Published: 09 December 2025
Online AM: 26 August 2025
Accepted: 19 July 2025
Revised: 13 June 2025
Received: 03 January 2025

RESEARCH-ARTICLE

Balancing the Unknown: Exploring Human Reliance on AI Advice under Aleatoric and Epistemic Uncertainty

JOSHUA HOLSTEIN, Karlsruhe Institute of Technology, Karlsruhe, Baden-Wurttemberg, Germany

LARS BÖCKING, University of Bayreuth, Bayreuth, Bayern, Germany

PHILIPP SPITZER, Karlsruhe Institute of Technology, Karlsruhe, Baden-Wurttemberg, Germany

NIKLAS KÜHL, University of Bayreuth, Bayreuth, Bayern, Germany

MICHAEL VÖSSING, Karlsruhe Institute of Technology, Karlsruhe, Baden-Wurttemberg, Germany

GERHARD SATZGER, Karlsruhe Institute of Technology, Karlsruhe, Baden-Wurttemberg, Germany

Open Access Support provided by:

Karlsruhe Institute of Technology

University of Bayreuth

ACM Transactions on Computer-Human Interaction, Volume 32, Issue 6 (December 2025)
https://doi.org/10.1145/3762813
EISSN: 1557-7325

Balancing the Unknown: Exploring Human Reliance on AI Advice under Aleatoric and Epistemic Uncertainty

JOSHUA HOLSTEIN, Institute for Information Systems, Karlsruhe Institute of Technology, Karlsruhe, Germany
LARS BÖCKING, Information Systems and Human-Centric Artificial Intelligence, University of Bayreuth, Bayreuth, Germany and Branch Business & Information Systems Engineering, Fraunhofer Institute for Applied Information Technology FIT, Sankt Augustin, Germany
PHILIPP SPITZER, Institute for Information Systems, Karlsruhe Institute of Technology, Karlsruhe, Germany
NIKLAS KÜHL, Information Systems and Human-Centric Artificial Intelligence, University of Bayreuth, Bayreuth, Germany and Branch Business & Information Systems Engineering, Fraunhofer Institute for Applied Information Technology FIT, Sankt Augustin, Germany
MICHAEL VÖSSING and GERHARD SATZGER, Institute for Information Systems, Karlsruhe Institute of Technology, Karlsruhe, Germany

Artificial intelligence (AI) systems increasingly support decision-making across a broad range of domains. The complexity of real-world tasks, however, introduces uncertainty into the prediction capabilities of these systems. This uncertainty can manifest as aleatoric uncertainty arising from inherent variability in outcomes or epistemic uncertainty stemming from limitations in the AI system's knowledge. While prior research has investigated uncertainty as a monolithic concept, the distinct effects of communicating aleatoric or epistemic uncertainty on humans and their reliance behavior remain unexplored. In this work, we present two behavioral experiments that systematically examine how participants rely on AI advice when faced with different types of uncertainty. While the first experiment manipulates the source of uncertainty, specifying it as either aleatoric or epistemic, the second decomposes uncertainty into its individual components, presenting aleatoric and epistemic uncertainty simultaneously. This work contributes to a deeper understanding of the multifaceted impact of different uncertainty types on human–AI interaction.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; • **Computing methodologies → Artificial intelligence**;

Lars Böcking and Philipp Spitzer contributed equally to this research.

Authors' Contact Information: Joshua Holstein (corresponding author), Institute for Information Systems, Karlsruhe Institute of Technology, Karlsruhe, Germany; e-mail: joshua.holstein@kit.edu; Lars Böcking, Information Systems and Human-Centric Artificial Intelligence, University of Bayreuth, Bayreuth, Germany and Branch Business & Information Systems Engineering, Fraunhofer Institute for Applied Information Technology FIT, Sankt Augustin, Germany; e-mail: lars.boecking@fim-rc.edu; Philipp Spitzer, Institute for Information Systems, Karlsruhe Institute of Technology, Karlsruhe, Germany; e-mail: philipp.spitzer@kit.edu; Niklas Kühl, Information Systems and Human-Centric Artificial Intelligence, University of Bayreuth, Bayreuth, Germany and Branch Business & Information Systems Engineering, Fraunhofer Institute for Applied Information Technology FIT, Sankt Augustin, Germany; e-mail: kuehl@uni-bayreuth.de; Michael Vössing, Institute for Information Systems, Karlsruhe Institute of Technology, Karlsruhe, Germany; e-mail: michael.voessing@kit.edu; Gerhard Satzger, Institute for Information Systems, Karlsruhe Institute of Technology, Karlsruhe, Germany; e-mail: gerhard.satzger@kit.de.

ACM 1557-7325/2025/12-ART64
https://doi.org/10.1145/3762813

Additional Key Words and Phrases: Reliance, Uncertainty, Epistemic Uncertainty, Aleatoric Uncertainty, Human-AI Collaboration, AI-Assisted Decision-Making

ACM Reference format:
Joshua Holstein, Lars Böcking, Philipp Spitzer, Niklas Kühl, Michael Vössing, and Gerhard Satzger. 2025. Balancing the Unknown: Exploring Human Reliance on AI Advice under Aleatoric and Epistemic Uncertainty. *ACM Trans. Comput.-Hum. Interact.* 32, 6, Article 64 (December 2025), 34 pages.
https://doi.org/10.1145/3762813

1 Introduction

Imagine you are about to purchase a luxury urban home and seek advice from your aunt, a real estate agent, who is familiar with pricing homes in rural areas. She expresses uncertainty about the value of the property you are interested in and provides a broad price range rather than a specific estimate. To what extent would you rely on your aunt's advice, given the uncertainty expressed via the price range? Would your reliance change if she attributed her uncertainty either to a lack of experience with luxury urban properties like yours or to the general variability in housing prices?

When making decisions subject to uncertainty, individuals frequently seek guidance from experts or other sources they deem knowledgeable [37, 81, 82], such as your aunt. The degree to which individuals then rely on the provided advice typically carries over from their perception of the advisor's reliability [13, 104]. This, in turn, is closely tied to the uncertainty (or confidence) expressed by the advisor [76]. Whereas this uncertainty is usually not further decomposed, it actually may arise from two different sources—from the limitations in the advisor's knowledge (epistemic uncertainty) and the inherent variability in the outcomes being predicted (aleatoric uncertainty) [10, 31, 39]. In the home-buying example, your aunt's lack of experience with luxury urban homes represents epistemic uncertainty, while the inherent variability in observable real estate prices introduces aleatoric uncertainty.

The growing capabilities of **artificial intelligence (AI)** across a broad range of tasks result in humans increasingly seeking advice from AI [81, 82]. As with human advisors, the reliability of AI advice can vary due to accompanying uncertainty. While insufficient training data similar to the current instance can result in epistemic uncertainty, the inherent variability of the observed process can introduce aleatoric uncertainty [39]. Therefore, quantifying the AI's uncertainties and presenting them alongside the AI advice may open up new possibilities for supporting human–AI decision-making processes [10, 34] by allowing humans to calibrate their reliance on AI advice [61, 82] and better interpret model outputs [98]. As a consequence, it is important to understand how different sources of uncertainty affect humans' reliance on AI advice.

However, reliance on advice under uncertainty is not solely determined by the source and degree of uncertainty expressed. Cognitive biases, such as prior beliefs, often influence how individuals process and incorporate uncertain information [43, 96]. These prior beliefs, i.e., individuals' existing knowledge, expectations, and judgments that are formed even before receiving advice, play a crucial role [2, 77]. In the housing example, an existing belief about the property's value may anchor the final estimate, potentially leading to varying reliance on the aunt's advice. While the influence of prior beliefs has been demonstrated across several AI-assisted decision-making scenarios [2, 16, 77], their interaction with uncertainty and its different sources remains underexplored. Yet, understanding this interaction is crucial for designing AI systems that can effectively communicate uncertainty while accounting for humans' cognitive biases.

Previous research in AI-assisted decision-making has already explored various aspects of uncertainty, including its impact on trust, reliance, and decision quality [6, 15, 26, 51, 56, 62, 63, 106, 107] as well as the effectiveness of different visualization techniques [62]. However, these studies have typically focused on presenting a "monolithic" view of uncertainty: Either aleatoric or epistemic uncertainty was considered in isolation [6, 15, 26, 106], or the overall uncertainty was not decomposed into its epistemic and aleatoric components [62, 63]. Thus, human decision-makers were left with incomplete information, limiting their ability to calibrate their reliance on AI advice [61, 82].

Despite the potential benefits of differentiating and presenting both sources of uncertainty, the distinct impacts of aleatoric and epistemic uncertainty on human reliance behavior remain underexplored [10, 34, 52]. The question of whether and how humans assimilate information about aleatoric and epistemic uncertainty and how this information might impact their reliance behavior remains unanswered. Given the importance of understanding how humans interpret and rely on different sources of uncertainty for designing effective AI systems [25, 34], addressing this gap is crucial. Therefore, we formulate the following **research questions (RQs)**:

RQ1: How does human reliance on uncertain AI advice change when the uncertainty is specified as either epistemic or aleatoric?

RQ2: How does human reliance on uncertain AI advice change when the uncertainty is decomposed into its aleatoric and epistemic components?

RQ3: How does human reliance on uncertain AI advice differ between aleatoric and epistemic uncertainty when the overall uncertainty is decomposed into its components?

RQ4: How do prior beliefs influence human reliance on uncertain AI advice across different sources and degrees of uncertainty (aleatoric vs. epistemic)?

To address these RQs, we conduct two behavioral experiments using a real estate price estimation task. Our first experiment examines how framing uncertainty as either aleatoric or epistemic influences human reliance on AI advice. Building on these findings, our second experiment explores the effects when uncertainty is decomposed into its aleatoric and epistemic components. Across both experiments, we investigate how prior beliefs interact with different sources and degrees of uncertainty to shape reliance behavior. Our study contributes to the field of **human–computer interaction (HCI)** and AI-assisted decision-making in four ways: First, we provide empirical evidence to understand how humans digest different sources of uncertainty when relying on AI advice; second, we explore the strategies adopted by decision-makers when relying on varying degrees and sources of uncertainty; third, we examine how prior beliefs interact with uncertainty source and degree to influence reliance behavior; and, fourth, we apply these insights to derive practical design implications for AI-assisted decision-making systems to communicate uncertainty effectively.

The remainder of this article is structured as follows: In Section 2, we establish the foundations of uncertainty and review related work on uncertain AI advice in human–AI decision-making. We then develop our research model in Section 3 and detail the procedures and results of our two experiments in Sections 4 and 5. Subsequently, we analyze and interpret our findings, discuss implications, and suggest future research directions in Section 6. Finally, we conclude in Section 7.

2 Background and Related Work

Next, we provide an overview of the foundations of uncertainty and its sources. Afterward, we cover the related work on uncertainty in human–AI decision-making.

2.1 Foundations of Uncertainty

Uncertainty is an inherent aspect of many decision-making contexts [76, 81] and refers to the lack of knowledge or comprehensive information about a situation or outcome [10]. This concept is closely related to, yet distinct from, confidence [76, 88, 90]. While confidence pertains to the degree of belief that a given prediction is correct, uncertainty encompasses beliefs about the range of possible outcomes [73].

The sources of uncertainty in advice can be categorized into two main types: aleatoric and epistemic uncertainty [10, 39]. *Aleatoric uncertainty*, also known as statistical or data uncertainty [31, 62, 97], stems from the inherent randomness or variability in the system being observed or predicted [10]. In the housing price prediction example, aleatoric uncertainty might arise from random variations (like market fluctuations or variations in buyer preferences) or from variability in factors that are not captured in the data (e.g., unique features of an individual house being priced). For a fixed set of variables, this source of uncertainty cannot be reduced by gathering more data, i.e., more examples [31]. However, adding additional relevant variables could potentially reduce this source of uncertainty [39]. *Epistemic uncertainty*, also referred to as systematic or knowledge uncertainty [30, 97], arises from insufficient training data (i.e., examples) or overly simplistic models that limit the ability to learn the underlying relationships. When data is scarce or models lack complexity, advice becomes less reliable due to gaps in knowledge [39]. In the context of real estate prices, this might include uncertainty due to limited historical data on similar houses or limitations in the model used, i.e., models with too many or too few parameters. Unlike aleatoric uncertainty, epistemic uncertainty can potentially be reduced by collecting more examples, but generally increases with the addition of variables [39].

The computation of epistemic and aleatoric uncertainty in AI models requires specialized techniques to provide accurate estimates of both components. Several methodological frameworks have emerged to address this challenge, ranging from classical probabilistic approaches to modern deep learning techniques [30, 50]. These methods typically involve explicitly modeling uncertainty distributions or employing architectural modifications that enable uncertainty estimation. Specific implementations vary by context: Classification models might leverage probability distributions over class predictions [33], regression models often estimate prediction intervals or output distributions [85], and Bayesian approaches model parameter uncertainties directly [33]. For neural networks, techniques include dropout-based sampling, ensemble methods, or specialized architecture designs with dedicated uncertainty estimation components [30, 33, 50]. However, raw uncertainty estimates often require additional calibration to ensure their reliability through post-processing methods or specialized loss functions during training [33, 85].

Understanding the concepts of uncertainty is critical for decision-making. By recognizing the presence and nature of uncertainty, decision-makers can make more informed choices by acknowledging the limitations of available information and the range of possible outcomes [10]. However, how humans perceive and process uncertainty can significantly impact their decision-making abilities.

2.2 Human Perception and Processing of Uncertainty

Human perception and processing of uncertainty are fundamentally constrained by cognitive limitations and systematic biases [96]. Individuals generally exhibit an aversion to uncertainty, preferring known probabilities over ambiguous ones [24, 28], and often struggle with interpreting statistical and probabilistic information [10, 32, 47]. This can manifest as either overconfidence or underconfidence depending on their perceived expertise [65].

A particularly influential factor in uncertainty processing is the role of prior beliefs—knowledge structures established through past experiences that serve as cognitive anchors when evaluating new information [2, 36, 77]. These prior beliefs create systematic biases in how individuals process uncertain information [67, 96]. When confronted with uncertainty, people tend to rely on these prior beliefs as reference points, often weighing information that confirms their existing beliefs more heavily while discounting contradictory evidence [67]. This manifests through mechanisms such as anchoring, where initial judgments disproportionately influence final decisions, and selective attention to confirming evidence [22, 67, 96]. These biases together often lead to selective interpretation of advice and consequentially suboptimal decisions [9, 65, 96].

To address these cognitive challenges, researchers have developed various approaches for communicating uncertainty more effectively. These include verbal descriptions [4, 12], visual representations such as probability distributions, confidence intervals, and graphical displays [20, 27, 40, 49, 58, 79], and hybrid approaches that combine multiple modalities [68]. The overarching goal of these methods is to facilitate informed decision-making and appropriately calibrate trust in prediction systems [10, 21, 70, 106].

Recent empirical work has begun to examine how these communication strategies interact with the outlined cognitive factors. Prabhudesai et al. [75] provided qualitative evidence that displaying uncertainty information can prompt humans to shift from heuristic-based to more analytical reasoning processes. Building on this, Cao et al. [16] demonstrated that presenting uncertainty through probability scores can help mitigate the influence of prior beliefs and reduce confirmation bias. However, this emerging research has primarily focused on monolithic representations of uncertainty. One interesting exception is from Padilla et al. [68], who demonstrated that humans can simultaneously process multiple sources of uncertainty, suggesting that humans may be capable of distinguishing between different sources of uncertainty when appropriately presented. This finding points toward important questions about how different uncertainty sources might interact with cognitive biases like those involving prior beliefs.

