Trust Calibration for Joint Human/AI Decision-making in Dynamic and Uncertain Contexts

Laura R. Marusich[1][0000−0002−3524−6110], Benjamin T. Files[1][0000−0002−1141−7886], Melanie Bancilhon[1][0000−0002−4834−6495], Justine C. Rawal[1][0000−0002−5232−1135], and Adrienne Raglin[1][0000−0003−2147−8938]

US Army DEVCOM Army Research Laboratory laura.m.cooper20.civ@army.mil

Abstract. Joint human/AI decision-making combines AI's ability to quickly process breathtaking amounts of data with human contextual understanding, adaptability and accountability. To achieve optimal performance, the human should have appropriately calibrated trust in the system, in which the amount of trust afforded to the system aligns with the trustworthiness of the system. Past work has explored several techniques to improve trust calibration, including transparency, explainability, and uncertainty visualization. Achieving trust calibration becomes even more difficult when the trustworthiness of the system is a moving target. In dynamic situations, the trustworthiness of AI systems can fluctuate wildly, demanding rapid updates to trust behaviors to achieve calibration. Accurate confidence or uncertainty measures have been proposed to help humans rapidly calibrate their trust in AI systems; however, this requires that accurate confidence measures exist and that humans can use them effectively. In this position paper, we join recent calls for research to improve confidence measures in AI systems, and we further emphasize the need to track and convey multidimensional confidence measures in the context of large, complex system-of-systems architectures. We discuss how these measures aid in establishing calibrated trust for AI systems even in the presence of uncertainty of information. Further, we highlight the opportunities for improved design in user interfaces that convey AI confidence to human users and for better preparing humans to optimally weight AI inputs against other sources of information, including their own judgment, to arrive at better results when making decisions under uncertainty in dynamic, complex environments.

Keywords: Trust calibration · AI uncertainty · Trust dynamics · Uncertainty of information · Uncertainty communication · Complex decision-making · Individual differences · Cognitive forcing functions · Cognitive biases.

1 Introduction

Artificial intelligence (AI) systems have the ability to rapidly synthesize and analyze vast amounts of information. In addition, recent advancements including large language models (LLMs) make interactions with AI systems increasingly intuitive and understandable to human users. These advantages mean that AI systems can potentially be enormously useful in terms of decision-making. Joint human/AI decision-making can combine the strengths of AI with those of humans in a way that leads to better performance than what can be attained by either alone. However, the benefits of joint human/AI decision making only manifest in domains that are too complex for an unaided human and too dynamic or uncertain for a pre-trained model to perform nearly perfectly. For optimal outcomes, human users must be able to assess when it is appropriate to rely on the AI system(s) and to what extent.

Well-studied issues of over- and under-reliance in human/AI decision-making [25, 37] illustrate that recovering from a mismatch between the person's reliance on the system and the system's reliability can be difficult, and this mismatch should be avoided. Many techniques to increase trust in AI have been examined, but over-trust can be just as much a problem as under-use [51]. Ideally, the degree of trust in the system should correspond to the system's reliability. This correspondence is known as trust calibration.

Trust calibration can be quite difficult to achieve even in static contexts, and a number of studies have explored interventions intended to promote it [54]. Critically, the type of complex domains in which joint human/AI decision-making can be most beneficial often involve information of varying uncertainty arriving from rapidly changing environments [48]. In such dynamic and uncertain contexts, the effective reliability of AI systems may also change rapidly, meaning that calibrating trust is not a one-and-done achievement. Rather, it must be a continuously updating process.

In the sections that follow, we first provide an overview of existing work on the dynamics of trust and trust calibration and potential gaps that may be productively explored. We then describe a number of relevant research findings and techniques from cognitive and computer science that suggest potentially useful directions to make progress in this area. Finally, we summarize these topics and provide recommendations for future work in trust calibration for dynamic joint human/AI decision-making.

2 Dynamics of Trust and Trust Calibration

Existing work on trust in automation/autonomy research does not tend to focus on the dynamics of trust over time, especially in the type of rapidly changing, multi-system contexts we describe here. Instead, much of the work assesses trust at a single point in time [55]. Some recent work has explored trust calibration with changing system reliabilities (e.g., [34]); however, most studies of trust dynamics have largely measured trust responses to discrete successes or failures from individual systems that have a constant reliability, rather than system(s) with changing reliabilities.

Nevertheless, results from these studies are informative about the ways that human decision-makers update their trust judgments over time. For example, there is evidence that human trust dynamics often exhibit three properties. One of these is continuity, meaning that trust is self-correlated: trust at the previous time step is highly predictive of trust at the current time step [24, 55]. A second property is negativity bias, in which humans have larger trust responses to system failures than they do to system successes [29, 57]. Finally, at least when interacting with systems with unchanging reliability levels, people demonstrate stabilization, meaning that over time and experience with the system, their trust adjustments will grow smaller [56, 57].

