plan on updating your dashboard in response to these heuristics? If so, would you be willing to send us an update?

We estimated the study would take approximately 45 minutes to complete. Participants were given three days to complete the study on their own time and were compensated with a $30 Amazon gift card. We recruited the expert participants (indicate by the notation [*E#*]) through a screening survey (included in supplementary material) posted on social media channels and distribution lists at a large software company. Participants were required to have experience (at least five years) designing or evaluating interactive dashboards using software like Tableau or PowerBI, notebook environments like Jupyter or Observable, or libraries like D3 or matplotlib. We also required participants to have a dashboard they were working on and that they were willing to share with us in some form (as a web link or a screenshot). We collected background information of the survey respondents that included a description of their current job role, years of experience designing dashboards, and a description of the topic and the audience of the interactive dashboard that they were designed for.

4.2.2 Assessing the Utility of the Heuristics

To assess the utility and comprehensibility of the heuristics, we reviewed participant responses for the following scenarios:

- **Heuristics indicated as 'does not apply'**. Instances where participants indicated that the heuristic was not relevant to the dashboard they were evaluating.

- **Misinterpreted or hard to understand heuristics**. Instances where participants misinterpreted a heuristic for another or simply did not understand them.

- **Heuristics marked as 'strong violation' / 'weak violation'**. Instances where participants indicated that their dashboard violated a given heuristic.

- **Heuristics marked as 'strong application' / 'weak application'**. Instances where participants indicated that their dashboard satisfied a given heuristic.

- **Duplicate or similar heuristics**. Instances where participants marked two or more heuristics as either duplicates or very similar.
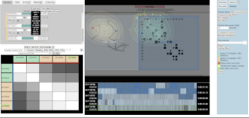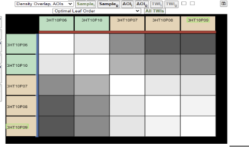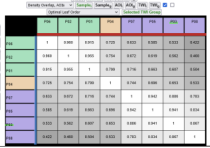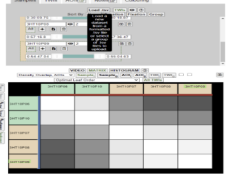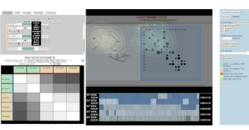
Heuristic	Rate the dashboard on a Likert scale ranging from strong violation <> weak violation <> weak application <> strong application, and qualify with additional details.	Before interaction	After interaction
Functional dead end: Does the dashboard freeze, crash, display errors, or otherwise unexpectedly interrupt the user?	**Weak application:** The developed prototype is not bug-free. At times it freezes and interrupt analysis. We currently complement this by suggesting users save a snapshot of their analysis by "save project". When the GUI hangs, reloading the saved project will allow to resume.		
Are charts within the dashboard at the appropriate level of detail to convey the intended message? If not, what can be done to change the dashboard to an appropriate level of detail?	**Strong application:** The charts currently convey reasonable messages to eye tracking analysts who evaluated the toolkit.		
Does the dashboard employ visual symbols and iconography to add meaning to the data being presented or help support the intended message? If no, are there opportunities to do so?	**Strong application:** Little icons made it easy for participants to show/hide the samples and time window of interest (TWIs) and create real-time visualistiaons to inspect their data.		
Is there a clear reading order and is it logical (e.g., top-down, bottom-up)?	**Strong application:** From our interview with customers we found all of them started with matrix (left) followed by either Space or Timelines (right).		

Fig. 2: Example of four heuristics used to evaluate a visual eye tracking dashboard by an expert. The table indicates the heuristic, its violation or applicability to the dashboard, and the states of the dashboard before and after the interaction, respectively.

4.2.3 Expert Responses

All co-authors inspected the 15 expert responses. The expert participants chose dashboards that they had authored for an audience that included either a client, a data visualization class, or sharing on Tableau Public. The themes of dashboards ranged from health monitoring, crime and violence, visual eye tracking analytics, sports, and finance. Figure 2 shows an example dashboard assessed by the heuristics. Here is an overall summary of how the heuristics were labeled:

- **Heuristics indicated as 'does not apply'**. 23 of the 53 heuristics were labeled as "does not apply" to their dashboards by at least one participant. For example, several participants marked the heuristic, *"Does iconography support or potentially replace repetitive text directives? If not, are there opportunities to do so?"* to be not applicable to the dashboards they were assessing.

- **Misinterpreted heuristics**. 28 of the heuristics were marked as difficult to interpret by at least one participant. For example, participants reported having trouble understanding heuristics that were rather vague: *"Does the dashboard support open-ended data exploration? If not, why?"* or contained jargon: *"Does the visualization disclose the provenance of the data?"*

- **Heuristics marked as ' strong violation' / 'weak violation'**. On average, 11 out of 53 heuristics were marked as being either strongly violated or weakly violated (min: 0, max: 27). For example, for the heuristic, *"Are vague concepts clarified if they exist within the data? (e.g., tall or high-performing) If no, which vague concepts should be clarified?"* was commonly marked as a 'strong violation'. E8 commented, *"This dashboard is meant for the public but uses many difficult terms like 'Case trajectory' instead of "number of people with covid" and 'wastewater concentration.' We should put the text through a plain language grader and improve the language for ease of understanding."*

- **Heuristics marked as 'strong application' / 'weak application'**. On average, 39 out of 53 heuristics were marked as being either strongly applicable or weakly applicable (min: 21, max: 49). We hypothesize that given that the dashboards are authored by

experts, a high number of heuristics were labeled as applicable to the dashboards. For instance, most participants (13 out of 15) stated that *"Is the dashboard interactive to support the user in completing a new analytical task or starting a new line of inquiry? Are there interactions that could be added to enhance the experience?"* strongly applied to their dashboards. E11 marked, *"If there is interaction, does the dashboard update as expected?"* as a 'weak application' and commented *"Filtering and hovering interactions update the story as expected. But the lack of instructions makes the user perceive filters as labels."*

- **Duplicate or similar heuristics**. Six sets of heuristics were marked as either being duplicates of one or more other heuristics or very similar. For example, under the theme, "Composition, layout, space, and sequencing", heuristics such as *"The layout, placement of charts, and the flow in the visualization should be easy to follow"*, *"There is a clear reading order within the dashboard and is it logical (e.g., top-down, bottom-up)"*, and *"The charts, text, and any other visuals are laid out in a way that is helpful for understanding the structure of the information being presented in the dashboard"* were identified to be similar.