These advances in uncertainty communication represent significant progress in bridging complex probabilistic information with human cognition. However, a critical gap remains in understanding how different sources of uncertainty (aleatoric versus epistemic) are processed and interact with humans' prior beliefs to influence reliance behavior, which warrants further exploration.

2.3 Uncertainty in AI-Assisted Decision Making

The discussed cognitive challenges and biases in uncertainty processing become particularly interesting in AI-assisted decision-making contexts, where the opacity of AI systems compounds existing human limitations. Central to these challenges is how humans exhibit reliance behavior, the observable action of following AI advice [82]. This behavior directly determines decision quality, yet humans often exhibit overreliance on AI advice, assuming higher accuracy than warranted, or under-reliance, dismissing correct AI advice due to mistrust or misunderstanding of the system's capabilities [82].

Understanding and measuring reliance requires domain-specific approaches. In classification tasks, reliance is often operationalized as binary agreement with AI recommendations or the frequency of switching initial decisions to align with AI advice [82]. For regression tasks, researchers measure the **weight of advice (WOA)**, which quantifies how much humans adjust their initial estimates toward AI predictions [38, 44, 46]. Prior research has revealed that reliance emerges from complex interactions between individual factors such as self-confidence [57, 94], contextual elements like task difficulty [89], and procedural aspects including whether decisions follow one-stage or two-stage paradigms [82, 84]. Particularly relevant are the prior beliefs discussed earlier, which create systematic biases through confirmation bias and anchoring effects, leading humans to more readily accept advice aligning with existing beliefs while discounting contradictory AI suggestions [6].

Recent research has begun exploring how different presentations of AI uncertainty can address these reliance challenges, examining effects on human confidence [17, 35], trust and reliance [5, 6, 15, 16, 18, 51, 56, 63, 78, 106, 107], and decision quality [26, 51, 62, 80, 91]. However, these studies have yielded mixed results, highlighting the complexity of uncertainty communication in AI systems. For instance, Cau et al. [17] showed that human uncertainty affects reliance on uncertain AI advice. Kim et al. [51] further found that displaying model uncertainty in large language models reduced participants' reliance on AI while improving their performance, suggesting that uncertainty information can promote more critical engagement with AI outputs. Similarly, Zhang et al. [106] demonstrated that showing AI confidence scores resulted in calibrated trust, though without significant performance improvements. In contrast, Marusich et al. [62] showed that providing uncertainty information significantly improved performance across three experiments, while Vasconcelos et al. [100] found that visualizing uncertainty through highlighting potentially erroneous tokens improved programmer efficiency, whereas highlighting based on generation uncertainty showed no benefits.

Despite these advances, a critical limitation persists: Most studies have focused on presenting only one source of uncertainty or aggregated uncertainty [62], potentially limiting humans' ability to make informed decisions. As established in the previous section, humans can process multiple sources of uncertainty simultaneously [68], yet current AI systems typically neglect the distinction between different uncertainty sources. Decomposing uncertainty into aleatoric and epistemic components could support decision-makers by enabling tailored reliance strategies [83], but research on simultaneously presenting both uncertainty sources and their impact on human reliance behavior remains limited [10, 34, 54]. This represents a critical gap in understanding how different sources of uncertainty interact with the cognitive biases and prior beliefs outlined earlier to influence reliance behavior in AI-assisted contexts.

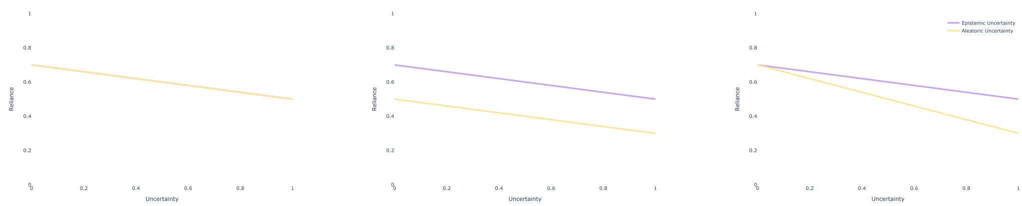
3 Theoretical Development

Building upon the foundations of uncertainty in decision-making and the challenges of communicating uncertainty, this study explores the effects of specifying uncertainty as either aleatoric or epistemic (*RQ1*), decomposing uncertainty into its sources (*RQ2* and *RQ3*), and how prior beliefs interact with different sources and degrees of uncertainty (*RQ4*) with regard to human reliance behavior.

The presence of uncertainty in AI advice may influence decision-makers' reliance on it, a behavior rooted in ambiguity aversion [28]. This principle, first demonstrated by Ellsberg [24] and further developed by Fox and Tversky [28], suggests that individuals prefer advice with known probabilities over that with ambiguous or uncertain ones. In the context of AI advice, ambiguity aversion leads decision-makers to favor their own judgment, with its familiar uncertainty, over the AI's advice. As uncertainty in AI advice increases, decision-makers perceive it as less reliable, as prior literature has shown [13, 87], further inclining them toward their own judgment. Therefore, we formulate the following hypothesis:

Hypothesis 1: Decision-makers rely less on AI advice as its uncertainty increases.

As prior research demonstrated that reliance on advice decreases with increasing uncertainty [13, 87], the varying effects of aleatoric versus epistemic uncertainty on reliance behavior remain unexplored [10]. The fundamental differences between these uncertainty sources may influence how decision-makers perceive and rely on them. Aleatoric uncertainty, being inherent to the problem and unavoidable or stemming from unmeasured variables, represents a natural limitation of predictability. In contrast, epistemic uncertainty reflects a gap in knowledge due to limited similar examples, indicating an incomplete understanding of the underlying relationships. These distinct characteristics in the nature and source of uncertainty could affect reliance behavior differently. To explore this potential difference systematically, we consider potential reliance



(a) Generic strategy.

(b) Uniform differentiation strategy.

(c) Level-dependent differentiation strategy.

Fig. 1. Three potential reliance strategies for decision-makers when faced with aleatoric (yellow) and epistemic (purple) uncertainty in AI-assisted decision-making: (a) Generic strategy: Both uncertainty sources lead to identical decreases in reliance as uncertainty increases. (b) Uniform differentiation strategy: The two uncertainty sources maintain parallel slopes but differ in baseline reliance levels throughout all uncertainty levels. (c) Level-dependent differentiation strategy: Both uncertainty sources start with similar reliance at low uncertainty levels, but diverge with different slopes as uncertainty increases, with one source showing steeper decline in reliance than the other.

strategies decision-makers might adopt when faced with aleatoric and epistemic uncertainty in AI-assisted decision-making:

- *Generic strategy*: Decision-makers might rely on aleatoric and epistemic uncertainty in the same way, reducing their reliance on AI advice as uncertainty increases (see Figure 1(a)). This strategy could be a result of decision-makers perceiving both sources of uncertainty as equally detrimental to the AI's reliability, not understanding the distinction between the two sources of uncertainty, or experiencing cognitive overload when processing the additional information of both sources of uncertainty [42, 92].
- *Uniform differentiation strategy*: Decision-makers might weigh aleatoric and epistemic uncertainty differently in their decision-making processes, as they represent different sources of uncertainty (see Figure 1(b)). This differentiation may manifest in varying reliance behaviors as decision-makers adjust their reliance based on their perception of the impact of the AI advice's uncertainty.
- *Level-dependent differentiation strategy*: Decision-makers might only differentiate between aleatoric and epistemic uncertainty as uncertainty increases (see Figure 1(c)). At lower levels of uncertainty, they may react similarly to both sources, but as uncertainty increases, they may adjust their reliance differently depending on the source of uncertainty present, resulting in a steeper decline in reliance for one source of uncertainty, as suggested by Fox and Ülkümen [29].

Given these potential strategies, we propose the following hypotheses:

Hypothesis 2: Decision-makers attribute different weights to varying sources of uncertainty (aleatoric vs. epistemic) when relying on AI advice.
Hypothesis 3: The source of uncertainty (aleatoric vs. epistemic) moderates the effect of increasing uncertainty on decision-makers' reliance.

Prior HCI research has emphasized the importance of considering humans' characteristics when collaborating with AI systems [14, 23]. In particular, researchers have investigated the role of cognitive biases in advice-taking [96, 105], demonstrating how prior beliefs shape decision-makers' reliance behavior. These prior beliefs remain relevant in the context of decision-support from AI systems and warrant further investigation [9, 36, 93, 102]. Of specific interest is how these prior beliefs interact with decision-makers' perception of uncertainty, potentially amplifying or

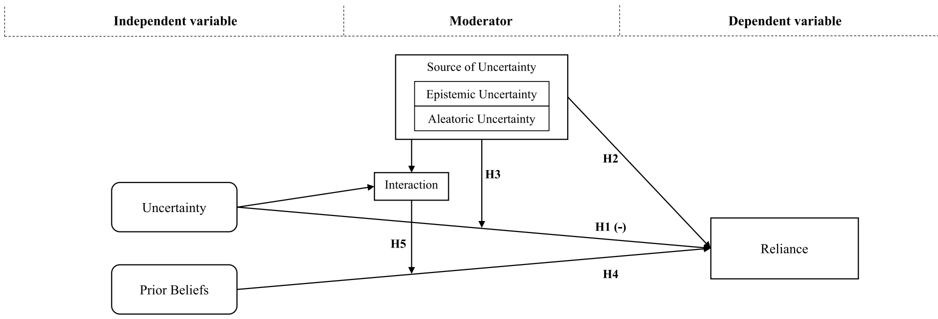


Fig. 2. The research model for our experiments.

mitigating their effects on reliance behavior. More specifically, the source of uncertainty itself may play a crucial role in this interaction, i.e., the distinction between aleatoric and epistemic uncertainty may have varying effects on how humans interpret and rely on AI-generated advice. Given these considerations, we propose the following hypotheses:

Hypothesis 4: Decision-makers' prior beliefs influence their reliance.
Hypothesis 5: The interaction between type and degree of uncertainty moderates the effect of decision-makers' prior beliefs on reliance.

These hypotheses constitute our research model (see Figure 2), which explores how decision-makers rely on uncertain AI advice stemming from varying sources, specifically aleatoric and epistemic. We investigate this through two complementary experiments: experiment 1 answers *RQ1* by examining how reliance behavior changes when uncertainty is classified as either aleatoric or epistemic. Building upon these findings, experiment 2 investigates the effects of decomposing overall uncertainty into these components to answer *RQ2* and *RQ3*. Both experiments then together help us answer *RQ4*. Through these experiments, we aim to understand how uncertainty specification and decomposition influence human reliance on AI advice.

4 Study 1—Understanding Reliance Behavior When Specifying Uncertainty as Aleatoric or Epistemic

In our first experiment, we aim to understand whether decision-makers' reliance behavior varies when the uncertainty of AI advice is specified as either aleatoric or epistemic to answer *RQ1* and how prior beliefs interact with these different uncertainty specifications (*RQ4*).

4.1 Research Method

4.1.1 Experimental Design. Next, we outline the core elements of our experimental design, including the selected task and dataset, how we simulate the AI advice, and our evaluation measure for reliance on AI advice.

Task and Dataset. We selected housing price estimation as our experimental task, as it naturally incorporates uncertainty and frequently prompts advice-seeking behavior while requiring no expert knowledge from participants [38]. This task is well-established in HCI research [11, 38, 74, 75] and inherently exhibits both aleatoric and epistemic uncertainty through factors such as incomplete property information (aleatoric) and varying expertise among real estate agents (epistemic). For our study, we utilized a US real estate dataset from Kaggle [53], which we simplified by reducing the number of features presented to participants, following previous HCI studies [11, 38]. Based on random forest feature importance scores, we selected four features that are interpretable without prior knowledge: number of bedrooms, bathrooms, living space, and median household income in the zip code area.

Advice Communication. We present AI advice using a structured text-based approach with prediction intervals, following the findings of Van Der Bles et al. [99] on effective uncertainty communication. Previous research has shown that this format reduces the effects of prior beliefs and promotes thoughtful consideration of advice [93], enabling us to systematically examine how participants react to varying levels of uncertainty. For each house, the AI provides a price estimate accompanied by a prediction interval representing its uncertainty (e.g., for a house valued at $100, 000, a 3% uncertainty results in a prediction interval of $97, 000 to $103, 000). The uncertainty is specified as either aleatoric or epistemic (see Figure A4(a) in Appendix A.3).

Simulating AI Advice. We simulate AI advice for a sub-sample of 120 houses from the dataset [53]. For each house, we generate price estimates with varying levels of uncertainty, ranging from 3% to 90%, in increments of 3%-pts, with four houses assigned to each uncertainty level. To simulate realistic AI behavior, we ensure predictions are not perfectly centered on true values by randomly shifting each interval's midpoint away from the ground truth. For each house, we randomly choose the shift direction and set its magnitude to 0.75 times the uncertainty level—a parameter chosen to create noticeable but plausible deviations in the AI's predictions. Finally, we impose a constraint that no advice can be less than 0 to maintain realistic predictions.

Treatments. We employ a between-subjects design with two conditions, where the uncertainty of advice is specified as either aleatoric or epistemic.

Evaluation Measure. To analyze participants' reliance on AI advice, we employ the judge-advisor setting [87], where participants first make an autonomous decision before receiving advice and having the opportunity to revise their estimate, a setting widely adopted in HCI research on human–AI decision making [38, 61, 66, 82, 106]. This allows us to observe participants' WOA, an established metric in HCI (e.g., [44, 45, 81]) to measure reliance in regression tasks [3, 13]:

$$\text{WOA} = \frac{\text{Revised Estimate} - \text{Initial Estimate}}{\text{Advisor's Estimate} - \text{Initial Estimate}}.$$

Given that participants receive an interval rather than a point estimate, we utilize the midpoint of this interval as the advisor's estimate to calculate WOA. Consequently, WOA represents the degree of human reliance toward the center of the interval advice. Consistent with prior research (e.g., Bailey et al. [3]), we constrain WOA to prevent the analysis from including scenarios where participants either move further away from the advice or excessively overshoot the advice, which likely represents random responses or outliers. While previous studies typically restricted WOA to the range of 0 to 1 when providing point estimates, we slightly expanded this range from the upper bound of WOA to 1.2 to account for minor overshoots, as participants are not explicitly given the midpoint of the intervals. Decision instances where WOA is outside of this interval are excluded from the analysis.

4.1.2 Procedure. The first experiment explores how humans adjust their reliance on uncertain AI advice when the uncertainty is specified as either epistemic or aleatoric. The experiment is approved by the university's institutional review board. An overview of the experiment's procedure is depicted in Figure 3.