Taken together, previous work examining the dynamics of trust indicates that the type of rapid updating of trust judgments required by complex and dynamic joint human/AI decision-making environments is not the default for human decision-makers. It appears that humans instead tend to demonstrate hysteresis, meaning that a change in their trust response can lag substantially behind a change in the system's trustworthiness [25]. In rapidly fluctuating contexts, a human trust response that is even only moderately out of sync with the AI system's behavior can potentially lead to snowballing trust miscalibration.

It is important to note additionally that studies of trust in autonomy and/or automation measure trust in different ways, including self-report assessments of subjective trust in the system, or inferring trust from observing participants' decisions to rely on the system. There is evidence that the dynamic responsiveness of subjective reporting of trust in automation may differ from that of behavioral measures of trust [2]. Future work in this area may benefit from assessing both self-report and behavioral measures of trust in AI systems.

Recent work indicates that providing cues to human decision-makers when they exhibit evidence of improperly calibrated trust (whether over- or under-reliance) can help to mitigate the inertia in trust adjustments [34]. The optimal way to represent both this type of cuing information, as well as other potential strategies for enabling humans to appropriately calibrate trust, represents a range of active research areas that we describe more fully in the sections below.

3 Cognitive Strategies for Appropriate Trust Calibration for AI

Prior work has shown that there exists several cognitive biases that can impact human-AI decision-making [5]. In this section, we consider interdisciplinary approaches, building upon the cognitive science, visualization and broader HCI literature to suggest solutions to foster effective trust calibration for AI. We provide guidelines for future research to account for individual differences in cognitive traits, mitigate cognitive biases and implement strategies for complex decision-making.

3.1 Individual Differences for Cognition and AI Trust Calibration

Prior work across various fields has shown that individual differences can impact reasoning and performance across a number of tasks. For example, Liu et al. [27] has reviewed 29 key articles in the visualization literature that altogether demonstrate the effect of 13 individual traits on metrics such as speed, accuracy, eye tracking, mouse data and subjective feedback when interacting with visualizations. When it comes to the effect of individual differences on human-AI interactions, prior work has shown evidence that individual traits can impact self-reported trust in AI [43]. There has also been evidence that traits such as trust in automation [6, 45], Need for Cognition (NFC) [47], locus of control [45], and neuroticism [47] can impact reliance on AI. Moreover, Swaroop et al. demonstrated in their recent work that over-reliance on a block could predict over-reliance on a subsequent block, thus introducing reliance on AI as a potential stable individual difference measure [47].

Appropriately calibrating trust for human-AI decision-making requires an understanding of individual risk perception and risk propensity, especially when the AI accuracy is a moving target. The fields of behavioral economics and cognitive science provide us with a broad understanding of how humans respond to decision-making under risk, typically following patterns of loss aversion, reference dependence, and probability weighting, as outlined in Prospect Theory [19]. These principles suggest that humans tend to overvalue potential losses relative to gains and often misinterpret probabilities, favoring certainty over ambiguity, even when the actual probabilities do not support such biases. Dual-process theory, another prominent decision-making theory, posits that people primarily rely on intuitive reasoning to make decisions, while logical reasoning demands substantial working memory resources [18]. In addition to these tendencies, individuals can have different baselines for risk, which can be measured by self-reported tests such as the general risk propensity scale (GRiPS) [58]. When interacting with AI systems, overly risk-averse individuals may be more likely to underutilize an AI system even when it demonstrates high accuracy, while those with higher risk tolerance might tend to over-trust a system prone to errors. This dynamic becomes even more complex when AI performance fluctuates or when the AI provides probabilistic outputs, as users may struggle to appropriately weigh uncertainty information.

There is a critical need for future research to explore the role of individual differences in human-AI decision-making, particularly in dynamic and uncertain contexts. Understanding how individual traits influence reliance and decision-making is essential for improving trust calibration. As mentioned above, in dynamic environments where uncertainty is high, miscalibrated trust can lead to over-reliance, underutilization, or inefficient collaboration, all of which can compromise performance and outcomes. By examining individual differences, researchers can identify stable traits that predict trust patterns, paving the way for personalized AI systems that adapt to users' cognitive profiles. This line of inquiry is essential to foster more effective, reliable, and resilient human-AI partnerships, ensuring that trust in AI systems is appropriately calibrated for the demands of real-world decision-making.

3.2 Mitigating Cognitive Biases for Joint Human/AI Decision-making

Feedback Mechanisms & Metacognition. Metacognition—the awareness and regulation of one's own cognitive processes—enables individuals to take control of their learning and decision-making across various tasks through self-awareness, self-reflection and self-monitoring [22]. Metacognition has primarily been studied in the context of education, where research has shown that individuals with strong metacognitive skills can outperform individuals with stronger aptitude in academic settings [40, 46]. In HCI and HAI, self-assessment has often been used to measure one's subjective experience during an experience or behavioral task. However, only a few researchers have incorporated metacognitive methods in a closed loop, which has the potential to enable significant improvements in decision-making. For instance, Wall et al. enabled participants in a controlled study to revise decisions after viewing interaction traces that showed how they allocated time and attention across the data, promoting conscious reflection [52].

