Robot-Assisted Decision-Making: Unveiling the Role of Uncertainty Visualisation and Embodiment

Sarah Schömbs
The University of Melbourne
Melbourne, Australia
sschombs@student.unimelb.edu.au

Saumya Pareek
The University of Melbourne
Melbourne, Australia
spareek@student.unimelb.edu.au

Jorge Goncalves
The University of Melbourne
Melbourne, Australia
jorge.goncalves@unimelb.edu.au

Wafa Johal
The University of Melbourne
Melbourne, Australia
wafa.johal@unimelb.edu.au

Figure 1: (left) The robot conveys low confidence in a pill box using a GUI before picking up the pill box. (right) The robot and the human pack their pill box into the package on the working station during a human-robot collaboration task.

ABSTRACT

Robots are embodied agents that act under several sources of uncertainty. When assisting humans in a collaborative task, robots need to communicate their uncertainty to help inform decisions. In this study, we examine the use of visualising a robot's uncertainty in a high-stakes assisted decision-making task. In particular, we explore how different modalities of uncertainty visualisations (graphical display vs. the robot's embodied behaviour) and confidence levels (low, high, 100%) conveyed by a robot affect the human decision-making and perception during a collaborative task. Our results show that these visualisations significantly impact how participants arrive to their decision as well as how they perceive the robot's transparency across the different confidence levels. We highlight potential trade-offs and offer implications for robot-assisted decision-making. Our work contributes empirical insights on how humans make use of uncertainty visualisations conveyed by a robot in a critical robot-assisted decision-making scenario.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642911

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

assisted decision-making, agency, uncertainty visualization, embodiment, risk communication

ACM Reference Format:
Sarah Schömbs, Saumya Pareek, Jorge Goncalves, and Wafa Johal. 2024. Robot-Assisted Decision-Making: Unveiling the Role of Uncertainty Visualisation and Embodiment. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3613904.3642911

1 INTRODUCTION

Robots are fueled by data. Data from their internal sensors to data collected by a thermal camera, lidar or proximity sensor. Data that allows a robot to work safely alongside humans, and humans to reason about a task and the robot itself [84]. Based on this data, robots can be used to make informed decisions and to accomplish tasks without sacrificing precision or experiencing fatigue. They can be deployed in environments that pose safety risks to humans or tasks that involve hazardous materials [40, 57, 87], while collecting data to support human decision-making. However, data is subject to uncertainty and thus robots have to act under several sources of uncertainty. From uncertainty caused by noisy sensors to uncertainty based on training sets of neural networks to enable robots to detect objects [34, 55, 80]. The communication of such uncertainty holds the potential to enhance transparency, to calibrate trust and to support assisted decision-making [9, 69]. Further, visualising a robot's confidence can be a means to reduce risks and overcome communication barriers between humans and robots in collaborative tasks, which poses a current limitation [40, 54]. Both assisted decision-making [8, 19] and uncertainty [12, 29, 81] have long been investigated by the HCI community. And yet, visualising and conveying a robot's uncertainty in an assisted decision-making task has rarely been addressed so far.

Past research on robots in assisted decision-making has predominantly focused on the context of communication strategies [86], social settings [37, 83] or service domains, e.g. a robot giving recommendations for financial planning [85, 95]. In addition, several studies have investigated these so-called robot-advisors without humans interacting with a *physically present* robot, but by e.g. using video clips of robots or framing [68, 85, 86, 95]. However, the presence [24, 90], appearance [71] and embodiment [72] of a *physically present* robot can affect people's perception and experience. When the robot is physically present, it interacts in a shared environment and becomes *someone*, responsible for a certain task and outcome. In this paper, we investigate robot-assisted decision-making with a physically present and non-anthropomorphic robot, implemented in a collaborative task. In addition, we examine robot-assisted decision-making through the lens of uncertainty visualisation to inform how a robot's confidence affects human judgement and decision-behavior. We further look at how the robot's conveyed uncertainty affects the perceived trust in the robot, since "the relationship between uncertainty estimates and trust in automation is a relatively unexplored idea" [9].

We hereby draw from best-practices in both, Information Visualisation (VIS) and Human-Robot Interaction (HRI), which allows us to compare the effect of different visualisation modalities and confidence levels. In order to address the aforementioned research objectives, we explore robot-assisted decision-making with a robot that conveys its confidence through either a graphical interface (GUI) or a robot's behaviour, in the following referred to as EMBODIED uncertainty visualisation. A physically embodied and autonomous agent that moves and interacts with humans introduces new possibilities to map data onto an agent's behaviour, without the need to e.g. rely on complex information visualisations [77]. It could open up new ways to reach audiences with limited literacy skills, to interpret data in an intuitive manner and to support understandability, since past research in AI-assisted decision-making has shown that providing users with explicit explanations is a double-edged sword [8, 15, 19]. Besides, robots are not always equipped with screen interfaces to display information. Robot arms, in particular, are often screenless and need to rely on other ways to map data to support the human decision-making.

Our work is guided by the following research questions:

- **RQ1** How does visualising a robot's uncertainty affect the user's decision-making in a robot-assisted decision task?
  - **RQ1.1** How do different visualisation modalities of a robot's uncertainty impact the ability to decide correctly in a high-stake scenario?
  - **RQ1.2** Do users alter their decisions in the decision-making process when confronted with different uncertainty visualisations?
- **RQ2** How does a robot's visualised level of confidence affect the user's trust?
  - **RQ2.1** How does a robot's visualised level of confidence influence users' perceived transparency of a robot?
  - **RQ2.2** How does a robot's visualised level of confidence influence users' perceived performance of a robot?

We present a 3×2 mixed design lab study with 36 participants, contextualized in a human-robot collaboration setting which resembled a packing station in a factory that produces pills. We consider two types of uncertainty visualisation, a GUI and an EMBODIED uncertainty visualisation. Based on our manipulation check and prior research in VIS, we investigate a part-to-whole icon array as an established visualisation technique for uncertainty [3, 60, 65]. For the embodied visualisation, we draw from previous research on hesitation gestures and non-verbal signals to guide our design [36, 62, 88]. This allows us to explore how users perceive physically embodied visualisations differently compared to traditional graphical displays and how they affect user's decision-making. Besides the type of uncertainty visualisation, we consider the degree of confidence as a within-subject factor (low, high, 100%) to make inferences about the human's understanding and perception of the magnitude of uncertainty that is visualised by the robot [60].

We summarize our key contributions as follows:

- Looking at HRI from a VIS perspective has recently been identified as a major gap in HRI research [84]. We bridge that gap by exploring uncertainty visualisations in *in-situ* robot-assisted decision-making, using a physically present robot and a real-world scenario.
- We gain empirical insights into people's understanding, perception and experience with a robot conveying its uncertainty and emphasise non-trivial trade-offs.
- We highlight the implications of our findings and propose fundamental considerations when visualising uncertainty in a robot-assisted decision-making task. Furthermore, we outline potential directions for future research to broaden the scope of uncertainty visualizations in the context of human-robot collaboration.

2 RELATED WORK

In this section, we explore relevant literature from AI-assisted decision-making, visualising uncertainty and behaviour based signals in HRI. We present an interdisciplinary review drawn from literature in HCI, HRI, Machine Learning and VIS. As a guideline, we define uncertainty as "the difference between observation and reality", according to Chung and Wark [21]'s categories of uncertainty for decision-support. However, we acknowledge the existence of other uncertainty types and definitions, varying form epistemological to stochastic uncertainties, which easily illustrates the complexity of the topic itself. Further, we define decision-making as a choice between two competing courses of action [7].

2.1 AI-assisted Decision Making and Robot-Advisors

AI-assisted decision-making has gained much attention based on its potential to integrate strengths from both parties - AI and human - to enhance human-AI team performance [17]. The latter becomes particularly important in critical scenarios that involve high risks, e.g. cancer diagnosis [28, 92], frequently referred to as high-stakes scenarios [8, 17, 93]. These scenarios often require high accuracy and are usually bounded by human traits like fatigue, biases, lack of attention during repetitive tasks, or lack of precision. Importantly, AI-assisted decision support can reduce users' cognitive load, the time to make an informed decision and the need to acquire mathematical expertise [69]. As a result, AI can be implemented as a decision-support tool to provide recommendations for human decision-making. Moreover, AI-assisted decision support offers an opportunity to assist lay decision-makers to make informed decisions even in high-stakes scenarios [69]. However, AI-based systems are not perfect, which is why it is important to keep humans in the loop and to avoid over-reliance, especially in critical task domains [15, 17, 69].

Interestingly, Prabhudesai et al. [69] highlight the need to make uncertainty explicit in AI-assisted decision support tools. The authors argue that withholding uncertainty oversimplifies the reality, and paints the wrong picture of a perfect system. Similarly, Bhatt et al. [9] stresses the importance of communicating uncertainty to enhance transparency and to support an appropriate formation of trust between the agent and the human. Making uncertainty explicit in a decision-making scenario has a long tradition in VIS, since "ignoring uncertainty and its associated risk may simplify the decision-making process, but (...) does not result in making informed decisions" [23].

Similar to AI-assisted decision-making tools, robots can also be implemented in high-stakes scenarios to support humans. They are increasingly used in healthcare, e.g. as surgery assistance [97] or caregivers [18], and industrial settings, e.g. as inspection robots or for assembly related task [30, 40]. Notably, research on robots as decision-support has predominantly centered around anthropomorphic robots, particularly in social settings and service domains, e.g. robots that give financial advice - often referred to as robot-advisors [37, 68, 86].

2.2 Graphical Uncertainty Visualisation

Visualising uncertainty is a domain-overlapping issue that has been investigated in domains varying from hazard and weather forecasting [10, 65], to map-based decision-making during helicopter landings [50] to everyday contexts like catching the next bus [46]. It aims to better communicate risk, support decision-making and to represent information as accurately as possible. However, visualising uncertainty does not follow a one-rule-applies-approach [64]. There are no universal guidelines that describes when to choose which kind of uncertainty visualisation for a given scenario. Rather, it is highly dependent on the context, the task itself, the data involved and what kind of uncertainty you aim to visualise. However, some visualisation methods are more suitable than others and have been established as best-practices in their domain.

A number of surveys provide a review of methodologies and classifications to support an adequate selection of the respective uncertainty visualisation [14, 39, 64]. In the context of decision-making, visualising uncertainty is a central and widely investigated concern. The way uncertainties are conveyed can impact how individuals interpret them and, consequently, the actions they may take. Uncertainty visualisations potentially open the door to increase users' understanding of the underlying data, to support trust [45, 47], to inform about the instability inherent to estimates [67], or to reduce biases [74]. However, visualising uncertainty can also present challenges such as a feeling of being overwhelmed or confusion [29, 46].

2.2.1 Visualising Risk and Probabilities. For the purpose of this study, we take a closer look at graphical techniques to visualise single probabilistic values. Probabilities can be displayed in a part-to-whole relationship, which conveys the ratio of a numerator, to its total, the denominator. Past research provides evidence that people tend to process part-to-whole relationships easier and faster compared to visualisations that solely display the numerator [3, 65]. A few examples of such visualisations include icon arrays, bar graphs, and grids.

Previous work has recognized the potential of icon arrays to communicate probabilities, particularly in risk communication and shared decision making environments [3, 64, 75, 98]. Icon arrays display the probability as a group of dots, figures or other icons, with the numerator and denominator often differing in transparency or colour. Interestingly, past research indicates that concrete icon types, particularly anthropomorphic ones e.g. person icons, tend to perform better compared to abstract icons such as rectangular blocks [98]. The symbolism of icon arrays can add contextual meaning to the numeric information and thus promote an intuitive understanding of the data. However, anthropomorphic icon arrays can also result in risk magnitude biases [75]. Interestingly, studies that investigated both icon arrays and bar graphs revealed contradicting findings with no clear preference for one or the other [60, 75]. Using metaphors or analogies to visualise risk can be particularly helpful to make numeric values accessible to different audiences. Past research has shown that a gauge is easy to understand and can be used to e.g. readout levels of chronic stress [4, 5]. Whereas gauge visualisations are well-suited for providing a general approximation, they are less suited for precise calculations of specific values [76].

