RESEARCH-ARTICLE

# Who is in Control? Understanding User Agency in AR-assisted Construction Assembly

**XILIU YANG**, University of Stuttgart, Stuttgart, Baden-Wurttemberg, Germany

**PRASANTH SASIKUMAR**, National University of Singapore, Singapore City, Singapore

**FELIX AMTSBERG**, University of Stuttgart, Stuttgart, Baden-Wurttemberg, Germany

**ACHIM MENGES**, University of Stuttgart, Stuttgart, Baden-Wurttemberg, Germany

**MICHAEL SEDLMAIR**, University of Stuttgart, Stuttgart, Baden-Wurttemberg, Germany

**SURANGA NANAYAKKARA**, National University of Singapore, Singapore City, Singapore

Open Access Support provided by:

University of Stuttgart

National University of Singapore



**Published:** 26 April 2025

Citation in BibTeX format

CHI 2025: CHI Conference on Human Factors in Computing Systems
*April 26 - May 1, 2025*
*Yokohama, Japan*

**Conference Sponsors:**
SIGCHI

CHI '25: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (April 2025)
https://doi.org/10.1145/3706598.3713765
ISBN: 9798400713941

# Who is in Control? Understanding User Agency in AR-assisted Construction Assembly

Xiliu Yang
Institute for Computational Design and Construction
University of Stuttgart
Stuttgart, Germany
xiliu.yang@icd.uni-stuttgart.de

Prasanth Sasikumar
Augmented Human Lab
National University of Singapore
Singapore, Singapore
prasanth.sasikumar@nus.edu.sg

Felix Amtsberg
Institute for Computational Design and Construction
University of Stuttgart
Stuttgart, Germany
felix.amtsberg@icd.uni-stuttgart.de

Achim Menges
Institute for Computational Design and Construction
University of Stuttgart
Stuttgart, Germany
achim.menges@icd.uni-stuttgart.de

Michael Sedlmair
Visualization Research Center
University of Stuttgart
Stuttgart, Germany
michael.sedlmair@visus.uni-stuttgart.de

Suranga Nanayakkara
Augmented Human Lab
National University of Singapore
Singapore, Singapore
suranga@ahlab.org

**Figure 1: Photos of the user study with eight carpenters: AR-assisted physical assembly tasks (left), AR-assisted cognitive assembly tasks (middle), and interview sessions (right).**

## Abstract

Adaptive AR assistance can automatically trigger content to support users based on their context. Such intelligent automation offers many benefits but also alters users' degree of control, which is seldom explored in existing research. In this paper, we compare high- and low-agency control in AR-assisted construction assembly to understand the role of user agency. We designed cognitive and physical assembly scenarios and conducted a lab study (N=24), showing that low-agency control reduced mental workloads and perceived autonomy in several tasks. A follow-up domain expert study with trained carpenters (N=8) contextualised these results in an ecologically valid setting. Through semi-structured interviews, we examined the carpenters' perspectives on AR support in their daily work and the trade-offs of automating interactions. Based on these findings, we summarise key design considerations to inform future adaptive AR designs in the context of timber construction.

## CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; **Mixed / augmented reality**; **User studies**.


*CHI '25, Yokohama, Japan*

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713765

## Keywords

Augmented Reality, Worker Assistance, User Agency, Construction Industry

**ACM Reference Format:**
Xiliu Yang, Prasanth Sasikumar, Felix Amtsberg, Achim Menges, Michael Sedlmair, and Suranga Nanayakkara. 2025. Who is in Control? Understanding User Agency in AR-assisted Construction Assembly. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3706598.3713765

## 1 Introduction

The construction industry is facing significant labour shortages as well as one of the highest rates of workplace injuries and fatalities [42]. Deployment of augmented reality (AR) applications can potentially contribute to mitigating these issues by aiding in the training of workers [51] and improving safety [30]. Context-aware adaptations of AR content can further enhance the performance of these systems [18].

To illustrate the use of AR assistance in construction assembly, imagine a construction worker equipped with an AR headset assembling a structure on site. The system automatically displays instructions when certain assembly states are detected [15] and dynamically adjusts the display content by monitoring their visual attention and cognitive workload [38]. While these automations reduce the burden of manually interacting with the interface, the worker does not have direct control over the instructions or display changes. Situations may arise where automation fails to align with their needs or makes choices confusing to them [7].

This scenario highlights an important question in adaptive AR systems: how to balance automation with preserving user control? Agency and autonomy are key design objectives in HCI, where systems should "support an internal locus of control" for users [47]. In recent years, the urgency of this issue is underscored by the rapid rise of intelligent systems and ever-closer integration of technologies and bodies [2].

The trade-off above is under-explored in recent work on AR-assisted assembly, which primarily focuses on the performance of adaptive systems as a whole [22, 38, 44]. Additionally, empirically evaluating this trade-off is difficult due to the interplay of two components in adaptive systems – (1) the selective *automation* of otherwise user-initiated actions, and (2) the design of intelligent *feedback* to determine when such automation should occur. Design-specific decisions on the feedback component introduces variability and complicates the evaluation of reduced user control.

To address this challenge in the context of construction assembly, the main objectives of this research are twofold. First, we empirically evaluate the trade-offs between high- and low-agency control for AR-assisted construction assembly. Second, we explore these trade-offs and evaluate the system with construction workers to contribute design insights suitable for construction applications.

For this purpose, we developed two operating modes for a head-mounted AR assistance system: (1) a *high-agency* interactive mode where users manually trigger the cues and (2) a *low-agency* automated mode where cues are triggered without user input. The *low-agency* mode runs on a fixed schedule, thus exposing all users to the same experience and eliminating the variability of feedback-based differences.

We first conducted a lab study with participants recruited from the university (N=24) to examine the differences in task performance, usability, perceived workload and psychological needs. We found lower autonomy and lower cognitive workloads in several tasks with the low-agency mode but no significant difference in performance or user experience.

To contextualise these results in an ecologically valid setting, we conducted an expert study with trained carpenters (N=8) at a construction company using the same system and tasks. We evaluated the task validity, AR system design, and carpenters' perspectives on user agency in their daily work. We summarise our findings to support future research on AR systems with higher degrees of system autonomy in construction assembly. In summary, our contributions are:

(1) Empirical evidence on the differences between high- and low-agency control modes in AR-assisted assembly;
(2) Insights from expert users on task design and AR system design for applications in the construction industry.
(3) Key design considerations for balancing user control and system autonomy in designing AR support.

## 2 Background and Related Work

First, we review the current research on adaptive AR applications and highlight the relevance of understanding the impact of reduced user control. We then define the contributions of our study in relation to broader literature in HCI that deals with user control and agency. Lastly, we review relevant AR task support strategies suitable for our application area in construction assembly, informing our study design.

### 2.1 Adaptive AR Assistance

Chiossi et al. outlined three groups of visual adaptations to the user – presentation, interaction, and content [6]. Our work focuses on content adaptations in assembly tasks, which can be categorised into four types of adaptations. **(1) A-priori task difficulty or user preferences** can be used to adapt AR work instructions, e.g., based on operator vision and content preferences [16] or assembly task difficulty [44]. **(2) Real-time sensing of the environment** can be used to detect the assembly state [15, 48], which reduces the need for human interaction and controls for potential human errors. **(3) Real-time sensing of the user** can detect motion, attention, and other mental and affective states through physiological signals, with eye-tracking reported as the most popular technique [50]. Lastly, **(4) multi-modal approaches** combines the above approaches, e.g., Huang et al. used headset and environment tracking to update levels of details (LOD) for embodied machine tasks [22].