We encourage researchers to implement real-time trust feedback mechanisms that allow users to monitor their trust alignment across multiple systems. For instance, interaction traces or confidence-level indicators can highlight how users allocate their trust among systems over time, prompting self-reflection on whether their reliance matches system accuracy. This metacognitive reflection can help users adapt their trust dynamically as AI performance changes.

Cognitive Forcing Functions. Cognitive Forcing Functions (CFFs) are interventions which, made at the time of decision-making, disrupt heuristic reasoning and instead encourage individuals to engage in analytical thinking [23]. Buçinca et al. applied this framework to human-AI decision-making, where CFFs may be used to encourage users to think critically rather than passively trust AI suggestions, which can lead to over-reliance [6]. They demonstrated that CFFs lead to better decision-making outcomes by fostering appropriate reliance on AI systems [6]. Some of the cognitive forcing strategies used in the human-AI decision-making literature include:

- Prompting the user to make a decision before seeing the AI's recommendation [14, 6]
- Delaying the presentation of the AI recommendation [38, 6]
- Displaying the AI recommendation on demand [12, 6].

CFFs have been applied to a wide range of practical applications and empirically examined using tasks ranging from generalizable to highly specific. For example, Kunar et al. has examined how to optimally display Computer-Aided Detection (CAD) information to foster better visual search performance, a task highly applicable to medical decision-making [20]. They found that delaying the presentation of the CAD recommendation until after the user has manually searched the display reduced miss errors [20].

In complex and dynamic contexts, metacognitive strategies can be used to help users evaluate AI recommendations critically and adapt their trust dynamically. They may be also be leveraged to foster appropriate trust by introducing cross-system comparison tasks, where users are required to assess and reconcile outputs from multiple AI systems before finalizing a decision. This approach encourages active consideration of discrepancies and system accuracy rather than passive acceptance. Additionally,confidence elicitation prompts can be implemented, asking users to report their confidence in a decision both before and after viewing AI outputs. This step would not only fosters critical thinking but also highlight mismatches between the user's confidence and system accuracy, aiding in trust recalibration. Cognitive forcing strategies have the potential to ensure that users engage thoughtfully with multiple AI systems, dynamically adapting their trust based on observed reliability. Delaying accuracy information or AI recommendations until after an initial decision can prevent over-reliance and encourage users to rely on their own judgment, particularly when interacting with systems of varying performance. When implementing these frameworks and strategies, it is important to note that individual differences and other cognitive biases may mediate interactions. For example, prior work has shown that that CFFs benefited individuals with higher Need for Cognition more compared to their lower Need for Cognition counterparts [6]. We encourage future research to consider joint approaches when evaluating the effect of cognitive factors on human-AI decisions.

3.3 Complex Decision-making

Decision-making is an essential aspect of operating in the world today. Depending on the environment and situation, decision-making can be considered complex. Decisions become complex when there are numerous options or alternatives to consider and where factors are uncertain and multifaceted [33]. Complex decision-making additionally can be time and risk critical, for example in healthcare environments or humanitarian relief efforts. Unlike simple day-to-day decision-making that generally have clear and predictable criteria and alternatives to consider, complex decision-making requires the decision maker to evaluate the risks, uncertainty, trade-offs, consequences, and the dynamicity of the environment.

There are many different algorithms and techniques that can aid in complex decision making, including Multi-Criteria Decision-Making (MCDM). MCDM allow for the evaluation, prioritization, and selection of an ideal alternative in considerations to the criteria that can effect that decision [42]. It is widely studied in operations research and has been applied too many different fields such as healthcare, engineering, and business [42]. Additionally, more research is evaluating hybrid approaches that combine two or more MCDM methods to leverage the strengths of the method while minimizing the limitations [7, 42, 50]. With utilizing MCDM for AI and ML applications, it is important to consider the aspects of complexity to ensure that the ideal solution is appropriate for the environment or situation the decision is taking place.

Trust has a direct impact on decision-making, and becomes even more dire when the levels of complexity and uncertainty are high. When presented with information, the decision-maker needs to determine their trust in the source and while also evaluating all of the factors that make up a decision. If the wrong decision is made (thus creating a consequential or less than ideal result), that can negatively affect the decision-makers trust in the source. Vice versa can be applied, and good results can positively affect the decision-makers trust. Ultimately, there needs to be a level of adaptability when it comes to assessing trust and uncertainty for complex decision-making.

4 Uncertainty Communication

In complex, real-world environments where joint human/AI decision-making is likely to be most beneficial, there is typically substantial uncertainty in the incoming data used by the AI systems, as well as uncertainty in the accuracy of the recommendations provided by the AI based on that data. Some work suggests that providing a representation of the uncertainty in a prediction to the human decision-maker may improve trust calibration (e.g., [48, 58]). In this section we describe in detail the concept of uncertainty of information and some of the challenges associated with representing uncertainty to human users.