Besides choosing an adequate uncertainty visualisation, the proportion or degree of uncertainty does also matter when it comes to how uncertainty is perceived and processed. McCaffery et al. [60] show that the numerator size of an icon array, hence risk size, affects whether or not an uncertainty visualisation supports accurate estimations in data tasks. Contrary to their assumptions, Sarma et al. [74] found that the proportion of missing values (either 30% or 50%) in datasets, hence the proportion of uncertainty, does affect people's biases depending on the data task they perform. In terms of probability values, findings by Jiang et al. [44] indicate that participants started to trust a system's recommendation at a conveyed confidence level of more than 75%, which implies an uncertainty of less than 25%. In addition, small probabilities are particularly difficult to interpret [3, 51].

2.3 Embodied Uncertainty Visualisation

A recent paper by Leusmann et al. [54] proposes the so-called *Human-Robot Uncertainty Loop*, a framework which identifies various forms of uncertainties in HRI and their need to be communicated. To depict uncertainty in HRI, a multitude of different approaches open up. From traditional graphical user interfaces to XAR to using the robot's physicality as an interface to convey data. For the purpose of this paper, we take a closer look at embodied, behaviour-based uncertainty signals. Whereas traditional VIS research has focused on explicit uncertainty visualisations [20], a robot's embodiment offers new ways to visualise uncertainty implicitly. Embodied visualisations enable a robot to display it's internal world without the need to use a graphical display or speech. This is especially important in industrial settings, in which (a) robotic arms are often not equipped with graphical displays, or (b) the environment is too noisy to implement speech responses. Besides, embodied uncertainty visualisations could potentially improve accessibility and facilitate peoples' intuitive understanding, similar to data physicalizations [6, 41]. Inspired by previous research on visualising a robot's motion intent [53, 66] and expressive motion behavior [26], we look at hesitation gestures to visualise a robot's uncertainty in an assisted-decision-making task.

2.3.1 Hesitation Gestures. Hesitation, as a known uncertainty signal from human-human interaction, has received much attention in the HRI community. As one of the first attempts to implement and investigate hesitation gestures, Moon et al. [62] examined human-human hesitation gestures and transferred the respective trajectories onto a robotic arm. Their findings confirm that participants were able to correctly infer and detect a robot's hesitation. Building upon their work, the authors recently presented an improved trajectory generator to implement hesitation gestures [61]. However, it is important to highlight that their research focuses on uncertainty in human-robot negotiations, i.e. the human and the robot reaching for the same object at the same time. Therefore, Moon et al. [63] define hesitation gestures as gestures, that express uncertainty about the "right or priority" to access a shared resource. Besides these efforts, past research also investigated pausing [96], trembling and repeated pull-back movements as cues for hesitation [88]. Importantly, a robot needs to be able to convey its degree of uncertainty. To do so, Hough and Schlangen [36] successfully developed a model to display a robot's confidence level, as in inverse to uncertainty, through motion-based behaviour. Here, the robot adjusts its speed and waiting time according to its internal confidence.

2.4 Summary

As human-robot collaboration continues to expand, the integration of uncertainty visualization in the decision-making process holds the potential to enhance transparency, performance and trust. In contrast to AI-assisted decision-support or framed robot-advisors, a *physically present* robot presents an entity which can be perceived as an agent itself - tangible and experienceable. Taking into account such Data-Agent Interplay [77], we aim to examine not just best-practices from VIS in the context of robot-assisted decision-making [84], but to open the design space of uncertainty visualisations by exploring embodied visualisations conveyed by a robot and how humans understand the aforementioned.

3 METHOD

We conducted a 3×2 mixed design lab study with *level of confidence* (low confidence vs. high confidence vs. 100% confidence) as a within-subjects factor and *type of visualisation* (graphical user interface vs. physically embodied) as a between-subjects factor. We thus manipulated the modality through which the robot conveyed its confidence and its confidence in the quality of the assessed pill box itself. This allows us to investigate the relationship between the type of visualisation and the level of confidence, their effect on the user's judgement, decision-behavior and perceived trust in the robot in the context of a critical, high-stakes scenario, i.e. shipping defective pills.

3.1 Experimental Task

Our decision to conduct a study on assisted-decision making with autonomous agents in this context was based on two main reasons. First, with the rise of collaborative robots that are designed to work safely alongside humans, we are experiencing a paradigm shift from automation to actual collaboration [91]. Collaborative robots can be equipped with various sensors and hence be used to accomplish repetitive and hazardous tasks and to inform decisions [40, 87]. Second, we were particularly interested in investigating a critical scenario that involves high risks. As such, we selected a healthcare scenario, which is widely used in assisted decision-making literature due to the importance of effectively communicating risks [60]. In our study, participants were told that they would examine pills from a factory in which a mishap has happened due to which some pills might be defective.

To support their decision-making, a robot was implemented to assess the content of the pill boxes and to visualise its confidence: During the task, the robot was in charge of (a) displaying its level of confidence in sensing that the pill boxes were defective, and (b) packing those risky pill boxes into a package. We implemented these robot activities through an autonomous pick-and-place task implemented via ROS2 [59]. Before the robot placed its pill box into the package, the robot conveyed its confidence to inform the participant with one of the modalities (GUI or EMBODIED). The participant on the other hand was responsible to pack the pill boxes known to be not affected by the mishap (marked with a ticked label - see Figure 2). In addition, the participant was in charge to decide whether each package should be placed to the *ready for shipment* or *needs additional testing* area, taking into account the confidence the robot conveyed beforehand.

Our experimental task allowed us to realise two dependencies between the robot and the human to increase the degree of collaboration: knowledge, since the participant needed the robot's assessment to form their decision, and time, since the task was performed sequentially, see Figure 2.

3.2 Experimental Manipulations

In order to answer our proposed research questions, we manipulated the robot's *type of visualisation* to convey its confidence and the *level of confidence* of the robot in the pill box not being defective. Next, we describe how we selected candidate visualisations and set confidence ranges.

Figure 2: Human-robot collaboration while packing pill boxes into packages.

3.2.1 Type of Visualisation. To investigate the potential effects of different types of uncertainty visualisations on the user's decision-making, perceived transparency and trust in the robot, we manipulated the way the robot conveyed its confidence. We hereby focused on two approaches. The first approach was to use the robot's physicality to embody and map the data onto the robot's movements to convey it's confidence. The robot used in this study was a robotic arm developed by Franka Emika [1]. This approach was inspired by past research efforts on hesitation gestures in HRI [36, 61–63]. Based on Hough and Schlangen [36], we implemented two functions to convey the robot's confidence through its behaviour. The first function describes the *waiting time*, i.e. how long the robot's gripper hovered above the target pill box. We implemented a waiting time inversely proportional to the robot's confidence. In this study, we set the maximum waiting time to 6 seconds with a degradation down to 0 (no waiting time in case of a 100% confidence). These bounds are scaled up from Hough and Schlangen [36] due to the size difference of the robot used in this study.

We additionally implemented a function that describes the time the robot needs after picking up the pill box to approaching the position to place the pill box into the package. This *travel time* was again inversely proportional to the robot's confidence with a maximum travel time of 10 seconds and a minimum travel time of 5 seconds, see exemplified values for both waiting and travel time for all three within factors in Table 1. In line with Hough and Schlangen [36], we assigned these bounds after a short pilot study and additionally validated their perception through a manipulation check (section 3.2.3).

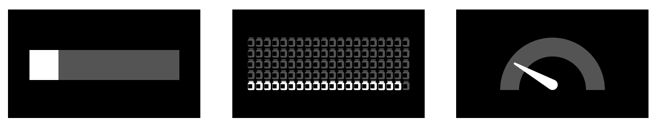
For the second approach we integrated a graphical user interface onto the robot's end-effector that enabled us to assess best practises from VIS during a human-robot collaborative scenario. In order to mount a digital screen onto the robot we designed and 3D printed a case to clip an iPhone 13 mini on to the robot, see Figure 1. The iPhone's screen was used to display the GUI on the browser. We developed a Flask webapp to display the GUI on the phone and to communicate with the robot via ROS2 [59]. Grounded on previous research described in section 2.2.1 we selected a bar graph, a gauge visualisation and an icon array to convey probability values on the robot's graphical display, shown in Figure 3. We used a part-to-whole relationship for all three visualisation types [3, 65]. Thus, all

Table 1: Waiting time and travel time of the robot's movement in the embodied uncertainty visualisation condition with exemplified values for each confidence level.

		Waiting Time (in sec.)	Travel Time (in sec.)
Low Confidence	16%	5.04	9.2
	19%	4.86	9.05
	23%	4.62	8.85
High Confidence	90%	0.6	5.5
	92%	0.48	5.4
	95%	0.3	5.25
100%	100%	0	5

GUI visualisations displayed the robot's confidence in the format "numerator out of a constant denominator of 100".

For the icon array, we decided to display the numerator and denominator as icons of pill boxes to improve intuitiveness and accuracy, instead of using abstract symbols like blocks or dots [58, 98]. We refrained from using colour to enhance accessibility [42] and to avoid biases [11, 79] or cultural confounds [43]. To improve visibility, we used a black background with a white numerator and a dark-grey denominator, see Figure 3. In line with Jiang et al. [44] we did not show labels and numbers to avoid any effect from the number itself.



(a) Bar graph as part-to-whole visualisation displaying low confidence.

(b) Icon array as part-to-whole visualisation, with pill box icons displaying low confidence.

(c) Gauge as part-to-whole visualisation displaying low confidence.

Figure 3: Example of GUI visualisations for the low confidence condition (here, 19% confidence).

3.2.2 Level of Confidence. To gain additional insights on the importance of the actual level of uncertainty and to better inform future implications, we manipulated the level of confidence as a within-subjects factor. This allowed us to make inferences about how the participants' decision-making, perception of the robot and trust varies depending on the robot's conveyed confidence. We defined the different levels of confidence through the lens of risk communication in healthcare. It is important to mention that we framed uncertainty as the inverse of confidence, consistent with previous research [9, 36, 38, 69] and hence translated the respective uncertainty values to confidence values and confidence levels.

In line with McCaffery et al. [60], we defined low uncertainty as probabilities < 9.9% and assigned them to the *High Confidence* condition and high uncertainty as probability values in the range between 75% - 85% and assigned them to the *Low Confidence* condition. As can be seen, the perception of a small and high uncertainty

Table 2: Accuracy for each type of visualisation tested as manipulation check before the main user study.

	Confidence Level	100% Confidence			High Confidence			Low Confidence			Total Accuracy
	Probability Value	100%	100%	100%	98%	94%	91%	23%	19%	16%	
GUI	Bar Graph	0.7	0.8	0.8	0.2	0.8	0.9	0.9	0.9	0.9	0.77
	Icon Array	0.7	0.8	0.8	0.8	0.7	0.8	0.9	0.8	0.8	0.79
	Gauge	0.8	0.8	0.8	0.3	0.8	0.6	0.8	0.8	0.7	0.71
EMBODIED	Waiting Time	0.8	0.7	0.6	0.3	0.2	0.8	0.8	0.8	0.9	0.66
	Travel Time	0.6	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.51
	Waiting and Travel Time	0.6	0.5	0.4	0.6	0.5	0.5	1	1	0.9	0.67

is not a dichotomous phenomenon. We were additionally interested in investigating the perception of a fully confident robot (100% confidence). Since each within-subject factor consisted of a trial with a batch of 10 pill boxes that needed to be assessed and packed by the robot, we randomly assigned 10 confidence values out of the respective range to each confidence level to sample the respective probability range. We furthermore introduced noise to the *High Confidence* and the *Low Confidence* batch, inspired by Sarma et al. [74]: 20% of the time, the robot conveyed probability values of the opposite condition. We fixed this noise as the pill box #4 and pill box #7. We refrained from introducing noise to the *100% Confidence* condition, since it would damage the intended scheme of a fully confident robot.