In the examples above, the user's control is mediated by both the sensing channels available in the system and the context-to-content associations defined by the designer (Figure 2). As such, user agency is diminished but cannot be distinguished from the specific design of the adaptive loop. In fact, user autonomy is rarely the focal point of evaluations in recent research on adaptive AR. This gap is also highlighted by recent reviews that examine user

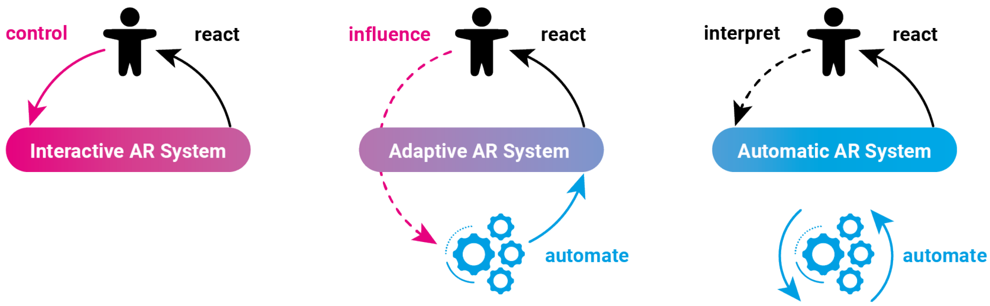



**Figure 2: Illustration of user autonomy in different systems. Interactive: users *control* the system and react to the results of their interactions, i.e., high-agency. Adaptive: users *influence* the system through certain sensing mechanisms and react to feedback generated by the system's control logic, i.e., user agency is dependent on the system parameters. Automatic: users do not have, or cannot perceive, their influence on the system and react to the system's behaviour on their own terms, i.e., low-agency.**

autonomy in HCI, where industrial task execution represents a little addressed application domain [2, 32].

Interactive and automatic systems, on the other hand, define the two ends of this spectrum and provide clearer insights into the trade-offs between user control and system autonomy. Our motivation for this study is grounded in the idea that interpolation between these two can lead to new insights for understanding the impact of autonomy and striking a productive balance in adaptive system designs for assembly task support.

## 2.2 User Control and Agency

Supporting agency and control is a widely accepted design objective in HCI, e.g., in Shneiderman and Plaisant's Eight Rules of Interface Design [47]. Its value is considered to have both intrinsic value and positive influences on technology satisfaction and user experience [34].

Based on a review of 32 years of HCI research on agency and autonomy, Bennett et al. outlined the diverse definitions, research approaches, and application domains that addressed this topic [2]. Building upon the agency dimensions highlighted in this review, we consider this term within the scope of this paper as the user's *independent and causal involvement in executing tasks and making decisions*. We focus on an *episodic* time-scale during interactions with AR support within construction assembly tasks, which does not address shorter (seconds) or longer (years) impacts.

Existing work with similar definitions has examined the balance of user control and automation in a variety of domains, e.g., proactive voice assistants [43], player control in video games [7], and autopilot aircraft supervisions [3]. In a recent study on human-LLM collaboration, Guo et al. conducted a study using two design probes with different interaction modes [19]. Using an open-ended, high-agency mode and a structured, low-agency mode, the authors examined differences in performance, user behaviour, and perceptions with 9 data scientists [19].

In a similar vein, our work contrasts two interaction modes for AR-assisted construction assembly. We combine an evaluation with domain experts, focusing on ecologically valid design insights, with a lab study that involves novice users, to allow a more comprehensive understanding of the effects of user control in AR-assisted construction assembly.

## 2.3 AR Support for Construction Assembly

The degree to which users value control in technology use is dependent on usage scenario and users' primary needs – e.g., Lukoff et al. noted different requirements for user agency between open-ended use (for diversionary needs) and goal-directed use (for informational needs) in YouTube video consumption [34].

To generate insights relevant to construction applications, we consider user needs under task scenarios specific to this context. Imagine a carpenter conducting an assembly task – they alternate between planning the assembly sequence and manipulating heavy materials to build the structure. These sub-tasks vary in cognitive and physical demands, which give rise to unique user needs and task support requirements.

The vast majority of AR applications in industrial settings currently focus on cognitively intensive tasks. They provide *task-oriented support* with the goal of enhancing performance or reducing errors. Physically demanding tasks are scarce in this context, but often addressed in exertion games and physical training [40]. Many physical tasks in construction are often repetitive in nature, for which *well-being-oriented support* that targets users' intrinsic motivation is particularly relevant.

*2.3.1 Task-oriented AR Support.* Task-related AR cues present information that directly impacts the users' ability to execute tasks, e.g., as overlaid assembly geometries or work instructions. These systems support users in cognitively demanding tasks such as high-variant product assembly, complex machine operations, and object identification [15, 22, 24]. While some studies showed that AR is more favourable than paper instructions in reducing errors and working memory demands [26], some also highlighted a loss of agency under AR conditions represented through users' hesitant behaviour [17]. Our study contributes to this body of work by comparing a user-driven, interactive mode and a low-agency, automatic mode for task-oriented AR support.

#### 2.3.2 *Well-being-oriented AR Support.*
In industry and workplace scenarios, well-being beyond physical ergonomics remains an under-explored area in HCI studies [12]. One influential theory in this context is Self-Determination Theory (SDT), which outlines three psychological needs (competence, autonomy, and relatedness) that lead to higher intrinsic motivation and well-being when satisfied [45]. Dhiman et al. demonstrated the use of SDT in designing projected instructions for woodworking, which led to higher perceived competence, intrinsic motivation, and task execution quality [13]. In the physical domain, i.e. sports and physical education, SDT has been implemented in coaching strategies, e.g., autonomy-supportive coaching (ASC), which shows positive effects on performance, particularly in the youth sport context [1, 8]. We add to this line of investigation by testing a coaching cue that provides acknowledgement and encouragement during repetitive, physical assembly tasks.

## 3 Task and AR System Design

We designed a user study for AR-assisted construction assembly to investigate the effects of high- v.s. low-agency control. The assembly tasks, AR system, and interaction conditions are repeated across the two studies, which we present below.

### 3.1 Tasks

First, we designed four assembly tasks that capture relevant conditions for timber construction assembly. This includes four tasks that vary in cognitive and physical demands.

#### 3.1.1 *Cognitively Intensive Task: Luban Lock.*
Cognitive tasks involve the assembly of two connector-free timber joints, which require dexterity but lower physical effort. These joints resemble those in traditional mortise-and-tenon timber construction [49], and each consists of 6 elements that can be assembled into a solid cross. Known also as the 6-piece Burr puzzle, each assembly has a *difficulty index* based on its combinatorial complexity. Six unique elements without rotational symmetry create a puzzle that has a *difficulty index* of 3840 ($2^5 \times 5!$) [9]. Excluding the "key" element, the two puzzles we selected have 2 and 5 unique elements for the easy and hard tasks respectively.

Since following step-by-step instructions does not necessarily induce high cognitive loads, we designed this to be a learning task. That is, the task includes a *practice* phase with AR support, followed by an *evaluation* phase where users solve the assembly independently. During practice, the AR system provided five scaffolding steps, shown in Figure 3.

#### 3.1.2 *Physically Intensive Task: Block Assembly.*
The physical tasks are modelled after a material handling and assembly process, where 31 x 27 x 26 cm wooden boxes are picked up, moved, and placed at target positions following AR holograms. The cognitive effort for aligning the boxes is low – if we appropriate the concept of a difficulty index, it would be 24 ($1 \times 4!$) for four possible positions with no rotational asymmetry. We designed four different assemblies, taking place on two tables 4 metres apart. We avoided assembly on the ground to minimise the risks of back injury. The boxes are unweighted (1kg) and weighted (6kg) in the easy and hard rounds, which stay within payload limits from ergonomic guidelines [28].