4.1 Uncertainty of Information

As stated previously, trust is key for HAI systems to work well and support tasks in a manner that exceeds the ability of the human or the AI alone. In [48] the authors explore AI supported decision making in military coalition operations. They state that "AI systems can help human teammates build suitable mental models by giving explanations of how their outputs were arrived at and estimates of the uncertainty in their outputs"[48]. In this use case the need for forming these human-AI teams may take place and change quickly. In these high risk situations, data collections may be impacted as well as the time for processing the data to make it actionable. Moreover, AI systems may be negatively impacted in these contested environment. Thus, the ability to have appropriately calibrated levels of trust even as the AI is adjusting to the rapid changing tasks and data sources is critical. This is one of the primary motivations of the uncertainty of information concept. Uncertainty of Information [41] is a concept that explores how uncertainty can be represented and communicated to humans and intelligent systems to enhance different tasks, especially decision making. For both the human and the intelligent system, decision making is performed with some level of uncertainty. For AI enabled systems that interact or partner with humans, having a computational model that leverages how humans might categorize uncertainty or what that uncertainty is associated with can aid in the communication and understanding between the two. Uncertainty may come from the model that the intelligent system relies on, usually referred to as epistemic. Uncertainty that may come from the data is usually referred to as aleatoric. These are broad categories which may not convey information of about the uncertainty that can help with the decision. With this in mind the uncertainty of information concept borrows ideas from the human information interaction domain. Specifically, Gershon's work related to the imperfect nature of information [13]. Uncertainty of information uses similar categories and leverages those categories as weight within a computational model to represent that idea.

The categories include corrupt, incomplete, inconsistent, questionable, imperfect, disjoint, imprecise, and inaccurate. Each category can be associated with various sources. For example, the source could come from text, imagery, networks, or devices. For more physical sources their these categories of uncertainty can represent limitations in the data generated or in the data of their state. In this concept the computational model can present and communicate uncertainty as a collection of the contributions from the sources by categories or for each individually. By using this concept, where the uncertainty lies can adjust or modify the outcome based on criteria relevant for the tasks. The uncertainty for specific categories and sources may have more impact or more risk when viewed from this perspective.

Recent work suggests that AI uncertainty quantification (UQ) is critical for joint decision-making [1, 17]. This is an active area of research, with numerous techniques developed for AI UQ, but also many remaining gaps and challenges to achieving reliable and unbiased measures of uncertainty in AI predictions [1]. By incorporating the uncertainty of information concept trust may be better described and understood, enhancing the performance of the human-AI system. This is one way to support the interpretability of AI systems at any stage in the tasks and state of processing. Below we discuss considerations related specifically to the communication and representation of uncertainty to the human user.

4.2 Representation of AI Uncertainty

Maintaining appropriate estimates of an AI or other model's uncertainties is a complex challenge, and as those challenges are addressed, an additional challenge arises: how to communicate these uncertainties to a human teammate. Uncertainty communication typically involves some combination of verbal, numeric, and visual expression [49].

Verbal and written communication of uncertainty often employs verbal probability expressions, such as "possibly" or "probably", but these words do not effectively or precisely convey quantitative uncertainty. A recent review of verbal probability expression and reception found disagreement and variability in how verbal probability statements are mapped to numeric probability, with variation between individuals and contexts [8]. Attempts have been made (e.g. [26, 15]) to provide standardized language to senders of messages about probabilistic events and evidence, but for precise communication, the receiver must understand those standardized expressions in the same way as the sender [28, 53]. When precise communication of uncertainty is important, numeric expressions might be more suitable than verbal expressions [8].

Although numeric expressions of uncertainty or probability have been found to be more precise, it can be difficult to comprehend many numbers at once along with their relationships. This is likely to be the circumstance in human/AI complex decision-making. In those cases, visual expression of uncertainty might be more suitable. Direct experimental comparisons of decision-making with numeric and visual representations of probability or uncertainty are relatively rare, numeric and visual expressions of the same data result in somewhat different decisions [31].

Visual expression of uncertainty of information, prediction, or inference can take many forms, with typical examples depicting a statistical summary, representative examples, or an entire probability distribution [35]. However, experts and non-experts alike can misunderstand uncertainty visualizations [4, 36], leading some authors to omit uncertainty entirely [16]. Generalizable theories of decision-making with visualized uncertainty and where it fails can point the way toward principles for designing and evaluating effective visualization of uncertainty [3, 36]. The idea of a single best approach to visualizing uncertainty might not be attainable [32, 44], as uncertainty visualization effectiveness depends on static and dynamic factors of the task, message, and individual receiver, such that the best visualization might change from task-to-task, moment-to-moment, and person-to-person.