3.2.3 Manipulation Check. We conducted a manipulation check to verify whether the type of visualisation and the level of confidence was perceived as intended. Past literature has shown that the performance of an icon array and a bar chart can depend on the numerator size, i.e. level of confidence [60]. The manipulation check was particularly important since we wanted to ensure that participants were able to detect and categorise the different levels of confidence in both visualisation conditions to make these conditions comparable. Furthermore, the manipulation check was used to select the most appropriate type of visualisation for each modality to later present to participant in our main lab study. To do so, we conducted a simple between-subjects manipulation check. Based on previous literature presented in section 2.2.1, we compared the accuracy of *icon array*, *bar chart*, and *gauge* visualisations for the graphical user interface and *waiting time*, *travel time*, *waiting and travel time* for the embodied visualisation.

We relied on independent raters and utilised the Prolific[1] crowdsourcing platform to recruit 60 participants who were based in the United States, native English speakers, to mitigate challenges with language proficiency [31] and who had approval ratings above 98%. All participants were compensated following prolific guidelines for payment [25]. Through the Qualtrics[2] survey platform, participants watched short video-clips of the robot (1.) assessing the content of a pill box, (2.) conveying its confidence based on the assigned visualisation type and (3.) placing the respective pill box into the package. We used videos of the same task setup of our in-person study. In a multiple choice question, participants were asked to categorise the level of confidence the robot conveyed. We measured the rater's accuracy to correctly categorise the conveyed confidence level. Table 2 outlines that the icon array has the highest accuracy. It emphasises that the icon array supports an accurate reading of the respective confidence level, which is why they are widely used as decision-aids, particularly in the field of healthcare [3, 32, 60]. It further shows that participants had difficulties correctly interpreting a 98% confidence in the gauge as well as the bar chart condition. Within the embodied visualisation types, the implementation of either waiting time or the combination of both waiting and travel time obtains a relatively high accuracy across the confidence levels. Since the total accuracy of waiting and travel time shows a slightly higher total accuracy score and has been shown to be correctly interpreted as a robot's confidence [36], we chose the latter as the embodied uncertainty visualisation for the in-person lab study.

[1]www.prolific.com
[2]www.qualtrics.com

3.3 Participants

Prior to the experiment, we conducted a power analysis to determine our sample size using *G*Power* [27]. The calculation was based on a within-between interaction with an alpha level of .05, a medium effect size of f = .25 and a power(1-β err prob) = .9, resulting in a targeted sample size of N = 36. We recruited 36 participants (25F, 11M) for our lab study using our university's notice board. This sample size is also in line with mixed-method studies previously published at CHI [16]. The study took roughly 45 minutes and participants were compensated with a $20 gift voucher. The study was approved by our University's Human Ethics Committee.

3.4 Study Procedure

The lab study consisted of five phases: (a) pre-questionnaire, (b) collaborative task and decision-making, (c) follow-up repeated measures, (d) procedure repeated for each confidence level and (e) open-ended questions, see Figure 4. Participants were randomly assigned an uncertainty visualisation type. Upon arrival to the usability lab, we provided each participant with a written plain language statement and collected their consent to participate in the study. After signing the consent form, the participant filled out a pre-survey to assess basic socio-demographics, and to gauge their previous

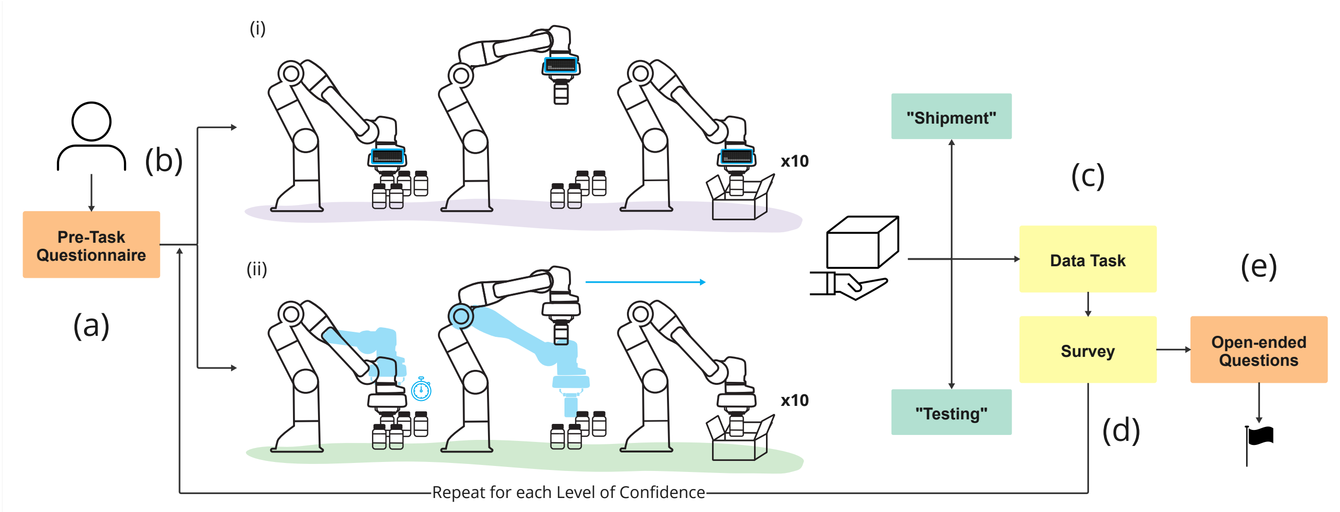


Figure 4: The experimental flow. (a): Pre-task questionnaire. (b): Collaborative task and decision-making (x10) for GUI or EMBODIED. (c): Data Task and repeated measures of trust. (d): Procedure repeated for all confidence levels, with confidence level as within factor (100% confidence, high confidence, low confidence). (e): Open-ended questions.

experience with robots and dispositional trust in new technology, see (a) in Figure 4. Secondly, the participant was provided with a written task description, including an infographic to visualise the experimental setup and the task at hand. After reading the document, the experimenter additionally verbalized the instructions and provided the participant with a preview video showing the robot conveying a 100% confidence depending on the visualisation modality the participant was assigned to. The preview video served to illustrate how the robot works, to ensure that the task is understood correctly, and to provide a baseline cue about the respective visualisation type. The next phase consisted of the actual task, followed by our repeated measures. Depending on the type of visualisation the participant was assigned to, they either experienced the robot conveying its confidence through the GUI or the EMBODIED condition, see (b) in Figure 4. The order in which the participant experienced the three different levels of confidence (100% Confidence, High Confidence, Low Confidence) was counterbalanced.

The procedure of the task itself was as follows:

1. Based on the first confidence level the participant was assigned to, the robot started to move to the first pill box, picked up the respective pill box and placed it into the package the participant positioned onto the packing station.
2. The robot conveyed its confidence either as an icon array via GUI on the robot's end-effector *or* through the EMBODIED visualisation - depending on the modality the participant was assigned to.
3. Subsequently, the participant's task was to pack their pill box into the package too, to close the package and to decide whether the package is "ready for shipment" or "needs additional testing" as their final judgement, see (b) in Figure 4.
4. After the participant's decision was made, the participant took a new, empty package, placed it onto the highlighted area of the packing station and repeated the steps above.

Each confidence level condition consisted of 10 pill boxes tested and packed by the robot and thus 10 pill boxes packed by the participant. After 10 packages had been successfully packed and put to either the "ready for shipment" or "needs additional testing" pile, the participant was asked to answer a follow-up survey, including a data task and the repeated measures. Following this, the participant was asked to wait in the resting room next door, until the next pill boxes were brought, and ready to be tested and packed. Upon completion of the participant's third session (third level of confidence), the participant was lastly asked to fill out the follow-up survey again, in addition to open-ended questions exploring their collaborative decision-making experience. Specifically, we posed questions on participants' trust, their assisted decision-making experience, their perception of the robot's confidence, and how these factors may have been mediated by the critical nature of the task scenario.

3.5 Measures

We combined validated questionnaires from assisted decision-making with autonomous agents with a classical data task from VIS. Besides questionnaires, we implemented behavioural measures to obtain more nuanced insights in people's experiences. We further measured peoples' dispositional trust in new technology [52] as a covariate, to avoid a confounding effect of participants' interindividual differences to trust technology.

3.5.1 Data Task: Average Estimation. The average estimation is a typical data task, frequently used in VIS research [2, 74, 84]. We asked participants to estimate the average confidence of the robot after each confidence level (*"Please estimate the average uncertainty (in %) of the robot:"*). Participants entered a numeric value from 0-100 in a text field. This allowed us to assess whether participants were able to correctly distinguish between low, high and 100% confidence.

3.5.2 Behavioral Measures: Decision-Making. We investigated a binary decision-making scenario to assess the participant's strategy and perception of risk in the context of (a) a robot visualising uncertainty and (b) a critical scenario, i.e. pills. Our scenario considered whether participants decided to put a package to "ready for shipment" or to "needs additional testing" after taking into account the robot's conveyed confidence in each pill box. Considering the context of the task, we defined a "correct" decision as the judgement to send the package off to shipment if the robot conveyed a 100% confidence. Otherwise, additional testing would be needed. We based this decision on the fact that we situated the task in a critical scenario with high risk, namely shipping medicine, i.e. pills, that could be defective and potentially may harm humans. We furthermore investigated if participants altered their judgement while making their final decision. We encoded "changing mind" as switching a package from "shipment" to "needs testing" and vice versa after the initial decision has been made, tracked in real-time by the experimenter. This served as an additional indicator of how confident participants were in their decision-making.

3.5.3 Subjective Measures: Trust, Transparency and Performance.

Overall Trust. First, we investigated participants' overall subjective trust in the robot, asking *"Do you trust the robot?"*, inspired by Yang et al. [94]. Participants were able to rate their trust on a bipolar slider, with *I don't trust the robot at all* on the left and *I trust the robot completely* on the right. To avoid biases caused by an anchor as a starting point on a slider, we provided participants with a bipolar slider without an anchor [78].

Multidimensional Trust Questionnaire. We furthermore captured participants perceived trust in the robot as repeated measures using the Multidimensional Trust Questionnaire (MTQ) [73]. The MQT is a validated questionnaire that particularly targets trust in automated systems and agents, while considering various subdimensions of trust to capture trust and trust dynamics in a nuanced manner. We selected the respective items related to transparency (*The way the robot works is clear to me*, *I am well informed how the robot works*, *I understand how the robot works*) and performance (*The robot works safely*, *The robot works well*, *The robot works accurately*) and asked participants to please rate the robot on a 4 point scale from *"Disagree"* to *"Agree"*.

4 RESULTS

Participants age ranged between 18 and 60 (M = 25.53, SD = 7.84); with 50% of the participants having no prior experience with robots and 50% reporting prior experience only with everyday robots, e.g. vacuum robots. Furthermore, participants generally trusted new technology (M = 5.48, SD = 1.13; 7-point Likert scale from (1) *"strongly disagree"* to (7) *"strongly agree"*).

