How Do We Measure Trust in Visual Data Communication?

Hamza Elhamdadi* UMass Amherst
Aimen Gaba† UMass Amherst
Yea-Seul Kim‡ University of Wisconsin-Madison
Cindy Xiong§ UMass Amherst

2022 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV) | 979-8-3503-9629-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/BELIV57783.2022.00014

ABSTRACT

Trust is fundamental to effective visual data communication between the visualization designer and the reader. Although personal experience and preference influence readers' trust in visualizations, visualization designers can leverage design techniques to create visualizations that evoke a "calibrated trust," at which readers arrive after critically evaluating the information presented. To systematically understand what drives readers to engage in "calibrated trust," we must first equip ourselves with reliable and valid methods for measuring trust. Computer science and data visualization researchers have not yet reached a consensus on a trust definition or metric, which are essential to building a comprehensive trust model in human-data interaction. On the other hand, social scientists and behavioral economists have developed and perfected metrics that can measure generalized and interpersonal trust, which the visualization community can reference, modify, and adapt for our needs. In this paper, we gather existing methods for evaluating trust from other disciplines and discuss how we might use them to measure, define, and model trust in data visualization research. Specifically, we discuss quantitative surveys from social sciences, trust games from behavioral economics, measuring trust through measuring belief updating, and measuring trust through perceptual methods. We assess the potential issues with these methods and consider how we can systematically apply them to visualization research.

Index Terms: Human-centered computing—Visualization—Visualization design and evaluation methods;

1 INTRODUCTION

Effective visual data communication is a dynamic exchange between a visualization (and its designers) and the information consumer. When designers create visualizations for communication, they make choices about encoding and design that they think will accurately and persuasively communicate their interpretation of the data. When information consumers interpret visualizations, they choose which patterns to focus on and what stories to take away. Visual data communication involves critical thinking and trust from both parties [17], making trust especially important to establish [42]. Human readers can learn to **trust** or distrust the conveyed information, and visualizations are judged more or less **trustworthy** depending on their design and delivery.

Ideally, we want human readers to engage in "calibrated trust" when interacting with data visualizations, which involves *critically evaluating the information, rather than unconditionally dismissing or accepting it*. When the reader does not blindly reject a visualization, they are exhibiting some trust that the visualization is worth at least contemplating. Similarly, when the reader does not blindly accept the conclusions of the visualization, they are exhibiting the capability to detect signs of mistrust that may be present in the visualization. This calibrated trust exists between a visualization reader and a visualization (human-to-visualization trust) but can also be extended as trust between the visualization reader and the visualization creator (human-to-human trust). For the purposes of this paper, we will cover only the human-to-visualization trust.

Social science research has demonstrated that by educating people to identify signs of trust, they can better calibrate themselves in identifying signs of mistrust [7]. As visualization researchers, we can investigate visualization components that elicit trust and mistrust to empower readers to identify signs of mistrust and guide designers to create trustworthy visualizations; thus, we optimize the effectiveness of visual data communication and data-driven decision making.

However, personal experience and preference can heavily influence trust [24]. For example, subjectively perceived transparency [73] and aesthetic preferences [2] can impact perceived trustworthiness. Trust is also very contextualized. Depending on the domain and application, the role of trust and peoples' criteria for trustworthiness might change [51, 63]. Therefore, it is unlikely that we will easily find a general set of rules to optimize trust across all scenarios of visual data communication. Instead, we may take a domain- or task-specific approach to look at the role of trust in different contexts. This bottom-up approach will allow us to model trust for visualization research by understanding what visual data communication trust rules generalize across domains and tasks.

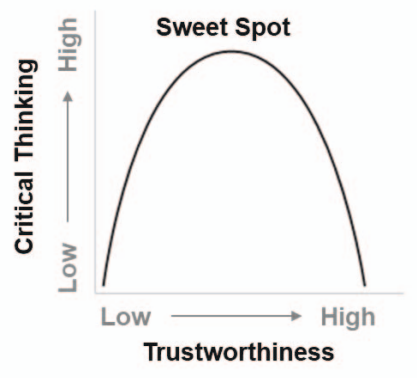

Figure 1: Extremely low and overwhelmingly high trust in visualized data might promote low critical thinking. Visualization authors should consider the trade-offs and aim for the sweet stop to maximize trust and enhance critical thinking.

To do so, we need to identify reliable and valid methods for measuring trust in visual data communication across different contexts and account for individual differences. Measuring trust with nuance and accuracy requires a multi-faceted approach. In this paper, we survey trust measurements used in existing research from social sciences, computer sciences, and behavioral economics, reflect on how these disciplines define and measure trust, and propose ways to use these metrics in visual data communication. This list is by no means exhaustive, but we hope that this collection will help the visualization research community better understand the complexity of trust and the need to measure it via a multi-faceted approach.

*e-mail: helhamdadi@umass.edu
†e-mail: agaba@umass.edu
‡e-mail: yeaseul.kim@cs.wisc.edu
§e-mail: cindy.xiong@cs.umass.edu

979-8-3503-9629-4/22/$31.00 ©2022 IEEE
DOI 10.1109/BELIV57783.2022.00014

2 Trust Measurement in Computer Science

Due to the complicated nature of trust, existing computer science research varies in the approach to measuring trust, and the definition of trust varies from paper to paper [4]. For example, Boyd et al. define trust based on whether people believe what they see [8]. Some define trust in terms of the entity relaying the information (i.e., is the source of this information actually who they claim to be?) [4]. Others define trust as "believing others in the absence of clear-cut reasons to disbelieve" [62].

Computer scientists have measured trust with multiple approaches. Jian et al. [34] developed a 12-item 7-point Likert scale to measure user trust in a movie recommendation app. To measure people's trust in a machine learning model, Yin et al. [76] used two metrics: how often the users agreed with the model's prediction and how frequently the users changed their prediction to be consistent with the model prediction. To measure the impact of computer-mediated communication on interpersonal trust, Zheng et al. [78] used a trust game referred to as the "Day-Trader Investment Task." Each participant is given a fixed amount of money in a hypothetical scenario and can choose to invest a portion. In this set-up, the amount of money a participant invests is a proxy for the amount of interpersonal trust they exhibit.