Research evaluating uncertainty visualization often focuses on the effectiveness of the visualization for a user who is assumed to have no special expertise with the visualization. This assumption is reasonable when the audience for a visualization is the general public, as in severe weather warnings, public health messages, or mass media. Although mere exposure to an uncertainty visualization does not necessarily foster understanding of that visualization, there is some evidence that deliberate practice [9] can lead to improvements in task performance when using visualized uncertainty. In an experiment evaluating different visualizations in the uncertainty of the arrival time of a bus in which participants made a choice based on the visualization and then received feedback on their performance, average performance improved over the course of the experiment [10]. Similarly, an experiment in which participants combined multiple visualized uncertain spatial estimates to select the most likely location, performance improved over time with feedback both on the practiced visualization as well as on unpracticed visualizations [21]. These two findings suggest that iterated attempts with feedback can improve task performance, but they leave unclear to what extent participants were learning to use an uncertainty visualization vs. improving their strategy or ability to reason over the relevant problem. Follow-up work on the spatial estimate combination work found that participants who practiced that task with one visualization performed another spatial estimate combination task using a much different visual context much faster (albeit at equivalent accuracy) relative to participants who had not practiced [11].

A cognitively inspired model of decision making with visualized uncertainty [39, 36] identifies multiple stages of information processing and reasoning in arriving at a decision. The first stages involve extracting information from the visual array by way of an instantiated graph schema, followed by formulating a conceptual question and inferring an answer based on information extracted from the graph schema. The effects of practice could improve any or all of these processes. The findings that practice with one task improve performance on another, closely related task using a very different visual array suggest that practice can, at least in some cases, improve steps beyond extracting information from the visual array. This is good news for the potential of uncertainty communication to help people appropriately rely on AI teammates, because although some practice might be required initially, that practice could accumulate skill in some generalized ability to use communicated uncertainty to weight AI input and understand how that uncertainty bears on a given decision.

Challenges to effectively conveying uncertainty are numerous. Research is needed to better understand and develop guidance for developers of AI decision aids to implement effective uncertainty communication. Although recent work suggests that adding visualizations of AI UQ information to AI predictions can improve both decision-making accuracy and confidence calibration [30], it might be too soon to judge whether uncertainty communication is sufficient to induce calibrated trust or appropriate reliance [48, 51]. As we shift focus from 'how much should I trust this AI' to 'to what extent can I rely on this AI in this situation relative to my other sources of information', additional research to understand how to effectively convey uncertainty alongside research to understand how individual differences, cognitive factors, and the role of training and practice in probabilistic decision making will be critical to enabling effective joint human/AI decision-making.

5 Conclusion

In this paper we describe one challenge associated with effective joint human/AI decision-making in complex and rapidly changing environments; namely, how can appropriate trust calibration be maintained among multiple dynamic AI systems. We review relevant literature on the dynamics of trust calibration, highlighting the previous focus on "snapshots" of trust calibration and the recent calls to assess trust calibration instead as a dynamic process. We discuss lessons from cognitive science that may be applied to the design of AI systems to take into account individual differences and to mitigate cognitive biases. Finally, we describe the concept of uncertainty of information and review research exploring how uncertainty might best be conveyed to human decision-makers.

From these discussions, we derive a number of recommendations for both design and research in this area. First, we recommend further empirical work on the dynamics of trust in AI that explores the human trust response to multiple systems of varying reliability; how people weight the input of the different systems based on observed performance, both in terms of their behavior/reliance on the systems, as well as their subjective reporting of trust levels. We also encourage research that considers individual differences in cognitive traits such as trust in automation and risk propensity, for personalized and adaptive AI systems that may more effectively promote trust calibration across different users. Similarly, it may be beneficial, where possible, to incorporate real-time trust feedback mechanisms that allow users to monitor their reliance behaviors and trust calibration across systems. Researchers should consider implementing cognitive forcing functions to promote continuous engagement and assessment of the reliability of AI systems, rather adopting a more fixed and passive trust strategy. Finally, we echo calls to improve uncertainty/confidence estimates in AI systems, and to take into consideration the tradeoffs of different representations of uncertainty for communicating this information to the human user. Moreover, we encourage evaluations that include uncertainty representations to make use of iterated practice with feedback to avoid the pitfall of mistaking participant inexperience with inability to reason with uncertainty.

As AI systems become more capable, they will be deployed in more complex and challenging contexts. To avoid deploying AI systems in challenging context would leave unaided humans at a potential disadvantage, but not one so disastrous as the disadvantage from using AI systems inappropriately. In this paper, we have laid out research directions we view as critical to unlocking a future in which AI systems can be used as part of joint human/AI decision-making in which the humans are able to appropriately rely on AI systems by maintaining well-calibrated trust in the face of dynamic and uncertain environments.