Uncertainty Tutorial and Comprehension Check. After giving consent to our study's terms, participants are randomly assigned to one of our two conditions. In the beginning, participants receive an introductory tutorial on a single source of uncertainty, either aleatoric or epistemic uncertainty. As the terms are difficult to understand, we use established translations of these terms for participants [31]: data uncertainty (describing aleatoric uncertainty) and knowledge uncertainty (describing

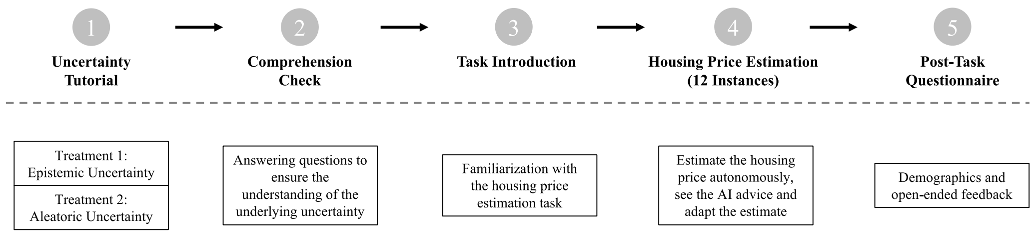


Fig. 3. Procedure of experiment 1.

epistemic uncertainty). To further ease the understanding of the concepts, we carefully design tutorials for each source of uncertainty that describes its source, how it can be reduced, and its consequences on the AI's advice. The interface of the tutorials for both treatments is shown in Figure A2 in Appendix A.3. After explaining the uncertainty sources, we employ a thorough comprehension check to ensure that participants understand the key characteristics of the uncertainty source (see Figure A3 in Appendix A.3). If participants answer incorrectly, the correct answer is explained to them, and they can repeat the comprehension check once. Only participants who answer all questions correctly can proceed to the main study.

Main Task. After the comprehension check, participants are introduced to the task. We explain each feature and mention the average house price. After two example instances, participants proceeded to the main part of our study, where they needed to estimate the prices of 12 houses (for the task interface see Figure A5(a) in Appendix A.3). The instances are randomly selected from the total set of 120 houses and presented in a randomized order to ensure that effects are neither based on the specific instances nor the order in which they are presented. Due to this randomization, individual participants may experience different levels of uncertainty (e.g., one participant might encounter houses with on average higher uncertainty than another). However, we ensured that the average uncertainty level was similar across both treatments (44% in the aleatoric treatment and 45% in the epistemic treatment).

Post-Task Questionnaire. After successfully estimating 12 housing prices, participants conclude with a post-task questionnaire, where we gather demographics and allow for additional feedback via open-ended questions.

4.1.3 Participants. We recruited participants through Prolific, a platform demonstrated to be a reliable source for research data [69, 71]. Participants were required to be US residents and fluent in English. Following Abbey and Meloy [1], we implemented two attention checks and excluded participants who failed either the attention or the comprehension check from the analysis. The final sample comprised 40 participants per treatment, totaling 80 participants (51.25% male; average age: 32.55 years, SD: 10.57; median time: 13.5 min). Participants received a base payment of $2 and the best-performing 10% earned an additional bonus payment of $1.3.

4.2 Analysis and Results

The next section presents a statistical analysis of our experimental findings on how aleatoric versus epistemic uncertainty specifications affect participants' reliance behavior in individual decision scenarios. This detailed analysis aims to uncover how specific variables—such as the level of uncertainty in AI advice, the participants' prior beliefs, and the framing of uncertainty—interact to shape reliance behavior, addressing both *RQ1* and *RQ4*.

To account for the repeated measures design of our study, we employ a mixed-effects regression model to test our hypotheses. The model includes a random effect for the participants to address

Table 1. Mixed Effects Model Analysis on Human Reliance on AI Advice

Dependent Variable	WOA		
	Coeff	SE	p-Value
Intercept	0.876***	0.038	<0.001
Treatment [Epistemic]	−0.003	0.051	0.95
Uncertainty	−0.338***	0.075	<0.001
Prior Beliefs Confirmed	−0.574***	0.046	<0.001
Uncertainty × Treatment [Epistemic]	−0.170	0.098	0.08
Uncertainty × Treatment [Epistemic] × Prior Beliefs Confirmed	0.444***	0.093	<0.001
Uncertainty × Treatment [Aleatoric] × Prior Beliefs Confirmed	0.372***	0.096	<0.001
Participant ID (random effect)	0.018	0.015	—
Log-Likelihood	−190.3258		
Scale	0.0815		
Converged	Yes		

***$p < 0.001$.

the nested structure of our data. As outlined earlier, we use reliance (measured by WOA) as the dependent variable. Our study incorporates three independent variables: the magnitude of uncertainty (operationalized as the width of the AI-generated advice interval); the prior beliefs (operationalized by a binary variable denoting whether the participant's initial estimate falls within the interval advised by the AI); and the source of uncertainty presented (aleatoric or epistemic). To capture the interplay among these factors, we include two interaction terms in our analysis as hypothesized in Section 3. First, we examine the interaction between treatment conditions and uncertainty magnitude to explore the potential differential effects of uncertainty across treatments. Second, we investigated a three-way interaction among uncertainty magnitude, treatment condition, and the binary variable of initial estimate placement. This interaction allows us to explore how the relationship between uncertainty and treatment might moderate the effect of prior beliefs on reliance behavior. A comprehensive overview of the regression model's results is presented in Table 1.

The model's intercept is positive and significant ($\beta = 0.876$, $p < 0.001$), indicating that when all other variables are at their reference levels or zero, participants, on average, rely on the AI advice for about 87.6% in their final estimate. Regarding the random effects, the model shows a group (participant) variance of $\sigma^2 = 0.015$, suggesting some individual differences in reliance behavior across participants.

Looking at the factors influencing WOA, the model reveals several significant predictors. We find a significant negative effect for uncertainty ($\beta = -0.338$, $p < 0.001$). This indicates that reliance on AI advice decreases as uncertainty grows. This finding aligns with the notion that decision-makers tend to perceive more precise information as more valuable and, thus, may place less weight on AI advice when associated with higher uncertainty. Therefore, we find support for Hypothesis 1.

Our analysis reveals no statistically significant effect for the treatment condition ($\beta = -0.003$, $p = 0.954$). This finding suggests that the source of uncertainty does not significantly influence the relationship between uncertainty and reliance. In other words, participants do not weigh aleatoric and epistemic uncertainty differently when relying on AI advice. Thus, we do not find support for Hypothesis 2.

Next, we observe a non-significant trend suggesting a potential interaction between the degree of uncertainty and the treatment condition ($\beta = -0.17$, $p = 0.081$). This interaction might indicate that the negative effect of uncertainty on reliance could be more pronounced when the uncertainty

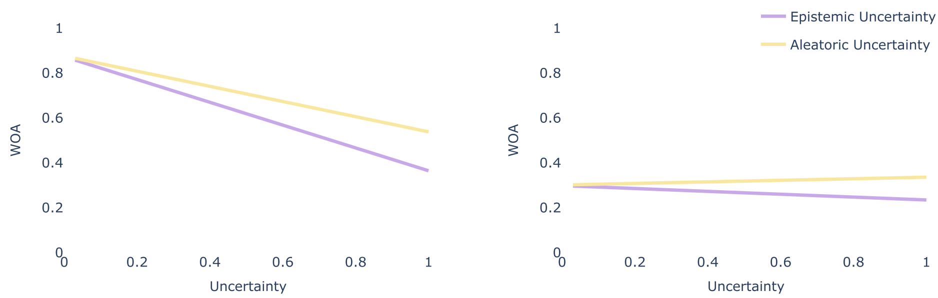


Fig. 4. Interaction effects between uncertainty level, uncertainty source (epistemic vs. aleatoric), and prior beliefs on reliance (WOA). The left plot shows the reliance behavior when the human's initial estimate falls outside the AI's advised range. The right plot depicts the reliance behavior when the human's initial estimate is confirmed through the AI's advised range.

is framed as epistemic rather than aleatoric uncertainty. While not statistically significant, this directional trend suggests participants may be more sensitive to increases in uncertainty when presented as a limitation in the AI's knowledge rather than as inherent randomness in the task or data. This trend aligns with the notion that epistemic uncertainty might be perceived as a more concerning limitation of the AI system, possibly leading to a stronger reduction in reliance on AI advice as uncertainty increases. However, we do not find statistical significance on an α-level of 0.05 to support Hypothesis 3.

Finally, the confirmation of prior beliefs—indicated by the participant's initial prediction falling within the AI's advice interval—significantly reduces WOA ($\beta = -0.574$, $p < 0.001$). This suggests that when participants' prior beliefs align with the AI's advice, they perceive less need to adjust their estimate, resulting in lower reliance. This finding supports Hypothesis 4. With regard to Hypothesis 5, we observe that the interaction between uncertainty degree and source moderates the effect of prior beliefs on reliance. Specifically, when prior beliefs are confirmed, the decrease in reliance with increasing uncertainty is present under both uncertainty sources, with interaction coefficients of $\beta = 0.444$ ($p < 0.001$) for epistemic and $\beta = 0.372$ ($p < 0.001$) for aleatoric uncertainty. This means that the effects of increasing uncertainty on reliance are partially offset, when prior beliefs are confirmed. However, a linear contrast test reveals no statistically significant difference between these interaction effects ($t = 0.89$, $p = 0.373$), indicating that both uncertainty sources do not show statistically different moderating effects. Therefore, the results provide partial support for Hypothesis 5: While the uncertainty degree moderates the effect of prior beliefs on reliance, this moderating effect does not differ significantly between uncertainty sources. These findings suggest that uncertainty level moderates how prior beliefs influence reliance, with this moderating relationship operating equivalently across both epistemic and aleatoric uncertainty framings (see Figure 4). When the human's initial estimate falls outside the AI's advised range, reliance decreases as uncertainty increases for both uncertainty sources. Conversely, when the initial estimate falls within the advised range, reliance on the AI advice is both lower and decreases less as uncertainty increases, again with similar patterns across uncertainty sources. These findings address *RQ4* by demonstrating that prior beliefs do influence reliance behavior, with this influence being moderated by the degree of uncertainty, though not differentially by uncertainty source (epistemic vs. aleatoric).

Summary. Our findings address *RQ1* and *RQ4* by revealing how uncertainty specification affects human reliance on AI advice and how prior beliefs interact with different sources and degrees of uncertainty. We find that reliance generally decreases as uncertainty increases, with a non-significant trend suggesting that this effect may be more pronounced when uncertainty is framed as epistemic rather than aleatoric. Participants also show significantly lower reliance when their initial estimates align with the AI advice, independent of the uncertainty source. Regarding the interaction between uncertainty and prior beliefs, we find that uncertainty degree moderates how prior beliefs influence reliance behavior. However, this moderating effect operates equivalently across both uncertainty sources, indicating that while prior belief confirmation consistently affects how people respond to increasing uncertainty, this relationship does not differ meaningfully between epistemic and aleatoric uncertainty framings. Taken together, these patterns suggest that participants may adopt a *level-dependent differentiation strategy* when exposed to a single source of uncertainty—showing a tendency toward greater sensitivity to increasing levels of epistemic uncertainty. However, when considering how prior beliefs interact with uncertainty, the differentiation between uncertainty sources becomes negligible, suggesting that belief confirmation effects operate similarly regardless of uncertainty framing.

5 Study 2—Understanding Reliance Behavior When Decomposing Uncertainty into Aleatoric and Epistemic

Building upon the insights gained from experiment 1, our second experiment examines the effects of decomposing uncertainty into its aleatoric and epistemic components, as opposed to providing a single monolithic measure. This extension addresses *RQ2* and provides a deeper understanding of how different representations of uncertainty impact decision-making in AI-assisted tasks. Furthermore, we examine how human reliance differs between aleatoric and epistemic uncertainty when the overall uncertainty is decomposed into its components to answer *RQ3*. Finally, we investigate how prior beliefs shape reliance based on different degrees and sources of uncertainty, thereby answering *RQ4*. This approach allows us to explore how decision-makers integrate and weigh multiple sources of uncertainty in their reliance strategies, potentially revealing more complex interactions between uncertainty sources and human cognitive processes.

5.1 Research Method

5.1.1 Experimental Design. Following, we outline the components of our experimental design, including the task structure, dataset selection, and the process for generating AI-assisted advice.

Task and Dataset. For our second experiment, we retain the housing price estimation task and dataset used in the first experiment.

Treatments. Our study employs a mixed between- and within-subjects design with two treatments to address our RQs. Between subjects, we compare overall reliance behavior when providing either a combined uncertainty estimate or decomposing the uncertainty into its aleatoric and epistemic components (*RQ2*). Within the decomposed treatment, we then analyze how these different sources of uncertainty influence individual decisions (*RQ3* and *RQ4*). This design allows us to examine both the overall effect of decomposing uncertainty and the specific influences of different uncertainty sources.

Advice Communication. To investigate the effects of different sources of uncertainty on advice-taking behavior, we simulate AI advice consistent with the textual approach in experiment 1. Participants in both treatments receive the AI's advice through a prediction interval representing the overall uncertainty. While participants in the first treatment only receive this interval, those in the second treatment additionally receive information about how this uncertainty is decomposed into its aleatoric and epistemic components. Following the framework of Van Der Bles et al. [99], we communicate each uncertainty source using categorical levels (very low, low, high, very high) based on quartiles rather than exact percentages or proportional contributions to total uncertainty. This categorical approach offers several advantages: It reduces cognitive load compared to numerical expressions and helps individuals with low numeracy better assess uncertainty [72]. However,

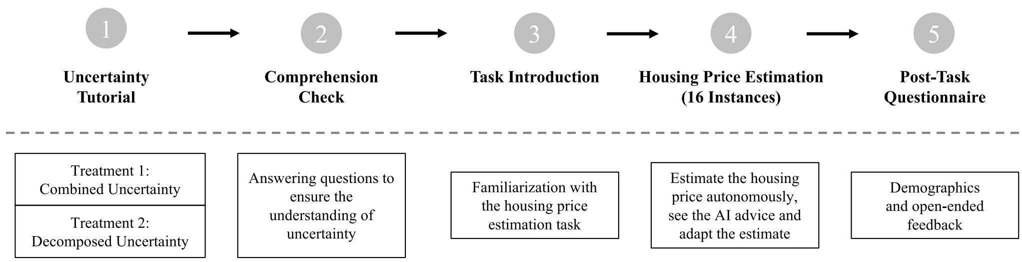


Fig. 5. Procedure of experiment 2.

this verbal communication approach also introduces challenges like interpretive variability among individuals, as uncertainty terms can carry different connotations for different people. Additionally, the categorical approach may influence how participants weigh different uncertainty sources in their decision-making process, potentially affecting the relative importance they assign to aleatoric versus epistemic information. Despite these potential effects, our design choice aligns with our research objective of understanding how humans respond to relative changes in model uncertainty. The categorical levels are based on the relative distribution within each uncertainty source rather than their proportional contribution to total uncertainty. In a well-calibrated AI system, these relative levels provide meaningful signals about the model's confidence and limitations within each uncertainty source, independent of their proportional contribution. For instance, "very high" epistemic uncertainty indicates a gap in the model's training data for that specific prediction context, providing actionable information regardless of how it compares proportionally to the aleatoric uncertainty.

Simulating AI Advice. Our simulated advice is based on an actual AI model trained on a real-world housing dataset (see Appendix A.1 for details), which grounds our simulation in realistic patterns of uncertainty while allowing for systematic variation of the advice characteristics. The technical details of the advice simulation procedure are provided in Appendix A.2, and the advice is illustrated in Figure A4(b) in Appendix A.3. In total, we generate 160 advice instances, with each combination of aleatoric and epistemic uncertainty levels represented equally. The total uncertainty of the simulated instances ranges from 0.05 to 0.9, and every level of each uncertainty source appears at least five times. This systematically generated set of simulated advice enables a controlled examination of the effects of different sources and levels of uncertainty on advice utilization.