However, researchers rarely test these trust metrics for reliability and validity. Psychologists often rigorously test their surveys to ensure the repeatability of their results and minimize harmful biases. We highlight a case study illustrating why this process is critical for measuring trust in computer science. In an experimental study regarding trust between humans and automated systems, Jian et al. developed a 12-item Likert scale (from 12 identified trust factors, six positive, six negative) for measuring human-computer trust [34]. Although Jian et al. noted that testing validation was necessary, the scale was used in 179 papers (in its original form in 100) as of 2019. Despite the proliferation of this scale, Gutzwiller et al. [27] ran an experimental study to test the scale's efficacy and found that the scale biased users' trust scores depending on the ordering of questions; if the positive items were given first, the trust scores were significantly higher than if the negative items were given first or if the items were given in a random order. This case study demonstrates that a lack of iteration on the trust metrics we use may lead to biased results; hence, a non-unified approach to measuring trust could hide a lack of repeatability or other unknown flaws with various existing trust measures.

In this paper, we propose multiple methods to measure trust grounded in theories of social sciences, behavioral economics, and cognitive sciences. But like the example illustrated, although these methods have been tested and perfected over time in these other communities as reliable and valid measures of trust, we must iterate and improve on them to ensure their reliability and validity in the context of measuring trust in computer science research. We also focus on quantitative trust measures in this paper; however, this does not mean we should neglect qualitative measures. Qualitative metrics, such as grounded theory [50], interviews, and open-text survey questions can reveal nuances and additional insights beyond quantitative methods. For example, Hong et al. measured trust via the qualitative coding of existing research papers in machine learning that mention trust [32]. Effectively, this approach considers the trust definition used by scientists working in machine learning. Similarly, Hu et al. measured users' trust in the output of an assembly code synthesizer via free-response questions that the users were encouraged to answer; Hut et al. evaluated user answers (e.g., "I'm involved enough in the process that I trust the results") qualitatively to assess general trust patterns [33].

3 Trust Measurement in Visualization Research

Visualization researchers recognize the complexity of trust and its measurement and leverage various metrics to measure trust. The most prominent being a Likert scale - e.g., "on a scale from 1 to 7, how much do you trust this visualization?" [73], or "on a scale of 1 to 100, how much do you trust that the base data is accurate? " [45], or "on a scale from 1 to 9, how much do you trust the visualized model predictions?" [79]. The Likert scale can vary in the number of discrete values offered (e.g., 0 to 100 [45], 1 to 5 [77], or 1 to 7 [34]), or the number of the Likert scale items (e.g., 3 items [49] or 5 items [73]).

Other approaches include using substitutions variables as a proxy of trust. For example, Xiong et al. asked participants to choose the best map visualization for a firetruck in a hypothetical emergency. They measured trust through the decision (i.e., participants likely trusted the chosen map more than the others), as well as ratings of perceived trust and transparency, which have four dimensions: the accuracy of the information, the clarity of the communicated information, the amount of information disclosed, and the extent to which the shared information is a thorough representation of the underlying data [73]. In [49], the researchers measured participants' trust in title-misaligned visualizations via the three proxies: perceived credibility, perceived bias, and appropriateness. In another example, [77] measured trust in human-generated versus algorithm-generated visualizations via perceived usefulness and preferences by asking participants to choose between different human-generated and algorithm-generated visualizations for a particular task.

3.1 Issues with Trust Measurement in Visualization

We discuss two issues with the existing trust metrics in visualization research: lack of objectivity and lack of reliability and validity.

The Likert scales used to measure trust are usually participants' self-reported ratings, which, like all self-report measures, are not always accurately reported nor representative of the participants' true inner state [14]. Furthermore, participants' interpretations of the Likert values may differ, sometimes causing two participants exhibiting the same level of trust to choose different values on the scale. This effect can compound when participants cannot opt-out [10] (e.g., via a N/A option). We can mitigate these shortcomings by supplementing the quantitative measures with a qualitative open-text response box so participants can further explain the rationale behind their ratings. However, low participant motivation to detail their inner thoughts and inability to articulate their mental representation of trust can stymie the effectiveness of these qualitative responses.

We must recognize that reliable and valid measures typically result from iterations of testing, experimentation, and revision [64, 70]. The current state, where trust is decomposed and defined by different researchers as different variables, it becomes difficult to evaluate whether the decomposition (and the subsequent measures) are valid, that they are actually measuring trust, or reliable, that it will yield consistent results across individuals sharing similar characteristics or within the same individual across time. For example, in [73], where both participants rated the perceived trustworthiness of maps and selected a map to follow, the researchers found inconsistent results from the two measures. However, Maps highly rated for perceived trust did not always coincide with the selected map, leaving questions about how these measures capture different trust components.

Furthermore, the different designs of the Likert scale, such as including different number of questions/items between studies (e.g., one [45], six [54], twelve [34]), different ranges of values (e.g., 0 to 100 [45], 1 to 5 [77], or 1 to 7 [34]), can make it difficult for researchers to iterate on one another and compare experimental effect sizes to improve the reliability of these measures. A single-item Likert scale approach may not be a reliable indication of a person's trust; ideally, the scale used would have more than one item, approaching the multi-faceted trust question from multiple angles to increase reliability and validity [19]. Additionally, while researchers frequently use scales with a higher number of items, the reasoning for choosing a particular scale and its efficacy are not

Original Evans and Revelle Item [19]	New Visualization-Appropriate Item
Anticipate the needs of others	Provides useful information
Have always been completely fair to others	Depicts balanced data
Stick to the rules	Follows best design practices
Value cooperation over competition	Uses unbiased data sources
Return extra change when a cashier makes a mistake	Issues notices and revises when mistakes are found
Would never cheat on my taxes	Does not falsify data
Finish what I start	Provides complete data

Table 1: Possible modifications for Evans and Revelle survey items for visualization purposes

often comprehensive. One of the most common (12-item) scales used [45] was shown to bias the responses of participants [27]. Visualization research should evaluate the existing n-item Likert scales and increase the reliability of the most promising scales via further iterative reliability testing on trust, changing the number of values, the number of questions, and even the wording of the questions when necessary.