References

1. Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., Makarenkov, V., Nahavandi, S.: A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Information Fusion **76**, 243–297 (2021). https://doi.org/10.1016/j.inffus.2021.05.008
2. Ayoub, J., Avetisyan, L., Makki, M., Zhou, F.: An investigation of drivers' dynamic situational trust in conditionally automated driving. IEEE Transactions on Human-Machine Systems **52**(3), 501–511 (2021). https://doi.org/10.1109/THMS.2021.3131676
3. Bancilhon, M., Padilla, L., Ottley, A.: Improving Evaluation Using Visualization Decision-Making Models: A Practical Guide. In: Albers Szafir, D., Borgo, R., Chen, M., Edwards, D.J., Fisher, B., Padilla, L. (eds.) Visualization Psychology, pp. 85–107. Springer International Publishing, Cham (2023). https://doi.org/10.1007/978-3-031-34738-2_4
4. Belia, S., Fidler, F., Williams, J., Cumming, G.: Researchers Misunderstand Confidence Intervals and Standard Error Bars. Psychological Methods **10**(4), 389–396 (2005). https://doi.org/10.1037/1082-989X.10.4.389
5. Bertrand, A., Belloum, R., Eagan, J.R., Maxwell, W.: How cognitive biases affect xai-assisted decision-making: A systematic review. In: Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society. pp. 78–91 (2022). https://doi.org/10.1145/3514094.353416
6. Buçinca, Z., Malaya, M.B., Gajos, K.Z.: To trust or to think: Cognitive forcing functions can reduce overreliance on ai in ai-assisted decision-making. Proc. ACM Hum.-Comput. Interact. **5**(CSCW1) (Apr 2021). https://doi.org/10.1145/3449287

7. Caylor, J.P., Hammell, R.J., Raglin, A.J.: Preliminary evaluation of multi-criteria decision-making methodology for emergency management. In: Nunes, I.L. (ed.) Advances in Human Factors and System Interactions. pp. 11–18. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-79816-1_2
8. Dhami, M.K., Mandel, D.R.: Communicating uncertainty using words and numbers. Trends in Cognitive Sciences **26**(6), 514–526 (Jun 2022). https://doi.org/10.1016/j.tics.2022.03.002
9. Ericsson, K.A.: Deliberate Practice and Acquisition of Expert Performance: A General Overview. Academic Emergency Medicine **15**(11), 988–994 (Nov 2008). https://doi.org/10.1111/j.1553-2712.2008.00227.x
10. Fernandes, M., Walls, L., Munson, S., Hullman, J., Kay, M.: Uncertainty Displays Using Quantile Dotplots or CDFs Improve Transit Decision-Making. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 1–12. ACM, Montreal QC Canada (Apr 2018). https://doi.org/10.1145/3173574.3173718
11. Files, B., Oiknine, A.H., Raber, T., Dalangin, B., Pollard, K.: Practice with uncertainty integration improves performance on a qualitatively different task and with new visualizations. Journal of Visualization and Interaction (Sep 2024). https://doi.org/10.54337/jovi.v1i1.7971
12. Fitzsimons, G.J., Lehmann, D.R.: Reactance to recommendations: When unsolicited advice yields contrary responses. Marketing Science **23**(1), 82–94 (2004). https://doi.org/10.1287/mksc.1030.0033
13. Gershon, N.: Visualization of an imperfect world. IEEE Computer Graphics and Applications **18**(4), 43–45 (1998). https://doi.org/10.1109/38.689662
14. Green, B., Chen, Y.: The principles and limits of algorithm-in-the-loop decision making. Proceedings of the ACM on Human-Computer Interaction **3**(CSCW), 1–24 (2019). https://doi.org/10.1145/3359152
15. Headquarters, Department of the Army: Army techniques publication 2-33.4: Intelligence analysis (Jan 2020), `https://armypubs.army.mil/ProductMaps/PubForm/Details_Printer.aspx?PUB_ID=1008410`
16. Hullman, J.: Why Authors Don't Visualize Uncertainty. IEEE Transactions on Visualization and Computer Graphics **26**(1), 130–139 (Jan 2020). https://doi.org/10.1109/TVCG.2019.2934287
17. Jalaian, B., Lee, M., Russell, S.: Uncertain context: Uncertainty quantification in machine learning. AI Magazine **40**(4), 40–49 (2019). https://doi.org/10.1609/aimag.v40i4.4812
18. Kahneman, D.: Thinking, Fast and Slow. Farrar, Straus and Giroux (2011)
19. Kahneman, D., Tversky, A.: Prospect theory: An analysis of decision under risk. Econometrica **47**(2), 263–291 (1979). https://doi.org/10.2307/1914185
20. Kunar, M.A.: The optimal use of computer aided detection to find low prevalence cancers. Cognitive Research: Principles and Implications **7**(1), 13 (2022). https://doi.org/10.1186/s41235-022-00361-1
21. Kusumastuti, S.A., Pollard, K.A., Oiknine, A.H., Dalangin, B., Raber, T.R., Files, B.T.: Practice Improves Performance of a 2D Uncertainty Integration Task Within and Across Visualizations. IEEE Transactions on Visualization and Computer Graphics **29**(9), 3949–3960 (Sep 2023). https://doi.org/10.1109/TVCG.2022.3173889
22. Lai, E.R.: Metacognition: A literature review. Always learning: Pearson research report **24**, 1–40 (2011)