Evaluation Measures. As with experiment 1, we analyze the reliance behavior by observing the participant's WOA and restrain WOA between 0 and 1.2.

5.1.2 Procedure. Experiment 2 investigates how humans adjust their reliance on AI advice when both epistemic and aleatoric uncertainty are present, either combined into a single estimate or decomposed into its sources. The university's institutional review board approved the experiment. Figure 5 provides an overview of the experimental procedure.

Uncertainty Tutorial and Comprehension Checks. After providing informed consent, participants are randomly assigned to one of the two conditions and receive an introductory tutorial on uncertainty in AI advice and its two sources, similar to experiment 1. Afterward, participants also receive guidance on integrating the information on uncertainty into their decision-making. To avoid providing too much information simultaneously, we scaffold the tutorial into four parts and employ comprehension checks after each tutorial to verify participants' understanding of the relevant concepts (see Figures A6 and A7 in Appendix A.1). Participants must answer all questions correctly to proceed. If participants answer incorrectly, they receive an additional explanation and another opportunity to complete the check. Failure to pass the comprehension check after this second attempt results in exclusion from the study.

Task. The main task involves estimating the prices of 16 apartments that are randomly drawn from the pool of simulated advice such that each combination of aleatoric and epistemic uncertainty is presented once (very low, low, high, very high). Again, we employ the judge-advisor paradigm [87] to measure WOA. During the task, participants can access the tutorials on uncertainty at any time (see the task interface in Figure A5 (b) in Appendix A.3).

Post-Task Questionnaire. After completing the price estimation task, we collect basic demographic information and provide space for any additional comments.

5.1.3 Participants. For experiment 2, we follow the same procedure as for experiment 1. After excluding participants who failed attention or comprehension checks, our final sample consisted of 80 participants (57.7% male; average age: 37.23 years, SD: 12.76; median time: 25.5 min). All participants were US residents fluent in English, recruited through Prolific. Participants received a base payment of $4, with the top 10% performers earning an additional $1.3 bonus.

5.2 Analysis and Results

The following presents a statistical analysis of our findings from experiment 2. We begin by examining the impact of uncertainty decomposition on participants' global reliance behavior to answer *RQ2* before delving into a more granular analysis of reliance across different levels of aleatoric and epistemic uncertainty to answer *RQ3* and *RQ4*.

5.2.1 Impact of Uncertainty Decomposition on Reliance Behavior. To assess whether decomposing uncertainty into aleatoric and epistemic components influences participants' average reliance on AI advice, we calculate the mean WOA for each participant across the decision instances and compare the two treatment groups. A Shapiro-Wilk test of normality indicates that the average reliance follows a Gaussian distribution in the combined uncertainty treatment (p = 0.988) but not in the decomposed uncertainty treatment (p = 0.032). Given these results, we proceed with a Mann-Whitney U test to compare the mean reliance between treatments. The test reveals no significant difference in average WOA between the combined uncertainty treatment (μ = 0.4752) and the decomposed uncertainty treatment (μ = 0.4464, p = 0.315). This finding suggests that decomposing uncertainty into its aleatoric and epistemic components does not significantly influence participants' overall reliance on AI advice.

Exploratory Analysis on the Role of Uncertainty Presentation in Advice Rejection. As an exploratory analysis, we investigate whether presenting uncertainty information in different formats affects how participants process and respond to high levels of uncertainty. This comparison allows us to understand whether making uncertainty sources explicit influences participants' decisions to reject AI advice. Our analysis focuses on high and very high uncertainty levels, where advice rejection is most relevant due to the increased risk these situations present. We employ Mann-Whitney U tests, appropriate for our non-normally distributed data, to compare rejection rates between the presentation formats. To control for multiple comparisons, we apply the Benjamini-Hochberg correction [8]. The results reveal that participants more frequently reject advice when high uncertainty is presented in its decomposed form compared to when it is combined, both for aleatoric (p_{adj} = 0.026) and epistemic uncertainty (p_{adj} = 0.053). This suggests that making uncertainty sources explicit enables participants to more critically evaluate the AI's advice, leading to more selective rejection when either source of uncertainty is high.

Table 2. Mixed Effects Model Analysis on Human Reliance on AI Advice

Dependent Variable	WOA		
	Coeff	SE	p-Value
Intercept	0.826***	0.046	<0.001
Prior Beliefs Confirmed	−0.384***	0.081	<0.001
Aleatoric [Low]	−0.134**	0.045	0.003
Aleatoric [High]	−0.217***	0.046	<0.001
Aleatoric [Very High]	−0.267***	0.067	<0.001
Epistemic [Low]	−0.033	0.046	0.471
Epistemic [High]	−0.179***	0.047	<0.001
Epistemic [Very High]	−0.244***	0.059	<0.001
Prior Beliefs Confirmed × Aleatoric [Low]	0.082	0.078	0.294
Prior Beliefs Confirmed × Aleatoric [High]	0.077	0.078	0.320
Prior Beliefs Confirmed × Aleatoric [Very High]	0.051	0.090	0.569
Prior Beliefs Confirmed × Epistemic [Low]	0.008	0.075	0.919
Prior Beliefs Confirmed × Epistemic [High]	0.126	0.075	0.095
Prior Beliefs Confirmed × Epistemic [Very High]	0.152	0.082	0.063
Participant_ID	0.028	0.028	—
Log-Likelihood	−165.9715		
Scale	0.0869		
Converged	Yes		

***p < 0.001. The table shows the results of the mixed effects model with reliance (WOA) as the dependent variable. Columns include coefficient, standard error, and exact p-value. Significance is indicated based on conventional thresholds.

5.2.2 Factors Influencing Individual Reliance Decisions. Building on our global analysis, we now examine how different levels of aleatoric and epistemic uncertainty influence participants' reliance on AI advice when explicitly presented with both components. Thus, we focus our analysis exclusively on the decomposed treatment, where participants received distinct information about aleatoric and epistemic uncertainty levels, to test our hypotheses and answer *RQ2, RQ3,* and *RQ4.*

Mixed Effects Model Analysis. To investigate these effects, we fit a mixed effects model similar to experiment 1. We include reliance, measured by WOA, as the dependent variable. The model includes fixed effects for aleatoric uncertainty level (very low, low, high, very high), epistemic uncertainty level (very low, low, high, very high), and a binary variable describing whether prior beliefs were confirmed, i.e., whether the participant's initial estimate fell within the AI's prediction interval. Further, we add an interaction between aleatoric and epistemic uncertainty and the binary variable of the placement of the initial estimate. Participants are included as a random effect to account for the repeated measures design (see Table 2).

Our analysis reveals significant main effects for aleatoric and epistemic uncertainty. For aleatoric uncertainty, compared to the baseline (Very Low), we observe progressively stronger coefficients as uncertainty increases: Low ($\beta = -0.134$, $p = 0.003$), High ($\beta = -0.217$, $p < 0.001$), and Very High ($\beta = -0.267$, $p < 0.001$). Similarly, for epistemic uncertainty, we find increasing negative coefficients: Low ($\beta = -0.033$, $p = 0.471$), High ($\beta = -0.179$, $p < 0.001$), and Very High ($\beta = -0.244$, $p < 0.001$). To explicitly test whether each incremental increase in uncertainty leads to a statistically significant reduction in reliance on AI advice, we further conduct pairwise comparisons between adjacent uncertainty levels. Given the violations of normality (Shapiro-Wilk tests, all $p < 0.001$) and homogeneity of variance (Levene's test for aleatoric: $p = 0.007$; for epistemic: $p = 0.004$), we perform Mann-Whitney U tests (see Table 3). For aleatoric uncertainty, we find significant decreases in reliance between all adjacent uncertainty levels: from Very Low to Low ($p < 0.001$), from Low to High ($p = 0.009$), and from High to Very High ($p < 0.001$). Similarly, for epistemic uncertainty, we observed significant decreases between all adjacent levels: from Very Low to Low ($p = 0.033$), from Low to High ($p = 0.027$), and from High to Very High ($p = 0.003$). These pairwise comparisons confirm that each incremental increase in uncertainty, whether aleatoric or epistemic, result in a statistically significant reduction in participants' reliance on AI advice. These findings provide support for Hypothesis 1, demonstrating that decision-makers rely progressively less on AI advice as its uncertainty increases.

Table 3. Pairwise Comparisons of Reliance on AI Advice between Adjacent Uncertainty Levels

Uncertainty Comparison (Adjacent Levels)	Aleatoric p-Value	Epistemic p-Value
Very Low to Low	$p < 0.001$	$p = 0.033$
Low to High	$p = 0.009$	$p = 0.027$
High to Very High	$p < 0.001$	$p = 0.003$

To evaluate Hypothesis 2, we conduct a comparative analysis of the mean WOA across the different levels of uncertainty. Specifically, we compare very low epistemic uncertainty instances with very low aleatoric uncertainty and extend this comparison across all uncertainty levels using post-hoc Tukey-HSD tests. The results, as presented in Table 4, indicate no statistically significant differences between the sources of uncertainty at any level examined, thus providing no support for Hypothesis 2.

Table 4. Tukey HSD Test Results for Different Uncertainty Levels

Level	**Group 1**	**Group 2**	**Mean Diff**	**p-Adj**	**Lower**	**Upper**
Very Low	Aleatoric	Epistemic	−0.0735	0.1130	−0.1646	0.0175
Low	Aleatoric	Epistemic	0.0029	0.9509	−0.0893	0.0951
High	Aleatoric	Epistemic	0.0249	0.5467	−0.0562	0.1059
Very High	Aleatoric	Epistemic	0.0515	0.2502	−0.0365	0.1395

The table shows the results of the post-hoc Tukey HSD test. In the columns, the groups, differences in mean, and adjusted p-values are presented.

Regarding Hypothesis 3, we examine whether aleatoric and epistemic uncertainty have different effects on reliance as uncertainty levels increase. For aleatoric uncertainty, compared to the baseline (Very Low), we observe progressively stronger negative coefficients: Low ($\beta = -0.134$, $p = 0.003$), High ($\beta = -0.217$, $p < 0.001$), and Very High ($\beta = -0.267$, $p < 0.001$). Similarly, for epistemic uncertainty: Low ($\beta = -0.033$, $p = 0.471$), High ($\beta = -0.179$, $p < 0.001$), and Very High ($\beta = -0.244$, $p < 0.001$). As the coefficients suggest potentially different magnitudes of effect on reliance behavior between uncertainty sources, we conduct linear contrast tests to directly compare the slopes between uncertainty sources across the ranges (see Table 5), i.e., from Very Low to Low, Very Low to High, and Very Low to Very High. These tests reveal no significant differences for any slope: Low ($t = -1.54$, $p = 0.125$), High ($t = -0.559$, $p = 0.576$), or Very High ($t = -0.259$, $p = 0.796$). This indicates that both uncertainty sources do not have statistically different moderating effects across the full range of uncertainty levels. Accordingly, we do not find support for 3.

Table 5. Linear Contrast Tests Comparing Slopes between Uncertainty Sources

Uncertainty Level Comparison	t-Value	p-Value
Very Low to Low	−1.54	0.125
Very Low to High	−0.559	0.576
Very Low to Very High	−0.259	0.796

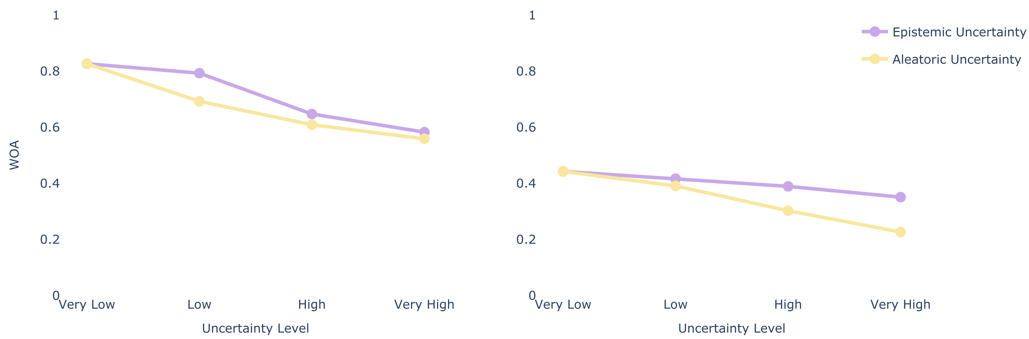


Fig. 6. Interaction effects between uncertainty level, uncertainty source (epistemic vs. aleatoric), and prior beliefs on reliance (WOA). The left panel shows reliance when human estimates fall outside the AI's advised range, while the right panel shows reliance when estimates fall within this range.

Addressing *RQ4*, our mixed-effects model reveals a significant negative effect when participants' prior beliefs align with the AI's advised range ($\beta = -0.384$, $p < 0.001$). This indicates that participants rely less on the AI advice when their initial estimate falls within the AI's prediction interval. This finding supports Hypothesis 4, suggesting that participants' prior beliefs shape their reliance behavior.

We further find non-significant trends between participants' prior beliefs and high levels of epistemic uncertainty (High: $\beta = 0.126$, p = 0.095; Very High: $\beta = 0.152$, p = 0.063; see Figure 6). These findings suggest that when prior beliefs are confirmed, high epistemic uncertainty may moderate and weaken the tendency for participants to favor their initial judgments, contrasting with the results from our first experiment. Interestingly, this leads to reliance on AI advice remaining nearly constant as epistemic uncertainty increases. However, no significant interactions were observed with aleatoric uncertainty. These results provide partial support for Hypothesis 5 and underscore the need for further investigation into the complex relationships between uncertainty source and degree, prior beliefs, and AI reliance.

Exploring Reliance across Uncertainty Levels. To further examine how different levels of aleatoric and epistemic uncertainty influence participants' reliance on AI advice, we calculate the mean and median WOA for each combination of uncertainty levels in the decomposed treatment.

We observe a general trend of decreasing mean WOA as the overall level of uncertainty increases (see Figure 7). The highest mean WOA (0.8826) occurs when the uncertainty is lowest, while the lowest mean WOA (0.1849) is observed when the uncertainty is highest. Upon closer inspection, we notice that the mean WOA decreases more rapidly as aleatoric uncertainty increases compared to epistemic uncertainty. This pattern indicates that participants may be more sensitive to changes in aleatoric uncertainty than epistemic uncertainty when the sources of uncertainty are explicitly presented alongside each other.

To get a more in-depth picture of reliance behavior across uncertainty levels, we next examine the median WOA. The right part of Figure 7 presents the median WOA for each combination of

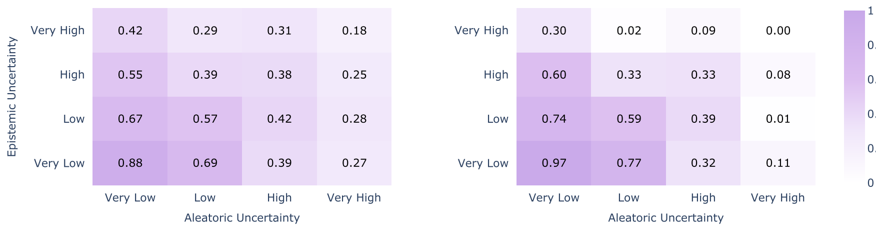


Fig. 7. Heatmaps depicting the mean (left) and median (right) WOA across levels of aleatoric and epistemic uncertainty in the decomposed treatment.

uncertainty levels in the decomposed treatment. The median WOA values follow a similar overall pattern to the mean values, with a general decrease in reliance as uncertainty increases. However, the median values reveal some additional insights into the distribution of reliance behavior. The median WOA exhibits a more dramatic drop as uncertainty levels increase compared to the mean WOA. This indicates that a significant proportion of participants relied very little on AI advice in high-uncertainty scenarios, pulling the median value down. In the most extreme case, when both aleatoric and epistemic uncertainty are very high, the median WOA equals 0, i.e., participants reject the advice entirely, while the mean WOA is 0.1849. This comparison shows that mean and median WOA values demonstrate similar decreasing trends as uncertainty increases, but differ more substantially at higher uncertainty levels. This suggests that reliance behavior becomes more skewed in uncertain conditions, with many participants showing very low reliance on AI advice.