One factor that might have contributed to the validity issue of trust measurement is that trust has been ill-defined in visual data communication and computer science research. For example, trust has been defined as strongly related to bias and credibility [49], integrity and deception and reliability [34]), or transparency [73]. But studies where participants rate trust directly (e.g., "on a scale of 1 to 100, how much do you trust that the base data is accurate" [45]), participants are usually not provided a clear definition of trust, and thus it remains unclear what kind of trust they are rating if it is trusted at all.

However, we recognize this is not a unique concern to the visualization research community. Social scientists, behavioral economists, and psychologists have operationalized trust differently [63]. Trust definitions include a "willingness to be vulnerable to exploitation within a social interaction" [53], reciprocity (i.e., willingness to give up something in return) [78], and reliance (e.g., on a machine learning model's recommendation [32].

Social and political research defines trust via top-down and bottom-up theories. Bottom-up theories propose that socio-political trust stems from individuals that initially trust until trauma or difficult decisions cause them to reevaluate. Top-down theories argue that trust results from societal or political factors like good government, general wealth, or happiness. Further theorizing has determined that social and political trust are also context-dependent (i.e., individuals evaluate trust in others on a case-by-case basis); for example, someone may trust their friend to take care of their pet but not to water their garden for a month [69].

4 Measuring Trust in Social Sciences

Much research has been done in the social sciences to define trust and create a robust trust metric. Trust has been measured as "Generalized Trust" in other people in the World Value Survey, which asks, via an open-text response box, 'Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" The responses are coded into two categories: those with high Generalized Trust who say that most people can be trusted, and those with low Generalized Trust who say one needs to be very careful in dealing with people [29]. The responses from the Generalized Trust survey questions positively correlate with other, more objective measures of trust, such as an "Investment Game" scenario (see Section 5) [36].

Although much of the research on trust in the social sciences are concerned with trusting behavior and general perspectives on trust and trustworthiness, similar studies have been aimed at capturing trust in a particular individual. Everett et al. measured how trusting people are of a particular leader using a 2-item 7-point Likert scale ("How trustworthy do you think this person is?" and "How likely would you be to trust this person's advice on other issues?"), and a voting task, where participants were asked to cast a vote between two possible leaders who had the opportunity to embezzle money from a donation to a charity [20]. However, trust was not explicitly defined in this survey. A more well-defined trust question in a different context could be "How much do you trust your physician to perform necessary medical tests and procedures regardless of cost?" from [24], as this question asks about a particular scenario (performing medical tests/procedures) rather than general trust.

Psychologists have argued that these single-item or two-item scales may be insufficient to reliably capture trust [19]. Evans and Revelle [19], for example, define trust as an enduring trait in people (rather than a transient one) that comprises the intention to accept vulnerability based upon the positive expectations of the intentions or behavior of another" [63]. They argue that because trust is an enduring trait, we should assess trust in similar ways as personality and study trust as a broad construct that complements personality models such as the Big Five [16, 60]. The Big Five Inventory is a 44-item scale that measures five dimensions of personality (extraversion, agreeableness, openness, conscientiousness, and neuroticism); each item describes a trait using different adjectives concerning a prototypical behavior, and the participant can rate via a 5-point scale indicating whether they disagree strongly (1) or agree strongly (5). For example, one item may state "I persevere until the task is finished" [35]. Based on this, Evans and Revelle created (and tested) a 21-item scale that measures interpersonal trust [19]. Sample items on the scale include "Listen to my conscience" and "Avoid contact with others."

Some other trust surveys include the one by Justwan et al., which uses the general trust question responses and 19 variables that correlate with social/political trust (e.g., political rights, corruption, income inequality, fertility rate) as factors in a Bayesian model, which measures trust in the context of international relations [37]. The World Values Survey has also made updates to now contain a set of questions (answered via a 4-point Likert scale) regarding trust/confidence in specific social groups (e.g., family, neighborhood) and organizations (e.g., churches, the press, television) [29] that provide more narrow focus on what elements of trust may be affecting generalized trust.

4.1 How we can do this in visualization research

We propose that visualization researchers can also modify interpersonal trust surveys, such as the one by Evans and Revelle [19], to be about trust in human-data interaction. For example, the survey item "[this person] stick[s] to the rules" measures trustworthiness and can be modified as "[this visualization] follows best design practices" to assess trust in human interaction with visualizations. See Table 1 for a more comprehensive list of modifications to the Evans and Revelle scale (Note: many items from the Revelle scale have no visualization counterpart).

For trust measures used by social scientists that measures general trust, while it may seem irrelevant in determining a reader's trust in a visualization, generalized trust can offer a baseline for measuring trust in visual data communication. In other words, we can compare

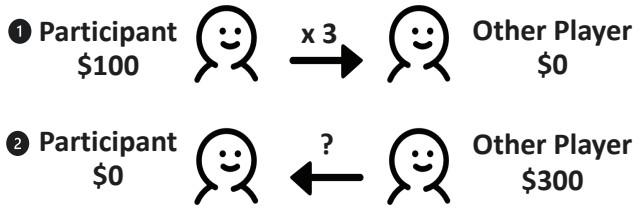

Figure 2: Trust game set-up in two steps. Top: The participant starts with some money and can decide on an amount to send to another player. The money sent gets tripled when it reaches the other player, and the other player can choose to return any amount back to the participant. The tripling is critical here because it is the lowest possible manipulation that enables the other player to share an amount of money that can profit for both players.

the generalized trust of the reader to their trust behavior regarding a visualization. A generally more trusting person may rate a visualization as more trustworthy than a generally less trusting person. But, we can approximate the trustworthiness of a visualization by computing the difference - how much trust does the visualization elicit above and beyond one's baseline level of trust?

We also recognize that trust is context-dependent. Previously, we demonstrated that adding more context to a trust question more effectively defines trust and helps participants better understand what they are rating [19]. [24] proposed to measure trust between physicians and patients by asking "How much do you trust your physician to perform necessary medical tests and procedures regardless of cost?". We can modify questions like this to be about visualizations. For example, "How trustworthy do you think the author of this visualization is?" or "How likely are you to rely on the information presented in this visualization to make future decisions?"