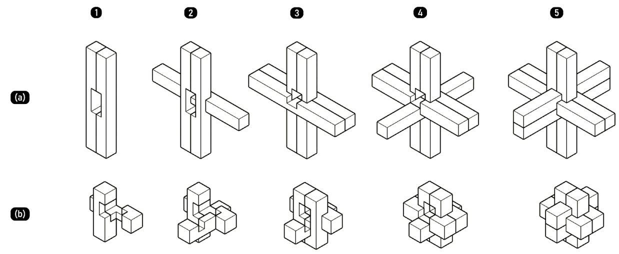



**Figure 3: Five scaffolding steps are shown in AR for each interlocking joint in the cognitive assembly task. (a) Easy: difficulty index 480 ($2^2 \times 5!$) (b) Hard: difficulty index 3840 ($2^5 \times 5!$).**

The task begins with all four boxes positioned on one table. The user transfers the boxes one at a time onto the other table to complete the first assembly. This action is then repeated, moving the boxes back onto the first table and so forth. Each box movement is visually represented by a hologram, and the users are asked to complete these tasks as fast as possible while ensuring safety.

### 3.2 AR System

Given the two types of task support reviewed in 2.3, we designed AR cues for the cognitive and physical assembly tasks respectively. Below, we describe the design rationale for the two cues followed by their implementations.

#### 3.2.1 *Task-oriented Cues.*
For the cognitive assembly tasks, the elements have complex interlocking relationships that require spatial reasoning and memory to be solved effectively. The AR cues are designed to scaffold the user sequentially, i.e., reducing the degree of freedom in the next action, with each step [21]. The first element is provided at the start and five scaffolding cues were included for the six-piece assembly. Each step includes one active element movable by near-hand gestures so users can examine the piece and its relation to neighbours in detail (Figure 4 left). At the start of the following step, the previously active piece turns grey and its position is reset and locked.

#### 3.2.2 *Well-being-oriented Cues.*
During physically demanding tasks, we focused on support for well-being. We designed an audio coaching cue using an autonomy-supportive coaching strategy: acknowledging users' feelings and perspectives [36]. During repetitive manual tasks, the system enquires whether the user experiences fatigue. We designed a simple audio question to gauge users' perceived fatigue: "*How tired are you right now?*" After the user responds verbally on a scale of 1–5 (1 = "not tired at all" and 5 = "so tired I'm almost giving up"), the system periodically provides coaching cues through one of the four pre-programmed responses: "*If you're tired, take a quick pause*", "*Don't stress, we're almost halfway through*", "*Take a breath if this is too heavy*", and "*This is hard work, but we're almost done*".


**Figure 4: Scaffolding cues were implemented for the cognitive tasks where users assemble interlocking joints (left). The image shows the active cue (green element), which has been moved away from the main structure, and the inactive piece in light grey. Coaching cues were implemented for physical tasks where users assemble box clusters (right). The image shows the location of the box placement in green, with a task panel to confirm the completion of the placement.**

## 3.3 Interaction Modes

Based on this AR system design, we created two interaction modes to study the effects of user agency with AR assistance. In the *high-agency* condition, users actively drive the system behaviour with direct inputs. In the *low-agency* condition, the system triggers changes automatically without such interactions. The timelines for these two conditions are summarised in Figure 5. One important design goal for the automated, low-agency system is to ensure an equal experience for all participants.

*3.3.1 High-Agency Condition.* The high agency condition supports users' independent and causal involvement in the system operation, i.e., the user can causally attribute the received AR content to their previous actions and decisions. During the cognitive assembly tasks, the *high-agency* mode allows users to view a subsequent assembly step by pressing a button on the interface. Users can carry out tasks at their own pace and move on when they decide to do so.

In the physical assembly tasks, the *high-agency* system responds to the user's fatigue rating, i.e., if the rating goes above 3, one of the pre-programmed acknowledgement cues is played via audio. One challenge is to account for individual differences in physical strength and personal threshold for expressing fatigue. If the user does not rate above three throughout the tasks, the system sends one cue before the last box cluster begins.

*3.3.2 Low-Agency Condition.* The low agency condition removes the causal link between user decisions and the AR content received. The automation also aims to ensure an equal experience for all participants. In the physical assembly tasks, the acknowledgement cues are given after every other cluster, regardless of user response.

Similarly in the cognitive assembly, each step progression is triggered on a fixed schedule. One challenge is to account for a fixed task duration caused by scheduled cues, which is not the case when users learn at their own pace. Therefore we pre-tested the puzzle on three users and included a time limit for the task – four minutes for the easy and eight minutes for the hard puzzle. The automatic instruction intervals were also set based on these results – the intervals are 60-60-60-30-30 seconds (easy) and 150-120-90-60-60 seconds (hard).

In both the high- and low-agency conditions, if the user cannot finish within this time, we provide an option to repeat the practice session with a reduced duration. The time reduction is set at 25 percent (1 or 2 minutes) to provide enough space for four repeated trials if needed. The scheduled intervals are also reduced proportionally.

## 3.4 Implementation

The assembly task geometries are generated from the Rhino 3D modelling environment (Version 7.0) and converted to HoloLens visualisations using a plugin in the visual scripting tool Grasshopper (Version 0.7.0046) [54]. After referencing the step-wise assembly geometries and specifying an anchor for their display, the plugin automatically generates the task content for the HoloLens.

The HoloLens application was programmed in Unity (2021) and implements an API that receives the AR task content generated from the plugin via WebSocket. This includes (1) textual instructions displayed on a UI panel and spoken via the TextToSpeech class with Zira's voice (Microsoft) and (2) 3D holograms with corresponding colliders to allow near-hand interactions. The display anchor specified in the plugin corresponds to the name of a marker in physical space to localise the elements.

A browser-based experiment control interface was used to control the experiment: the task-oriented assembly instructions are triggered based on a count-down timer implemented in Javascript, and the coaching cues are triggered by the researcher based on

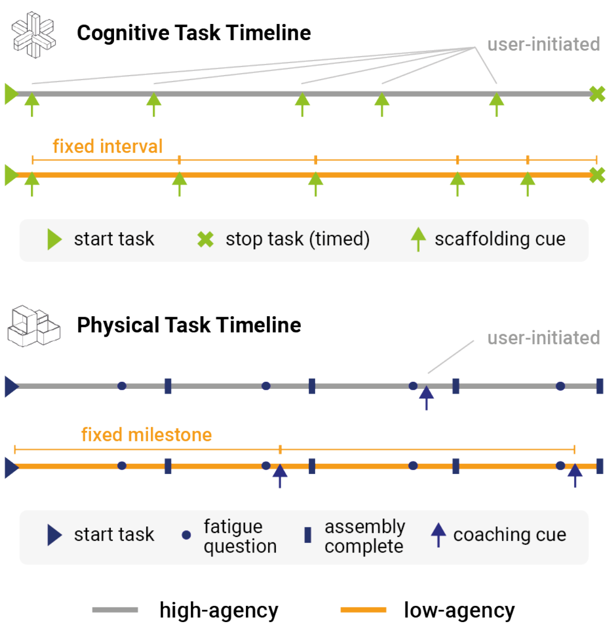



**Figure 5: Timeline of events under the high- and low-agency modes are illustrated above in grey and orange. During cognitive tasks, high-agency mode users manually trigger the cues when they choose to, whereas low-agency mode users receive these cues on a fixed interval. In the physical tasks, there are four assemblies per task and the fatigue questions are asked per assembly. High-agency mode users receive coaching cues in response to their fatigue rating, and low-agency mode users receive them in every other assembly.**

the conditions described above. A server for receiving and logging study data was run on the same PC as the instruction server. Figure 6 depicts the overall system setup.

# 4 Lab Study on the Impact of Reduced User Control

We first conducted a lab study to investigate how the high- and low-agency AR system influences usability, task performance, psychological needs, and workload.