Generally, participants found the dashboard reflection exercise to be helpful. E14 said, " *The heuristics and themes are very helpful in understanding many of the considerations that need to be made while designing a dashboard such as (1) Multi-modal conversational support, (2) Integrating text with visual information for communication, and (3) Visual scaffolding for helping with conversation clarity.*" Participants also found that the reflection inspired them to consider dashboard design in new ways. E05 said, " *Multimodal interactivity and NLI provide a new way of thinking. It would be exciting to integrate this in an eye-tracking analysis tool for improving the sense-making loop.*"

After reviewing the experts' reflections, we clarified heuristics that were unclear and ambiguous as well as consolidated redundant ones, resulting in a total of 46 heuristics. For example, we removed redundant heuristics such as *"The quantitative units are clearly defined or specified."* as we already included the heuristic, *"Concepts or metrics are either easily understandable or clearly defined in the dashboard."*

and added heuristics suggested by the participants, such as *"Is there adequate evidence that the dashboard is truthful? Is the dashboard able to convince the key takeaway through credibility and trustworthiness?"*

4.3 Phase 3: Author Reflections and Final Iterations

We further reflected on the set of heuristics, given that the goal was to evaluate them with a student population to assess how the students would apply and critique the conversational nature of dashboards. We reformulated the remaining heuristics to follow a clear and consistent format and to clarify issues identified by the expert evaluators. Specifically, we further iterated on the heuristics based on the criteria:

- Reworded heuristics posed as questions to be imperative guidelines of what the dashboard ought to support. For example, heuristic *"Is the text in the dashboard legible, easy to read, and useful? Are the different parts of the chart (e.g., titles, captions, or narration) well-described?"* was rephrased as *"There are text and visual elements to frame or guide salient information."*

- Ensured that the heuristics were understandable without technical jargon where heuristics such as *"Do starting points for interactivity align with user experience and expectations?"* were reworded as *"The dashboard is interactive and supports the user in completing a new analytical task or starting a new line of inquiry."*

- Made sure that each heuristic could be clearly validated for whether it was applied or violated in an interactive dashboard. To that end, any conjunctions, if present, were removed to prevent the inclusion of multiple guidelines within a single heuristic.

Finally, after winnowing down the heuristics to 39, we found that they could be reorganized thematically into the 5 basic conversational states: initiation, grounding, turn-taking, repair & refinement, and close. While the final set of heuristics provides an initial framework for assessing cooperative conversation in interactive dashboards, we do not guarantee completeness; rather, we sought to assess their utility and identify opportunities to further improve and refine them. The next section describes how the heuristics were applied by students in a visualization education setting. The final table of conversational dashboard heuristics is shown in Table 1, and its various iterations leading to the final set are included in the supplementary material.

5 USE OF HEURISTICS IN EDUCATION PRACTICE

To evaluate the utility of the heuristics, we provided two opt-in homework exercises with visualization learners in a post-graduate data visualization class at a university. Part A was a heuristics reflection exercise on pre-authored interactive dashboards, while Part B was an exercise to apply the heuristics to improve the conversational nature of an existing dashboard. Both exercises were not graded to mitigate any biases when students provided feedback. The university review board granted formal approval to conduct the exercises. We include class exercise material and evaluations as supplementary material.

5.1 Part A: Heuristics Reflection Exercise

The goals of the heuristics reflection exercise were to 1) assess the heuristics' value in supporting visualization learners and 2) gain feedback on the heuristics for iterative improvement. Since our development phases involved visualization experts, we focused on learners to ensure the heuristics were understandable by a less experienced population.

The 52 participants were master's students (with backgrounds in Computer Science or Engineering). We use the notation [P#] when referring to participants in this heuristics evaluation. We refer to particular heuristics from our final list as [H#].

The homework exercise was conducted similarly to the reflection exercise described in section 4.2.1 but with the updated heuristics table (Table 1, organized by the five conversational states). The exercise was introduced during class by the class instructor and then completed as a homework assignment over a week. To ensure that students remained engaged when applying the heuristics to evaluate dashboards, we provided a list of 18 dashboards and asked students to describe which dashboard they chose and why. Four dashboards were not picked from the list, with the highest number of students (six) choosing a renewable

energy consumption dashboard. The complete list of dashboards and the frequency of choices is included in the supplementary material. The actual reflection exercise was the same as in section 4.2.1; it involved assigning the dashboard a rating ('strong application,' 'weak application,' 'weak violation,' or 'strong violation') for each heuristic with written commentary and screenshots to justify the ratings and then answering the reflection questions. The students additionally gave an in-class presentation of their findings from the homework exercise.

We conducted a thematic analysis of the heuristic reflection responses and survey answers. We looked for feedback on the heuristics, interesting examples of how the heuristics were applied to the dashboards, and insights that were revealed. We also examined frequency data on how dashboards were ranked across the different heuristics.

5.1.1 Rating Frequencies

Conversational State	Application	Violation
Initiation	74	23
Grounding	76	26
Turn-Taking	41	68
Repair & Refinement	38	63
Close	40	95

Table 2: Part A heatmap showing the frequency of dashboards applying or violating the heuristics, grouped by conversational state. Frequencies are averaged across participants and normalized by the number of heuristics per conversational state on a 0 to 100 scale.