4.1 Quantitative Results

4.1.1 Average Estimation. We summarized participants' performance on the average estimation task in Figure 5. Although the estimation of the robot's confidence was not employed as the primary manipulation check, nor the main focus of this work, we did observe that participants successfully distinguished between low, high and 100% confidence and thus perceived our manipulation as intended. Across both conditions, the GUI and the EMBODIED visualisation type, participants correctly estimated the average of the low confidence condition < high confidence < 100% confidence. Our results show an average estimation of M = 48.28% (SD = 32.96) for the low confidence condition, M = 72.78% (SD = 26.98) for the high confidence condition, and M = 75.61% (SD = 34.75) for the 100% condition for participants experiencing the GUI, see Figure 5 A. Demonstrating a comparable pattern, our findings show an average estimation of M = 54.44% (SD = 28.69) for the low confidence condition, M = 80.28% (SD = 16.49) for the high confidence condition, and M = 87.89% (SD = 13.71) for the 100% confidence condition for participants experiencing the EMBODIED visualisation, see Figure 5 B.

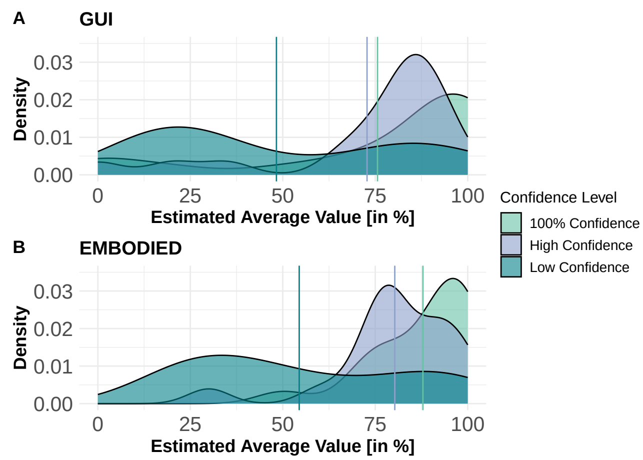
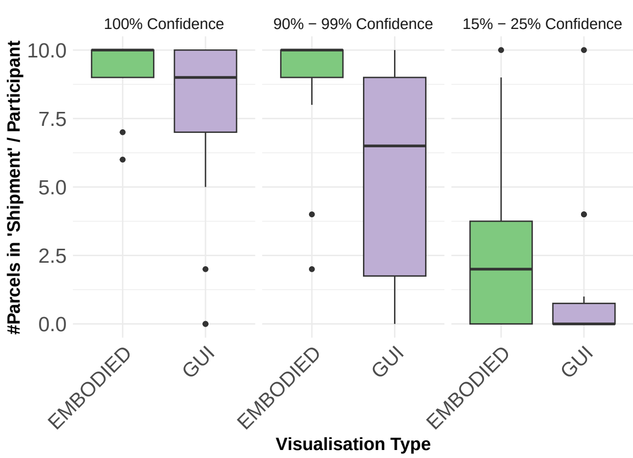


Figure 5: Average estimations shown for each visualisation type and confidence level.

4.1.2 Decision-Making: Shipment. Next, we calculated participants' decision accuracy to get insights on RQ1.1. To do so, we investigated how often participants decided to send packages off to shipment, depending on the experienced type of uncertainty visualisation and confidence level. In the 100% confidence condition, 12 out of 18 participants in the EMBODIED group correctly decided to put all 10 packages into the ready for shipment pile. Contrary to only 7 out of 18 participants out of GUI group, see Figure 6a. We further conducted an analysis on how many packages participants decided to ship depending on the actual probability value the robot conveyed for each pill box and the type of visualisation. For this analysis, we therefore considered the noise packages and examined the probability range of the respective pill box (100% vs. 90%-99% confidence vs. 15%-25% confidence). The results are shown in Figure 6b.

We performed a mixed analysis of covariance (ANCOVA) test after controlling for the dispositional trust in new technology. Our analysis reveals a significant main effect of the type of visualisation on the decision to send packages off to shipment ($F(1, 33) = 11.304$, $p = .002$, $\eta^2 = 0.13$). A post-hoc Tukey pairwise comparison shows that participants who experienced the EMBODIED visualisation (M = 7.0, SE = 1.73) placed significantly more packages to the "ready for shipment" pile compared to participants who experienced the GUI ($M = 4.7, SE = 1.80$), $p < .005$. The results show a significant main effect of the confidence level on the decision to send packages off to shipment ($F(2, 68) = 62.80$, $p < .000$, $\eta^2 = 0.56$). A post-hoc Tukey pairwise comparison reveals that participants in the 100% confidence condition ($M = 8.38$, $SE = 1.77$) sent significantly more packages off to shipment, compared to the low confidence condition ($M = 1.93$, $SE = 1.77$); $p < .0001$. Additionally, participants in the high confidence ($M = 7.24$, $SE = 1.77$) condition sent significantly more packages off to shipment compared to the low confidence condition; $p < .0001$. There was no significant difference between the 100% and high confidence condition; $p = .160$.

(a) Total count of participants who correctly decided to place all 10 packages to shipment in the 100% confidence condition.



(b) Total count of packages placed to shipment/participant for each confidence range of the respective pill boxes.

Figure 6: Participants' decisions to send packages off to shipment.

We additionally assessed if participants changed their mind during their decision-making process to get insights on RQ1.2. Results are shown in Figure 7. Interestingly, 6 out of 18 participants in the GUI condition changed their mind in the 100% confidence condition compared to 0 out of 18 participants in the EMBODIED condition. In both, the high and low confidence condition, the amount of participants who changed their mind was ≤ 4.

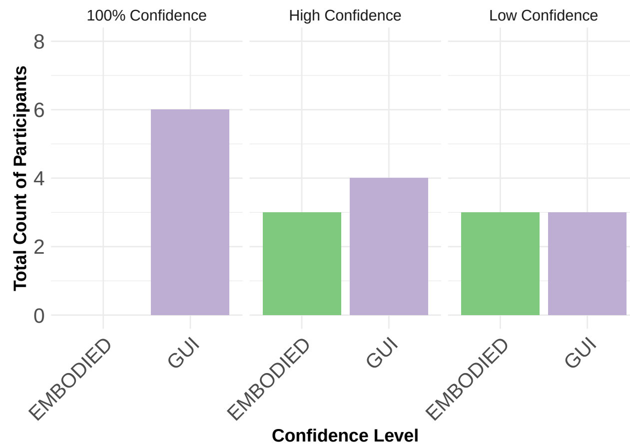


Figure 7: Total count of participants who changed their mind during their decision-making process for each confidence level.

4.1.3 *Overall Trust in the Robot.*

Guided by RQ2, we explored the relationship between visualising uncertainty and trust in autonomous agents. Participants' overall subjective trust in the robot was descriptively high across conditions, see Figure 8a. Notably, a mixed ANCOVA shows no significant main effect of the confidence level ($F(2, 68) = 2.356$, $p = .102$) or the visualisation type ($F(2, 68) = 2.716$, $p = .109$), nor does it show an interaction effect between the type of uncertainty visualisation and the level of confidence the participants experienced ($F(2, 68) = 1.080$, $p = .345$). Interestingly, on a descriptive level, participants in the GUI condition rated their overall trust in the robot the lowest in the 100% confidence condition, shown in Figure 8a.

4.1.4 *Trust-related Transparency and Performance.*

To obtain more nuanced insights regarding trust in robots, particularly in high-risk assisted decision-making scenarios, we examined two subdimensions of the MTQ [73] in addition to our single-item: transparency (RQ2.1) and performance (RQ2.2). The Cronbach's α for the transparency ($\alpha = 0.87 - 0.91$) and the performance ($\alpha = 0.77 - 0.89$) subdimension indicates a strong internal consistency among the items. Interestingly, the results of a mixed ANCOVA show no significant effect between the GUI ($M = 1.68$, $SE = 0.746$) and the EMBODIED visualisation ($M = 1.68$, $SE = 0.714$) on perceived transparency; $F(1, 33) = 0.031$, $p = .862$. This is surprising, considering the items that measure transparency as a subdimension (*The way the robot works is clear to me*, *I am well informed how the robot works*, *I understand how the robot works*). The results are shown in Figure 8b. Further, there was no interaction effect between the type of uncertainty visualisation and the confidence level ($F(2, 68) = 0.765$, $p = .469$). However, we observed a significant effect of confidence level on perceived transparency; $F(2, 68) = 3.895$, $p = .025$, $\eta^2 = 0.02$. A post-hoc Tukey pairwise comparison revealed that participants in the high confidence condition ($M = 1.83$, $SE = 0.719$)

perceived the robot as significantly more transparent compared to the low confidence condition (M = 1.54, SE = 0.719), $p < 0.05$.

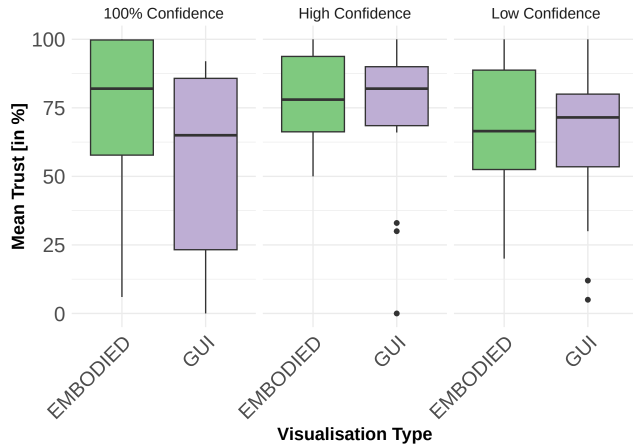
Next, we investigated the subdimension that covers trust related to the perceived performance of the automated agent, in our context, a robotic arm. However, the results of a mixed ANCOVA reveal no significant effect of either the visualisation type ($F(1, 33)$ = 0.810, p = .375) or confidence level ($F(1.5, 49.48)$ = 1.020, p = .349), nor an interaction between those two ($F(2, 68)$ = 1.023, p = .365). Lastly, we investigated the overall score of both subdimensions to infer insights on participant's trust in the robot. Here too, our results reveal no significant effect of either the visualisation type ($F(1, 33)$ = 0.077, p = .784) or confidence level ($F(2, 68)$ = 2.939, p = .0597), nor an interaction between those two ($F(1.37, 45.17)$ = 0.710, p = .446). However, descriptively, our results show that participants in the high confidence condition (M = 2.67, SE = 0.568) rated their trust higher compared to the 100% confidence condition (M = 2.53, SE = 0.568) and the low confidence condition (M = 2.45, SE = 0.568).

4.2 Qualitative Results

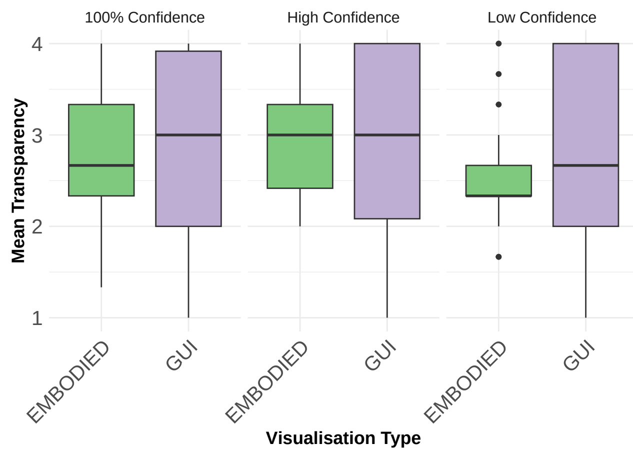
Lastly, we analysed the open-ended questions asked at the end of the experiment. We aimed to gain insights into people's decision-making process, trust and their experience with a robot that is assisting them in a critical task. We followed a deductive thematic analysis approach to systematically code participants' responses [13]. To perform a thematic analysis, we constructed a coding framework rooted in pre-established themes. Each theme was based on past literature and corresponded to our research objectives. First, we gained a holistic understanding of the data by reviewing all the responses provided by the participants. Subsequently, we annotated sections of participants' responses that aligned with our predetermined themes and assigned these quotes to the respective themes.