## 4.1 Hypotheses

We outline four hypotheses when comparing the two modes:

**H1 (Psychological Needs)**: We believed the high-agency mode would lead to greater psychological needs satisfaction. Self Determination Theory (SDT) is a frequently used theoretical framework in recent research on agency and autonomy [2]. The theory is based on three Basic Psychological Needs (autonomy, competence, relatedness) [45]. Since higher degrees of automation have been linked with a reduced sense of competence and autonomy [3, 14], we hypothesised that the high-agency mode would result in higher needs satisfaction in these two dimensions.

**H2 (Workload)**: We also believed the high-agency mode would result in higher workloads. This is due partly to the intuition that automation reduces the efforts needed for actively interacting with the system (both cognitively and physically). In addition, prior work on adaptive AR systems has reported reduced cognitive workloads during assembly tasks where workers were guided via a projected in-situ assembly instruction system [26]. As such, we expected a similar reduction in workload when using automated cues from the low-agency system.

**H3 (User Experience)**: We expected the high-agency mode to positively influence user experience. Though user agency has been treated as both a component and an antecedent of good UX in HCI research [2], there is currently no clear model linking the two constructs. Some studies have reported that agency positively correlates with how users perceive the hedonic qualities of the system [4].

**H4 (Performance)**: We anticipated better task performance under the high-agency mode. Prior work has shown that when users have greater control over stimuli, they are more efficient at processing information, which leads to better performance in visual search tasks [31]. In addition, since confusion can arise when the system behaviour deviates from user actions and expectations, intuitively, we believe that users would better focus on tasks in the high-agency condition, leading to better task performance.

## 4.2 Participants

We recruited 24 participants, 11 female and 13 male, aged between 20 and 45 ($\mu = 31.6, \sigma = 7.6$). The recruitment was conducted through word-of-mouth with students and employees at the university. There was no monetary compensation for participation. Participants were asked about their experience with the HoloLens during the appointment scheduling, and only 2 participants reported having extensive experience. The study was approved by the Institutional Review Board of the university, and informed consent was obtained from every participant.

## 4.3 Materials and Apparatus

The study used a HoloLens 2 device as the AR interface. Two puzzles and four wooden boxes were prepared for the assembly tasks. The experiment was conducted in a room at the university, and the setup is shown in Figure 7.

Two armband sensors – Emotibit from OpenBCI and Polar Verity Sense – collected physiological data during the tasks. The primary motivation is to validate existing methods [35] for understanding user task loads with physiological correlates. However, this objective is secondary to the main focus of this paper, so we list the collected measurements for transparency but leave the detailed analysis to future work.

## 4.4 Procedure

We conducted the study using a between-subject design. The participants were divided into two groups. Each group used one *interaction mode*, i.e., the high-agency users carried out all tasks with interactive cues, and low-agency users all received cues on a fixed interval as described in 3.3.

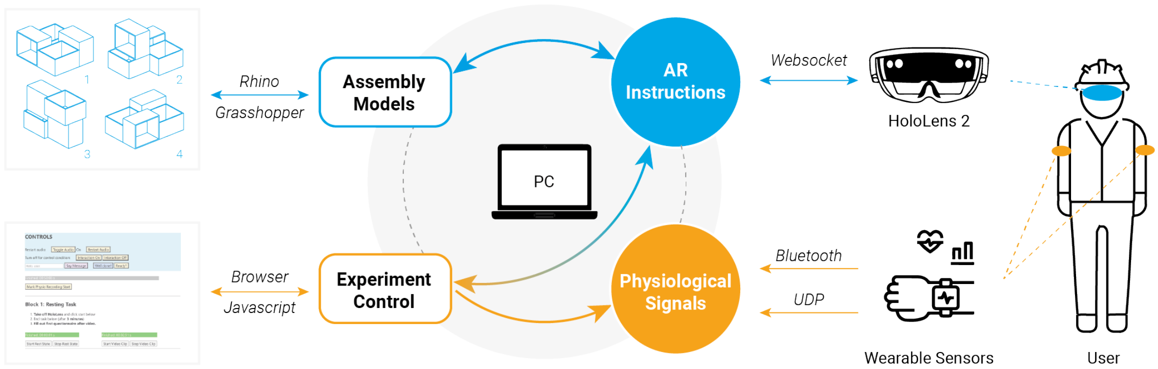



**Figure 6: System Diagram. Task models are created in the Rhino 3D modelling environment, and instructions are dispatched to the AR headset through a websocket connection. Data from the polar armband is streamed to a Python server via Bluetooth. A browser interface is implemented to control the data logging and experiment procedure.**

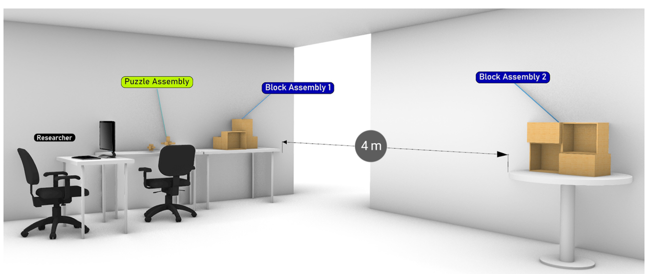



**Figure 7: This figure illustrates the study setup. The block assembly tasks (blue) take place on two tables four metres apart. The puzzle assembly tasks (green), the resting period and the interviews take place on a third table (green). The researcher conducting the study is seated on the side behind a monitor (black).**

The study procedure is summarised in Figure 8. After briefing and signing the consent forms, users first conducted a system tutorial by completing two pre-tasks using the AR interface: (1) assembling three interlocking elements of a different puzzle and (2) placing one weighted box between two tables four times. We asked the Single Ease Question [46] afterwards to verify that all users were able to complete the tasks before proceeding.

Users then put on the two armband sensors and completed two baseline tasks (resting with nature sounds and video watching), each three minutes long. The sequence of the following two blocks - cognitive and physical assembly - was randomly assigned to control for order effects. Within each block, users first completed the easy task and then the difficult one. After each task, users completed the questionnaires and were briefed on the next task. The breaks between tasks ranged from 2-4 minutes, and the length of the two main task blocks ranged from 40 to 80 minutes. Afterwards, users completed the SUS and demographic questions. The session then concluded with a short interview lasting 3-5 minutes.

## 4.5 Measurements

Our measurements include task performance, psychological needs, task loads, and overall user experience. Task duration is the primary performance measure. Secondary measures are included for the two cognitive tasks: (1) number of learning trials needed, (2) whether the user succeeds in assembling the joint independently during evaluation, and (3) number of hints needed, in case the user did not manage independently.

User experience of the overall system is measured through the System Usability Scale (SUS) [5]. Task-related usability is measured through the Single Ease Question (SEQ) [46] on a scale of 1 (very difficult) to 7 (very easy). Psychological needs are measured through a questionnaire on the perception of competence and autonomy, adapted from the Basic Psychological Need Satisfaction at Work Scale (BPNSS) [11]. NASA-TLX is used as the task load measure [20], and the collected physiological correlates include photoplethysmography, electrodermal activity, skin temperature, heart rate, and accelerometer data.

Demographics questions are provided at the end of the study, including age and self-rating on fitness, strength, and past experience with timber assembly tasks on a scale of 1 to 7. The purpose of these questions is to screen for outliers and ensure that the users' prior capability with physically demanding and timber assembly tasks is on par between groups.