Relative frequencies of heuristic applications and violations, as rated by participants for their chosen dashboard, are summarized in Table 2. Because the conversational state categories contain different numbers of heuristics, we used a normalized metric rather than raw counts. To compute these scores, we first combined strong and weak application ratings, and similarly combined strong and weak violation ratings. We averaged the number of ratings across participants and normalized the result by the number of heuristics in each state on a $0 - 100$ scale. Note that these are not exactly percentages because a participant could identify multiple applications and/or violations of a single heuristic.

Table 2 shows that the rate of violations increased for later conversation phases, and the rate of applications decreased. This observation was consistent across participants. It suggests that today's dashboards offer reasonable support for initiation and grounding, but are progressively less supportive as human-data conversations get into turn-taking, repair, and close activities. For instance, turn-taking repeatedly showed up as a challenge, where dashboard inflexibility or awkward interactions made it difficult for users to complete analytical workflows.

5.1.2 Use of Heuristics Across Conversational States

Next, we examine themes and interesting examples of how the heuristics were used across the conversational states.

Initiation. Heuristics in the initiation state ($H1 - H14$) were often marked as either strong or weak application (normalized application frequency of 74 in Table 2). Participants noted that dashboards initiated the conversation by including instructions on how to use the dashboard, making it easier to explore the data. The reading order, encodings, and formatting conventions used were often easy to understand and follow. P13 stated *"the color combination used by the chart maker keeps the reader attentive and focuses the attention at the right regions."* However, dashboards did have violations in revealing the provenance of their data. P8 stated, *"strong violation as the dataset source hyperlink they tried to give doesn't work and the data preparation is not mentioned."*

Grounding. Similar to the initiation state, heuristics in this conversation state ($H15 - H24$) were often marked as either strong or weak application (normalized application score of 76). Many of the dashboards (38 out of 52) were described as having a clear presentation of context and level of detail. P38 commented, *"Yes ordering is logical, It's sorted in highest to lowest expense. First row shows line chart and next row shows details of breakdown."*
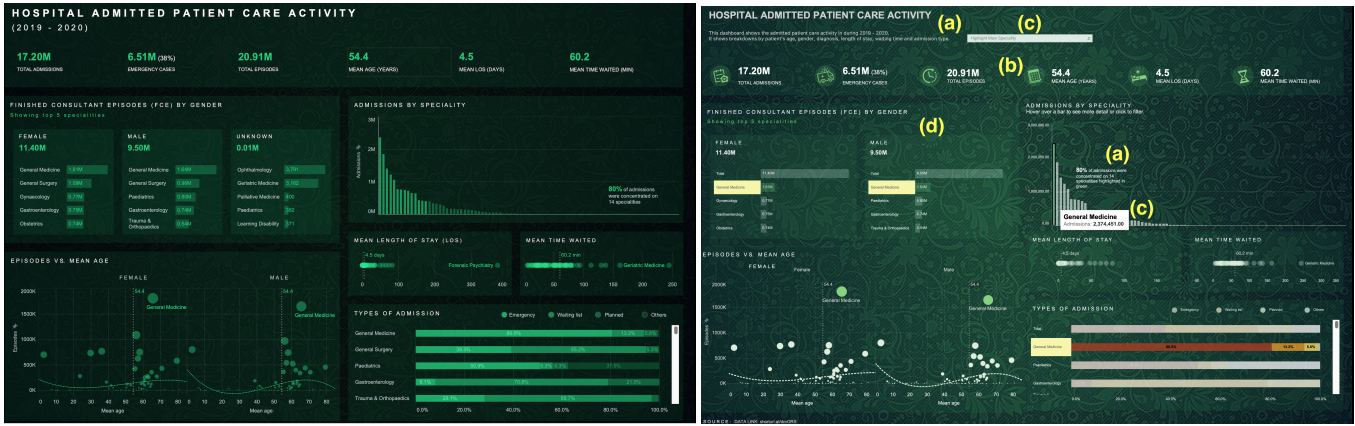
Fig. 3: Dashboard used in Part B classroom exercise. Left: Original dashboard. Right: Modified dashboard. Updated dashboard annotated in yellow based on heuristics from the conversational states, indicated by *H*#: (a) Initiation: The text in the dashboard is updated to provide more context (*H*6). (b) Grounding: Dashboard contains iconography to add meaning to the data being presented (*H*17). (c) Turn-taking: The dashboard is updated to add filtering across views (*H*28). (d) Close: The "Unknown" category was removed from the original dashboard to further clarify the takeaway of the dashboard (*H*37). Note that students made aesthetic changes (not always an improvement), such as modifying the dashboard's background in addition to applying the heuristics. (Permission granted to use original and modified dashboards).

Turn-taking. For this conversational state, there was a lower frequency of application ratings (41) and a higher frequency of violations (68), indicating the limited interactivity (*H*25) that the dashboards provided. *P*20 stated, *"The dashboard should update its view based on what is selected, highlighted, or filtered by the user. As there are no filters and update options available in the dashboard."* Participants also noticed some friction when interacting with the dashboard (*H*26). *P*3: *"The process of zooming in is clunky and disrupts continuity."* and guiding to the next step. *P*3: *"Very little visual warning/cueing to accompany changes to graphs, particularly in the side panel. Some changes are initially off screen and have to be scrolled to."*

Repair & Refinement. Participants (46 out of the 52 students) often found that the dashboards violated the functional and navigational deadends (*H*32 and *H*33) (63 violations per 100 cases). *P*28 commented, *"The dashboard doesn't provide interactivity at all. It has no filters or searches. Just a basic static visual. Just looking at the graph doesn't make any sense unless we hover over it."* Further multi-modal support (*H*34) was violated in many cases as the interactions were limited to selecting filters in the drop-down, for example. *P*41 said, *"There are no filters. Filters could have helped a lot when analyzing certain time periods but are given as only two values between year ranges."*

Close. This category had the highest frequency of violations (95 violations per 100 cases). For several dashboards, it was not apparent what the key takeaway was to close the conversation (*H*36 − *H*38). *P*36 commented, *"Weak Violation. Just by looking at this dashboard, one cannot conclude something; the user has to gather data from each hexagon, then analyse it and only then something can be concluded."* Other violations concerned around trust (*H*39). *P*5 said, *"Strong violation: though there's no reason the believe the dashboard is lying, without key context a user with no additional information could easily come away with a confused message, or even the wrong idea entirely."* Similarly *P*45 said, *"Weak Violation. The source of the data is nowhere mentioned, which would have increased the credibility of the dashboard."*

In summary, we found that while dashboards tend to be effective at initiation and grounding of the conversation, they struggle with other aspects of conversation that include turn-taking, repair & refinement.