Summary. Our findings address three RQs about uncertainty decomposition and human reliance on AI advice. Regarding *RQ2*, we find that decomposing uncertainty into its components does not significantly affect overall reliance levels compared to presenting combined uncertainty. However, our exploratory analysis revealed that decomposition leads to more frequent advice rejection when either uncertainty source is high, suggesting more critical evaluation of AI advice. Addressing *RQ3*, we find that both uncertainty sources negatively impact reliance as they increase. However, we observe no statistically significant differences between how humans respond to aleatoric versus epistemic uncertainty, indicating a generic strategy that contrasts with the findings from experiment 1. Finally, we find that prior beliefs play a significant role in shaping reliance behavior thereby addressing *RQ4*. Participants exhibit lower reliance when their prior beliefs align with the AI's prediction interval, whereas we find a non-significant trend that high epistemic uncertainty (but not aleatoric) moderates this effect. These findings emphasize that while decomposing uncertainty may not alter overall reliance levels, it can influence how humans process and reject advice in high-uncertainty situations, highlighting the importance of thoughtful uncertainty communication design in AI systems.

6 Discussion

Our study contributes to the field of HCI and, more specifically, to the growing body of research on uncertainty in AI-assisted decision-making. As motivated in our related work section, previous studies have primarily focused on monolithic uncertainty, presenting a single source or aggregated uncertainty measures (e.g., [51, 62, 106]). While these works have provided valuable insights into the effects of uncertainty on human perceptions, trust, and decision quality, they have not investigated the complexity of the uncertainty's distinct sources [10, 54]. Our research addresses this gap by examining how humans respond to uncertainty in AI advice—investigating responses to individual uncertainty sources (*RQ 1*), the effects of decomposing uncertainty into its components (*RQ 2*), potential differences in reliance between uncertainty sources when presented together (*RQ 3*),

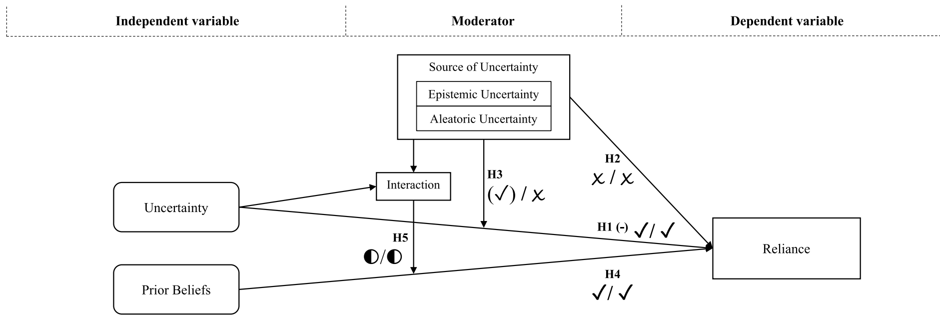


Fig. 8. Overview of hypothesis support across studies. *Study 1*/*Study 2*, where ✓ indicates statistical significance at $p < 0.05$; (✓) denotes a non-significant trend ($0.05 < p < 0.1$); ◐ denotes partial support; x indicates no support ($p > 0.1$).

and how prior beliefs interact with different uncertainty sources and levels (*RQ 4*). While not statistically significant, we observe a trend that suggests that humans adopt a level-dependent strategy when presented with either source alone, but switch to a generic strategy when faced with both sources of uncertainty simultaneously. These insights provide a more nuanced understanding of how humans process and integrate different sources of uncertainty in their decision-making, contributing to the limited body of work examining the distinct effects of aleatoric and epistemic uncertainty in HCI research [10, 54, 68].

6.1 Empirical Findings on the Influence of Aleatoric and Epistemic Uncertainty on Human Reliance Behavior

Through our two-experiment study, we uncover complex patterns in how humans interpret and utilize different sources of uncertainty information in AI-assisted decision-making, revealing distinct reliance strategies depending on how uncertainty is presented. Figure 8 provides a consolidated overview of our hypothesized relationships and the corresponding empirical support across both studies. As shown, some effects were robust across both studies (H1 and H4), while others showed partial (H3 and H5) or no support (H2), highlighting the nuanced role of uncertainty source and prior beliefs in shaping reliance behavior.

The Role of Uncertainty. Our study examines the complex relationship between uncertainty presentation and human reliance behavior in AI-assisted decision-making. Across both experiments, we analyze the effects of specifying and decomposing uncertainty into its sources to address *RQ1–3*. This builds upon previous research showing that uncertainty presentation can affect human perceptions, trust, and decision quality [6, 15, 26, 51, 56, 62, 63, 106, 107]. However, these studies have primarily focused on the effects of monolithic uncertainty or a single source of uncertainty rather than considering the potential differences between aleatoric and epistemic uncertainty. Our study goes further by examining the distinct effects of these uncertainty sources, addressing a relevant gap in current HCI research [10, 54, 68].

While we find a trend that participants demonstrate distinct behaviors when presented separately with aleatoric and epistemic uncertainty—aligning with the *level-dependent differentiation strategy*—this distinction seems to diminish when both sources of uncertainty are jointly presented, resulting in a *generic strategy*. This shift in strategy between experiments highlights the critical role of the way uncertainty information is presented in shaping human reliance on AI systems. When epistemic uncertainty is presented alone, as in our first experiment, it seems to be interpreted as a direct indicator of the AI system's knowledge and capabilities. This, in turn, appears to lead to a steeper decline in reliance on AI advice as uncertainty increases. In contrast, the simultaneous presentation of both uncertainty sources in our second experiment seems to diminish the perceived significance of epistemic uncertainty. Here, participants may view epistemic uncertainty as just another component of overall uncertainty rather than a distinct signal about the AI's limitations. As a result, they may respond similarly to both sources of uncertainty when making their decisions.

These findings advance the field of HCI by providing a deeper understanding of how humans process complex uncertainty information in AI-assisted decision-making contexts. While previous HCI research, such as studies by Bansal et al. [6] and Buçinca et al. [15], has shown that uncertainty presentation can affect trust and reliance, our work extends this knowledge by demonstrating how the specification and contextualization of different uncertainty sources can lead to distinct reliance strategies highlighting the need for more sophisticated approaches to uncertainty communication in human–AI collaboration.

The Interplay between Prior Beliefs and Uncertainty Levels. A key aspect of our investigation (*RQ4*) concerns how prior beliefs interact with different sources and degrees of uncertainty to influence reliance behavior. Across both experiments, we find evidence of prior beliefs shaping participants' decision-making processes, aligning with existing research on how prior beliefs can lead to selective interpretation of uncertain predictions [9, 36, 93, 102]. Interestingly, the manifestation of these prior beliefs interacts with the degree of uncertainty presented, but not with its source. Linear contrast tests in the first experiment reveal that while uncertainty degree moderates the effect of prior beliefs on reliance, this effect does not differ significantly between epistemic and aleatoric uncertainty sources. In the second experiment, we observe only marginal trends between participants' prior beliefs and high levels of epistemic uncertainty, with no significant interactions for aleatoric uncertainty.

This interaction between prior beliefs and uncertainty levels creates complex dynamics in reliance behavior that have important implications for human–AI decision-making. As uncertainty increases, the AI's prediction interval widens, increasing the likelihood of alignment with the human's initial estimate. This alignment can paradoxically lead to both overreliance on the overall AI assessment (due to perceived confirmation of the human's estimate) despite expressing greater uncertainty and under-reliance on the specific advice within the AI's prediction interval. Both experiments reveal a consistent pattern where participants do not meaningfully differentiate between uncertainty sources when integrating their prior beliefs with AI advice. This indicates that cognitive processes underlying reliance decisions may be driven more by the overall magnitude of uncertainty rather than its conceptual source. These findings challenge the assumption that theoretical distinctions between uncertainty sources translate into practical differences in human–AI interaction and highlight the need for more sophisticated approaches to uncertainty communication that account for these complex psychological dynamics. For HCI practitioners, these findings suggest that uncertainty quantification, while essential for informed decision-making, can inadvertently amplify the effects of prior beliefs, requiring careful consideration of how uncertainty visualizations might reinforce existing cognitive biases rather than mitigate them.

Synthesizing Insights across Experiments. By designing our second experiment to extend the findings of the first, we uncover interesting patterns in human interpretation of uncertainty information. The progression from isolated presentation of uncertainty sources to combined presentation reveals how the context of uncertainty information shapes its interpretation and use in decision-making. This progression in our experimental design allows us to observe how humans adapt their decision strategies based on the complexity and presentation of uncertainty information. These findings collectively demonstrate the need for careful consideration of how uncertainty is communicated in AI systems and suggest that the effectiveness of uncertainty communication may depend not just on the source of uncertainty present but also on how it is presented alongside other information.

6.2 Implications for Designing Systems for AI-Assisted Decision Making

Our findings have implications for the design of AI-assisted decision support systems, particularly in terms of how uncertainty information is communicated to humans and how these systems can be improved over time. Based on our results, we propose the following implications for system design and implementation:

Present Uncertainty Information Carefully. The way uncertainty is presented and contextualized significantly influences human perception of its relevance and subsequent reliance on AI advice. Our experiments reveal that humans adopt different strategies when faced with isolated versus combined presentations of uncertainty sources. This suggests that system designers need to carefully consider not just *what* uncertainty information to present, but also *how* to frame it within the broader context of the decision-making task.

Our study further demonstrates that the contextualization of uncertainty plays a key role in shaping human reliance on AI systems. In our first experiment, participants utilized information about the uncertainty's source and adapted their reliance behavior accordingly, effectively adopting an *level-dependent strategy*. However, when we extend the uncertainty setting in our second experiment by decomposing uncertainty into its sources, participants shift to a *generic strategy*. They no longer distinguish between the two sources of uncertainty and, instead, rely similarly to AI advice from a single, unspecified source. The only difference in participants' reliance behavior when uncertainty was decomposed was their reduced reliance when any source of uncertainty was high or very high—information not available without decomposing uncertainty into its components. This finding highlights the need for two critical areas of further research: (1) investigating the underlying reasons why decision-makers alter their reliance strategies when presented with different forms of uncertainty and (2) developing effective methods for communicating multiple sources of uncertainty simultaneously that preserve their distinct interpretation.

Given these observations, designers should carefully evaluate the potential consequences of presenting epistemic uncertainty in isolation or in combination with aleatoric uncertainty. In cases where a holistic overview of possible outcomes is unnecessary or where there are high risks when facing cases unknown to the model, presenting epistemic uncertainty alone might be preferable. For example, in medical diagnosis systems, presenting only epistemic uncertainty might be preferable when the AI encounters a rare condition in which it has only limited training data on [7, 19, 59]. This approach would alert clinicians to the model's knowledge gaps, prompting them not to rely on AI advice and, instead, consult other experts. Conversely, in real estate pricing, presenting both aleatoric and epistemic uncertainty could be relevant [55] as the aleatoric uncertainty would capture inherent market volatility or missing information, while epistemic uncertainty would indicate the model's uncertainty in its predictions based on historical data.

One approach to balancing these considerations might be to use a layered method, where humans are first presented with high-level uncertainty information and can then drill down into more specific sources of uncertainty as needed [86]. This could help maintain the salience of epistemic uncertainty as an indicator of system reliability while still providing detailed uncertainty information when required. By helping humans understand these different sources of uncertainty and their implications, designers can support more informed decisions about when and how to rely on AI recommendations.

Mitigate Effects of Prior Beliefs through Design. Our findings on the interaction between prior beliefs and uncertainty levels underscore the need for systems that actively encourage humans to challenge their assumptions and consider alternative perspectives. This is particularly relevant in high-uncertainty scenarios, where humans are more prone to overreliance on their initial judgments. Therefore, designers should incorporate features that emphasize the increased range of possible outcomes and prompt humans to explore multiple scenarios before making final decisions.

Interestingly, this consideration is closely tied to the challenge of uncertainty presentation. By presenting uncertainty information in a way that highlights both the AI's limitations (through epistemic uncertainty) and the inherent variability in the data (through aleatoric uncertainty), designers can create systems that naturally encourage humans to think more critically about their decisions. For example, a system could use visual cues to highlight discrepancies between a human's initial estimate and the AI's advice [21, 101], especially when the human's estimate falls within a high-uncertainty range. Interactive visualizations could allow humans to explore how varying features affect AI advice and, consequently, levels of epistemic and aleatoric uncertainty [75], thereby increasing decision-makers' engagement with the full spectrum of possibilities. These approaches not only provide transparency but also actively encourage humans to grapple with the complexities of AI-assisted decision-making. Other works have adapted the uncertainty presented to human decision-makers to account for existing biases [103].

To summarize, our findings emphasize the need for careful selection and presentation of uncertainty in AI-assisted decision-making systems. By carefully considering how to present different sources of uncertainty, and encouraging critical thinking, designers and organizations can create more effective and trustworthy human–AI collaborative environments.

6.3 Limitations and Future Research

While our study provides valuable insights into the effects of uncertainty decomposition on human–AI decision-making, it is essential to acknowledge the limitations of our work and identify avenues for future research.

Presentation of Uncertainty Information. While our study used textual presentation of uncertainty to make it more accessible, HCI research has explored various graphical methods for communicating uncertainty [20, 27, 40, 49, 58]. Future research should explore alternative methods of communicating uncertainty, such as visualizations, which may improve participants' understanding and interpretation of the uncertainty associated with AI advice. For example, graphical representations of uncertainty, such as confidence intervals or probability distributions [40, 41, 68], could be used to provide a more intuitive depiction of the uncertainty while emphasizing the range of possible outcomes. Besides the actual communication method, understanding of uncertainties could also be supported by visualizations that resemble SHAP-like explanations [60], attributing uncertainty to specific features. This approach would allow humans to see not only the AI's confidence in its predictions but also how uncertainty results from individual feature values and influences the final decision. These visualizations could complement the framing of epistemic uncertainty as the AI's confidence in its predictions, potentially enhancing participants' comprehension of the distinct nature of epistemic uncertainty.

Appropriate Reliance and Performance. A limitation of our study is its focus on reliance behavior without examining whether this reliance was appropriate or led to improved decision-making outcomes. While previous HCI researchers have investigated the concept of "appropriateness of reliance" in human–AI collaboration [6, 61, 82], our study specifically focused on understanding how uncertainty decomposition affects reliance patterns. This focus on reliance behavior provides valuable insights into how humans cognitively process and respond to different presentations of uncertainty information. Our findings suggest that decomposed uncertainty may be particularly valuable in high-uncertainty situations, where we observed increased advice rejection compared to combined uncertainty presentation. Building on these behavioral insights, future work should investigate how uncertainty decomposition influences the appropriateness of reliance and overall decision-making performance. Such research would help determine whether and how different approaches to presenting decomposed uncertainty can actually improve the quality of human–AI collaborative decision-making.