Although we proposed multiple ways to evaluate trust in visual data communication referencing survey items from social sciences, we emphasize that future researchers who wish to measure trust using similar scales should empirically test, iterate, and improve upon them. The wording of each question might have a biasing effect on the survey taker. Moreover, if responses to these questions use a Likert scale, we should consider that, if given the option to select "Not Applicable," people will opt for this answer in some scenarios [10]. Failing to provide this option forces the participant to answer the question, and when the subject feels the question does not apply, the answer may not be indicative of what they believe, which could skew the results.

5 Trust Games in Behavioral Economics

Although there is no clear cross-discipline definition of trust, most behavioral economists [22, 80] adopt James S. Coleman's definition from "Foundations of Social Theory" [11]. According to this definition, trust occurs when resources are given by a trustor to a trustee with no enforceable commitment from the trustee. For example, a trustor can decide to lend their car to a trustee, knowing the risk that the trustee may not give it back.

A subgroup of economists, particularly neoclassical economists, argue that people always seek to maximize their monetary profits and personal satisfaction without considering how their choices might affect others [23]. According to this view of rationality, trusting others (i.e., not thinking of maximizing one's profits) is irrational because the risks involved usually work against the person's material interests. Following this assumption, there is no reason to take the "irrational" risk of trusting someone else, and even simple exchanges in the real world (e.g., ordering clothes online) would require some form of complex litigation. However, from our real-world experiences, we know this is not the case. We trust others and hope for others' trust in us.

To understand what motivates trust, behavioral economics research has employed trust games. As shown in Figure 2, a typical trust game involves two anonymously paired participants; one participant (the trustor) is given an amount of money from which they can choose to give a portion or none to the other participant (the trustee); the experimenter triples the trustee's money, and the trustee can give some portion or none of the resulting amount back to the trustor [6]. Note, a researcher can conduct a trust game with one participant, where the trustee is played by a computer program [3]. By giving money to the trustee, the trustor is willing to put themselves in a vulnerable position; hence, the implication is that a high amount of money indicates high trust. Another possible implication is that the trustor is being altruistic and, thus, not engaging in trust. However, Brülhart and Usunier rejected altruism as the cause of "trust-like" decisions by [9].

As measuring trust via interactive games has led to multiple versions of the trust game, we discuss some of the prominent/relevant ones here. Kreps et al. introduced a simpler version of the trust game by providing participants with only two options: to trust or not to trust [52]. When a trustor decides to trust, the trustee has the option to either honor it leading to equal payoffs of $10 or dishonor it leading to a payoff of $15 for the trustee and -$5 for the trustor. If the trustor decides not to trust, both the trustor and the trustee are left with $0.

Guth et al. (1982) introduced a "take-it-or-leave-it" variation of the game, often referred to as *The Ultimatum Game*. It consists of two players given an endowment they must divide amongst themselves. Player One has the option to offer a part of the endowment to Player Two, which Player Two can either accept or reject. Acceptance of the proposal meant enforcement of the proposal, and rejection meant neither of the two received anything. The results showed that Player One made substantial offers, and Player Two rejected small but positive offers. Similar results occurred in a "take-it" variation of this game called *The Dictator Game* [21]. In this case, Player One chooses how to allocate the endowment and Player Two could benefit from the allocation but had no power to accept or reject the decision. The findings showed Player One (the dictator) granted surprisingly positive amounts to the other player. There is also *the game of trust* [28] in which the first mover can decide between non-cooperation and cooperation, whereas the second mover, who decides only when cooperating, must choose whether to share the fruits of cooperation equally or not leave anything to the first mover who put the trust by cooperating in the first place.

Zheng et al. explore how trust varies by playing an interactive game called *The Day-Trader Investment Task* [78]. The game is played in pairs in multiple trials, and the participants took part in a variant of a Prisoner's Dilemma task, which has a history of testing group cooperation and trust [48]. Each participant received $40 per day and imagined being a day-trader during a multi-day investment period. Out of that amount, they could invest all or some of it. The day-trader had two choices for investing the money: invest in a common pool whose payoff depended on how much the other partner invested in it, or keep it in an individual account. In order to conceal what each partner exactly contributed, the researcher added a random factor between -10 and +10 to represent stock market fluctuations. After including the random factor, each participant's investment with the group was doubled and split evenly between the two participants. At the end of the 'day', the participants had the amount they chose to keep in their individual account, or the amount they gained from the common pool after the payout. The authors observed higher trust (i.e., more cooperation) in face-to-face pairs compared with pairs that never met one another.

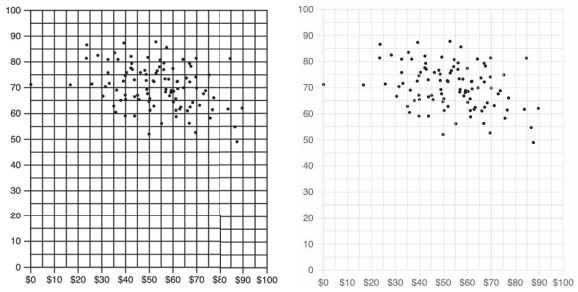

Figure 3: (left) a bad use of Grid Lines, α=1, that is intrusive and causes the visualization to be dis-fluent, (right) a better use of Grid Lines, α=0.45, that is not intrusive and is relatively fluent, following best practice recommendation from [5].

5.1 How we can do this in visualization research

In [80], the researchers compared the trustworthiness of easy- versus difficult-to-articulate names using the trust game set-up. Participants uniformly played with computer programs disguised as real people whose usernames varied in pronounceability. The amount of money the participants invested represented the level of trust, and the researchers found that participants tend to trust 'other players' with easier-to-pronounce usernames.

To our knowledge, visualization research has not used trust games to measure trust, despite their providing a metric less influenced by subjective trust cognition. Here are some methods that visualization researchers can use to leverage trust games in their research. We can modify the trust game to replace the second player with a visualization. Then, we can manipulate different design elements in a visualization and compare how the amount of trust changes depending on these design choices. This set-up parallels existing findings from behavioral economics where researchers find aesthetics (or superficial) features of the other player can impact trust and will likely enable us to observe similar effects of visualization design on trust. For example, the trust game was able to demonstrate that perceived attractiveness [72], age [67], and the width of the face [66] impact trust. Similarly, we can expect the trust game to be able to identify the effect various visualization design have on perceived trust.