4.2.1 Perception of Confidence. The majority of the participants who experienced the EMBODIED visualisation interpreted *"the time it took the robot to pick up the pills"* (P12, EMBODIED) as a cue for the robot's confidence. In line with previous research in HRI, the majority of participants identified the robot's waiting time as *hesitation* [61–63]; *"It would hesitate for quite a while whenever it was not confident, and it wouldn't stop at all if it was confident."* (P30, EMBODIED). Besides the robot's waiting time, participants also perceived the robot's travel time as an additional cue for uncertainty; *"Its movements were very smooth when it packed bottles it knew were not defective."* (P2, EMBODIED). Hence, the robot's smoothness, also referred to as motion fluency [88], was recognized. Interestingly, their retrospective perception of the robot's confidence was perceived rather binary (confident or not confident), depending on the presence of these cues; *"It appeared confident to me when it picked up the pill box quickly, because it seemed like it was not hesitating to put it into the parcel."* (P36, EMBODIED).

As expected, the majority of participants in the GUI condition correctly interpreted the graphical display as the robot's confidence. However, their perception of the robot's confidence was more varied compared to participants in the EMBODIED condition, resulting in the following insights.



(a) Participants' overall trust in the robot [in %] for each condition.



(b) Participants' perceived transparency of the robot, using the transparency subdimension of the MTQ [73].

Figure 8: Overall trust and transparency as a trust-related construct.

4.2.2 Risk Threshold and Awareness. The majority of participants in the GUI condition applied a personal risk threshold to make a decision; *"I just kind of set the threshold to be 95%."* (P1, GUI) and *"If it was above 95% I would place it in the ready to be shipped section, otherwise it [...] needs additional testing."* (P31, GUI). Notably, several participants indicated a high risk-awareness of the setting in which they collaboratively executed the task with the robot. Participants elaborated *"I didn't want to risk it"* (P3, GUI) and explained that *"above 90% is still not enough assurance of someone taking a bad pill."* (P21, GUI). Further, this participant stated that *"I wouldn't give these pills to my children"* (P21, GUI). Besides setting a defined threshold, some participants reported that their risk threshold changed over time, depending on the confidence the robot conveyed and the frame of reference they experienced; *"As some boxes were predicted to have around 20%, I moved the 90% confidence boxes to the ready for shipment area."* (P25, GUI). Further, several participants struggled to apply a threshold that could be realistically applied; *"I originally *planned to assign all of the boxes that did not achieve 100% confidence for additional testing. But I was just not sure if the number can really be achieved in a real situation, so I set a threshold for the last several boxes."* (P1, GUI).

4.2.3 Fully Confident vs. Overconfident. Surprisingly, several participants in the GUI condition interpreted the 100% confidence condition as the robot being too confident. This perception made participants feel uneasy and caused them to distrust the robot; *"In the third session [100% Confidence], the robot seemed too confident, for that I can not trust it very well."* (P11, GUI) and *"I think something went wrong with it in trial 2 [100% Confidence] causing it to be overconfident."* (P29, GUI). Additionally, some participants expressed that the 100% confidence condition made them question the robot's reliability; *"When I saw all the packages are all 100% confidence, I am [sic] unsure about the reliability of the robot."* (P7, GUI). Conversely, some participants did believe that a fully confident robot was plausible; *"100% did make me question its reliability, but I still thought it was a plausible result and would still trust the robot's judgement."* (P31, GUI). In the embodied uncertainty visualisation condition, only a few participants experienced a distrust in a fully confident robot; *"The trust changed from positive to negative extremely after doing the second session [100% Confidence] in which I got a sense that the robot didn't work properly as it showed no hesitation on picking and putting all bottles in the boxes."* (P22, EMBODIED).

A few participants considered the robot's *behavioural consistency* as an influencing factor in their assessment of the robot, unveiling contrasting statements. Some participants perceived consistent exhibitions of confidence to be an indicator of poor ability; *"The results must be varied to show me that the robot is really working."* (P1, GUI). In contrast, some participants regarded a lack of consistency as an indication of the robot's reduced confidence; *"Since it did not show consistency, I felt that the robot [sic] behavior is not confident."* (P15, GUI).

4.2.4 A Need for Additional Explanation. Interestingly, some participants in the GUI condition expressed concerns regarding a lack of information and their need for further explanation, since they did not know *"the mechanism behind the confidence"* (P7, GUI). *"I would just wish I had an explanation for what it is checking for and how. How it is coming up with its confidence level."* (P3, GUI). Further, participants explained that the lack of additional information impacted the robot's credibility. *"I need more information to be able to believe the robot's assessment"* (P15, GUI).

4.2.5 Responsibility. Our qualitative analysis also reveals that participants had a different idea of who should be in charge for the final call of the decision-making task. Only a few participants stated that the final judgement should be up to the user; *"Workers can make their own choices about if they believe the pills are ok to be shipped or not."* (P16, EMBODIED). Conversely, several participants indicated that they were willing to default the decision-making to the robot; *"I gave up my decision making capability to the robot's confidence. That's how far it affected my expectations."* (P22, EMBODIED). Further, they expressed that they would have preferred the robot to explicitly tell them which package to ship or test; ' *'Tell me if the pills are ready for shipment or additional packaging is required."* (P9, GUI). Participants also explained that they *"assumed that if it [the robot] was fully unconfident, it would not place that particular pill box in the [shipping] box."* (P32, EMBODIED), which also suggests a tendency to rely on the robot's judgement in the decision-making process, and the willingness to hand over responsibility.

5 DISCUSSION

Both the EMBODIED and GUI visualisation can potentially serve as *communicative actions* [33] to make humans understand what robots do and think and to inform decision-making in a collaborative setting. We discuss a potential trade-off between EMBODIED uncertainty visualisations, that enable users to make intuitive decisions, with the downside of information loss, and GUI visualisations, that allow users to make fully informed decisions, with the pitfall of distrust in robots. To do so, we highlight our insights on how using uncertainty visualisations in robot-assisted decision-making affect people's judgement, their perception of the robot and the risk itself. Further, we outline fundamental implications to inform future design of uncertainty visualisations in robot-assisted decision tasks.

5.1 Visualising Uncertainty affects the Perception of the Robot

The results from our manipulation check and average estimation show that participants across conditions were able to successfully perceive the robot conveying a low, high, and 100% confidence. Moreover, participants in the EMBODIED condition recognized the manipulated *waiting and travel time* as *hesitation* and an indicator for the robot's confidence. These results are in line with previous research on non-verbal cues in HRI and show that people apply known anthropomorphic gestures and characteristics to a robot's behavior [36, 62, 96]. In the GUI condition, participants also correctly interpreted the uncertainty visualisation as the robot's conveyed confidence. Further, our results show that an icon array, a classical approach to visualise risk [60], is an appropriate approach to visualise a robot's confidence. Our qualitative analysis additionally suggests that the graphical display was interpreted as a robot's extension, and not as an artefact separate from the robot.

5.2 Robot-Assisted Decision-Making

We defined decision as a binary variable: Participants were able to either send a package "off to shipment" or place it in the "needs additional testing" area. Our results reveal a general risk-awareness for both uncertainty visualisation conditions, with a clear distinction between high and low risks. Participants in the 100% confidence and the high confidence condition sent significantly more packages off to shipment compared to the low confidence condition. Based on the task setup and the severity of a false-positive, i.e. to ship defective pills, we defined a *correct decision* as the judgement to send a package off to shipment *if* the robot conveys a 100% confidence. Surprisingly, our results show that in the 100% confidence condition, the majority of participants in the EMBODIED condition correctly sent packages off to shipment; compared to only 7 out of the participants in the GUI condition. This is non-trivial, since it indicates that participants in the GUI condition were more willing to override the robot's information suggesting that the pills are not defective with the utmost confidence.

5.2.1 *The Curse of Perfection.*

Interestingly, our qualitative analysis reveals that participants in the GUI condition felt somewhat uncomfortable if the robot was fully confident: the majority of participants in the GUI condition perceived the robot as "over-confident", which made them distrust the robot and question its "reliability." Thus, participants in the GUI condition showed reservations about the robot's abilities despite it exhibiting 100% confidence. We posit that the perceptional differences of a fully confident robot (100% confidence condition) could have occurred based on three reasons: First, the robot in the EMBODIED condition neither *waited* nor *adjusted its speed*, since both trajectory functions take the respective confidence value as the variable that determines its waiting and travel time. As a result, participants interacted with a robot that, for them, displayed an expected default movement, never deviating from it. Second, the graphical display mounted on the robot's end-effector could have resulted in a higher information salience, which might have made participants in the GUI condition more cautious and more engaged with the displayed information. In a recent think-aloud study, Prabhudesai et al. [69] show that "communicating uncertainty about ML predictions forced people to slow down and think analytically about their decision." The authors conclude that visualising uncertainty made people more aware and considerate, which reduced their over-reliance in AI-assisted decision-support systems. The latter is also reflected in the amount of alterations in participants' opinions, shown in Figure 7. If the robot was fully confident, 6 out of 18 participants in the GUI condition changed their mind during their decision-making, compared to 0 participants in the EMBODIED condition. We expect that an alteration of an opinion can be an indicator for an ongoing reflection and hence can be interpreted as being more considerate. Some participants in the GUI condition even specifically asked for additional explanations and wished to understand the robot's functionalities better. Third, participants in the GUI condition could have been risk-aware based on the fact that the robot repetitively conveyed 100% and might have been taken by surprise by the robot's behavioral consistency, compared to participants in the EMBODIED condition who were satisfied with the expected default movement of the robot.

5.2.2 *Threshold vs. General Heuristic.*

Besides peoples' distrust in a robot that conveys 100% confidence, our qualitative results show that the majority of participants in the GUI condition expressed to have applied a personal risk-threshold in the range of 90% - 95% to inform their decision-making. This risk threshold is higher compared to e.g. a recently conducted gluten detection task, in which people applied an average threshold of 75% to trust a scanner [44]. Participants hereby based their personal threshold on the fact that shipping defect pills poses a critical risk. In addition, several participants explained to have changed their risk-threshold over time, which also indicates a more considered approach to form their decision.

Conversely, participants in the EMBODIED condition retrospectively perceived the robot as rather confident or not confident and did not express any thoughts on thresholds. In addition, our quantitative analysis reveals that participants in the EMBODIED condition placed significantly more packages to the "ready for shipment" pile compared to participants in the GUI condition, which we interpret as an indicator for less caution and risk-awareness. The embodied uncertainty visualisation might have encouraged participants to apply a more general and intuitive heuristic to assess the robot's confidence, instead of carefully checking and comparing the robot's speed and waiting time. This assumption is motivated by Buçinca et al. [15], who proposes that people are *intrinsically* not motivated to approach a decision analytically in an AI-assisted task. The authors suggest that people rather develop an overall rule-of-thumb to generally assess the competence of an AI [15]. We postulate that the same theory can be applied to robot-assisted decision-making, if the uncertainty visualisation allows such a shortcut due to a lack of salience.

5.2.3 *The Willingness to Hand Over Responsibility.*

In our setup, the robot was responsible to assess whether pills are defect or not and to visualise its confidence respectively. As a result, participants had to take into account the robot's assessment and recommendation. Surprisingly, our qualitative analysis reveals that several participants were willing to give up their responsibility in the decision-making process and hand over the final judgement to the robot. This might be due to individual differences in risk perception. Huang et al. [37] highlight that individual risk perceptions might affect the extent to which people are willing to follow a robot in an assisted-decision making task. People with an overall low-risk awareness might be more willing to hand over the final judgement. Further, this result is in line with previous research on AI-assisted decision-making that identified over-reliance as a crucial issue and supports the theory that people tend to avoid cognitive effort if another shortcut presents itself [15].