## 4.6 Data Analysis

For all questionnaire responses and task performance data, we selected statistical tests based on whether the data met normality assumptions using Shapiro-Wilk tests. If the majority of items within a questionnaire violated this assumption, we applied Mann-Whitney U tests to all items in that questionnaire for consistency. Otherwise, two-sample T-tests were conducted, with adjustments for equality of variance determined using Levene's test. We note the effect sizes using Cliff's delta (small 0.15 – 0.33, medium 0.33 – 0.47,

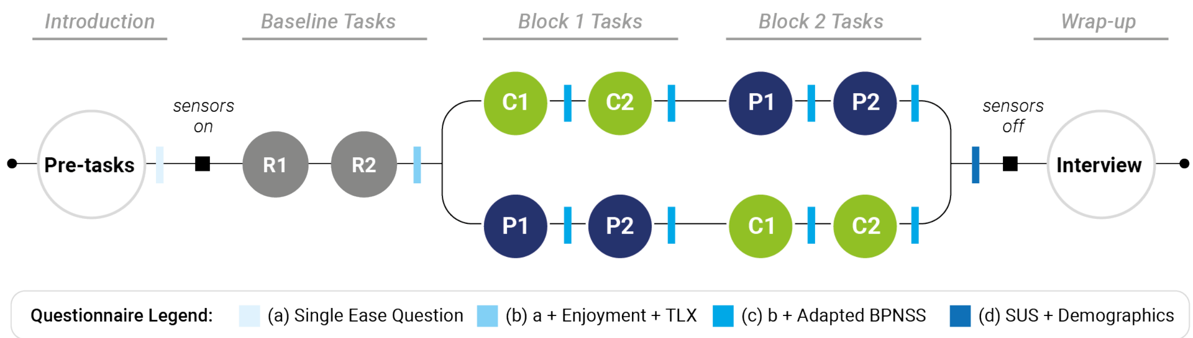



**Figure 8: The study procedure includes five blocks of activities. *Introduction* includes two pre-tasks as a system tutorial. The *baseline tasks* (R1, R2) are two low-workload activities (resting and video watching) to provide a baseline measurement for the physiological signals. The cognitive (C1, C2) and physical (P1, P2) assembly tasks are randomised to control for order effects (noted as *block 1* and *block 2*). A short interview gathers user feedback on the AR system during *wrap-up*. Questionnaires throughout the experiment are noted in blue boxes.**

large $\geq 0.47$ effects) [37]. The two conditions are noted with subscript l and h for low- and high-agency.

## 4.7 Results

First, we evaluated the assembly experience, fitness, and strength self-reports between two groups, which were normally distributed (Shapiro-Wilk: $p > 0.1$) with no difference between high- and low-agency groups (Assembly Experience: $\mu_h = 3.0, \mu_l = 3.5, t(22) = 0.555, p = 0.584$, Fitness: $\mu_h = 4.1, \mu_l = 4.5, t(22) = 0.800, p = 0.432$, Strength: $\mu_h = 4.3, \mu_l = 4.2, t(22) = 0.151, p = 0.880$). 2 of the 24 participants had extensive experience using HoloLens before the study, with one in each group.

*4.7.1 Psychological Needs.* We collated the five questions on autonomy into a single score [11] and found that high-agency system users' ratings were higher on average for all tasks, but the difference was not significant. In the individual autonomy questions however, the high-agency group rated two questions significantly higher: (1) "*My needs are taken into consideration*" in the difficult cognitive assembly ($^*p = 0.034, \delta = 0.44$) and (2) "*I feel like I am free to decide for myself how the task was done*" in the weighted physical assembly ($^*p = 0.034, \delta = 0.44$). Results on the second question showed a similar trend in the simpler cognitive task ($p = 0.09, \delta = 0.31$) but was insignificant.

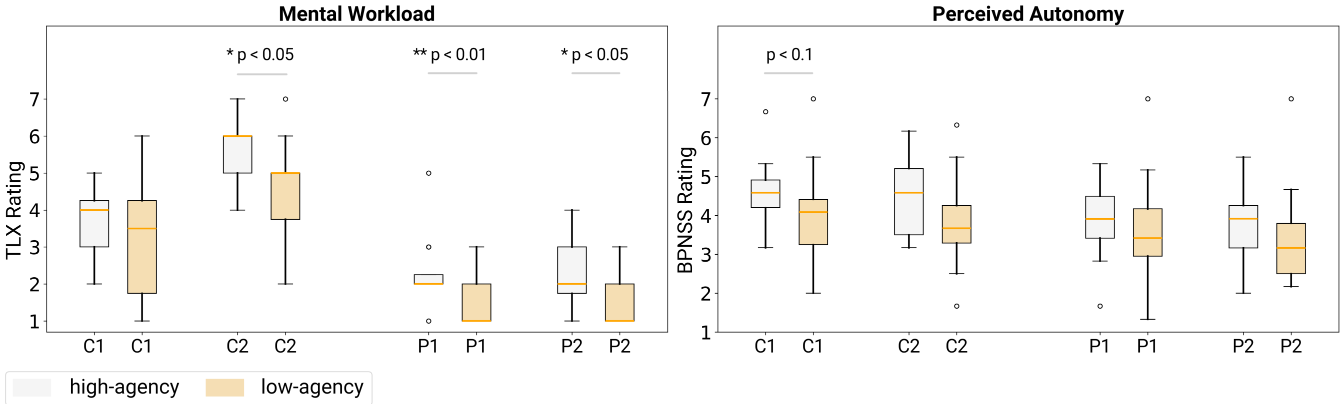
*4.7.2 NASA-TLX.* The distribution of mental workload across different tasks is shown in Figure 9a. In both physical tasks, *high-agency* users rated higher mental workloads (easy: $^{**}p = 0.004, \delta = 0.60$, hard: $^*p = 0.032, \delta = 0.42$). In the difficult joint assembly, *high-agency* users rated higher mental workload ($^*p = 0.021, \delta = 0.47$) as well as increased efforts ($^*p = 0.029, \delta = 0.45$). The simpler joint assembly showed higher but insignificant differences in mental workload ($p = 0.23, \delta = 0.17$) and efforts ($p = 0.10, \delta = 0.30$). We found no difference in physical workloads in any tasks.

*4.7.3 Task Performance.* The two groups achieved similar results on all task performance measures (Figure 9b). In the easy cognitive task, all users took only one learning trial, and 23 out of 24 users completed the evaluation without hints. In the difficult cognitive task, both groups used an average of 1.75 learning trials, with 7 and 5 users completing without hints in the high- and low-agency group respectively. Though the two groups were given the same practice duration, the first run-through of all instruction was completed faster by the high-agency group, since many moved on immediately after completing one step. This duration is 2 minutes 44 seconds (02:44) compared to 04:00 for the easy, and 06:39 compared to 08:00 for the hard task.

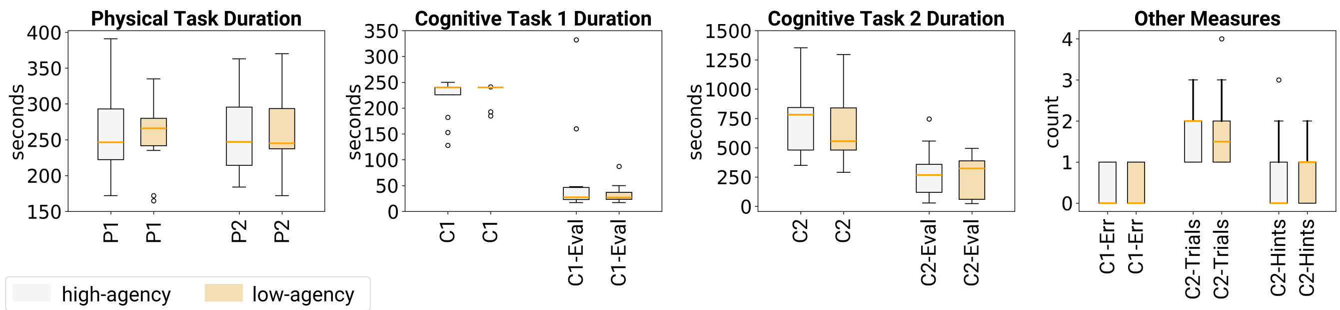
*4.7.4 Observation.* We took notes of participants' behaviours during the study. In the cognitive tasks, we categorised how the *low-agency* mode users dealt with the time between system actions into four types – (1) *idle*: user often started by just waiting for the system to respond (2) *review*: user examined the AR visualisations and/or the physical element from different angles, possibly to memorise the configuration; (3) *practice*: user deconstructed and reconstructed existing steps; (4) *explore*: user explored their own solutions without system support until the next cue provided the definitive solution. We also noted how users reacted to the coaching messages during the physical tasks. While many did not react directly, some users smiled when hearing the message, and some verbally responded to the system, e.g., "*ah thanks, it's not so bad*".