Simulated AI Advice. While our simulated AI advice was grounded in a real AI model trained on housing data (see Appendix A.1), we systematically varied the uncertainty levels to examine their effects on reliance behavior. This controlled approach allowed us to systematically study human behavior across different uncertainty combinations, which would be difficult with real-world model outputs. However, a limitation that arises from this design decision is that both uncertainty sources are independent in our presentation. In practice, these uncertainty sources are often interdependent: when a model encounters data points that significantly differ from its training distribution (i.e., having high epistemic uncertainty), its ability to accurately estimate aleatoric uncertainty may also become less reliable, as limited model knowledge can impact the ability to accurately capture the variability of data. Future work should, therefore, use real-world AI models with analytically computed uncertainties to investigate both how natural uncertainty patterns influence reliance behavior and how the interaction between different sources of uncertainty affects human trust and decision-making.

Task Selection, Domain Expertise, and Confidence. The task of estimating housing prices is well established in HCI research and well suited for studying human reliance on AI under uncertainty. This setup allowed for a controlled investigation of uncertainty communication effects; however, it represents only one source of AI-assisted decision-making scenario. The generalizability of our findings to domains with different risk profiles, cognitive demands, or decision consequences—such as medical diagnosis, legal judgment, or financial investing—remains an open question. Additionally, we did not control for participants' domain expertise in housing price estimation, which may have influenced how they interpreted and responded to uncertainty information. We focused exclusively on the behavioral measure of reliance (WOA) and deliberately excluded subjective measures of confidence and perceived expertise due to the study's considerable cognitive load: participants were introduced to the complex concepts of aleatoric and epistemic uncertainty and processed numerous decision instances. Future work could explore how domain-specific factors as well as participants' self-assessed uncertainty or expertise interact with uncertainty communication strategies across various high-stakes contexts.

To summarize, while our study provides valuable insights into the effects of uncertainty decomposition on reliance, further research is needed to address the limitations of our work and explore alternative methods of communicating uncertainty to identify effective strategies for integrating uncertainty information into human–AI decision-making processes.

7 Conclusion

This study investigates the effects of uncertainty framing and decomposition on human reliance in AI-assisted decision-making through two experiments using a housing price estimation task. Our findings yield several insights into human–AI collaboration under uncertainty, demonstrating that the framing and presentation of uncertainty influence humans' reliance behavior.

First, we observe a general uncertainty effect, where decision-makers consistently rely less on AI advice as uncertainty increases. Second, we find a *level-dependent differentiation strategy* when presenting a single source of uncertainty: participants show a more pronounced decrease in reliance for epistemic uncertainty compared to aleatoric uncertainty. Surprisingly, this distinction diminishes when both uncertainty sources are presented simultaneously, resulting in a *generic differentiation strategy*. We also observe a complex interplay between the degree and source of uncertainty and prior beliefs, creating a tradeoff between confirming prior beliefs and decreased reliance on specific AI advice as uncertainty increases.

These findings have important implications for the design of AI-assisted decision-making systems, emphasizing the need for thoughtful uncertainty communication strategies. Our research suggests that system designers must carefully consider how to present different sources of uncertainty and mitigate the effects of prior beliefs. Future research should explore alternative methods of jointly communicating aleatoric and epistemic uncertainty and examine how uncertainty decomposition affects decision quality and performance outcomes. By addressing these challenges and opportunities, we may create more effective human–AI collaborative environments that support informed decision-making across domains.

Acknowledgments

Generative AI tools were utilized throughout this work. Specifically, ChatGPT, Claude, and GitHub Copilot were employed to generate code for visualizations. Additionally, ChatGPT, Claude, DeepL Write, and Grammarly were used to enhance the writing quality of tutorials and explanations provided to participants during the experiments, as well as to improve the language across all sections of this article.

References

[1] James D. Abbey and Margaret G. Meloy. 2017. Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management* 53 (2017), 63–70.

[2] Ilke Aydogan. 2021. Prior beliefs and ambiguity attitudes in decision from experience. *Management Science* 67, 11 (2021), 6934–6945.

[3] Phoebe E. Bailey, Tarren Leon, Natalie C. Ebner, Ahmed A. Moustafa, and Gabrielle Weidemann. 2023. A meta-analysis of the weight of advice in decision-making. *Current Psychology* 42, 28 (2023), 24516–24541.

[4] Howard Balshem, Mark Helfand, Holger J. Schünemann, Andrew D. Oxman, Regina Kunz, Jan Brozek, Gunn E. Vist, Yngve Falck-Ytter, Joerg Meerpohl, Susan Norris, et al. 2011. GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology* 64, 4 (2011), 401–406.

[5] Nikola Banovic, Zhuoran Yang, Aditya Ramesh, and Alice Liu. 2023. Being trustworthy is not enough: How untrustworthy artificial intelligence (AI) can deceive the end-users and gain their trust. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–17.

[6] Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. 2021. Does the whole exceed its parts? The effect of AI explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–16.

[7] Edmon Begoli, Tanmoy Bhattacharya, and Dimitri Kusnezov. 2019. The need for uncertainty quantification in machine-assisted medical decision making. *Nature Machine Intelligence* 1, 1 (2019), 20–23.

[8] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: series B (Methodological)* 57, 1 (1995), 289–300.

[9] Astrid Bertrand, Rafik Belloum, James R. Eagan, and Winston Maxwell. 2022. How cognitive biases affect XAI-assisted decision-making: A systematic review. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 78–91.

[10] Umang Bhatt, Javier Antorán, Yunfeng Zhang, Q. Vera Liao, Prasanna Sattigeri, Riccardo Fogliato, Gabrielle Melançon, Ranganath Krishnan, Jason Stanley, Omesh Tickoo, et al. 2021. Uncertainty as a form of transparency: Measuring, communicating, and using uncertainty. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 401–413.

[11] Jessica Y. Bo, Pan Hao, and Brian Y. Lim. 2024. Incremental XAI: Memorable understanding of AI with incremental explanations. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–17.

[12] Robin Bodenberger and Kirsten Thommes. 2025. Words or numbers? How framing uncertainties affects risk assessment and decision-making. arXiv:2502.06241. Retrieved from https://arxiv.org/abs/2502.06241

[13] Silvia Bonaccio and Reeshad S. Dalal. 2006. Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes* 101, 2 (2006), 127–151.

[14] Darrell L. Brown and Donald R. Jones. 1998. Factors that influence reliance on decision aids: A model and an experiment. *Journal of Information Systems* 12, 2 (1998).

[15] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–21.

[16] Shiye Cao, Anqi Liu, and Chien-Ming Huang. 2024. Designing for appropriate reliance: The roles of ai uncertainty presentation, initial user decision, and user demographics in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–32.

[17] Federico Maria Cau, Hanna Hauptmann, Lucio Davide Spano, and Nava Tintarev. 2023. Effects of AI and logic-style explanations on users' decisions under different levels of uncertainty. *ACM Transactions on Interactive Intelligent Systems* 13, 4 (2023), 1–42.

[18] Federico Maria Cau, Hanna Hauptmann, Lucio Davide Spano, and Nava Tintarev. 2023. Supporting high-uncertainty decisions through AI and logic-style explanations. In *Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23)*. ACM, New York, NY, 251–263.

[19] Michelle Chua, Doyun Kim, Jongmun Choi, Nahyoung G. Lee, Vikram Deshpande, Joseph Schwab, Michael H. Lev, Ramon G. Gonzalez, Michael S. Gee, and Synho Do. 2023. Tackling prediction uncertainty in machine learning for healthcare. *Nature Biomedical Engineering* 7, 6 (2023), 711–718.

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

[21] Michael Correll, Dominik Moritz, and Jeffrey Heer. 2018. Value-suppressing uncertainty palettes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–11.

[22] Jessica Maria Echterhoff, Matin Yarmand, and Julian McAuley. 2022. AI-moderated decision-making: Capturing and balancing anchoring bias in sequential decision tasks. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–9.

[23] Sven Eckhardt, Niklas Kühl, Mateusz Dolata, and Gerhard Schwabe. 2024. A survey of AI reliance. arXiv:2408.03948. Retrieved from https://arxiv.org/abs/2408.03948

[24] Daniel Ellsberg. 1961. Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics* 75, 4 (1961), 643–669.

[25] Martin Eppler and Jeanne Mengis. 2004. The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The Information Society* 20 (Nov. 2004), 325–344.

[26] Shi Feng and Jordan Boyd-Graber. 2019. What can AI do for me? Evaluating machine learning interpretations in cooperative play. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 229–239.

[27] Michael Fernandes, Logan Walls, Sean Munson, Jessica Hullman, and Matthew Kay. 2018. Uncertainty displays using quantile dotplots or CDFs improve transit decision-making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, 1–12.

[28] Craig R. Fox and Amos Tversky. 1995. Ambiguity aversion and comparative ignorance. *The Quarterly Journal of Economics* 110, 3 (1995), 585–603.

[29] Craig R. Fox and Gülden Ülkümen. 2011. Distinguishing two dimensions of uncertainty. In *Essays in Judgment and Decision Making*. W. Brun, G. Kirkebøen, and H. Montgomery (Eds.), Universitetsforlaget.

[30] Yarin Gal. 2016. *Uncertainty in Deep Learning*. Ph.D. Dissertation. University of Cambridge, Cambridge.

[31] Yarin Gal, Petros Koumoutsakos, Francois Lanusse, Gilles Louppe, and Costas Papadimitriou. 2022. Bayesian uncertainty quantification for machine-learned models in physics. *Nature Reviews Physics* 4, 9 (2022), 573–577.

[32] Mirta Galesic and Rocio Garcia-Retamero. 2010. Statistical numeracy for health: A cross-cultural comparison with probabilistic national samples. *Archives of Internal Medicine* 170, 5 (2010), 462–468.

[33] Jakob Gawlikowski, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, et al. 2023. A survey of uncertainty in deep neural networks. *Artificial Intelligence Review* 56, Suppl 1 (2023), 1513–1589.

[34] Miriam Greis, Jessica Hullman, Michael Correll, Matthew Kay, and Orit Shaer. 2017. Designing for uncertainty in HCI: When does uncertainty help? In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 593–600.

[35] Shunan Guo, Fan Du, Sana Malik, Eunyee Koh, Sungchul Kim, Zhicheng Liu, Donghyun Kim, Hongyuan Zha, and Nan Cao. 2019. Visualizing uncertainty and alternatives in event sequence predictions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12.

[36] Taehyun Ha and Sangyeon Kim. 2024. Improving trust in AI with mitigating confirmation bias: Effects of explanation type and debiasing strategy for decision-making with explainable AI. *International Journal of Human–Computer Interaction* 40, 24 (2024), 8562–8573.

[37] Uriel Haran, Asaf Mazar, Mordechai Hurwitz, and Simone Moran. 2022. Confidently at your service: Advisors alter their stated confidence to be helpful. *Organizational Behavior and Human Decision Processes* 171 (2022), 104154.

[38] Patrick Hemmer, Max Schemmer, Niklas Kühl, Michael Vössing, and Gerhard Satzger. 2024. Complementarity in human-ai collaboration: Concept, sources, and evidence. arXiv:2404.00029. Retrieved from https://arxiv.org/abs/2404.00029

[39] Eyke Hüllermeier and Willem Waegeman. 2021. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning* 110, 3 (2021), 457–506.

[40] Jessica Hullman, Xiaoli Qiao, Michael Correll, Alex Kale, and Matthew Kay. 2018. In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 903–913.
[41] Jessica Hullman, Paul Resnick, and Eytan Adar. 2015. Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PLoS One* 10, 11 (2015), e0142444.
[42] Eric J. Johnson and John W. Payne. 1985. Effort and accuracy in choice. *Management Science* 31, 4 (1985), 395–414.
[43] Daniel Kahneman and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47, 2 (1979), 263–291.
[44] Patricia K. Kahr, Gerrit Rooks, Chris Snijders, and Martijn C. Willemsen. 2024. The trust recovery journey. The effect of timing of errors on the willingness to follow AI advice. In *Proceedings of the 29th International Conference on Intelligent User Interfaces (IUI '24)*. ACM, New York, NY, 609–622. DOI: https://doi.org/10.1145/3640543.3645167
[45] Patricia K. Kahr, Gerrit Rooks, Martijn C. Willemsen, and Chris C. P. Snijders. 2023. It seems smart, but it acts stupid: Development of trust in AI advice in a repeated legal decision-making task. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 528–539.
[46] Patricia K. Kahr, Gerrit Rooks, Martijn C. Willemsen, and Chris C. P. Snijders. 2024. Understanding trust and reliance development in AI advice: Assessing model accuracy, model explanations, and experiences from previous interactions. *ACM Transactions on Interactive Intelligent Systems* 14, 4 (2024), 1–30.
[47] Alex Kale, Matthew Kay, and Jessica Hullman. 2019. Decision-making under uncertainty in research synthesis: Designing for the garden of forking paths. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14.
[48] Graham Kalton, Julie Roberts, and David Holt. 1980. The effects of offering a middle response option with opinion questions. *Journal of the Royal Statistical Society: Series D (The Statistician)* 29, 1 (1980), 65–78.
[49] Matthew Kay, Tara Kola, Jessica R. Hullman, and Sean A. Munson. 2016. When (ish) is my bus? User-centered visualizations of uncertainty in everyday, mobile predictive systems. In *Proceedings of the 2016 Chi Conference on Human Factors in Computing Systems*, 5092–5103.
[50] Alex Kendall and Yarin Gal. 2017. What uncertainties do we need in Bayesian deep learning for computer vision? In *Proceedings of the Advances in Neural Information Processing Systems*, Vol. 30, 2017.
[51] Sunnie S. Y. Kim, Q. Vera Liao, Mihaela Vorvoreanu, Stephanie Ballard, and Jennifer Wortman Vaughan. 2024. "I'm not sure, but...": Examining the impact of large language models' uncertainty expression on user reliance and trust. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 822–835.
[52] Vivian Lai, Chacha Chen, Q. Vera Liao, Alison Smith-Renner, and Chenhao Tan. 2021. Towards a science of human-AI decision making: A survey of empirical studies. arXiv:2112.11471. Retrieved from https://arxiv.org/abs/2112.11471
[53] Jeremy Larcher. 2023. American House Prices. Retrieved from https://www.kaggle.com/dsv/7162651
[54] Thao Le. 2023. Explaining the uncertainty in AI-assisted decision making. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37, 16119–16120.
[55] Changro Lee and Keith Key-Ho Park. 2020. Representing uncertainty in property valuation through a Bayesian deep learning approach. *Real Estate Management and Valuation* 28, 4 (2020), 15–23.
[56] Dirk Leffrang and Oliver Müller. 2021. Should I follow this model? The effect of uncertainty visualization on the acceptance of time series forecasts. In *Proceedings of the 2021 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX)*. IEEE, 20–26.
[57] Jingshu Li, Yitian Yang, Q. Vera Liao, Junti Zhang, and Yi-Chieh Lee. 2025. As confidence aligns: Understanding the effect of AI confidence on human self-confidence in human-AI decision making. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, New York, NY, Article 1111, 16 pages. DOI: https://doi.org/10.1145/3706598.3713336
[58] Le Liu, Lace Padilla, Sarah H. Creem-Regehr, and Donald H. House. 2019. Visualizing uncertain tropical cyclone predictions using representative samples from ensembles of forecast tracks. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 882–891.
[59] Tyler J. Loftus, Benjamin Shickel, Matthew M. Ruppert, Jeremy A. Balch, Tezcan Ozrazgat-Baslanti, Patrick J. Tighe, Philip A. Efron, William R. Hogan, Parisa Rashidi, Gilbert R. Upchurch, Jr, et al. 2022. Uncertainty-aware deep learning in healthcare: A scoping review. *PLoS Digital Health* 1, 8 (2022), e0000085.
[60] Scott M. Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS '17)*. Curran Associates Inc., Red Hook, NY, 4768–4777.
[61] Shuai Ma, Xinru Wang, Ying Lei, Chuhan Shi, Ming Yin, and Xiaojuan Ma. 2024. "Are you really sure?" Understanding the effects of human self-confidence calibration in AI-assisted decision making. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, New York, NY, Article 840, 20 pages.
[62] Laura R. Marusich, Jonathan Z. Bakdash, Yan Zhou, and Murat Kantarcioglu. 2023. Using AI uncertainty quantification to improve human decision-making. arXiv:2309.10852. Retrieved from https://arxiv.org/abs/2309.10852

[63] Sean McGrath, Parth Mehta, Alexandra Zytek, Isaac Lage, and Himabindu Lakkaraju. 2023. When does uncertainty matter? Understanding the impact of predictive uncertainty in ML assisted decision making. *Transactions on Machine Learning Research* (2023). Retrieved from https://openreview.net/forum?id=pbs22kJmEO

[64] Natalja Menold and Kathrin Bogner. 2016. Design of rating scales in questionnaires. *GESIS Survey Guidelines* 4 (2016), 2.