5.3 Trust, Performance and Transparency

Neither the robot's uncertainty visualisation nor the conveyed confidence level had a significant effect on the perceived overall trust in the robot or its perceived performance. A potential explanation could be the fact that the robot conveyed its uncertainty to inform the participants' decision, without explicitly providing a recommendation of whether a pill box *should* be shipped or tested. Hence, the participant holds the agency to decide. Further, we did not offer the opportunity to verify the accuracy of the robot and to factor in human-robot agreement [56]. We acknowledge the latter in our limitations, see Section 5.4.

Furthermore, transparency has been identified as a highly relevant trust construct when interacting with an automated system or agent [70, 73]. A recent paper by Leusmann et al. [54] proposes that a robot needs to be able to convey its own uncertainty in order to support understandability. We assumed that a graphical display would provide a higher salience and context-awareness, based on the symbolism of an icon array [58] and hence would support transparency. However, contrary to our assumptions, there was no difference between the visualisation groups regarding transparency. We argue that this result is a strong indication for the power of embodied visualisations and their intuitive understanding. If we interpret the EMBODIED condition as an anthropomorphic visualisation and the GUI as a technical visualisation, our results are also not confirming recent findings by Roesler [70] on perceived transparency.

Notably, we observe a significant effect of the robot’s conveyed confidence level on perceived transparency. Participants in the high confidence condition perceived the robot as significantly more transparent compared to the low confidence condition. Interestingly, the items we assessed to determine the perceived transparency focus on how well people *understood* the robot and its functionality (*I understand how the robot works*) [73]. Hence, our results suggest that participants in the low confidence condition felt as if they understood the robot less. Research in psychology has investigated the perception of probabilities. An early study by Kunreuther et al. [51] shows that low probability-high consequence events are particularly difficult to interpret, which is why these settings require rich context and comparison scenarios. Interestingly, our results indicate that (a) the level of confidence a robot conveys affects the user’s *understanding* of the robot and (b) that we need to find ways to enhance users understanding in low-probability scenarios. It is important to highlight that transparency and performance are only two out of many trust-related constructs. Future work should investigate other related constructs such as distrust, reliance and compliance, that have been identified as important trust factors in assisted decision-making [89].

5.3.1 Implications for Robot-assisted Decision-making. Along with the aforementioned findings, we outline fundamental implications and design considerations when visualising uncertainty in a robot-assisted decision-making task.

1. *Uncertainty visualised by a robot affects humans’ decision-making.* Our results show that the type of uncertainty visualisation conveyed by a robot has a significant impact on participants’ decision-making in a high-stake scenario. Future applications have to carefully consider the type of visualisation and the expected confidence range to best support human decision-making in robot-assisted decision tasks.
2. *Caution is needed when visualising a fully confident robot.* In detail, our study shows that repetitively visualising 100% confidence in a robot-assisted decision-task can be a double-edged sword. Participants in the GUI condition perceived the robot as too confident, which caused them to incorrectly decide to test pill boxes. These non-trivial findings suggest that in cases of repetitive 100% probability values, the robot’s confidence should be conveyed using an implicit EMBODIED visualisation. Future research should investigate whether these findings are based on the particular probability value (100%) or based on the repetition itself and hence the behavioural consistency of the robot. Our results also open up the question whether, in case of a 100% confidence, confidence should be visualised at all.
3. *Visualising uncertainty is highly context-dependent.* The aforementioned GUI induced skepticism of a robot conveying 100% confidence can be interpreted as a hindrance or an asset, depending on the application context and goal of the task. A robot using a GUI as a means to convey its confidence enables users to assess a risk in detail, to define or to apply a risk threshold. However, this approach may lead to incorrect decisions caused by overthinking, distrust, or fluctuations in judgment, potentially impacting decision speed. Conversely, our results show that a robot conveying its confidence using an EMBODIED visualisation is perceived as either confident or not confident, with the pitfall of less information granularity and a less detailed risk assessment. We believe that an EMBODIED visualisation could nevertheless be useful for low-stakes scenarios and decisions that require an intuitive understanding of the situation. Further, people with low literacy skills could benefit from such implicit visualisations.
4. *Low probabilities should be explained.* Participants perceived the robot as significantly less transparent in the low confidence condition compared to the high confidence condition. Taking into account our qualitative findings and literature on the perception of probabilities [48, 51], we recommend that low risks displayed by a robot should be further explained and contextualized to support user understanding.
5. *Conveying uncertainty does not affect trust in a robot.* Notably, participants’ overall trust in the robot was high across conditions. This finding raises the question of whether visualising a robot’s uncertainty could potentially prevent negative effects after e.g. a robot’s failure [22] and act as a buffer, cushioning user trust. However, more research is needed to assess the latter in detail.

5.4 Limitations

We acknowledge a number of limitations in our study. First, we situated our study in a high-stake decision-making scenario. Therefore, our findings may not be applicable to more mundane scenarios. However, we were particularly interested in robot-assisted decision-making for critical applications, since robots can be ideally used to be implemented in hazardous environments or to work with hazardous materials. Hence, it is important to understand how humans perceive and assess the robot, and the risks in case of severe consequences. Interestingly, even though we did not assign a cost to “additional testing”, participants did not blindly decided to put packages into the “needs additional testing” area and over-reported the safe decision, as previous studies have shown [44]. However, future research should investigate how putting a cost to testing could affect human’s decision-behaviour and the perception of the robot.

Furthermore, sending a package to either “shipment” or “testing” represents a binary decision, a very common daily life experience [44]. However, we acknowledge the fact that there are more complex decision types, like tactical decision-making [35], and that our results cannot be generalized to other decision categories. Moreover, we acknowledge that our research does not cover the entire range of probability values, but focuses on three distinct categories (low, high and 100%) based on McCaffery et al. [60]. Addressing uncertainty values outside our categories opens up new possibilities for future research. To increase the internal validity of our study, we further conducted our experiment as a lab study instead of a pharmaceutical warehouse. Nevertheless, we aimed to enhance the transferability of our study by designing our setup as realistically as possible. For instance, we prepared the working station to resemble a packing station. We further provided participants with 30 different packages and separate pill boxes, to ensure that they understand that the robot is assessing different and not yet detected pill boxes during each consecutive trial.

We did not find significant difference in participants' trust and perceived performance of the robot. This could be based on the fact that we did not manipulate the robot's competence or error frequency, which has proven to affect people's perceptions negatively [82]. In our study setup, the robot conveyed its confidence in its pill detection, since this study's focus is on the impact of uncertainty visualisations and confidence levels. Future work should aim to address the relationship between trust formation and failure of a robot that conveys its confidence in a robot-assisted decision-making. We believe that visualising a robot's confidence could potentially serve as a failure justification [49], and improve trust and error-prone robot-assisted decision-making.

6 CONCLUSION

We conducted a lab study with 36 participants who collaborated with a robot arm in an assisted decision-making task. Our main objectives were twofold: First, to investigate how a robot that conveys its uncertainty affects the human's decision-making process. Second, to explore how visualising a robot's uncertainty affects the human's perception of the robot itself. We further situated the study in a high-stake scenario. Our findings reveal that the type of visualisation significantly affects users' judgement and decision-behaviour. We show that the robot's conveyed confidence level significantly affects a user's perceived transparency of the robot. Besides, we shed light on a potential trade-off between making an intuitive or an informed decision in a robot-assisted decision-making scenario and how visualising a fully confident robot can be a double-edged sword. Based on our findings, we identify new avenues for future research and provide implications for uncertainty visualisations in critical robot-assisted decision-making tasks. With this study, we take a step closer towards better understanding how data visualisation can be employed in collaborative setting between autonomous embodied agents and humans to support human decision-making.

ACKNOWLEDGMENTS

We wish to acknowledge the contribution of the Melbourne Research Scholarship funded by the University of Melbourne and the Rowden White Scholarship established with a donation from Sir Alfred Edward Rowden White. This research is partially supported by the Australian Research Council Discovery Early Career Research Award (Grant No. DE210100858). Further, we thank Jiahe Pan, Muhammad Bilal and Antony Chacon Salas for their support with the technical implementation.

REFERENCES

[1] FRANKA EMIKA GmbH 2023. *The new FRANKA RESEARCH 3 - The platform of choice for cutting edge AI & Robotics research.* FRANKA EMIKA GmbH. Retrieved September 06, 2023 from https://www.franka.de/research/

[2] Danielle Albers, Michael Correll, and Michael Gleicher. 2014. Task-Driven Evaluation of Aggregation in Time Series Visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 551–560. https://doi.org/10.1145/2556288.2557200

[3] Jessica S. Ancker, Yalini Senathirajah, Rita Kukafka, and Justin B. Starren. 2006. Design Features of Graphs in Health Risk Communication: A Systematic Review. *Journal of the American Medical Informatics Association* 13, 6 (11 2006), 608–618. https://doi.org/10.1197/jamia.M2115

[4] Adriana Arcia, Michael E Bales, William Brown III, Manuel C Co Jr, Melinda Gilmore, Young Ji Lee, Chin S Park, Jennifer Prey, Mark Velez, Janet Woollen, et al. 2013. Method for the development of data visualizations for community members with varying levels of health literacy. In *AMIA Annual Symposium Proceedings*, Vol. 2013. American Medical Informatics Association, 51.

[5] Adriana Arcia, Niurka Suero-Tejeda, Michael E Bales, Jacqueline A Merrill, Sunmoo Yoon, Janet Woollen, and Suzanne Bakken. 2015. Sometimes more is more: iterative participatory design of infographics for engagement of community members with varying levels of health literacy. *Journal of the American Medical Informatics Association* 23, 1 (07 2015), 174–183. https://doi.org/10.1093/jamia/ocv079

[6] S. Sandra Bae, Clement Zheng, Mary Etta West, Ellen Yi-Luen Do, Samuel Huron, and Danielle Albers Szafir. 2022. Making Data Tangible: A Cross-Disciplinary Design Space for Data Physicalization. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. ACM, New York, NY, USA, Article 81, 18 pages. https://doi.org/10.1145/3491102.3501939

[7] Bernard W. Balleine. 2007. The Neural Basis of Choice and Decision Making. *Journal of Neuroscience* 27, 31 (2007), 8159–8160. https://doi.org/10.1523/JNEUROSCI.1939-07.2007

[8] 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* (Yokohama, Japan) *(CHI '21)*. ACM, New York, NY, USA, Article 81, 16 pages. https://doi.org/10.1145/3411764.3445717

[9] Umang Bhatt, Javier Antorán, Yunfeng Zhang, Q. Vera Liao, Prasanna Sattigeri, Riccardo Fogliato, Gabrielle Melançon, Ranganath Krishnan, Jason Stanley, Omesh Tickoo, Lama Nachman, Rumi Chunara, Madhulika Srikumar, Adrian Weller, and Alice Xiang. 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* (Virtual Event, USA) *(AIES '21)*. ACM, New York, NY, USA, 401–413. https://doi.org/10.1145/3461702.3462571

[10] Melissa Bica, Julie L. Demuth, James E. Dykes, and Leysia Palen. 2019. Communicating Hurricane Risks: Multi-Method Examination of Risk Imagery Diffusion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300545

[11] Michelle Borkin, Krzysztof Gajos, Amanda Peters, Dimitrios Mitsouras, Simone Melchionna, Frank Rybicki, Charles Feldman, and Hanspeter Pfister. 2011. Evaluation of Artery Visualizations for Heart Disease Diagnosis. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2479–2488. https://doi.org/10.1109/TVCG.2011.192

[12] Nadia Boukhelifa and David John Duke. 2009. Uncertainty Visualization: Why Might It Fail?. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI EA '09)*. ACM, New York, NY, USA, 4051–4056. https://doi.org/10.1145/1520340.1520616

[13] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[14] Ken Brodlie, Rodolfo Allendes Osorio, and Adriano Lopes. 2012. A review of uncertainty in data visualization. *Expanding the frontiers of visual analytics and visualization* (2012), 81–109.