5.2 Part B: Update or Create Dashboards Using Heuristics

The 52 students self-organized into groups of three or four, forming a total of 15 groups where they applied the heuristics to update an existing dashboard from Tableau Public (4 out of 15 groups) or create new dashboards from a Kaggle dataset [40] (11 out of 15 groups). Students completed the exercise over a week and rated the dashboard with the same set of heuristics (Table 1), providing commentary and screenshots. For exercises involving updating an existing dashboard, students rated the dashboard before and after the update.

We conducted a thematic analysis of the heuristic reflection responses. We looked for feedback on the heuristics, interesting examples of how the heuristics were applied to the dashboards, and insights that were revealed. We also examined frequency data on how dashboards were ranked across the different heuristics.

5.2.1 Application of Heuristics Across Conversational States

Conversational State	Application		Violation		Conversational State	Application	Violation
	Before	After	Before	After			
Initiation	77	100	23	0	Initiation	93	7
Grounding	78	98	23	3	Grounding	91	9
Turn-Taking	57	96	43	4	Turn-Taking	88	12
Repair & Refinement	44	94	56	6	Repair & Refinement	89	11
Close	56	94	44	13	Close	95	5
(a) Updated dashboards − before and after					(b) Newly created dashboards		

Table 3: Part B heatmaps showing the frequency of applying or violating the conversational heuristics, grouped by conversational state. Frequencies are averaged across participants and normalized by the number of heuristics per state on a 0 to 100 scale.

Similar to Part A (Section 5.1.2), we computed frequencies of heuristic applications and violations, as reported by students (Table 3). In cases where students modified an existing dashboard, students reported an overall increase in the rate of applications and a decrease in the rate of violations of heuristics across all conversation states. In particular, we saw a higher rate of decrease in violations for 'turn taking,' 'repair & refinement,' and 'close'; states that had fewer application rates in general during Part A's exercise. However, note that students performed these self-evaluations, which could contribute to a higher rate of applications. While the class instructor reviewed the self-reflection ratings, future work should consider an external reviewer to validate these ratings. Figure 3 shows an example of an original dashboard on hospital admittances (left) with a corresponding modified version (right). Updates include supporting better turn-taking by adding interactivity and multi-view coordination, along with a search bar to navigate to a specific medical department specialty. Iconography was added to better convey the meaning of the information being presented to the user, along with additional descriptive text to ground the conversation.

We found that having done Part A, students were familiar with the heuristics and focused specifically on addressing heuristics for turn-taking, repair & refinement, and close. *P*18 stated, *"I was more observant of how the dashboard behaved when I interacted with it. I*

focused on making sure there were no dead-ends when I clicked on the widgets and all the views updated appropriately." In both the updated and newly created dashboards, we observed a greater prevalence of text to help ground contextual information alongside the visualizations (applying heuristics, Initiation - H6, Grounding - H16, H21, H23, Turn-taking - H29, and Close - H36). *P*4 remarked, *"For each category, the heuristics reminded me that text plays a vital role with the charts for communicating the key ideas."* Some students reported that they sacrificed visual style for clearer communication with the user. *P*37 stated, *"Although visual style get [sic] little disturbed in the color part, it looks necessary to make dashboard more easy to understand."* Future work should further explore how these heuristics, alongside visual design guidelines can support the dashboard authoring process.

5.3 Feedback on the Heuristics

Now, we summarize the various qualitative themes of feedback on the heuristics across both exercises.

Heuristics were helpful and understandable. Participants found the heuristics to be useful for understanding the structure and flow of dashboards as part of an analytical conversation, as well as for authoring new ones. *P*8 commented, *"The dashboard communicates a certain style or mood to the user, and there are clear strategies employed in the dashboard to mark charts or marks more prominently to encourage a user to interact with them, as well as de-emphasize items not relevant to the conversation.".* *P*4 stated, *"all heuristics were explained clearly, and I did not encounter any confusion while completing the form."*

Unique and unexpected heuristics. Others found some heuristics to be rather unique and unexpected when considering dashboard design. For example, participants found heuristics, *H*11, *H*17 − 19 on visual symbols and iconography to be helpful - *"The use of semiotics for symbolic communication as well as the exposition sections stood out to me in particular [P4].* *P*15 found *H*26 to be useful when thinking of evaluating friction in dashboards: *"I was not able to find or look at all the cities at once, and it was difficult to click on the small bubbles."* *P*13 was intrigued by the heuristic on navigational dead-ends (*H*33) and said, *"This made me think to make visualization work in every case, whenever the user selects or searches anything on the dashboard, to navigate easily."* *P*23 found *H*5 concerning logical reading order to be insightful - *"Before this assignment, I never thought that the placement of charts should have a logical sequence. It makes perfect sense, and I will apply this in my future dashboards."*

Confusing and missing heuristics. Participants found the heuristics on bias (*H*13) and mood (*H*24) to be vague and not very actionable. *P*11 stated, *"This [the bias] heuristic was confusing as I did not understand the biases in the dashboard very well. Mood communication in terms of the dashboard was somewhat confusing to me."* There were suggestions for considering adding animation as part of the analytical conversation (*P*24, *P*27, *P*36) and further helping users recognize, diagnose, and recover from errors when interacting with the dashboards (*P*47).

6 DISCUSSION AND FUTURE WORK

While existing dashboard guidelines capture visual design issues such as legibility and complexity, our development and evaluation of heuristics from the lens of analytical conversations suggest that there are ways that dashboard design can succeed (or fail), which are not captured by existing recommendations or pedagogy, and so are often overlooked.