*4.7.5 User Experience.* We split the SUS results into two dimensions [29] and found both results higher on average in the high-agency group, with a larger difference in usability ($\mu_h = 81.6, \mu_l = 76.0$) than learnability ($\mu_h = 82.5, \mu_l = 80.5$). Neither difference was significant. However, the high-agency group rated the question "*I found the system very cumbersome to use*" better ($^*p = 0.04, \delta = 0.40$). There were no significant differences in the SEQ, but the high-agency group rated all tasks more difficult on average. The average scores of the four tasks are: easy cognitive $\mu = 3.83, \sigma = 0.89$; hard cognitive $\mu = 1.79, \sigma = 0.81$; easy physical $\mu = 4.41, \sigma = 0.57$; hard physical $\mu = 3.29, \sigma = 1.01$ (1=very difficult).

During the short interviews, we asked the low-agency mode users whether they would have liked to control the pace of the instructions in the cognitive task, 10 out of 12 said yes. One user




**(a) Mental workload (TLX) and perceived autonomy (collated score) between groups.**




**(b) Task performance between groups. Errors in the easy cognitive task, trials and hints needed for the hard cognitive task are included in *other measures*.**

**Figure 9: Results on mental workload, perceived autonomy, and task performance between groups. The easy and hard cognitive tasks are denoted C1, C2 and physical tasks denoted P1, P2 in the figure. (Boxes: inter-quartile range, Whiskers: maximum and minimum, Lines: median, Outliers: 1.5 times above or below Q3 and Q1)**

said they did not mind the system controlling the pace – "*It (the time spent waiting for the system) actually gave me some time to reflect*". Another user commented – "*Since I am a novice at this task, I actually don't mind following the system. But if I am not (a novice user), I would prefer to control it myself*". For the coaching cues during physical tasks, participants in the low-agency group did not mind hearing the messages even though they were not tired at that point – "*sure, it was nice to hear I can do what I want*".

## 4.8 Reflections on Results

The results from our lab study revealed several trade-offs between high- and low-agency AR systems. First, in three of the four tasks, high-agency system users reported higher mental workloads. We therefore accept **H2** in terms of mental workload as observed in all but the simple cognitive task. Next, we found negligible task performance differences between the two groups, so we reject **H4**. We were also surprised to see that low-agency system users rated lower effort in the hard cognitive task (a trend in the easy task), especially given many instances of the system progressing too slowly or quickly. Intuitively, these frictions would result in more effort, not less.

For usability and perception of autonomy, the collated scores did not show significant differences but the high-agency users gave higher ratings for one usability question overall and two autonomy questions for the two difficult tasks. We therefore cannot accept **H1** and **H3** as a whole, but note that the low-agency system is perceived to be (1) more "cumbersome" to the user, and (2) less considerate of user needs and limiting freedom in deciding how the task was done in the difficult tasks. To better understand these observations, we dig deeper into qualitative insights in the following study.

Lastly, we verified that our task design was able to induce different mental and physical workloads from the TLX responses ($^{***}p < 0.001$). However, the weighted physical tasks were considered not tiring enough to lead to high fatigue ratings for some. To ensure that carpenters, who are likely more accustomed to manual tasks, do not find the physical assembly too easy, we increased the physical assembly difficulty in the following study.

## 5 Domain Expert Study

Building upon our findings above, we conducted an expert study with users trained in timber construction tasks. We focused on generating qualitative insights through semi-structured interviews to understand their perception and preference for low- v.s. high-agency control of AR systems. We also included questions to evaluate the task and AR system design. This aims to ensure that the study conditions and subsequent findings can reflect situations workers encounter in their daily work.

## 5.1 Participants

We recruited 9 employees from a timber construction company in Southern Germany. We requested participation from female users and had one female participant. However, she did primarily office work and had no construction experience, so we excluded the data. The remaining 8 workers were male and aged between 18 - 65 ($\mu = 37.3, \sigma = 14.5$). Their age, role, and work experience are summarised in Table 1 (exact age reported in ranges for privacy considerations).

## 5.2 Materials and Apparatus

We used the same HoloLens 2 headsets and wooden puzzles for the study. A conference room at the company was chosen as the study site. The room had a much larger width and length compared to that of the previous study, which allowed modifications to the assembly table layout described further below.

## 5.3 Procedure

Each session with the carpenter consisted of the same study procedure as described in 4.4, followed by a more extended semi-structured interview. To ensure the physical assembly tasks were sufficiently challenging, we extended the distance between two tables from 4 to 7 m and added obstacles (chairs), requiring users to travel in a slalom with two abrupt turns. We also included a "bar code station" which required the user to hold the boxes at chest height for three seconds until a "pass" signal is given before placing the boxes.

## 5.4 Measurements

We translated the study material into German and designed a semi-structured interview with each worker after they completed the tasks. The interviews were structured around four topics:

(1) **Evaluation of Task Design**: comparison between the cognitive/physical assembly tasks we designed and the carpenters' daily work tasks.
(2) **Evaluation of AR Cues**: feedback on the AR cues during the two different task types.
(3) **Control and Autonomy**: opinions and preferences for low-agency (automated) v.s. high-agency (interactive) system use.
(4) **Overall Impression and Further Applications**: carpenters' reception of AR headsets and potential of AR supporting other tasks in their daily work.

## 5.5 Results

Given our focus on qualitative findings and the limited sample size, we did not conduct an in-depth statistical analysis as in Study 1. The overall SUS scores were on par with the participants in study 1, with the learnability dimension slightly higher ($\mu_h = 83.3, \mu_l = 87.5$) and usability slightly lower ($\mu_h = 80.3, \mu_l = 75.8$) in the automatic group. Below, we present the interview results grouped by the four main topics.

*5.5.1 Task Design Evaluation.* Most workers (5/8) remarked that the cognitive assembly was more complex than their usual assembly tasks. U1 commented that complex joinery is used in renovation but is rare in modern construction, which is the majority of projects at the company. U3 worked predominantly with timber cutting processes and thought the assembly task was similar in complexity in that there are "*many unique pieces and you always need to think ahead*". U2 and U4 both commented that most tasks they handle are less complex. However, they note that for certain projects with many distinct elements or needs for protecting surface finish, a lot of forethought and planning are required.

All workers rated the study tasks as less demanding than their daily work. One aspect of physical demand comes from various body postures and motions (U4), such as kneeling and bending forward. The other comes from needing to walk up and down the stairs in on-site construction (U7). The most frequently mentioned task characteristic which was not captured in our design is the **dimensions of the elements** (U2, U3, U7, U8). U3 commented that *"The panels are mostly between 2.5 and 6 metres long, and I need to move them with a vacuum crane."* Though machinery can hold weight during travel, alignments and adjustments still require manual power. U7 commented that "*the crane is very slow ... if I can I just rather carry things by hand*" (U7).