One issue to resolve between the traditional set-up of having two players (despite one being a computer program) and the proposed set-up, which involves only one player and visualization, is the *why* the participant player should invest money into a visualization. Alternatively, if the game lasts multiple rounds, we will need some method to return the participant player their investment in a justifiable way, to create interaction between the player and the visualization, and potentially build (or diminish) trust.

We propose two scenarios to make the trust game set-up between a human player and a visualization more realistic. In one scenario, the participant can make a decision on how much money to invest based on the information presented in the visualization. In this case, the more they trust the visualization, the more money they will invest. In the second scenario, the participant can bet on whether the visualization will lead another hypothetical viewer to make a decision. If they bet correctly, they will receive some returns on their money. If not, they lose the money. This approach tests to what extent the participant sees the visualization as trustworthy to others. From related work on the theory of mind [75], we can infer that the participant's prediction of a visualization's general trustworthiness reflects how trustworthy they themselves perceive the visualization to be. Future work can further test these two scenarios, so the story can be iterated and improved to be comparable, reliable and valid as the traditional trust game set-up in behavioral economics research.

6 Measuring Trust via Perceptual Fluency

Trust has been linked to the halo effect, a phenomenon coined by Edward Thorndike in which a single positive quality of a person is extrapolated to a generally positive assessment of the person in other areas [68]. This effect is quite significant and surprisingly is not reduced when participants are made aware of it in an experimental context [71]. Beauty or aesthetics commonly produces a strong positive bias on people's impression of others. People or objects that are perceived as beautiful are also categorized as more functional and trustworthy [38]. Extrapolating this effect to visualizations, we might expect beautiful visualizations would be perceived as more trustworthy.

However, a person's judgment of another's trustworthiness is formed by myriad causes, including past experiences, knowledge, culture, and even their current mood. Psychology researchers have found a strong correlation between the perceived beauty of a subject and perceptual fluency, the ease with which a stimulus is perceived [61]. For example, stocks with aesthetically pleasing, fluent (read: easier-to-pronounce) names and codes performed better than those with aesthetically displeasing, dis-fluent (read: hard-to-pronounce) names/codes shortly after the stocks were released to the general public when controlled for the size and industry of the companies [1]. Similarly, faces that are easier to categorize socially and racially are perceived as fluent and judged more attractive [30]. Hence, we would expect that more fluently-perceived people or objects are also perceived as more trustworthy. As expected, stimuli that are more perceptually fluent (i.e., easier to perceive) are often considered more beautiful or judged more positively, leading to a positive affective association and higher trust in the stimulus [25]. This relationship between perceptual fluency and perceived beauty or satisfaction also works in the converse direction; stimuli that are perceptually dis-fluent (visually cluttered) appear less satisfying and less trustworthy [65].

Fluency also plays a role in the trust judgment of faces. For example, Olszanowski et al. found that participants rated faces that expressed a "pure" emotion (e.g., only one emotion, sad, happy) are easier to read, processed more fluently, and seen as more trustworthy than "mixed" emotions (e.g., emotions that were hard to categorize as a single emotion, happy-sad) [58].

It is important to note that the effect of fluency on trust does not always occur. Oppenheimer notes people interpret fluency via naive theories regarding the cause of the fluency; in other words, people do not immediately judge fluent stimuli as fluent, but rather only after other plausible reasons for the fluency experience are not found [59]. For example, in the Olszanowski et al. study, when the emotions varied in dominance (e.g., angry vs. sad), fluency (i.e., "pure" or "mixed" emotions) did not impact the trust ratings given by participants [58]. Olszanowski et al. postulated that people take fluency into account for emotions that vary in valence (pleasant vs. unpleasant; attractive vs. unattractive) because fluency can be attributed to this variance (as fluency is a sort of valence metric); however, with emotions that vary in motivation, fluency is discounted as not having an effect on the level of motivation (because both sad and angry are negative valence emotions). Hence, in an experimental setting, researchers should control for possible factors that participants might perceive as causing the fluency effect.

6.1 How we can do this in visualization research

More fluently processed stimuli can be perceived as more trustworthy. This phenomenon implies that it is possible for us to measure the perceived trustworthiness of a visualization by measuring how fluently can a reader perceptually process the visualization. Existing work has begun to show some evidence that perceptual fluency has the potential to positively influence trust in visualizations [55]. For example, Elhamdadi et al. measured participants' trust of fluent and dis-fluent visualizations. They found that more perceptually fluent

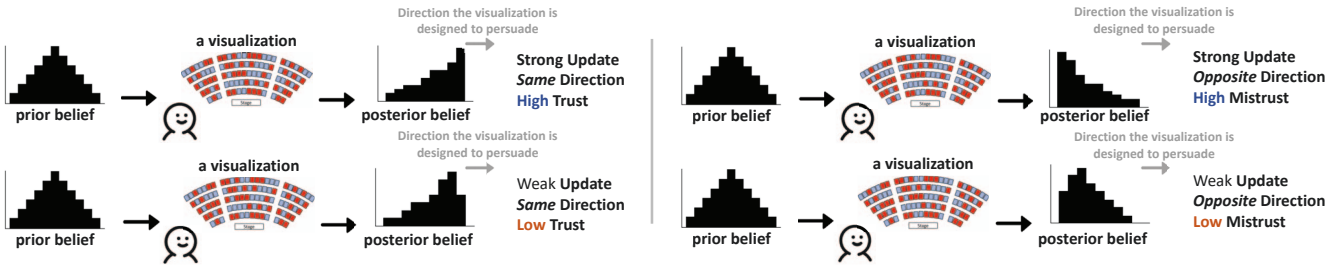

Figure 4: Examples of potential relationship between belief updating and trust.

visualizations (e.g., visualizations without heavy grid lines) are rated more trustworthy vis-a-vis trust games and subjective ratings of trust on a Likert scale [18]. Although it warrants further investigation, we posit that once we establish a more comprehensive model of the extent to which perceptual fluency can impact trust, we can predict how trustworthy a visualization is by measuring how fluently people perceive it.