Dashboards struggle with turn-taking, repair & refinement. Participants pointed out many violations of heuristics in the turn-taking and repair & refinement phases, suggesting that today's dashboards may be weak in these aspects. We strongly encourage future work that makes dashboards more flexible, cooperative conversational partners. Future dashboards could enable users to more easily pivot between analytical goals (e.g., via flexible construction so the end user can change dashboard metrics, field ordering, chart type, etc.) and could employ predictive analytics to anticipate a user's upcoming information needs.

Interpreting heuristics for guidance and mitigation strategies. By their nature, heuristics offer guidance rather than prescriptive solutions. They should be considered in the context of the designer's expert knowledge of the domain, design goals, and audience. Heuristics may, at times, contradict each other or suggest design directions that are counter to specific communication goals or domain conventions. For example, in Sarikaya et al.'s [66] framework of dashboard types, dashboards for learning may need greater emphasis on grounding (e.g., contextual information) than dashboards for ongoing awareness of well-understood metrics. We envision that designers will use the heuristics to inspire ideas and identify potential gaps and flaws while thoughtfully discarding less suitable suggestions. Utilizing the heuristics to provide in-situ mitigation strategies in dashboard authoring tools is an area of future research. For example, tools could flag warnings if the interactions have errors or there are no graceful fallbacks for preventing functional dead-ends. Other guidelines can support authors with progressive disclosure of content through interaction and templates for adding text to prevent cognitive overload during conversational initiation.

Developing heuristics for conversations around data. We encourage the revision and extension of the heuristics themselves, as academics and practitioners use and adopt them. For example, we introduced heuristics to guide dashboard design and evaluation, with the lens of dashboards as a medium to enable conversations *with* data. Dashboards also support the important role of human-human communication *around* data [86], including discussing and circulating the data within an organization. A future extension to the heuristics could focus on dashboard characteristics to support circulation or persuasion.

Assessing utility of heuristics during dashboard design. Our evaluation focuses on applying heuristics to critique or improve existing dashboards. An acid test of heuristics' utility is their ability to productively shape the design process: we would ideally see how mindfulness of our heuristics impacts the final design of dashboards or the iterative process of choosing design alternatives. While we do note examples of participants saying that they would, as per *P*25, " *apply this [heuristic] in my future dashboards*," we leave this longitudinal assessment to future work. Additional future work is the connection of our heuristics to other forms of evaluation. For instance, does a (re-) designed cooperative dashboard result in benefits to user performance or satisfaction?

Extending the cooperative principles to other analytical workflows. We believe that *cooperative dashboards* represent a new perspective on visual analytics and potentially an emerging genre of visualization design. While it has been long understood that analytics is a multi-stage process (e.g., the Pirolli/Card sensemaking loop [61]), there has been less work on visual analytics tools that operate *across* stages. We consider dashboards to be useful testbeds for learning about the structure of analytical conversations, and for testing novel designs to support users. Cooperative dashboards allow a wide range of potential design or technique work for topics like mixed-initiative systems, NLIs, and rhetoric. Beyond dashboards, we also wish to apply these cooperative principles to other related forms (such as data stories) and media (such as designing visualizations for mobile or wearable devices).

7 CONCLUSION

In this paper, we explore the design of interactive dashboards as artifacts that support analytical conversations with their users. In particular, we explore how the role of language pragmatics and cooperative conversation can support data exploration, interaction, and reasoning. Inspired by existing models of conversational implicature and its states, we proposed and evaluated 39 heuristics for helping guide the design of analytical conversation with interactive dashboards. These heuristics were iteratively validated with 16 visualization practitioners and subsequently evaluated by 52 students to assess how useful they are for effectively authoring dashboards. Through the evaluation of these heuristics, we found that while dashboards tend to be effective at initiation and grounding of the conversation with the user, they struggle with other aspects of conversation that include turn-taking, repair & refinement, and close. We hope that this work inspires the broader research and practitioner communities to explore new design and interaction paradigms for authoring more cooperative dashboard conversations.

REFERENCES

[1] Washington State COVID-19 Summary Dashboard. https://doh.wa.gov/emergencies/covid-19/data-dashboard, 2022. 5

[2] E. Adamopoulou and L. Moussiades. Chatbots: History, Technology, and Applications. *Machine Learning with Applications*, 2:100006, 2020. 2

[3] S. Agarwal, B. Agarwal, and R. Gupta. Chatbots and Virtual Assistants: a Bibliometric Analysis. *Library Hi Tech*, 40(4):1013–1030, 2023/03/23 2022. 2

[4] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, et al. Guidelines for Human-AI Interaction. In *CHI*, pp. 1–13, 2019. 2

[5] B. Bach, E. Freeman, A. Abdul-Rahman, C. Turkay, S. Khan, Y. Fan, and M. Chen. Dashboard Design Patterns. *IEEE TVCG*, 2022. 2

[6] L. Bartram, M. Correll, and M. Tory. Untidy Data: The Unreasonable Effectiveness of Tables. *arXiv preprint arXiv:2106.15005*, 2021. 4

[7] S. Beebe. *Interpersonal Communication: Relating to Others*. Pearson Education Canada, 2004. 2

[8] E. Bertini, S. Gabrielli, S. Kimani, T. Catarci, and G. Santucci. Appropriating and Assessing Heuristics for Mobile Computing. In *Advanced Visual Interfaces*, pp. 119–126, 2006. 2

[9] M. Brehmer and R. Kosara. From Jam Session to Recital: Synchronous Communication and Collaboration Around Data in Organizations. *IEEE TVCG*, 28(1):1139–1149, 2022. 4

[10] J. Cassell. Embodied Conversational Agents: Representation and Intelligence in User Interfaces. *AI Mag.*, 22:67–84, 2001. 4

[11] A. P. Chaves and M. A. Gerosa. How Should My Chatbot Interact? A Survey on Social Characteristics in Human–chatbot Interaction Design. *International Journal of Human–Computer Interaction*, 37(8):729–758, 2021. 2