[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. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 188 (apr 2021), 21 pages. https://doi.org/10.1145/3449287

[16] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 981–992. https://doi.org/10.1145/2858036.2858498

[17] Shiye Cao and Chien-Ming Huang. 2022. Understanding User Reliance on AI in Assisted Decision-Making. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 471 (nov 2022), 23 pages. https://doi.org/10.1145/3555572

[18] Felix Carros, Johanna Meurer, Diana Löffler, David Unbehaun, Sarah Matthies, Inga Koch, Rainer Wieching, Dave Randall, Marc Hassenzahl, and Volker Wulf. 2020. Exploring Human-Robot Interaction with the Elderly: Results from a Ten-Week Case Study in a Care Home. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376402

[19] Hao-Fei Cheng, Ruotong Wang, Zheng Zhang, Fiona O'Connell, Terrance Gray, F. Maxwell Harper, and Haiyi Zhu. 2019. Explaining Decision-Making Algorithms through UI: Strategies to Help Non-Expert Stakeholders. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300789

[20] Alan M. MacEachren Christoph Kinkeldey and Jochen Schiewe. 2014. How to Assess Visual Communication of Uncertainty? A Systematic Review of Geospatial Uncertainty Visualisation User Studies. *The Cartographic Journal* 51, 4 (2014), 372–386. https://doi.org/10.1179/1743277414Y.0000000099

[21] Jae Chung and Steven Wark. 2016. Visualising uncertainty for decision support. *Defence Science and Technology Group* (2016).

[22] Filipa Correia, Carla Guerra, Samuel Mascarenhas, Francisco S. Melo, and Ana Paiva. 2018. Exploring the Impact of Fault Justification in Human-Robot Trust. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems* (Stockholm, Sweden) *(AAMAS '18)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 507–513.

[23] Mohammad Daradkeh, Clare Churcher, and Alan McKinnon. 2013. Supporting Informed Decision-Making under Uncertainty and Risk through Interactive Visualisation. In *Proceedings of the Fourteenth Australasian User Interface Conference - Volume 139* (Melbourne, Australia) *(AUIC '13)*. Australian Computer Society, Inc., AUS, 23–32.

[24] Eric Deng, Bilge Mutlu, and Maja J Mataric. 2019. Embodiment in Socially Interactive Robots. *Foundations and Trends® in Robotics* 7, 4 (2019), 251–356. https://doi.org/10.1561/2300000056

[25] George Denison. 2023. *How much should you pay research participants?* Prolific. Retrieved August 14, 2023 from https://www.prolific.co/blog/how-much-should-you-pay-research-participants

[26] Ruta Desai, Fraser Anderson, Justin Matejka, Stelian Coros, James McCann, George Fitzmaurice, and Tovi Grossman. 2019. Geppetto: Enabling Semantic Design of Expressive Robot Behaviors. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300599

[27] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39, 2 (2007), 175–191. https://doi.org/10.3758/BF03193146

[28] S Larry Goldenberg, Guy Nir, and Septimiu E Salcudean. 2019. A new era: artificial intelligence and machine learning in prostate cancer. *Nature Reviews Urology* 16, 7 (2019), 391–403. https://doi.org/10.1038/s41585-019-0193-3

[29] 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* (Denver, Colorado, USA) *(CHI EA '17)*. ACM, New York, NY, USA, 593–600. https://doi.org/10.1145/3027063.3027091

[30] Jian Guo, Xiaodong Zhang, Haiming Shen, Haibo Xu, and Liangliang Li. 2020. Design on Control System of Bionic Crawling Inspection Robot. In *Proceedings of the 2019 2nd International Conference on Control and Robot Technology* (Jeju, Republic of Korea) *(ICCRT 2019)*. ACM, New York, NY, USA, 24–29. https://doi.org/10.1145/3387304.3387312

[31] Zixuan Guo and Tomoo Inoue. 2019. Using a Conversational Agent to Facilitate Non-Native Speaker's Active Participation in Conversation. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. ACM, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3313075

[32] Sarah T. Hawley, Brian Zikmund-Fisher, Peter Ubel, Aleksandra Jancovic, Todd Lucas, and Angela Fagerlin. 2008. The impact of the format of graphical presentation on health-related knowledge and treatment choices. *Patient Education and Counseling* 73, 3 (2008), 448–455. https://doi.org/10.1016/j.pec.2008.07.023 4th International Conference on Shared Decision Making.

[33] Thomas Hellström and Suna Bensch. 2018. Understandable robots - What, Why, and How. *Paladyn, Journal of Behavioral Robotics* 9, 1 (2018), 110–123. https://doi.org/doi:10.1515/pjbr-2018-0009

[34] Alejandra C. Hernández, Clara Gómez, Jonathan Crespo, and Ramón Barber. 2017. Adding Uncertainty to an Object Detection System for Mobile Robots. In *2017 6th International Conference on Space Mission Challenges for Information Technology (SMC-IT)*. 7–12.

[35] Carl-Johan Hoel, Katherine Driggs-Campbell, Krister Wolff, Leo Laine, and Mykel J. Kochenderfer. 2020. Combining Planning and Deep Reinforcement Learning in Tactical Decision Making for Autonomous Driving. *IEEE Transactions on Intelligent Vehicles* 5, 2 (2020), 294–305.

[36] Julian Hough and David Schlangen. 2017. It's Not What You Do, It's How You Do It: Grounding Uncertainty for a Simple Robot. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna, Austria) *(HRI '17)*. ACM, New York, NY, USA, 274–282. https://doi.org/10.1145/2909824.3020214

[37] Hanjing Huang, Pei-Luen Patrick Rau, and Liang Ma. 2021. Will you listen to a robot? Effects of robot ability, task complexity, and risk on human decision-making. *Advanced Robotics* 35, 19 (2021), 1156–1166. https://doi.org/10.1080/01691864.2021.1974940

[38] Jessica Hullman. 2016. Why Evaluating Uncertainty Visualization is Error Prone. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization* (Baltimore, MD, USA) *(BELIV '16)*. ACM, New York, NY, USA, 143–151.

[39] Jessica Hullman, Xiaoli Qiao, Michael Correll, Alex Kale, and Matthew Kay. 2019. In Pursuit of Error: A Survey of Uncertainty Visualization Evaluation. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 903–913. https://doi.org/10.1109/TVCG.2018.2864889

[40] Anil Kumar Inkulu, MVA Raju Bahubalendruni, and Ashok Dara. 2022. Challenges and opportunities in human robot collaboration context of Industry 4.0-a state of the art review. *Industrial Robot: the international journal of robotics research and application* 49, 2 (2022), 226–239.

[41] Yvonne Jansen, Pierre Dragicevic, Petra Isenberg, Jason Alexander, Abhijit Karnik, Johan Kildal, Sriram Subramanian, and Kasper Hornbæk. 2015. Opportunities and Challenges for Data Physicalization. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 3227–3236. https://doi.org/10.1145/2702123.2702180

[42] Luke Jefferson and Richard Harvey. 2007. An Interface to Support Color Blind Computer Users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. ACM, New York, NY, USA, 1535–1538. https://doi.org/10.1145/1240624.1240855

[43] Feng Jiang, Su Lu, Xiang Yao, Xiaodong Yue, and Wing tung Au. 2014. Up or Down? How Culture and Color Affect Judgments. *Journal of Behavioral Decision Making* 27, 3 (2014), 226–234. https://doi.org/10.1002/bdm.1800

[44] Weiwei Jiang, Zhanna Sarsenbayeva, Niels van Berkel, Chaofan Wang, Difeng Yu, Jing Wei, Jorge Goncalves, and Vassilis Kostakos. 2021. User Trust in Assisted Decision-Making Using Miniaturized Near-Infrared Spectroscopy. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. ACM, New York, NY, USA, Article 153, 16 pages. https://doi.org/10.1145/3411764.3445710

[45] Susan Joslyn and Jared LeClerc. 2013. Decisions With Uncertainty: The Glass Half Full. *Current Directions in Psychological Science* 22, 4 (2013), 308–315. https://doi.org/10.1177/0963721413481473

[46] 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* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 5092–5103. https://doi.org/10.1145/2858036.2858558

[47] Matthew Kay, Dan Morris, mc schraefel, and Julie A. Kientz. 2013. There's No Such Thing as Gaining a Pound: Reconsidering the Bathroom Scale User Interface. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Zurich, Switzerland) *(UbiComp '13)*. ACM, New York, NY, USA, 401–410. https://doi.org/10.1145/2493432.2493456

[48] Mel W. Khaw, Luminita Stevens, and Michael Woodford. 2021. Individual differences in the perception of probability. *PLOS Computational Biology* 17, 4 (04 2021), 1–25. https://doi.org/10.1371/journal.pcbi.1008871

[49] Ross A Knepper, Stefanie Tellex, Adrian Li, Nicholas Roy, and Daniela Rus. 2015. Recovering from failure by asking for help. *Autonomous Robots* 39 (2015), 347–362. https://doi.org/10.1007/s10514-015-9460-1

[50] Michelle Korporaal, Ian T. Ruginski, and Sara Irina Fabrikant. 2020. Effects of Uncertainty Visualization on Map-Based Decision Making Under Time Pressure. *Frontiers in Computer Science* 2 (2020). https://doi.org/10.3389/fcomp.2020.00032

[51] Howard Kunreuther, Nathan Novemsky, and Daniel Kahneman. 2001. Making low probabilities useful. *Journal of risk and uncertainty* 23 (2001), 103–120.

[52] Nancy K Lankton, D Harrison McKnight, and John Tripp. 2015. Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems* 16, 10 (2015), 1.

[53] Gregory Lemasurier, Gal Bejerano, Victoria Albanese, Jenna Parrillo, Holly A. Yanco, Nicholas Amerson, Rebecca Hetrick, and Elizabeth Phillips. 2021. Methods for Expressing Robot Intent for Human–Robot Collaboration in Shared Workspaces. *J. Hum.-Robot Interact.* 10, 4, Article 40 (sep 2021), 27 pages. https://doi.org/10.1145/3472223

[54] Jan Leusmann, Chao Wang, Michael Gienger, Albrecht Schmidt, and Sven Mayer. 2023. Understanding the Uncertainty Loop of Human-Robot Interaction. *arXiv preprint arXiv:2303.07889* (2023).

[55] Antonio Loquercio, Mattia Segu, and Davide Scaramuzza. 2020. A General Framework for Uncertainty Estimation in Deep Learning. *IEEE Robotics and Automation Letters* 5, 2 (2020), 3153–3160. https://doi.org/10.1109/LRA.2020.2974682

[56] Zhuoran Lu and Ming Yin. 2021. Human Reliance on Machine Learning Models When Performance Feedback is Limited: Heuristics and Risks. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, New York, NY, USA, Article 78, 16 pages. https://doi.org/10.1145/3411764.3445562

[57] Bing L. Luk, David S. Cooke, Stuart Galt, Arthur A. Collie, and Sheng Chen. 2005. Intelligent legged climbing service robot for remote maintenance applications in hazardous environments. *Robotics and Autonomous Systems* 53, 2 (2005), 142–152. https://doi.org/10.1016/j.robot.2005.06.004

[58] Alan M. MacEachren, Robert E. Roth, James O'Brien, Bonan Li, Derek Swingley, and Mark Gahegan. 2012. Visual Semiotics & Uncertainty Visualization: An Empirical Study. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2496–2505. https://doi.org/10.1109/TVCG.2012.279

[59] Steven Macenski, Tully Foote, Brian Gerkey, Chris Lalancette, and William Woodall. 2022. Robot Operating System 2: Design, architecture, and uses in the wild. *Science Robotics* 7, 66 (2022), eabm6074. https://doi.org/10.1126/scirobotics.abm6074

[60] Kirsten J. McCaffery, Ann Dixon, Andrew Hayen, Jesse Jansen, Sian Smith, and Judy M. Simpson. 2012. The Influence of Graphic Display Format on the Interpretations of Quantitative Risk Information among Adults with Lower Education

and Literacy: A Randomized Experimental Study. *Medical Decision Making* 32, 4 (2012), 532–544. https://doi.org/10.1177/0272989X11424926 PMID: 22074912.