Existing work at the intersection between data visualization and human perception has contributed concrete metrics on design factors in a visualization that can increase or decrease its processing fluency. For example, Bartram et al. discovered a range of alpha values for the transparency of grid lines on scatter plots that ensure the lines don't intrude on a reader's ability to interpret a visualization. This range of alpha values is 0.1 to 0.45, with lower alpha values preferred for sparse scatter plots and higher for dense plots [5]. Before the paper by Bartram et al., it was not uncommon to see alpha values of 1.0 as the default in visualization tools; heavy lines (with high alpha values) obscure the data and may cause the visualization to be dis-fluent. Because heavy grid lines create a dis-fluency effect that is unlikely to be attributed to another causal relationship, it is reasonable to expect that the dis-fluency effect will not be discounted and may cause distrust in the visualization.

As a community, we should continue to explore guidelines that can increase perceptual fluency in visualization designs and systematically test their effectiveness in improving trust in visual data communication. Following that, we will be able to measure the likelihood a reader will trust a visualization by proxy of perceptual fluency.

7 Measuring Trust through Belief Updating

In [76], the authors measured people's willingness to trust a machine learning model by the frequency with which people updated their predictions to match the model's output. This approach inspires us to approximate individual trust as defined by "the likelihood to update one's belief based on the given information." The more strongly someone trusts the information presented, the more likely they are to incorporate it in updating their beliefs.

The direction of the belief updating can indicate whether trust or distrust was involved. As shown in Figure 4, if the information consumer updates their beliefs in the same direction as what the information presented, such as more firmly believing that a vaccine works after seeing information about increased vaccination rates and decreased infection rates, we can say they trust the information. If the information consumer updates their belief in the opposite direction, such as becoming more skeptical of the vaccine after seeing the same information, we can infer that they distrust. These results belied polarization, where two people update their beliefs in opposite directions when responding to the same evidence. Although it may seem irrational and violate the normative Bayesian, this effect occurs when interacting with controversial topics such as climate change as a product of distrust [12].

This approach provides the potential to quantify trust. Suppose we calculate the difference between one's updated beliefs and their prior beliefs with a measure (e.g., Kullback–Leibler divergence or Earth Mover's Distance). The difference can be decomposed by the signal the person interprets from the visualization and weights (i.e., trust) that this person weighs the interpreted information to incorporate into their beliefs. While modeling these two components are empirical questions that require many controlled experiments, the lens of using belief updating to approximate trust offers the potential to further observe the degree to which the person trusts the data.

7.1 How we can do this in visualization research

Some visualization techniques facilitate belief-updating and, by our definition, has the potential to improve the trustworthiness of the visualization. For example, effectively showing uncertainty in data using quantile dot plots, gradient plots, or hypothetical outcome plots can help people more appropriately update their beliefs [13, 39, 40, 44]. However, it has not yet been empirically tested whether these techniques that facilitate adequate belief-updating are also perceived as more trustworthy. Future work should measure the ability of a visualization to update people's beliefs and the amount of trust they have in the visualization (via the other methods introduced in this paper) to model the exact relationship between belief updating and trust.

It is also important to note that people can be biased by their knowledge and past experience when making sense of new information [15, 38], such as exhibiting confirmation bias to overweigh information that is congruent with their prior beliefs and under weigh information that is incongruent [47]. Therefore when attempting to measure trust via measuring belief updating, it is critical for us to consider valid and reliable methods of belief elicitation. We can capture people's beliefs by asking people to manipulate the slope of a line in a chart [41, 46, 56], indicating the range of their belief on a Likert scale [43], or to make predictions [26]. Future work can compare these channels of belief measurements with each other to refine metrics that directly map with trustworthiness.

Existing work has demonstrated that people can update their belief on the relationship between two entities after seeing just one instance of a visualization [74]. So while the amount of change between prior and posterior beliefs can likely be used as a proxy to determine how much people trust the new data they saw, we should also consider the possibility of people updating their beliefs due to mere exposure effects [31], and account for them when we create quantitative models of belief updating and trust.

8 Conclusion

Trust is a complex topic. How trust impacts human-data interactions (and by extension, the trust relationship between the reader and the visualization designer) remains mostly under-explored. The visualization community does not yet have a systematic understanding of what factors impact trust in visual data communication nor a formalized model for establishing trust between humans and data.

In this paper, we presented a multidisciplinary, but by no means exhaustive, collection of trust measurements from existing work in computer science, psychology, behavioral economics, and other social sciences. We call for visualization researchers to consider these methods, adapt them for visualization research, and, as a community, iteratively improve them for reliability and validity. This paper is our first step in an ongoing process toward defining and modeling trust in visual data communication, the future result of which we hope will be a set of design guidelines for enhancing trust.

With these trust measurements, we can develop a better understanding of trust in human-data interactions. Establishing this understanding can further inform visualization practices in many sub-fields of computer science. For example, visualization is increasingly used in the field of explainable artificial intelligence to describe the inner workings of machine learning algorithms, and efforts such as the TRust and EXpertise in Visual Analytics (TREX) Workshop [57] continue to promote interdisciplinary science to support human-centered AI research and practice. Presenting data visualizations in a trustworthy manner can enhance understanding and engagement, preventing potential confusion or misuse of the algorithms and AI tools.

REFERENCES

[1] A. L. Alter and D. M. Oppenheimer. Predicting short-term stock fluctuations by using processing fluency. *Proceedings of the National Academy of Sciences*, 103(24):9369–9372, 2006. doi: 10.1073/pnas.0601071103

[2] A. L. Alter and D. M. Oppenheimer. Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, 13(3):219–235, 2009.