[12] X. Chen, W. Zeng, Y. Lin, H. M. Ai-Maneea, J. Roberts, and R. Chang. Composition and Configuration Patterns in Multiple-view Visualizations. *IEEE TVCG*, 27(2):1514–1524, 2020. 2

[13] H. H. Clark and S. Brennan. Grounding in Communication. In *Perspectives on Socially Shared Cognition*, 1991. 3

[14] L. Coppock. Politeness Strategies in Conversation Closings, 01 2005. 4

[15] A. Cuttone, M. K. Petersen, and J. E. Larsen. Four Data Visualization Heuristics to Facilitate Reflection in Personal Informatics. In *International Conference on Universal Access in Human-Computer Interaction*, pp. 541–552. Springer, 2014. 2

[16] W. Davis. Implicature. In E. N. Zalta, ed., *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2019. 5

[17] D. Deng, W. Cui, X. Meng, M. Xu, Y. Liao, H. Zhang, and Y. Wu. Revisiting the Design Patterns of Composite Visualizations. *arXiv preprint arXiv:2203.10476*, 2022. 2

[18] V. Dhanoa, C. Walchshofer, A. Hinterreiter, H. Stitz, E. Groeller, and M. Streit. A Process Model for Dashboard Onboarding. *Computer Graphics Forum*, 41(3):501–513, 2022. 3

[19] E. Dimara and C. Perin. What is Interaction for Data Visualization? *IEEE TVCG*, 26(1):119–129, 2019. 1

[20] E. Dimara, H. Zhang, M. Tory, and S. Franconeri. The Unmet Data Visualization Needs of Decision Makers within Organizations. *IEEE TVCG*, 2021. 2

[21] D. Dowding and J. A. Merrill. The Development of Heuristics for Evaluation of Dashboard Visualizations. *Applied Clinical Informatics*, 9(03):511–518, 2018. 2

[22] P. Dragicevic, Y. Jansen, A. Sarma, M. Kay, and F. Chevalier. Increasing the Transparency of Research Papers with Explorable Multiverse Analyses. In *CHI*, CHI '19, p. 1–15. ACM, 2019. 3

[23] L. Q. En and S. S. Lan. The Applicability of Gricean maxims in Social Robotics Polite Dialogue. *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 195–196, 2011. 4

[24] T. C. Endsley, K. A. Sprehn, R. M. Brill, K. J. Ryan, E. C. Vincent, and J. M. Martin. Augmented Reality Design Heuristics: Designing for Dynamic Interactions. *Proceedings of the human factors and ergonomics society annual meeting*, 61(1):2100–2104, 2017. 2

[25] S. Epley. Best States to Retire in the United States. https://public.tableau.com/app/profile/sam.epley/viz/BestStatestoRetireintheUnitedStates/BestStatestoRetireintheUnitedStates, 2022. 5

[26] M. Feng, C. Deng, E. M. Peck, and L. Harrison. HindSight: Encouraging Exploration through Direct Encoding of Personal Interaction History. *IEEE TVCG*, 23(1):351–360, 2017. 4

[27] S. Few. *Information Dashboard Design: The Effective Visual Communication of Data*. O'Reilly Media, Inc., 2006. 1, 2

[28] B. Fiore-Gartland and G. Neff. Communication, Mediation, and the Expectations of Data: Data Valences Across Health and Wellness Communities. *International Journal of Communication*, 9:19, 2015. 2

[29] C. Forsell and J. Johansson. A Heuristic Set for Evaluation in Information Visualization. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pp. 199–206, 2010. 2

[30] L. Fulfagar, A. Gupta, A. Mathur, and A. Shrivastava. Development and Evaluation of Usability Heuristics for Voice User Interfaces. In *International Conference on Research into Design*, pp. 375–385. Springer, 2021. 2

[31] C. Furmanski, R. Azuma, and M. Daily. Augmented-Reality Visualizations Guided by Cognition: Perceptual Heuristics for Combining Visible and Obscured Information. In *Proceedings. International Symposium on Mixed and Augmented Reality*, pp. 215–320. IEEE, 2002. 2

[32] T. Gao, M. Dontcheva, E. Adar, Z. Liu, and K. G. Karahalios. Data-Tone: Managing Ambiguity in Natural Language Interfaces for Data Visualization. In *UIST*, UIST 2015, pp. 489–500. ACM, 2015. 4

[33] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. D. III, and K. Crawford. Datasheets for Datasets. *Commun. ACM*, 64(12):86–92, 2021. 3

[34] H. P. Grice. Logic and conversation. In M. Ezcurdia and R. J. Stainton, eds., *The Semantics-Pragmatics Boundary in Philosophy*, p. 47. Broadview Press, 1975. 2, 3, 5

[35] J. Heer. Agency Plus Automation: Designing Artificial Intelligence into Interactive Systems. *Proceedings of the National Academy of Sciences*, 116(6):1844–1850, 2019. 3

[36] S. Höhn and K. Bongard-Blanchy. Heuristic Evaluation of COVID-19 Chatbots. In *International Workshop on Chatbot Research and Design*, pp. 131–144. Springer, 2020. 2

[37] E. Hoque, V. Setlur, M. Tory, and I. Dykeman. Applying Pragmatics Principles for Interaction with Visual Analytics. *IEEE TVCG*, 24(1):309–318, 2017. 4

[38] J. Hullman and A. Gelman. Designing for Interactive Exploratory Data Analysis Requires Theories of Graphical Inference. *Harvard Data Science Review*, 3(3), 2021. 3

[39] IBM Watson Analytics. http://www.ibm.com/analytics/watson-analytics, 2021. 4

[40] Kaggle. https://www.kaggle.com/datasets, 2023. 8

[41] A. Key, B. Howe, D. Perry, and C. Aragon. Vizdeck: Self-organizing Dashboards for Visual Analytics. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, pp. 681–684, 2012. 2