[65] Don A. Moore and Paul J. Healy. 2008. The trouble with overconfidence. *Psychological Review* 115, 2 (2008), 502.

[66] Katelyn Morrison, Philipp Spitzer, Violet Turri, Michelle Feng, Niklas Kühl, and Adam Perer. 2024. The impact of imperfect XAI on human-AI decision-making. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–39.

[67] Raymond S. Nickerson. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* 2, 2 (1998), 175–220.

[68] Lace M. K. Padilla, Maia Powell, Matthew Kay, and Jessica Hullman. 2021. Uncertain about uncertainty: How qualitative expressions of forecaster confidence impact decision-making with uncertainty visualizations. *Frontiers in Psychology* 11 (2021), 579267.

[69] Stefan Palan and Christian Schitter. 2018. Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance* 17 (2018), 22–27.

[70] Raja Parasuraman and Christopher A. Miller. 2004. Trust and etiquette in high-criticality automated systems. *Communications of the ACM* 47, 4 (2004), 51–55.

[71] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology* 70 (2017), 153–163.

[72] Ellen Peters, Judith Hibbard, Paul Slovic, and Nathan Dieckmann. 2007. Numeracy skill and the communication, comprehension, and use of risk-benefit information. *Health Affairs (Project Hope)* 26, 3 (2007), 741–748.

[73] Dane K. Peterson and Gordon F. Pitz. 1988. Confidence, uncertainty, and the use of information. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14, 1 (1988), 85.

[74] Forough Poursabzi-Sangdeh, Daniel G. Goldstein, Jake M. Hofman, Jennifer Wortman Wortman Vaughan, and Hanna Wallach. 2021. Manipulating and measuring model interpretability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–52.

[75] Snehal Prabhudesai, Leyao Yang, Sumit Asthana, Xun Huan, Q. Vera Liao, and Nikola Banovic. 2023. Understanding uncertainty: How lay decision-makers perceive and interpret uncertainty in human-AI decision making. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 379–396.

[76] Svetlozar Todorov Rachev, Stoyan V. Stoyanov, and Frank J. Fabozzi. 2008. *Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization: The Ideal Risk, Uncertainty, and Performance Measures.* John Wiley & Sons, Hoboken.

[77] Emely Rosbach, Jonas Ammeling, Sebastian Krügel, Angelika Kießig, Alexis Fritz, Jonathan Ganz, Chloé Puget, Taryn Donovan, Andrea Klang, Maximilian C. Köller, et al. 2025. "When two wrongs don't make a right"—Examining confirmation bias and the role of time pressure during human-AI collaboration in computational pathology. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25).* ACM, New York, NY, Article 528, 18 pages.

[78] Sara Salimzadeh, Gaole He, and Ujwal Gadiraju. 2024. Dealing with uncertainty: Understanding the impact of prognostic versus diagnostic tasks on trust and reliance in human-AI decision making. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–17.

[79] Abhraneel Sarma, Maryam Hedayati, and Matthew Kay. 2025. More forecasts, more (decision) problems: How uncertainty representations for multiple forecasts impact decision-making. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–14.

[80] Abhraneel Sarma, Xiaoying Pu, Yuan Cui, Michael Correll, Eli T. Brown, and Matthew Kay. 2024. Odds and insights: Decision quality in exploratory data analysis under uncertainty. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–14.

[81] Aaron Schecter, Eric Bogert, and Nina Lauharatanahirun. 2023. Algorithmic appreciation or aversion? The moderating effects of uncertainty on algorithmic decision making. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–8.

[82] Max Schemmer, Niklas Kuehl, Carina Benz, Andrea Bartos, and Gerhard Satzger. 2023. Appropriate reliance on AI advice: Conceptualization and the effect of explanations. In *Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23).* ACM, New York, NY, 410–422.

[83] Björna Schmitz, Gerhard Satzger, and Ralf Gitzel. 2017. More observations, more variables or more quality? Data acquisition strategies to enhance uncertainty analytics for industrial service contracting. In *Exploring Services Science*. Stefano Za, Monica Drăgoicea, and Maurizio Cavallari (Eds.), Springer International Publishing, Cham, 159–172.

[84] Jakob Schoeffer, Maria De-Arteaga, and Niklas Kühl. 2024. Explanations, fairness, and appropriate reliance in human-AI decision-making. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24).* ACM, New York, NY, Article 836, 18 pages. DOI: https://doi.org/10.1145/3613904.3642621

[85] Maximilian Seitzer, Arash Tavakoli, Dimitrije Antic, and Georg Martius. 2022. On the pitfalls of heteroscedastic uncertainty estimation with probabilistic neural networks. In *Proceedings of the International Conference on Learning Representations*.

[86] Ben Shneiderman. 2003. The eyes have it: A task by data type taxonomy for information visualizations. In *The Craft of Information Visualization*. Benjamin B. Bederson and Ben Shneiderman (Eds.), Morgan Kaufmann, 364–371.

[87] Janet A. Sniezek and Timothy Buckley. 1995. Cueing and cognitive conflict in judge-advisor decision making. *Organizational Behavior and Human Decision Processes* 62, 2 (1995), 159–174.

[88] Robert Soden, Laura Devendorf, Richmond Y. Wong, Lydia B. Chilton, Ann Light, and Yoko Akama. 2020. Embracing uncertainty in HCI. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–8.

[89] Philipp Spitzer, Joshua Holstein, Patrick Hemmer, Michael Vössing, Niklas Kühl, Dominik Martin, and Gerhard Satzger. 2025. Human delegation behavior in human-AI collaboration: The effect of contextual information. *Proceedings of the ACM on Human-Computer Interaction* 9, 2 (2025), 1–28.

[90] Philipp Spitzer, Joshua Holstein, Michael Vössing, and Niklas Kühl. 2023. On the perception of difficulty: Differences between humans and AI. arXiv:2304.09803. Retrieved from https://arxiv.org/abs/2304.09803

[91] Kimberly Stowers, Nicholas Kasdaglis, Michael Rupp, Jessie Chen, Daniel Barber, and Michael Barnes. 2017. Insights into human-agent teaming: Intelligent agent transparency and uncertainty. In *Advances in Human Factors in Robots and Unmanned Systems: Proceedings of the AHFE 2016 International Conference on Human Factors in Robots and Unmanned Systems*. Springer, 149–160.

[92] John Sweller. 1994. Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction* 4, 4 (1994), 295–312.

[93] Maxwell Szymanski, Martijn Millecamp, and Katrien Verbert. 2021. Visual, textual or hybrid: The effect of user expertise on different explanations. In *Proceedings of the 26th International Conference on Intelligent User Interfaces*, 109–119.

[94] Anna Taudien, Dominik David Walzner, Andreas Fuegener, Alok Gupta, and Wolfgang Ketter. 2024. Know thyself: The relationship between metacognition and human-AI collaboration. In *ICIS 2024 Proceedings*, 7 pages.

[95] Alexander Treiss, Jannis Walk, and Niklas Kühl. 2021. An uncertainty-based human-in-the-loop system for industrial tool wear analysis. In *Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track*. Yuxiao Dong, Georgiana Ifrim, Dunja Mladenić, Craig Saunders, and Sofie Van Hoecke (Eds.), Springer International Publishing, Cham, 85–100.

[96] Amos Tversky, Daniel Kahneman, and Paul Slovic. 1982. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge.

[97] Matias Valdenegro-Toro and Daniel Saromo Mori. 2022. A deeper look into aleatoric and epistemic uncertainty disentanglement. In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 1508–1516.

[98] Niels Van Berkel, Maura Bellio, Mikael B. Skov, and Ann Blandford. 2023. Measurements, algorithms, and presentations of reality: Framing interactions with AI-enabled decision support. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–33.

[99] Anne Marthe Van Der Bles, Sander Van Der Linden, Alexandra L. J. Freeman, James Mitchell, Ana B. Galvao, Lisa Zaval, and David J. Spiegelhalter. 2019. Communicating uncertainty about facts, numbers and science. *Royal Society Open Science* 6, 5 (2019), 181870.

[100] Helena Vasconcelos, Gagan Bansal, Adam Fourney, Q. Vera Liao, and Jennifer Wortman Vaughan. 2025. Generation probabilities are not enough: Uncertainty highlighting in AI code completions. *ACM Transactions on Computer-Human Interaction* 32, 1 (Feb. 2025), 30.

[101] Emily Wall, John Stasko, and Alex Endert. 2019. Toward a design space for mitigating cognitive bias in vis. In *Proceedings of the 2019 IEEE Visualization Conference (VIS)*. IEEE, 111–115.

[102] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15.

[103] Fumeng Yang, Maryam Hedayati, and Matthew Kay. 2023. Subjective probability correction for uncertainty representations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–17.

[104] Ilan Yaniv. 2004. Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes* 93, 1 (2004), 1–13.

[105] Ilan Yaniv and Eli Kleinberger. 2000. Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes* 83, 2 (2000), 260–281.

[106] Yunfeng Zhang, Q. Vera Liao, and Rachel K. E. Bellamy. 2020. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 295–305.

[107] Jieqiong Zhao, Yixuan Wang, Michelle V. Mancenido, Erin K. Chiou, and Ross Maciejewski. 2024. Evaluating the impact of uncertainty visualization on model reliance. *IEEE Transactions on Visualization and Computer Graphics* 30, 7 (Jul. 2024), 4093–4107.

A Appendix

A.1 Training of AI Model for Advice Generation

To simulate the AI advice, we train an AI model and base our recommendations on the model's predictions and uncertainty estimates. The data preprocessing and feature selection are performed differently compared to features shown to participants to enable the estimation of meaningful uncertainties.

Data Preprocessing. First, we preprocess the data by creating a new feature, rooms, which combines the beds and baths features. Additionally, we use the city as a categorical variable and convert it into binary features using one-hot encoding. In contrast to the advice, we omit the living space feature to facilitate the calculation of meaningful uncertainties. The dataset is split into **in-distribution (IID)** and **out-of-distribution (OOD)** subsets based on the number of rooms, with data points having more than five rooms considered OOD. The IID data is further divided into train (80%), validation (10%), and test sets (10%), while the OOD data is added to the test set to evaluate the model's performance on unseen data. Numerical features are scaled to ensure consistent scaling across the datasets.

Model Training and Evaluation. The model training process involves exploring various hyperparameter configurations, including the network architecture, regularization techniques, and learning rates. The model's performance is evaluated using the R2-score during training and on the IID and OOD test sets during evaluation.

The best-performing model for each hyperparameter configuration is saved and analyzed to inform the advice. We design a custom neural network architecture with varying depths, consisting of fully connected layers and dropout for being able to calculate epistemic uncertainty. To quantify aleatoric uncertainty, we employ a distribution layer at the end of the network that predicts the standard deviation alongside the mean of a normal distribution. As a loss function, we use the beta-NLL loss as proposed by [85], which improves the stability of standard deviation estimates. To finally estimate aleatoric and epistemic uncertainty, we perform 25 forward passes while randomly dropping neurons in each pass. The mean of the standard deviations then represents the aleatoric uncertainty while the standard deviation of the means resembles the epistemic uncertainty [50]. We find that epistemic uncertainty is usually higher for instances that are rare and are further away from the majority of instances, i.e., small houses in high-income areas or large houses in low-income areas (see Figure A1). At the same time, aleatoric uncertainty is more equally distributed across instances compared to epistemic uncertainty.

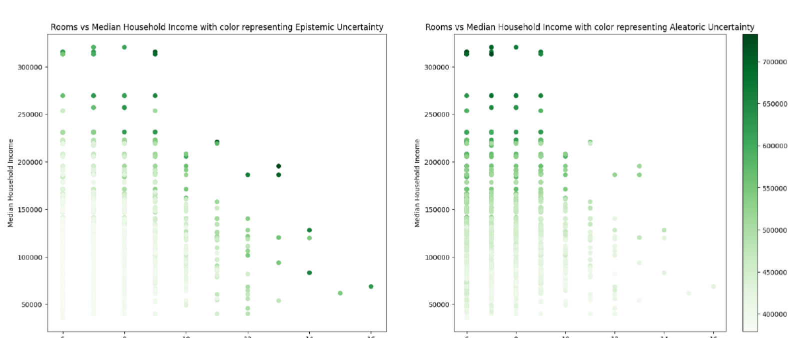


Fig. A1. Distribution of aleatoric and epistemic uncertainty across income and number of rooms.