When asked about the repetitive aspects of physical tasks, three users (U2, U6, U7) rated the study tasks as similar to their daily work, while most rated them as more repetitive. U7 stated that the repetitive nature of the functions was the same as what he does day-to-day. For U2 – "*There are phases you have a lot of repetitions like this, but throughout one day, there is more variety.*" U6 gave a similar example where "*You repeat one thing for 20 mins then the next for 30 mins, and in 2-3 hours you start again.*" A summary of these comparisons is shown in Table 2, illustrating the number of participants who rated the study tasks to be more / less / similar in demand compared to their daily work tasks in the cognitive and physical dimensions.

*5.5.2 Evaluation of AR Cues.* The second group of questions focused on the usability and perception of the cues provided through the HoloLens during the two different assembly task types.

**Scaffolding Cues**: All users found the AR guidance helpful for assembling elements. We grouped the reasons they cited into three categories. The first category cites enhanced **understanding of spatial relationships** with holographic elements: "*You can see everything better in 3D*" (U7), "*it helps that you can move the pieces and learn how each one is*" (U8). The second reason cites the ease with which one can **identify unique elements**: "*when there are many pieces you can see exactly which ones you need*" (U3).

The third category mentions being able to access information in AR to reduce the need for **communication and checking**: "*it's useful when I need to make something for the first time, and there is no need to ask a colleague"* (U6). Similarly, this applies to communication between the factory and office: *"sometimes a part is so complex that the (2D) plan doesn't show it any more ... this is*

**Table 1: Summary of the eight carpenters (age, role, work experience, and AR interaction mode used) in the study.**

| Code | Group | Role | Experience | Age |
|---|---|---|---|---|
| U1 | High-Agency | Carpenter | Worked for 48 years in timber construction at the same company. | 55-65 |
| U2 | High-Agency | Carpenter | Worked for 26 years in timber construction at the same company. | 35-45 |
| U3 | Low-Agency | Carpenter | Primarily worked with timber cutting processes at the current company since 10 years. 10 years experience in saw mills prior. | 35-45 |
| U4 | Low-Agency | Carpenter | Worked in timber construction for 7 years and window assembly for 7 years prior. | 35-45 |
| U5 | High-Agency | Carpenter | Worked in timber construction for 5 years at the same company. | 25-35 |
| U6 | High-Agency | Carpenter's Assistant | Worked in timber construction for 1 year and concrete construction for 4 years prior. | 35-45 |
| U7 | Low-Agency | Apprentice | Apprentice since 1 year, mainly assisting more senior carpenters. | 18-25 |
| U8 | Low-Agency | Apprentice | Started apprenticeship training at the company since 2 weeks. | 25-35 |

**Table 2: Summary of comparison between the study tasks and carpenters' daily work tasks in the cognitive and physical dimensions**

| Comparison | Physical | | Cognitive |
|---|---|---|---|
| | Repetition | Difficulty | Difficulty |
| Study ≈ Daily Tasks | 3 | 0 | 3 |
| Study > Daily Tasks | 5 | 0 | 5 |
| Study < Daily Tasks | 0 | 8 | 0 |

*very useful when it can save me a trip to the office to print out details of the part".*

**Coaching Cues**: 7 out of 8 users perceived these cues favourably, but only 3 users found such cues *useful* for their work. U1 mentioned *"it is definitely useful when a system can advise you to slow down or speed up when you should"*, which he believed would be helpful for new workers who are not familiar with working on site. Two users cited working in hot weather as a context for having such support: *"when you are working in hot weather you can get a reminder when you should take a break or drink water"* (U8) or *"sometimes you are working on the roof for a long time and should be reminded to go inside and do something else"* (U2).

Only one user perceived such cues negatively – *"I find it annoying - I work independently and will decide if I need a break for myself."* (U7) Most users who found the cues harmless but were not convinced of their necessity cited similar reasons around the independence of work: *"sometimes if I'm excited I just want to work longer; if it is not exciting, I will take a break"* (U5). Many users also made comments along the lines of *"nobody ever told me that at work ... I would prefer information over motivation"* (U3). U4 put it as – "*Usually no complaints is already good enough, but if you receive flowers that is of course nice to have*".

*5.5.3 Control and Autonomy.* For all users who experienced the automatic condition, they stated they would prefer the option to control the system. Only one user, who was relatively new to timber construction (U8), followed this preference with an alternative proposal – "*I can either control it myself, or the system can just run a bit faster*".

We then asked the workers whether they find low-agency, automatic systems acceptable for use in their work and whether they may be preferable to high-agency, interactive ones. To anchor the discussion, we described automation as a system "automated to achieve optimal results for the tasks" but did not give explicit rationale for how the system may do so.

All users stated that they are *open* to use such automatic systems but also stated a preference for having full control. When prompted to provide reasons for such decisions, some contested how much intelligence the systems can have – "*sometimes you are excited about your work and sometimes you don't enjoy it very much, how would a system know the difference*" (U5). Others mentioned the need to limit the degree of control – "*it's OK if the system decides, but it shouldn't be too much. Trusting the system might be good but controlling it is better*" (U4). A few users echoed comments on the independent nature of their work in response to the coaching cues – "*even if the system is more intelligent, I still prefer to decide for myself*" (U2).

*5.5.4 Applications of AR-HMDs in Timber Construction.* Workers gave mixed responses on the feasibility of wearing the head-mounted device during work tasks – "*it is helpful for some things but I do not want to wear it all day*" (U8). On the other hand, one worker who had experience using HoloLens before commented that "*the first time I used it it was really tiring, but now it is getting faster and easier*" (U4). Two users noted issues specifically during physical tasks – "*I find the headset a bit distracting and this can be dangerous*" (U7). The opposite opinions cite that "*it is helpful when the system can show you where to place the materials...and help organise the workspace while you work*" (U1).

In addition to the two cues implemented for the study, we asked the carpenters what other types of applications AR could be useful for in their daily tasks. The most commonly cited area was logistics: "*It would be useful for finding how to load a truck in the most efficient way and seeing exactly where to put the packages*"(U5), and "*when I am driving a forklift to pick up materials, it would be useful to see which one to pick*" (U2).

## 5.6 Reflections on Results

Findings from this domain expert study provided insight into (1) the overall applicability and validity of the study design for professionals in timber construction and (2) a more nuanced and ecologically valid perspective on how these users perceive control and autonomy in AR support.

User feedback on the task design shed light on important cognitive and physical dimensions of the carpenters' work. Additionally, key aspects of AR assistance that were found to be useful and appealing during these tasks were identified. In this context, we discussed the workers' perception of intelligent automation that may diminish direct control; we found that although all participants were open to such systems, subtle differences exist where individual attitudes and broader work cultures play a role.

In discussion, we connect these results with the quantitative findings from the lab study to provide a richer understanding of the impact of user control in AR assistance and how such impact is contextualised for users in timber construction.

## 6 Discussion and Key Takeaways

This research investigated the impact of user control and agency for AR-assisted assembly tasks. Our results addressed two evaluation dimensions: (1) empirical differences between high- and low-agency control modes and (2) more ecologically valid understanding considering the needs of workers in construction.

Below, we first discuss our findings from the two studies and summarise how users perceived and behaved under reduced levels of control. Connecting this with specific conditions in timber construction, we then highlight two challenges to inform future designs.

### 6.1 Reduced User Control in AR-assisted Assembly

Based on users' NASA-TLX ratings, we found that the low-agency mode reduced cognitive workloads in both physical assembly tasks and the difficult cognitive task (see 4.7.2). These effects arose even though the low-agency system disregarded individual needs, leading to friction such as misaligned progress in the cognitive tasks, misinformed fatigue responses in the physical tasks, and a lower average SUS score. These results indicate that **automatic systems reduced users' mental workloads in a way that is robust to friction in user experience**.

*6.1.1 User Behaviours under Reduced Control.* We found negligible performance differences between the two interaction modes in the cognitive tasks, contradicting our hypothesis (H4). Based on the task logs and observations, we believe two aspects may have contributed to this effect.