[42] Y.-S. Kim, K. Reinecke, and J. Hullman. Explaining the Gap: Visualizing One's Predictions Improves Recall and Comprehension of Data. In *CHI*, CHI '17, p. 1375–1386. ACM, 2017. 3

[43] Y.-S. Kim, K. Reinecke, and J. Hullman. Data Through Others' Eyes: The Impact of Visualizing Others' Expectations on Visualization Interpretation. *IEEE TVCG*, 24(1):760–769, 2018. 3, 4

[44] C. Knaflic. *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley, 2015. 4

[45] Y. S. Kristiansen, L. Garrison, and S. Bruckner. Semantic Snapping for Guided Multi-view Visualization Design. *IEEE TVCG*, 28(1):43–53, 2021. 2

[46] H. Lam, M. Tory, and T. Munzner. Bridging from Goals to Tasks with Design Study Analysis Reports. *IEEE TVCG*, 24(1):435–445, 2017. 2

[47] R. Langevin, R. J. Lordon, T. Avrahami, B. R. Cowan, T. Hirsch, and G. Hsieh. Heuristic Evaluation of Conversational Agents. In *CHI*, pp. 1–15, 2021. 2

[48] D. J.-L. Lee, V. Setlur, M. Tory, K. Karahalios, and A. Parameswaran. Deconstructing Categorization in Visualization Recommendation: A Taxonomy and Comparative Study. *IEEE TVCG*, 28(12):4225–4239, 2022. 3

[49] E. Lee-Robbins and E. Adar. Affective Learning Objectives for Communicative Visualizations. *IEEE TVCG*, 2022. 2

[50] H. Lin, D. Akbaba, M. Meyer, and A. Lex. Data Hunches: Incorporating Personal Knowledge into Visualizations. *IEEE TVCG*, 29(1):504–514, 2023. doi: 10.1109/TVCG.2022.3209451 3

[51] Y. Lin, H. Li, A. Wu, Y. Wang, and H. Qu. Dashboard Design Mining and Recommendation. *IEEE TVCG*, pp. 1–15, 2023. 2

[52] B. Luo, R. Y. Lau, C. Li, and Y.-W. Si. A Critical Review of State-of-the-art Chatbot Designs and Applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(1):e1434, 2022. 2

[53] M. Maguire. Development of a Heuristic Evaluation Tool for Voice User Interfaces. In *International Conference on Human-Computer Interaction*, pp. 212–225. Springer, 2019. 2

[54] Merriam-Webster. Heuristic, 2022. 5

[55] M. Muller, I. Lange, D. Wang, D. Piorkowski, J. Tsay, Q. V. Liao, C. Dugan, and T. Erickson. How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. In *CHI*, pp. 1–15, 2019. 2

[56] J. Nielsen. Enhancing the Explanatory Power of Usability Heuristics. In *CHI*, pp. 152–158, 1994. 2

[57] J. Nielsen. Ten Usability Heuristics. http://www.nngroup.com/articles/ten-usability-heuristics/, 2005. 2

[58] C. Nowacki, A. Gordeeva, and A.-H. Lizé. Improving the Usability of Voice User Interfaces: A New Set of Ergonomic Criteria. In *International Conference on HCI*, pp. 117–133. Springer, 2020. 2

[59] A. Pandey, A. Srinivasan, and V. Setlur. MEDLEY: Intent-Based Recommendations to Support Dashboard Composition. *arXiv preprint arXiv:2208.03175*, 2022. 2

[60] L. Panfili, S. Duman, A. Nave, K. Ridgeway, N. Eversole, and R. Sarikaya. Human-AI Interactions through a Gricean Lens. *Proceedings of the Linguistic Society of America*, 6:288, 03 2021. 4

[61] P. Pirolli and S. Card. The Sensemaking Process and Leverage Points for Analyst Technology as Identified through Cognitive Task Analysis. *Proceedings of International Conference on Intelligence Analysis*, 25:2–4, 01 2005. 9

[62] Microsoft Q&A. https://powerbi.microsoft.com/en-us/documentation/powerbi-service-q-and-a, 2021. 4

[63] Z. Qu and J. Hullman. Keeping Multiple Views Consistent: Constraints, Validations, and Exceptions in Visualization Authoring. *IEEE TVCG*, 24(1):468–477, 2017. 2

[64] J. C. Roberts. State of the Art: Coordinated & Multiple Views in Exploratory Visualization. In *Conference on Coordinated and Multiple Views in Exploratory Visualization*, pp. 61–71. IEEE, 2007. 2

[65] H. Sacks, E. A. Schegloff, and G. Jefferson. A Simplest Systematics for the Organization of Turn-taking for Conversation. *Language*, 50(4):696–735, 1974. 2, 3

[66] A. Sarikaya, M. Correll, L. Bartram, M. Tory, and D. Fisher. What Do We Talk About When We Talk About Dashboards? *IEEE TVCG*, 25(1):682–692, 2018. 1, 2, 3, 9

[67] A. P. Saygin and I. Cicekli. Pragmatics in Human-Computer Conversations. *Journal of Pragmatics*, 34:227–258, 03 2002. 1

[68] E. Schegloff, G. Jefferson, and H. Sacks. The Preference for Self-correction in the Organization of Repair in Conversation. *Language*, 53:361–382, 06 1977. 3

[69] V. Setlur, S. E. Battersby, M. Tory, R. Gossweiler, and A. X. Chang. Eviza: A Natural Language Interface for Visual Analysis. In *UIST*, pp. 365–377, 2016. 4

[70] V. Setlur and B. Cogley. *Functional Aesthetics for Data Visualization*. Wiley, 2022. 4

[71] V. Setlur, E. Hoque, D. H. Kim, and A. X. Chang. Sneak Pique: Exploring Autocompletion as a Data Discovery Scaffold for Supporting Visual Analysis. In *UIST*, UIST '20, p. 966–978. ACM, 2020. 3