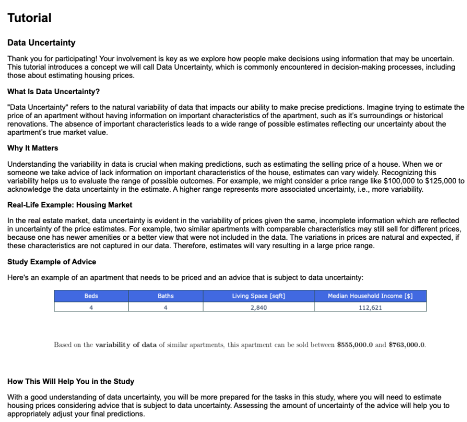
A.2 Advice Generation Procedure

For our second experiment, the AI advice presented to participants is generated based on the trained AI model described in Appendix A.1. However, we simulate the advice to allow for a systematic exploration of how different uncertainty levels and decomposition affect user reliance on AI recommendations. The simulation process follows these steps:

— The proportion of total uncertainty attributable to epistemic versus aleatoric sources is sampled from a truncated ([0,1]) normal distribution $\mathcal{N}(0.5, 0.2)$. This ensures that the proportion is normally distributed with a mean of 0.5. The complementary proportion is assigned to aleatoric uncertainty.
— Total uncertainty levels are generated from 0.05 to 0.9 in increments of 0.05, covering a range from very low to very high uncertainty. The absolute magnitude of each uncertainty source is calculated by multiplying the total uncertainty by the proportion attributable to that source.
— Four levels (very low, low, high, very high) are defined for each uncertainty source based on the quartiles of its magnitude distribution. Given the unipolar scale and the ambiguity of midpoints observed in questionnaire items, we omit the middle level, i.e., medium uncertainty [48, 64]. This ensures that all possible combinations of uncertainty levels for each source are represented in the simulation.
— To mirror the patterns observed in the trained model, epistemic uncertainty is assigned to be higher for OOD instances (e.g., very high-income areas or large houses in low-income neighborhoods), while aleatoric uncertainty is evenly distributed across features.
— To induce a realistic correlation between epistemic uncertainty and error magnitude, the AI estimate is shifted by a random amount drawn from $\mathcal{N}(0, 0.1)$ plus an additional term that scales with epistemic uncertainty by a correlation factor [95]. This also avoids the midpoint of the advised interval always reflecting the true price.
— The simulation generates 160 total instances, with each combination of aleatoric and epistemic uncertainty levels equally represented (10 instances each).

Based on these instances, we then generate the visualization for both treatments.

A.3 Tutorials, Comprehension Checks, and Advice



Tutorial

Data Uncertainty

Thank you for participating! Your involvement is key as we explore how people make decisions using information that may be uncertain. This tutorial introduces a concept we will call Data Uncertainty, which is commonly encountered in decision-making processes, including those about estimating housing prices.

What Is Data Uncertainty?

"Data Uncertainty" refers to the natural variability of data that impacts our ability to make precise predictions. Imagine trying to estimate the price of an apartment without having information on important characteristics of the apartment, such as it's surroundings or historical renovations. The absence of important characteristics leads to a wide range of possible estimates reflecting our uncertainty about the apartment's true market value.

Why It Matters

Understanding the variability in data is crucial when making predictions, such as estimating the selling price of a house. When we or someone we take advice of lack information on important characteristics of the house, estimates can vary widely. Recognizing this variability helps us to evaluate the range of possible outcomes. For example, we might consider a price range like $100,000 to $125,000 to acknowledge the data uncertainty in the estimate. A higher range represents more associated uncertainty, i.e., more variability.

Real-Life Example: Housing Market

In the real estate market, data uncertainty is evident in the variability of prices given the same, incomplete information which are reflected in uncertainty of the price estimates. For example, two similar apartments with comparable characteristics may still sell for different prices, because one has newer amenities or a better view that were not included in the data. The variations in prices are natural and expected, if these characteristics are not captured in our data. Therefore, estimates will vary resulting in a large price range.

Study Example of Advice

Here's an example of an apartment that needs to be priced and an advice that is subject to data uncertainty:

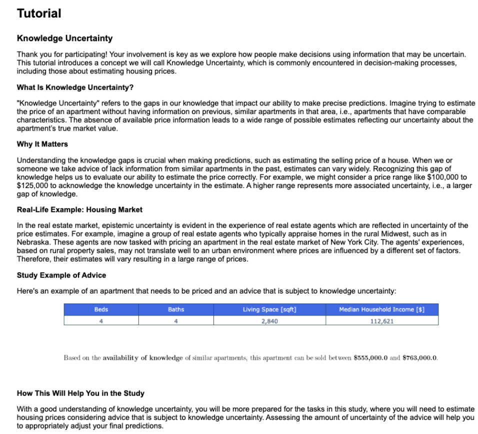
Beds	Baths	Living Space [sqft]	Median Household Income [$]
4	4	2,840	112,621

Based on the **variability of data** of similar apartments, this apartment can be sold between **$555,000.0** and **$763,000.0**.

How This Will Help You in the Study

With a good understanding of data uncertainty, you will be more prepared for the tasks in this study, where you will need to estimate housing prices considering advice that is subject to data uncertainty. Assessing the amount of uncertainty of the advice will help you to appropriately adjust your final predictions.

(a) Tutorial used in the first experiment to explain the concept of aleatoric uncertainty.



Tutorial

Knowledge Uncertainty

Thank you for participating! Your involvement is key as we explore how people make decisions using information that may be uncertain. This tutorial introduces a concept we will call Knowledge Uncertainty, which is commonly encountered in decision-making processes, including those about estimating housing prices.

What Is Knowledge Uncertainty?

"Knowledge Uncertainty" refers to the gaps in our knowledge that impact our ability to make precise predictions. Imagine trying to estimate the price of an apartment without having information on previous, similar apartments in that area, i.e., apartments that have comparable characteristics. The absence of available price information leads to a wide range of possible estimates reflecting our uncertainty about the apartment's true market value.

Why It Matters

Understanding the knowledge gaps is crucial when making predictions, such as estimating the selling price of a house. When we or someone we take advice of lack information from similar apartments in the past, estimates can vary widely. Recognizing this gap of knowledge helps us to evaluate our ability to estimate the price correctly. For example, we might consider a price range like $100,000 to $125,000 to acknowledge the knowledge uncertainty in the estimate. A higher range represents more associated uncertainty, i.e., a larger gap of knowledge.

Real-Life Example: Housing Market

In the real estate market, epistemic uncertainty is evident in the experience of real estate agents which are reflected in uncertainty of the price estimates. For example, imagine a group of real estate agents who typically appraise homes in the rural Midwest, such as in Nebraska. These agents are now tasked with pricing an apartment in the real estate market of New York City. The agents' experiences, based on rural property sales, may not translate well to an urban environment where prices are influenced by a different set of factors. Therefore, their estimates will vary resulting in a large range of prices.

Study Example of Advice

Here's an example of an apartment that needs to be priced and an advice that is subject to knowledge uncertainty:

Beds	Baths	Living Space [sqft]	Median Household Income [$]
4	4	2,840	112,621

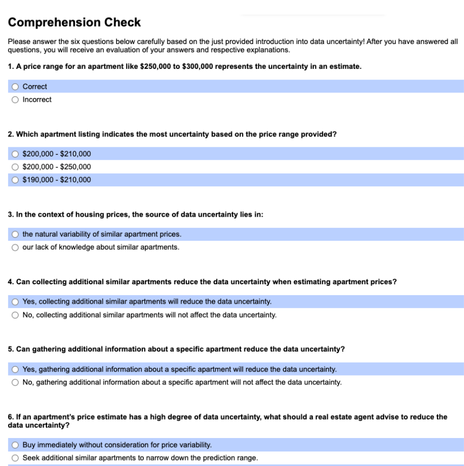
Based on the **availability of knowledge** of similar apartments, this apartment can be sold between **$555,000.0** and **$763,000.0**.

How This Will Help You in the Study

With a good understanding of knowledge uncertainty, you will be more prepared for the tasks in this study, where you will need to estimate housing prices considering advice that is subject to knowledge uncertainty. Assessing the amount of uncertainty of the advice will help you to appropriately adjust your final predictions.

(b) Tutorial used in the first experiment to explain the concept of epistemic uncertainty.

Fig. A2. Tutorials and comprehension checks utilized in the first experiment.



Comprehension Check

Please answer the six questions below carefully based on the just provided introduction into data uncertainty! After you have answered all questions, you will receive an evaluation of your answers and respective explanations.

1. A price range for an apartment like $250,000 to $300,000 represents the uncertainty in an estimate.

- Correct
- Incorrect

2. Which apartment listing indicates the most uncertainty based on the price range provided?

- $200,000 - $210,000
- $200,000 - $250,000
- $190,000 - $210,000

3. In the context of housing prices, the source of data uncertainty lies in:

- the natural variability of similar apartment prices.
- our lack of knowledge about similar apartments.

4. Can collecting additional similar apartments reduce the data uncertainty when estimating apartment prices?

- Yes, collecting additional similar apartments will reduce the data uncertainty.
- No, collecting additional similar apartments will not affect the data uncertainty.

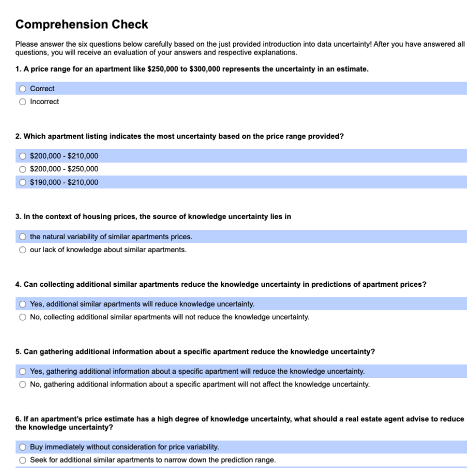
5. Can gathering additional information about a specific apartment reduce the data uncertainty?

- Yes, gathering additional information about a specific apartment will reduce the data uncertainty.
- No, gathering additional information about a specific apartment will not affect the data uncertainty.

6. If an apartment's price estimate has a high degree of data uncertainty, what should a real estate agent advise to reduce the data uncertainty?

- Buy immediately without consideration for price variability.
- Seek additional similar apartments to narrow down the prediction range.
- Seek additional information on the apartment to narrow down the prediction range.

(a) Comprehension check employed in the first experiment to test the comprehension of aleatoric uncertainty.



Comprehension Check

Please answer the six questions below carefully based on the just provided introduction into data uncertainty! After you have answered all questions, you will receive an evaluation of your answers and respective explanations.

1. A price range for an apartment like $250,000 to $300,000 represents the uncertainty in an estimate.

- Correct
- Incorrect

2. Which apartment listing indicates the most uncertainty based on the price range provided?

- $200,000 - $210,000
- $200,000 - $250,000
- $190,000 - $210,000

3. In the context of housing prices, the source of knowledge uncertainty lies in

- the natural variability of similar apartments prices.
- our lack of knowledge about similar apartments.

4. Can collecting additional similar apartments reduce the knowledge uncertainty in predictions of apartment prices?

- Yes, additional similar apartments will reduce knowledge uncertainty.
- No, collecting additional similar apartments will not reduce the knowledge uncertainty.

5. Can gathering additional information about a specific apartment reduce the knowledge uncertainty?

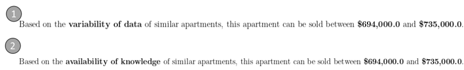
- Yes, gathering additional information about a specific apartment will reduce the knowledge uncertainty.
- No, gathering additional information about a specific apartment will not affect the knowledge uncertainty.

6. If an apartment's price estimate has a high degree of knowledge uncertainty, what should a real estate agent advise to reduce the knowledge uncertainty?

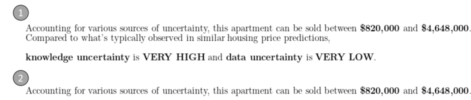
- Buy immediately without consideration for price variability.
- Seek for additional similar apartments to narrow down the prediction range.
- Seek additional information on the apartment to narrow down the prediction range.

(b) Comprehension check employed in the first experiment to test the comprehension of epistemic uncertainty.

Fig. A3. Comprehension checks utilized in the first experiment.

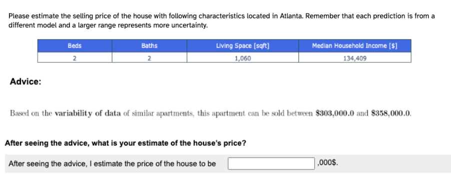


(a) Advice presented to participants in experiment 1. Top shows advice for aleatoric uncertainty, bottom shows advice for epistemic uncertainty.

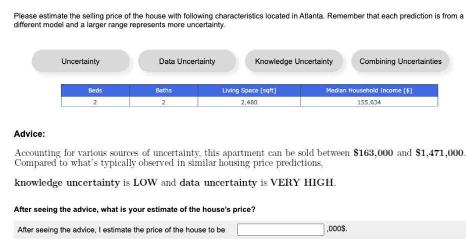


(b) Advice presented to participants in experiment 2. Top shows advice for decomposed treatment while the bottom shows advice for the combined treatment.

Fig. A4. Advice presented to participants across experiments.

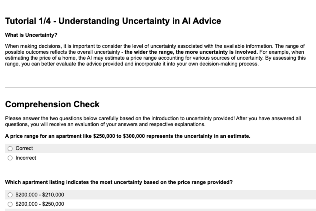


(a) Task interface for aleatoric treatment in experiment 1.

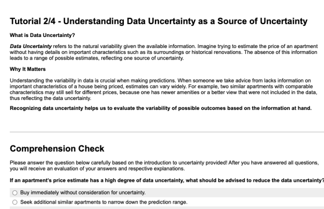


(b) Task interface for decomposed treatment in experiment 2.

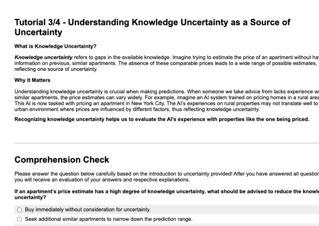
Fig. A5. Interface presented to participants across experiments.



(a) Step 1

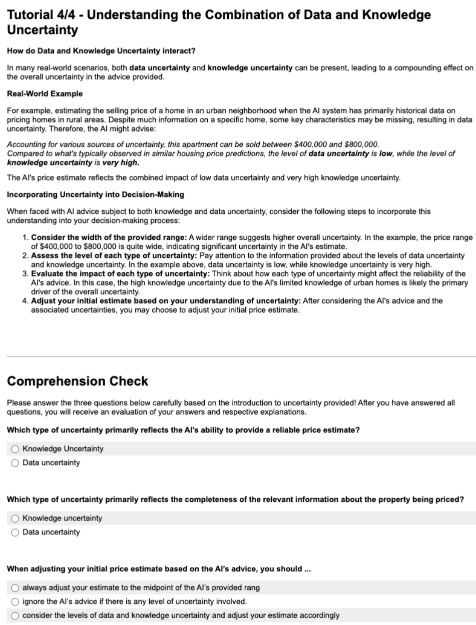


(b) Step 2

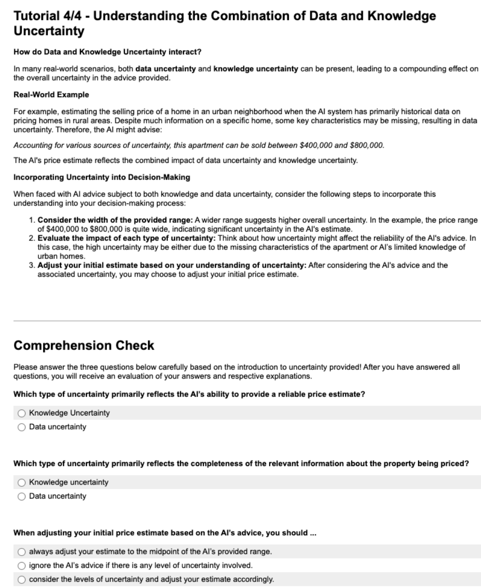


(c) Step 3

Fig. A6. The first three steps of the scaffolded tutorial on uncertainty and its sources, together with the associated comprehension checks employed in the second experiment. These steps are shared for both treatment conditions.



(a) Step 4 for the decomposed treatment.



(b) Step 3 for the combined treatment.

Fig. A7. The final step of the scaffolded tutorial on uncertainty and its sources, together with the comprehension checks. This step differs between treatments.

Received 3 January 2025; revised 13 June 2025; accepted 19 July 2025