[3] V. Anderhub, D. Engelmann, and W. Güth. An experimental study of the repeated trust game with incomplete information. *Journal of Economic Behavior Organization*, 48(2):197–216, 2002. doi: 10.1016/S0167-2681(01)00216-5

[4] D. Artz and Y. Gil. A survey of trust in computer science and the semantic web. *Journal of Web Semantics*, 5(2):58–71, 2007. Software Engineering and the Semantic Web. doi: 10.1016/j.websem.2007.03.002

[5] L. Bartram and M. C. Stone. Whisper, don't scream: Grids and transparency. *IEEE Transactions on Visualization and Computer Graphics*, 17(10):1444–1458, 2010.

[6] J. Berg, J. Dickhaut, and K. McCabe. Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1):122–142, 1995. doi: 10.1006/game.1995.1027

[7] A. Böckler-Raettig. *Theory of mind*. utb GmbH, 2019.

[8] S. Boyd Davis, O. Vane, and F. Kräutli. Can i believe what i see? data visualization and trust in the humanities. *Interdisciplinary Science Reviews*, 46(4):522–546, 2021.

[9] M. Brülhart and J.-C. Usunier. Does the trust game measure trust? *Economics Letters*, 115(1):20–23, 2012. doi: 10.1016/j.econlet.2011.11.039

[10] M. Chita-Tegmark, T. Law, N. Rabb, and M. Scheutz. Can you trust your trust measure? In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 92–100, 2021.

[11] J. Coleman and American Council of Learned Societies. *Foundations of Social Theory*. ACLS Humanities E-Book. Belknap Press of Harvard University Press, 1990.

[12] J. Cook and S. Lewandowsky. Rational irrationality: Modeling climate change belief polarization using bayesian networks. *Topics in Cognitive Science*, 8(1):160–179, 2016.

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

[14] P. C. Cozby, S. Bates, C. Krageloh, P. Lacherez, and D. Van Rooy. *Methods in behavioral research*. McGraw-Hill New York, NY, 2012.

[15] N. F. Dieckmann, R. Gregory, E. Peters, and R. Hartman. Seeing what you want to see: How imprecise uncertainty ranges enhance motivated reasoning. *Risk Analysis*, 37(3):471–486, 2017.

[16] J. M. Digman. Higher-order factors of the big five. *Journal of Personality and Social Psychology*, 73(6):1246, 1997.

[17] M. Dörk, P. Feng, C. Collins, and S. Carpendale. Critical infovis: Exploring the politics of visualization. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, p. 2189–2198. Association for Computing Machinery, New York, NY, USA, 2013.

[18] H. Elhamdadi, L. Padilla, and C. Xiong. Processing fluency improves trust in scatterplot visualizations. *IEEE VIS 2022 Posters*, 2022.

[19] A. M. Evans and W. Revelle. Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality*, 42(6):1585–1593, 2008.

[20] J. A. Everett, C. Colombatto, E. Awad, P. Boggio, B. Bos, W. J. Brady, M. Chawla, V. Chituc, D. Chung, M. A. Drupp, et al. Moral dilemmas and trust in leaders during a global health crisis. *Nature Human Behaviour*, 5(8):1074–1088, 2021.

[21] R. Forsythe, J. L. Horowitz, N. E. Savin, and M. Sefton. Fairness in simple bargaining experiments. *Games and Economic behavior*, 6(3):347–369, 1994.

[22] E. L. Glaeser, D. I. Laibson, J. A. Scheinkman, and C. L. Soutter. Measuring trust. *The Quarterly Journal of Economics*, 115(3):811–846, 2000.

[23] R. Goodland and G. Ledec. Neoclassical economics and principles of sustainable development. *Ecological Modelling*, 38(1):19–46, 1987. Ecological Economics. doi: 10.1016/0304-3800(87)90043-3

[24] J. Goudge and L. Gilson. How can trust be investigated? drawing lessons from past experience. *Social Science & Medicine*, 61(7):1439–1451, 2005.

[25] L. K. Graf and J. R. Landwehr. A dual-process perspective on fluency-based aesthetics: The pleasure-interest model of aesthetic liking. *Personality and Social Psychology Review*, 19(4):395–410, 2015.

[26] T. L. Griffiths and J. B. Tenenbaum. Optimal predictions in everyday cognition. *Psychological Science*, 17(9):767–773, 2006.

[27] R. S. Gutzwiller, E. K. Chiou, S. D. Craig, C. M. Lewis, G. J. Lematta, and C.-P. Hsiung. Positive bias in the 'trust in automated systems survey'? an examination of the jian et al.(2000) scale. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 63, pp. 217–221. SAGE Publications Sage CA: Los Angeles, CA, 2019.

[28] W. Güth, P. Ockenfels, and M. Wendel. Cooperation based on trust. an experimental investigation. *Journal of Economic Psychology*, 18(1):15–43, 1997. doi: 10.1016/S0167-4870(96)00045-1

[29] C. Haerpfer, R. Inglehart, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin, and B. Puranen. World values survey wave 7 (2017-2020) cross-national data-set, 2020. doi: 10.14281/18241.1

[30] J. Halberstadt and P. Winkielman. Easy on the eyes, or hard to categorize: Classification difficulty decreases the appeal of facial blends. *Journal of Experimental Social Psychology*, 50:175–183, 2014.

[31] A. A. Harrison. Mere exposure. In *Advances in Experimental Social Psychology*, vol. 10, pp. 39–83. Elsevier, 1977.

[32] S. R. Hong, J. Hullman, and E. Bertini. Human factors in model interpretability: Industry practices, challenges, and needs. *ACM Conference On Computer-Supported Cooperative Work And Social Computing*, May 2020. doi: 10.1145/3392878

[33] J. Hu, P. Vaithilingam, S. Chong, M. Seltzer, and E. L. Glassman. Assuage: Assembly synthesis using a guided exploration. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pp. 134–148, 2021.

[34] J.-Y. Jian, A. M. Bisantz, and C. G. Drury. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4(1):53–71, 2000.

[35] O. P. John, S. Srivastava, et al. *The Big-Five trait taxonomy: History, measurement, and theoretical perspectives*. University of California Berkeley, 1999.

[36] N. D. Johnson and A. Mislin. How much should we trust the world values survey trust question? *Economics Letters*, 116(2):210–212, 2012.

[37] F. Justwan, R. Bakker, and J. D. Berejikian. Measuring social trust and trusting the measure. *The Social Science Journal*, 55(2):149–159, 2018.