[61] Ajung Moon, Maneezhay Hashmi, H. F. Machiel Van Der Loos, Elizabeth A. Croft, and Aude Billard. 2021. Design of Hesitation Gestures for Nonverbal Human-Robot Negotiation of Conflicts. *J. Hum.-Robot Interact.* 10, 3, Article 24 (jul 2021), 25 pages. https://doi.org/10.1145/3418302

[62] AJung Moon, Chris A. C. Parker, Elizabeth A. Croft, and H. F. Machiel Van der Loos. 2011. Did you see it hesitate? - Empirically grounded design of hesitation trajectories for collaborative robots. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 1994–1999. https://doi.org/10.1109/IROS.2011.6094605

[63] AJung Moon, Chris A. C. Parker, Elizabeth A. Croft, and H. F. Machiel Van der Loos. 2013. Design and Impact of Hesitation Gestures during Human-Robot Resource Conflicts. *J. Hum.-Robot Interact.* 2, 3 (sep 2013), 18–40. https://doi.org/10.5898/JHRI.2.3.Moon

[64] Lace Padilla, Matthew Kay, and Jessica Hullman. 2020. Uncertainty visualization. (2020).

[65] Lace M Padilla, Ian T Ruginski, and Sarah H Creem-Regehr. 2017. Effects of ensemble and summary displays on interpretations of geospatial uncertainty data. *Cognitive research: principles and implications* 2 (2017), 1–16.

[66] Max Pascher, Uwe Gruenefeld, Stefan Schneegass, and Jens Gerken. 2023. How to Communicate Robot Motion Intent: A Scoping Review. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. ACM, New York, NY, USA, Article 409, 17 pages. https://doi.org/10.1145/3544548.3580857

[67] Ameya Patil, Gaëlle Richer, Christopher Jermaine, Dominik Moritz, and Jean-Daniel Fekete. 2023. Studying Early Decision Making with Progressive Bar Charts. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 407–417. https://doi.org/10.1109/TVCG.2022.3209426

[68] Aaron Powers and Sara Kiesler. 2006. The Advisor Robot: Tracing People's Mental Model from a Robot's Physical Attributes. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction* (Salt Lake City, Utah, USA) *(HRI '06)*. ACM, New York, NY, USA, 218–225. https://doi.org/10.1145/1121241.1121280

[69] 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* (Sydney, NSW, Australia) *(IUI '23)*. ACM, New York, NY, USA, 379–396. https://doi.org/10.1145/3581641.3584033

[70] Eileen Roesler. 2023. Anthropomorphic framing and failure comprehensibility influence different facets of trust towards industrial robots. *Frontiers in Robotics and AI* 10 (2023). https://doi.org/10.3389/frobt.2023.1235017

[71] E. Roesler, D. Manzey, and L. Onnasch. 2021. A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. *Science Robotics* 6, 58 (2021), eabj5425. https://doi.org/10.1126/scirobotics.abj5425

[72] Eileen Roesler, Dietrich Manzey, and Linda Onnasch. 2023. Embodiment Matters in Social HRI Research: Effectiveness of Anthropomorphism on Subjective and Objective Outcomes. *J. Hum.-Robot Interact.* 12, 1, Article 7 (feb 2023), 9 pages. https://doi.org/10.1145/3555812

[73] Eileen Roesler*, Tobias Rieger*, and Dietrich Manzey. 2022. Trust towards Human vs. Automated Agents: Using a Multidimensional Trust Questionnaire to Assess The Role of Performance, Utility, Purpose, and Transparency. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 66, 1 (2022), 2047–2051. https://doi.org/10.1177/1071181322661065

[74] Abhraneel Sarma, Shunan Guo, Jane Hoffswell, Ryan Rossi, Fan Du, Eunyee Koh, and Matthew Kay. 2023. Evaluating the Use of Uncertainty Visualisations for Imputations of Data Missing At Random in Scatterplots. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 602–612. https://doi.org/10.1109/TVCG.2022.3209348

[75] Marilyn M. Schapira, Ann B. Nattinger, and Colleen A. McHorney. 2001. Frequency or Probability? A Qualitative Study of Risk Communication Formats Used in Health Care. *Medical Decision Making* 21, 6 (2001), 459–467. https://doi.org/10.1177/0272989X0102100604 PMID: 11760103.

[76] Jonathan Schwabish. 2021. *4. COMPARING CATEGORIES*. Columbia University Press, New York Chichester, West Sussex, 67–132. https://doi.org/doi:10.7312/schw19310-005

[77] Sarah Schömbs, Jorge Goncalves, and Wafa Johal. 2024. Exploring Data Agency and Autonomous Agents as Embodied Data Visualizations. arXiv:2402.04598 [cs.HC]

[78] Ron Sellers. 2013. How sliders bias survey data. *MRA's Alert* 53, 3 (2013), 56–57.

[79] Samuel Silva, Beatriz Sousa Santos, and Joaquim Madeira. 2011. Using color in visualization: A survey. *Computers & Graphics* 35, 2 (2011), 320–333. Virtual Reality in Brazil Visual Computing in Biology and Medicine Semantic 3D media and content Cultural Heritage.

[80] Meredith Skeels, Bongshin Lee, Greg Smith, and George Robertson. 2008. Revealing Uncertainty for Information Visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (Napoli, Italy) *(AVI '08)*. ACM, New York, NY, USA, 376–379.

[81] 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* (Honolulu, HI, USA) *(CHI EA '20)*. ACM, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3375177

[82] Maia Stiber and Chien-Ming Huang. 2021. Not All Errors Are Created Equal: Exploring Human Responses to Robot Errors with Varying Severity. In *Companion Publication of the 2020 International Conference on Multimodal Interaction* (Virtual Event, Netherlands) *(ICMI '20 Companion)*. ACM, New York, NY, USA, 97–101. https://doi.org/10.1145/3395035.3425245

[83] Megan Strait, Cody Canning, and Matthias Scheutz. 2014. Let Me Tell You! Investigating the Effects of Robot Communication Strategies in Advice-Giving Situations Based on Robot Appearance, Interaction Modality and Distance. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (Bielefeld, Germany) *(HRI '14)*. ACM, New York, NY, USA, 479–486. https://doi.org/10.1145/2559636.2559670

[84] Daniel Szafir and Danielle Albers Szafir. 2021. Connecting Human-Robot Interaction and Data Visualization. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '21)*. ACM, New York, NY, USA, 281–292.

[85] Christoph Tauchert, Neda Mesbah, et al. 2019. Following the Robot? Investigating Users' Utilization of Advice from Robo-Advisors.. In *ICIS*.

[86] Cristen Torrey, Susan R. Fussell, and Sara Kiesler. 2013. How a robot should give advice. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 275–282. https://doi.org/10.1109/HRI.2013.6483599

[87] James Trevelyan, William R. Hamel, and Sung-Chul Kang. 2016. *Robotics in Hazardous Applications*. Springer International Publishing, Cham, 1521–1548. https://doi.org/10.1007/978-3-319-32552-1_58

[88] Rik van den Brule, Ron Dotsch, Gijsbert Bijlstra, Daniel Wigboldus, and Pim Haselager. 2014. Do Robot Performance and Behavioral Style affect Human Trust?: A Multi-Method Approach. *International Journal of Social Robotics* 6 (03 2014), 519–531. https://doi.org/10.1007/s12369-014-0231-5

[89] Oleksandra Vereschak, Gilles Bailly, and Baptiste Caramiaux. 2021. How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 327 (oct 2021), 39 pages. https://doi.org/10.1145/3476068

[90] Joshua Wainer, David J. Feil-Seifer, Dylan A. Shell, and Maja J. Mataric. 2007. Embodiment and Human-Robot Interaction: A Task-Based Perspective. In *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*. 872–877.

[91] Astrid Weiss, Ann-Kathrin Wortmeier, and Bettina Kubicek. 2021. Cobots in Industry 4.0: A Roadmap for Future Practice Studies on Human–Robot Collaboration. *IEEE Transactions on Human-Machine Systems* 51, 4 (2021), 335–345. https://doi.org/10.1109/THMS.2021.3092684

[92] Nan Wu, Jason Phang, Jungkyu Park, Yiqiu Shen, Zhe Huang, Masha Zorin, Stanisław Jastrzębski, Thibault Févry, Joe Katsnelson, Eric Kim, Stacey Wolfson, Ujas Parikh, Sushma Gaddam, Leng Leng Young Lin, Kara Ho, Joshua D. Weinstein, Beatriu Reig, Yiming Gao, Hildegard Toth, Kristine Pysarenko, Alana Lewin, Jiyon Lee, Krystal Airola, Eralda Mema, Stephanie Chung, Esther Hwang, Naziya Samreen, S. Gene Kim, Laura Heacock, Linda Moy, Kyunghyun Cho, and Krzysztof J. Geras. 2020. Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. *IEEE Transactions on Medical Imaging* 39, 4 (2020), 1184–1194. https://doi.org/10.1109/TMI.2019.2945514

[93] Qian Yang, Aaron Steinfeld, and John Zimmerman. 2019. Unremarkable AI: Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes. *CoRR* abs/1904.09612 (2019). http://arxiv.org/abs/1904.09612

[94] X. Jessie Yang, Vaibhav V. Unhelkar, Kevin Li, and Julie A. Shah. 2017. Evaluating Effects of User Experience and System Transparency on Trust in Automation. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna, Austria) *(HRI '17)*. ACM, New York, NY, USA, 408–416. https://doi.org/10.1145/2909824.3020230

[95] Lixuan Zhang, Iryna Pentina, and Yuhong Fan. 2021. Who do you choose? Comparing perceptions of human vs robo-advisor in the context of financial services. *Journal of Services Marketing* 35, 5 (2021), 634–646.

[96] Allan Zhou, Dylan Hadfield-Menell, Anusha Nagabandi, and Anca D. Dragan. 2017. Expressive Robot Motion Timing. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna, Austria) *(HRI '17)*. ACM, New York, NY, USA, 22–31.

[97] Tian Zhou, Jackie S. Cha, Glebys Gonzalez, Juan P. Wachs, Chandru P. Sundaram, and Denny Yu. 2020. Multimodal Physiological Signals for Workload Prediction in Robot-Assisted Surgery. *J. Hum.-Robot Interact.* 9, 2, Article 12 (jan 2020), 26 pages. https://doi.org/10.1145/3368589

[98] Brian J. Zikmund-Fisher, Holly O. Witteman, Mark Dickson, Andrea Fuhrel-Forbis, Valerie C. Kahn, Nicole L. Exe, Melissa Valerio, Lisa G. Holtzman, Laura D. Scherer, and Angela Fagerlin. 2014. Blocks, Ovals, or People? Icon Type Affects Risk Perceptions and Recall of Pictographs. *Medical Decision Making* 34, 4 (2014), 443–453. https://doi.org/10.1177/0272989X13511706 PMID: 24246564.