[72] V. Setlur and A. Kumar. Sentifiers: Interpreting Vague Intent Modifiers in Visual Analysis using Word Co-occurrence and Sentiment Analysis. *2020 IEEE Visualization Conference (VIS)*, pp. 216–220, 2020. 3

[73] V. Setlur and M. Tory. How Do You Converse with an Analytical Chatbot? Revisiting Gricean Maxims for Designing Analytical Conversational Behavior. In *CHI*. ACM, 2022. 2, 3, 4

[74] V. Setlur, M. Tory, and A. Djalali. Inferencing Underspecified Natural Language Utterances in Visual Analysis. IUI '19, pp. 40–51. ACM, 2019. 3

[75] L. Shen, E. Shen, Y. Luo, X. Yang, X. Hu, X. Zhang, Z. Tai, and J. Wang. Towards Natural Language Interfaces for Data Visualization: A Survey. *IEEE TVCG*, PP, 2022. 2

[76] Y. Shi, T. Gao, X. Jiao, and N. Cao. Breaking the Fourth Wall of Data Stories through Interaction. *IEEE TVCG*, pp. 1–11, 2022. 3

[77] B. Shneiderman, C. Plaisant, M. S. Cohen, S. Jacobs, N. Elmqvist, and N. Diakopoulos. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson, 2016. 2

[78] P. Singh. World Indicators Dashboard. https://public.tableau.com/app/profile/prabhasis.singh/viz/WorldIndicators-TableauGeneralExample_16258338789240/Story2Techassignment1, 2022. 5

[79] A. Srinivasan, S. M. Drucker, A. Endert, and J. Stasko. Augmenting Visualizations with Interactive Data Facts to Facilitate Interpretation and Communication. *IEEE TVCG*, 25(1):672–681, 2018. 3

[80] A. Srinivasan and V. Setlur. BOLT: A Natural Language Interface for Dashboard Authoring. In T. Hoellt, W. Aigner, and B. Wang, eds., *EuroVis 2023 - Short Papers*. The Eurographics Association, 2023. doi: 10.2312/evs.20231035 2

[81] A. Srinivasan and J. Stasko. Orko: Facilitating Multimodal Interaction for Visual Exploration and Analysis of Networks. *IEEE TVCG*, 24(1):511–521, 2017. 4

[82] K. Sugisaki and A. Bleiker. Usability Guidelines and Evaluation Criteria for Conversational User Interfaces: A Heuristic and Linguistic Approach. In *Proceedings of the Conference on Mensch und Computer*, pp. 309–319, 2020. 2

[83] A. Tarrell, A. Fruhling, R. Borgo, C. Forsell, G. Grinstein, and J. Scholtz. Toward Visualization-Specific Heuristic Evaluation. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, pp. 110–117, 2014. 2

[84] ThoughtSpot. http://www.thoughtspot.com, 2021. 4

[85] C. Tominski. Interaction for Visualization. *Synthesis Lectures on Visualization*, 3(1):1–107, 2015. 1

[86] M. Tory, L. Bartram, B. Fiore-Gartland, and A. Crisan. Finding Their Data Voice: Practices and Challenges of Dashboard Users. *IEEE Computer Graphics and Applications*, 2021. 2, 3, 4, 9

[87] M. Tory and T. Moller. Evaluating Visualizations: Do Expert Reviews Work? *IEEE CG & A*, 25(5):8–11, 2005. 1, 2

[88] M. Tory and V. Setlur. Do What I Mean, Not What I Say! Design Considerations for Supporting Intent and Context in Analytical Conversation. In *IEEE TVCG*, 2019. 1, 3

[89] H. Väätäjä, J. Varsaluoma, T. Heimonen, K. Tiitinen, J. Hakulinen, M. Turunen, H. Nieminen, and P. Ihantola. Information Visualization Heuristics in Practical Expert Evaluation. In *BELIV Workshop*, pp. 36–43, 2016. 2

[90] E. Wall, L. M. Blaha, L. Franklin, and A. Endert. Warning, Bias May Occur: A Proposed Approach to Detecting Cognitive Bias in Interactive Visual Analytics. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 104–115, 2017. 4

[91] E. Wall, A. Narechania, A. Coscia, J. Paden, and A. Endert. Left, Right, and Gender: Exploring Interaction Traces to Mitigate Human Biases. *IEEE TVCG*, 28(1):966–975, 2022. 4

[92] E. Wall, J. Stasko, and A. Endert. Toward a Design Space for Mitigating Cognitive Bias in Vis. In *IEEE TVCG*, pp. 111–115, 2019. 4

[93] C. Ware. *Visual Thinking: For Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2008. 2

[94] Z. Wei and J. A. Landay. Evaluating Speech-Based Smart Devices using New Usability Heuristics. *IEEE Pervasive Computing*, 17(2):84–96, 2018. 2

[95] S. Wexler, J. Shaffer, and A. Cotgreave. *The Big Book of Dashboards: Visualizing your Data using Real-world Business Scenarios*. John Wiley & Sons, 2017. 1, 2

[96] B. Whitworth. Polite Computing. *Behaviour & Information Technology*, 24(5):353–363, 2005. 3

[97] A. Wu, Y. Wang, M. Zhou, X. He, H. Zhang, H. Qu, and D. Zhang. Multivision: Designing Analytical Dashboards with Deep Learning Based Recommendation. *IEEE TVCG*, 28(1):162–172, 2021. 2

[98] O. M. Yigitbasioglu and O. Velcu. A Review of Dashboards in Performance Management: Implications for Design and Research. *International Journal of Accounting Information Systems*, 13(1):41–59, 2012. 2

[99] Y. Zhang, Y. Sun, J. D. Gaggiano, N. Kumar, C. Andris, and A. G. Parker. Visualization Design Practices in a Crisis: Behind the Scenes with COVID-19 Dashboard Creators. *IEEE TVCG*, 2022. 2

[100] T. Zuk, L. Schlesier, P. Neumann, M. S. Hancock, and S. Carpendale. Heuristics for Information Visualization Evaluation. In *Proceedings of the 2006 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization*, pp. 1–6, 2006. 2