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

[39] A. Kale, M. Kay, and J. Hullman. Visual reasoning strategies for effect size judgments and decisions. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):272–282, 2020.

[40] A. Kale, F. Nguyen, M. Kay, and J. Hullman. Hypothetical outcome plots help untrained observers judge trends in ambiguous data. *IEEE Transactions on Visualization and Computer Graphics*, 2018.

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

[42] K. Kelton, K. R. Fleischmann, and W. A. Wallace. Trust in digital information. *Journal of the American Society for Information Science and Technology*, 59(3):363–374, 2008.

[43] M. Kim and K. Liu. A bayesian deep learning framework for interval estimation of remaining useful life in complex systems by incorporating general degradation characteristics. *IISE Transactions*, 53(3):326–340, 2020.

[44] Y. S. Kim. *Designing Belief-driven Interactions with Data*. University of Washington, 2020.

[45] Y. S. Kim, K. Reinecke, and J. Hullman. Data through others' eyes: The impact of visualizing others' expectations on visualization interpretation. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):760–769, 2017.

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

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

[48] S. S. Komorita. *Social dilemmas*. Routledge, 2019.

[49] H.-K. Kong, Z. Liu, and K. Karahalios. Trust and recall of information across varying degrees of title-visualization misalignment. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–13, 2019.

[50] R. V. Kozinets. *Netnography: The essential guide to qualitative social media research*. Sage, 2019.

[51] R. M. Kramer. Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology*, 50:569, 1999.

[52] D. Kreps. *Corporate culture and economic theory. in alt, J., Shepsle, K.(Eds.). Perspectives on positive political economy*. New York: Cambridge University press, 1990.

[53] E. E. Levine, T. B. Bitterly, T. R. Cohen, and M. E. Schweitzer. Who is trustworthy? predicting trustworthy intentions and behavior. *Journal of Personality and Social Psychology*, 115(3):468, 2018.

[54] M. Liao, S. S. Sundar, and J. B. Walther. User trust in recommendation systems: A comparison of content-based, collaborative and demographic filtering. In *CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2022.

[55] C. Lin and M. A. Thornton. Fooled by beautiful data: Visualization aesthetics bias trust in science, news, and social media. 2021.

[56] P. Mantri, H. Subramonyam, A. L. Michal, and C. Xiong. How do viewers synthesize conflicting information from data visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[57] M. Nourani, E. Wall, E. Ragan, and A. Karduni. Trust and expertise in visual analytics. *IEEE VIS Workshop*, 2022, https://trexvis.github.io/Workshop2021/.

[58] M. Olszanowski, O. K. Kaminska, and P. Winkielman. Mixed matters: fluency impacts trust ratings when faces range on valence but not on motivational implications. *Cognition and Emotion*, 32(5):1032–1051, 2018.

[59] D. M. Oppenheimer. The secret life of fluency. *Trends in Cognitive Sciences*, 12(6):237–241, 2008.

[60] B. Rammstedt and O. P. John. Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of Research in Personality*, 41(1):203–212, 2007.

[61] R. Reber, N. Schwarz, and P. Winkielman. Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, 8(4):364–382, 2004.

[62] J. B. Rotter. Interpersonal trust, trustworthiness, and gullibility. *American Psychologist*, 35(1):1, 1980.

[63] D. M. Rousseau, S. B. Sitkin, R. S. Burt, and C. Camerer. Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3):393–404, 1998.

[64] P. E. Shrout and S. P. Lane. *Psychometrics*. The Guilford Press, 2012.

[65] S. Sohn. Consumer processing of mobile online stores: Sources and effects of processing fluency. *Journal of Retailing and Consumer Services*, 36:137–147, 2017.

[66] M. Stirrat and D. I. Perrett. Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, 21(3):349–354, 2010.

[67] M. Sutter and M. G. Kocher. Trust and trustworthiness across different age groups. *Games and Economic behavior*, 59(2):364–382, 2007.

[68] E. L. Thorndike. A constant error in psychological ratings. *Journal of Applied Psychology*, 4(1):25, 1920.

[69] E. M. Uslaner. *The Oxford handbook of social and political trust*. Oxford University Press, 2018.

[70] E. Wall, C. Xiong, and Y. S. Kim. Vishikers' guide to evaluation: Competing considerations in study design. *IEEE Computer Graphics and Applications*, 42(3):29–38, 2022.

[71] C. G. Wetzel, T. D. Wilson, and J. Kort. The halo effect revisited: Forewarned is not forearmed. *Journal of Experimental Social Psychology*, 17(4):427–439, 1981.

[72] R. K. Wilson and C. C. Eckel. Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly*, 59(2):189–202, 2006.

[73] C. Xiong, L. Padilla, K. Grayson, and S. Franconeri. *Examining the components of trust in map-based visualizations*. The Eurographics Association, 2019.

[74] C. Xiong, C. Stokes, Y. S. Kim, and S. Franconeri. Seeing what you believe or believing what you see? belief biases correlation estimation. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[75] C. Xiong, L. van Weelden, and S. Franconeri. The curse of knowledge in visual data communication. *IEEE Transactions on Visualization and Computer Graphics*, 2019.

[76] M. Yin, J. Wortman Vaughan, and H. Wallach. Understanding the effect of accuracy on trust in machine learning models. In *Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.

[77] R. Zehrung, A. Singhal, M. Correll, and L. Battle. Vis ex machina: An analysis of trust in human versus algorithmically generated visualization recommendations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2021.

[78] J. Zheng, E. Veinott, N. Bos, J. S. Olson, and G. M. Olson. Trust without touch: Jumpstarting long-distance trust with initial social activities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, p. 141–146. Association for Computing Machinery, New York, NY, USA, 2002. doi: 10.1145/503376.503402

[79] J. Zhou, Z. Li, H. Hu, K. Yu, F. Chen, Z. Li, and Y. Wang. Effects of influence on user trust in predictive decision making. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–6, 2019.

[80] M. Zürn and S. Topolinski. When trust comes easy: Articulatory fluency increases transfers in the trust game. *Journal of Economic Psychology*, 61:74–86, 2017.