23. Lambe, K.A., O'Reilly, G., Kelly, B.D., Curristan, S.: Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review. BMJ quality & safety **25**(10), 808–820 (2016). https://doi.org/10.1136/bmjqs-2015-004417
24. Lee, J., Moray, N.: Trust, control strategies and allocation of function in human-machine systems. Ergonomics **35**(10), 1243–1270 (1992). https://doi.org/10.1080/00140139208967392
25. Lee, J.D., See, K.A.: Trust in automation: Designing for appropriate reliance. Human Factors **46**(1), 50–80 (2004). https://doi.org/10.1518/hfes.46.1.50_30392
26. Lewis, S.C., King, A.D., Perkins-Kirkpatrick, S.E., Wehner, M.F.: Toward Calibrated Language for Effectively Communicating the Results of Extreme Event Attribution Studies. Earth's Future **7**(9), 1020–1026 (Sep 2019). https://doi.org/10.1029/2019EF001273
27. Liu, Z., Crouser, R.J., Ottley, A.: Survey on individual differences in visualization. In: Computer Graphics Forum. vol. 39, pp. 693–712. Wiley Online Library (2020). https://doi.org/10.1111/cgf.14033
28. Mandel, D.R., Irwin, D.: On measuring agreement with numerically bounded linguistic probability schemes: A re-analysis of data from Wintle, Fraser, Wills, Nicholson, and Fidler (2019). PLOS ONE **16**(3), e0248424 (Mar 2021). https://doi.org/10.1371/journal.pone.0248424
29. Manzey, D., Reichenbach, J., Onnasch, L.: Human performance consequences of automated decision aids: The impact of degree of automation and system experience. Journal of Cognitive Engineering and Decision Making **6**(1), 57–87 (2012). https://doi.org/10.1177/1555343411433844
30. Marusich, L.R., Bakdash, J.Z., Zhou, Y., Kantarcioglu, M.: Using ai uncertainty quantification to improve human decision-making. In: Proceedings of the 41st International Conference on Machine Learning. ICML'24, JMLR.org (2024). https://doi.org/10.31234/osf.io/6syfv
31. Matzen, L.E., Howell, B.C., Trumbo, M.C.S., Divis, K.M.: Numerical and Visual Representations of Uncertainty Lead to Different Patterns of Decision Making. IEEE Computer Graphics and Applications **43**(5), 72–82 (Sep 2023). https://doi.org/10.1109/MCG.2023.3299875
32. Nobre, C., Zhu, K., Mörth, E., Pfister, H., Beyer, J.: Reading Between the Pixels: Investigating the Barriers to Visualization Literacy. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. pp. 1–17. ACM, Honolulu HI USA (May 2024). https://doi.org/10.1145/3613904.3642760
33. Nyarko, K., Taiwo, P., Nwachukwu, K., Rawal, J., Richardson, J., Raglin, A.: A generalized framework for multicriteria decision making under uncertainty. In: Schwartz, P.J., Jensen, B., Hohil, M.E. (eds.) Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications VI. vol. 13051, p. 130510L. International Society for Optics and Photonics, SPIE (2024). https://doi.org/10.1117/12.3015949
34. Okamura, K., Yamada, S.: Empirical evaluations of framework for adaptive trust calibration in human-ai cooperation. IEEE Access **8**, 220335–220351 (2020). https://doi.org/10.1109/ACCESS.2020.3042556
35. Padilla, L., Kay, M., Hullman, J.: Uncertainty Visualization. In: Piegorsch, W.W., Levine, R.A., Zhang, H.H., Lee, T.C.M. (eds.) Computational statistics in data science, pp. 405–421. Wiley, Hoboken, NJ (2022). https://doi.org/10.1002/9781118445112.stat08296
36. Padilla, L.M., Creem-Regehr, S.H., Hegarty, M., Stefanucci, J.K.: Decision making with visualizations: a cognitive framework across disciplines. Cognitive Research:

Principles and Implications **3**(1), 29 (Jul 2018). https://doi.org/10.1186/s41235-018-0120-9
37. Parasuraman, R., Riley, V.: Humans and Automation: Use, Misuse, Disuse, Abuse. Human Factors: The Journal of the Human Factors and Ergonomics Society **39**(2), 230–253 (Jun 1997). https://doi.org/10.1518/001872097778543886
38. Park, J.S., Barber, R., Kirlik, A., Karahalios, K.: A slow algorithm improves users' assessments of the algorithm's accuracy. Proceedings of the ACM on Human-Computer Interaction **3**(CSCW), 1–15 (2019). https://doi.org/10.1145/3359204
39. Pinker, S.: A theory of graph comprehension. In: Artificial intelligence and the future of testing., pp. 73–126. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US (1990)
40. Prins, F.J., Veenman, M.V., Elshout, J.J.: The impact of intellectual ability and metacognition on learning: New support for the threshold of problematicity theory. Learning and Instruction **16**(4), 374–387 (2006). https://doi.org/10.1016/j.learninstruc.2006.07.008, https://doi.org/10.1016/j.learninstruc.2006.07.008
41. Raglin, A., Metu, S., Lott, D.: Challenges of simulating uncertainty of information. In: International Conference on Human-Computer Interaction. pp. 255–261. Springer (2020). https://doi.org/10.1007/978-3-030-60700-5_33
42. Rawal, A., Rawal, J., Raglin, A.: Multi-criteria decision making (mcdm) with causal reasoning for ai/ml applications – a survey. In: Degen, H., Ntoa, S. (eds.) Artificial Intelligence in HCI. pp. 439–448. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-60611-3_30
43. Riedl, R.: Is trust in artificial intelligence systems related to user personality? review of empirical evidence and future research directions. Electronic Markets **32**(4), 2021–2051 (2022). https://doi.org/10.1007/s12525-022-00594-4
44. Sarma, A., Long, S., Correll, M., Kay, M.: Tasks and Telephones: Threats to Experimental Validity due to Misunderstandings of Visualisation Tasks and Strategies Position Paper. In: 2024 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV). pp. 33–40. IEEE, St Pete Beach, FL, USA (Oct 2024). https://doi.org/10.1109/BELIV64461.2024.00009
45. Sharan, N.N., Romano, D.M.: The effects of personality and locus of control on trust in humans versus artificial intelligence. Heliyon **6**(8), e04572 (2020). https://doi.org/10.1016/j.heliyon.2020.e04572
46. Swanson, H.L.: Influence of metacognitive knowledge and aptitude on problem solving. Journal of educational psychology **82**(2), 306 (1990). https://doi.org/10.1037//0022-0663.82.2.306
47. Swaroop, S., Buçinca, Z., Gajos, K.Z., Doshi-Velez, F.: Accuracy-time tradeoffs in ai-assisted decision making under time pressure. In: Proceedings of the 29th International Conference on Intelligent User Interfaces. p. 138–154. IUI '24, Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3640543.3645206
48. Tomsett, R., Preece, A., Braines, D., Cerutti, F., Chakraborty, S., Srivastava, M., Pearson, G., Kaplan, L.: Rapid Trust Calibration through Interpretable and Uncertainty-Aware AI. Patterns **1**(4) (Jul 2020). https://doi.org/10.1016/j.patter.2020.100049
49. Van Der Bles, A.M., Van Der Linden, S., Freeman, A.L.J., Mitchell, J., Galvao, A.B., Zaval, L., Spiegelhalter, D.J.: Communicating uncertainty about facts, numbers and science. Royal Society Open Science **6**(5), 181870 (May 2019). https://doi.org/10.1098/rsos.181870

50. Velasquez, M., Hester, P.: An analysis of multi-criteria decision making methods. International Journal of Operations Research **10**(2), 56–66 (2013)
51. Wall, E., Matzen, L., El-Assady, M., Masters, P., Hosseinpour, H., Endert, A., Borgo, R., Chau, P., Perer, A., Schupp, H., Strobelt, H., Padilla, L.: Trust Junk and Evil Knobs: Calibrating Trust in AI Visualization. In: 2024 IEEE 17th Pacific Visualization Conference (PacificVis). pp. 22–31. IEEE, Tokyo, Japan (Apr 2024). https://doi.org/10.1109/PacificVis60374.2024.00012
52. Wall, E., Narechania, A., Coscia, A., Paden, J., Endert, A.: Left, Right, and Gender: Exploring Interaction Traces to Mitigate Human Biases . IEEE Transactions on Visualization & Computer Graphics **28**(01), 966–975 (Jan 2022). https://doi.org/10.1109/TVCG.2021.3114862
53. Wintle, B.C., Fraser, H., Wills, B.C., Nicholson, A.E., Fidler, F.: Verbal probabilities: Very likely to be somewhat more confusing than numbers. PLOS ONE **14**(4), e0213522 (Apr 2019). https://doi.org/10.1371/journal.pone.0213522
54. Wischnewski, M., Krämer, N., Müller, E.: Measuring and understanding trust calibrations for automated systems: A survey of the state-of-the-art and future directions. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. CHI '23, Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3544548.3581197
55. Yang, X.J., Guo, Y., Schemanske, C.: From trust to trust dynamics: Combining empirical and computational approaches to model and predict trust dynamics in human-autonomy interaction. In: Duffy, V.G., Landry, S.J., Lee, J.D., Stanton, N. (eds.) Human-Automation Interaction: Transportation, pp. 253–265. Springer International Publishing, Cham (2023). https://doi.org/10.1007/978-3-031-10784-9_15
56. Yang, X.J., Unhelkar, V.V., Li, K., Shah, J.A.: Evaluating effects of user experience and system transparency on trust in automation. In: Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction. pp. 408–416 (2017). https://doi.org/10.1145/2909824.3020230
57. Yu, K., Berkovsky, S., Taib, R., Conway, D., Zhou, J., Chen, F.: User trust dynamics: An investigation driven by differences in system performance. In: Proceedings of the 22nd International Conference on Intelligent User Interfaces. p. 307–317. IUI '17, Association for Computing Machinery, New York, NY, USA (2017). https://doi.org/10.1145/3025171.3025219
58. Zhang, D.C., Highhouse, S., Nye, C.D.: Development and validation of the general risk propensity scale (grips). Journal of Behavioral Decision Making **32**(2), 152–167 (2019). https://doi.org/10.1002/bdm.2102