First, **human adaptation strategies bridged the performance gap**. The four patterns of behaviours identified in section 4.7.4 showed diverse approaches as users in the low-agency mode devised strategies to deal with the lack of control. The time between each instruction step required users to spend longer with the assembly in the intermediate stages. The difference in task times reported in section 4.7.3 showed that high-agency system users went through the assembly steps well before the allotted deadline during practice. Because the system allowed users to progress as soon as the previous step was completed, most made use of this freedom and only realised the complexity of the assembly once the structure was complete.

Though this does not necessarily support conclusions on performance differences between groups, it reveals an interesting side-effect of the reduction of control. While interactive, high-agency systems are always subject to the desires of the user, automation allows the injection of influences which may elude the user's short-term desires and benefit long-term task goals. Therein lies a **design opportunity to consider long-term goals in adaptive systems**, e.g., explicitly encouraging *review*, *practice*, and *explore* actions, instead of focusing on fast-feedback adaptations oriented around efficiency, e.g., eliminating *idle* time.

*6.1.2 User Perception of Reduced Control.* The carpenters were unanimously open to working with low-agency systems but expressed a preference for having control. U1 commented that "*This is a generational thing – the younger generation is controlled by electronics already and would do better with this, but I am open to it.*" The main reasons in favour of control were twofold: (1) **lack of belief in that a system can truly comprehend their needs** (U5), which can be complex, individual, and personal, and (2) construction work is **a highly goal-driven and independent profession**, which means that users are accustomed to dictating their own pace work (U4, U5, U7).

The low-agency condition showed lower average SUS scores and a significant increase in the system being considered "cumbersome" (see 4.7.5). Additionally, the two autonomy questions were significantly lower in the low-agency group for more difficult cognitive and physical tasks respectively (see 4.7.1). During the interviews after the study, most low-agency system users from both studies preferred to have more control (see 4.7.5 and 5.5.3). Therefore, we report that **users stated an overall preference for higher user control**.

The exceptions were found with two novice users who remarked that they did not mind giving the system control, given that it provided a guided pace, which was helpful when they were new to the task and allowed them more headspace to reflect on their actions. The only carpenter user who proposed an alternative to having control – "*make the system run faster*" – had only started working in timber construction. In existing work, there is similar evidence that novice users benefited more from AR guidance [15], and we observe a similar trend that **novice users may be more tolerant to reduced control, trusting that automation "knows better".**

### 6.2 Design Considerations for AR-assisted Construction

Lastly, we integrate the findings above into design considerations for future work. We first outline four productive scenarios where designers can consider low-agency systems in AR-assisted construction. We then highlight two challenges designers should consider when dealing with user agency in construction.

*6.2.1 Productive Scenarios for Low-agency Systems.* In previous work, Lukoff et al. found that higher user agency was not *always* desirable, i.e., users preferred less control when using technology for diversionary needs and more for informational needs [34]. Similarly, Guo et al. found that in human-LLM collaboration, lower agency at the execution level did not result in dissatisfaction, but participants showed greater caution in planning-level activities [19]. Contributing to this line of work, we summarise four scenarios where low-agency interfaces may present benefits in construction assembly:

**The application targets novice users**. During interviews, we enquired whether users would prefer more control over the instructions – though most said yes, some cited that being a novice made the lack of control more acceptable. This is in line with prior studies on adaptive AR assistance, which noted positive effects with novice users and more adverse ones with experienced users [15]. This type of adaptation is widely explored in educational scenarios, e.g., adaptively increasing difficulty based on cognitive workloads to support learning [55]. Our study further supports the use of adaptive AR, especially in the context of training in the AEC industry [51].

**The situation demands to manage mental workload**. We observed lower mental workloads and effort under the low-agency mode during the difficult cognitive assembly (see 4.7.2). In certain situations, the importance of managing cognitive workload may trump usability drawbacks, e.g., safety-critical applications such as operating heavy and dangerous machinery. These situations already motivate current work on adaptive AR displays [53], where the system needs to provide timely and relevant information for the operator while preventing information overload and distractions.

**The task requires consideration of objectives not immediately apparent to users**. During the cognitive puzzle assembly, we found that users of the low-agency system made use of the waiting time to review, practice and explore, while high-agency system users often moved on quickly (see 4.7.3). Systems that temporarily reduce user agency to curtail myopic behaviours can promote long-term benefits, e.g., digital well-being features that limit screen time and promote healthier device usage [39]. Provided these intentions are transparent and have obtained consent, this provides a productive opportunity for system-directed decisions.

**The interface delivers automated health and safety information**. During the interviews with carpenters, the users who found the coaching cues useful mostly cited physical health and safety reasons (U2, U8) (see 5.5.2). Additionally, a few workers did not believe the benefits from AR warranted the discomfort of wearing a headset during physical tasks (see 5.5.4). These comments point to the relevance of health and safety support using more lightweight, smaller-footprint devices. These technologies exist in consumer products, e.g., smart watches reminding users to exercise or sending an alert when abnormal bio-signals are detected [25, 27]. Not requiring explicit interactions here allows users to focus on their primary tasks. This is further supported by the fact that users reported lower cognitive workloads when interacting with the automatic coaching cues in both physical tasks (see 4.7.2).

*6.2.2 Challenges in Accounting for User Agency in Construction.* The results of our study also highlighted two challenges that should be considered when studying user agency in construction task settings:

**Work culture might govern potential acceptance.** Though all workers held open attitudes towards alternative degrees of autonomy, they also stated a preference for full control. Notably, many workers cited the **independence** of the work culture as a reason for favouring control (see 5.5.3), e.g., "*even if the system is more intelligent, I still prefer to decide for myself*" (U2). The independent work culture in construction has been noted in prior work, e.g., Lwstedt et al. highlighted that the identity of the construction site manager was "*ideologically crafted around a proclivity for free and independent work*" [33]. This leads to the question – even with hypothetically "*perfect*" adaptive systems, which accurately understand user needs and can ensure optimal outcomes, are such systems suitable for these users?

Another aspect prevalent in construction work culture is the notion of **masculinity**, which can impact worker perceptions on health and safety topics [23]. This is echoed by workers' feedback on the coaching cues, which indicated that the messages clashed with their culture and expectations, e.g., "*nobody ever told me that at work*" (U3). This aspect is a critical consideration for designing automated systems that support health, safety, and well-being.

**Effectiveness may be task dependent.** We designed four tasks that varied in cognitive and physical difficulties to reflect scenarios encountered by construction workers. This introduced complexities in the study design but also shed light on the role of user control in contexts most relevant to construction.

The high- and low-agency control modes had different impacts on our measurements across these tasks. For instance, the difference in autonomy was only noted in the more difficult cognitive and physical tasks (see 4.7.1), and the reduction in cognitive workload was not significant for the simple cognitive task (see 4.7.2). Though we cannot conclude the relationship between task type and the effects observed, this highlights the challenge of accounting for diverse tasks when evaluating agency in construction applications.

Furthermore, we highlight two aspects of carpenters' tasks that we did not account for in this study (see 5.5.1). The first is large-scale element manipulation (U2, U3, U7, U8), which is the primary source of physical demand. The second is communication support scenarios (U6), which relate to existing work on remote collaboration [52]. Both aspects may involve collaboration between multiple users, which presents an interesting but challenging context for evaluating individual user agency.

## 6.3 Limitations and Future Work

With the increasing adoption of AR technologies in the architecture, engineering and construction (AEC) industry [41], we conducted a study to examine the impact of user control in AR assistance and contextualised the results with inputs from construction workers.

Our sample size (N=24) is relatively small, though it aligns with prior studies on user agency [10]. The evaluation with trained carpenters (N=8) anchors our findings with users familiar with timber construction tasks. This, however, included workers from only one construction company, while the construction industry encompasses a wide range of work environments. Future work could validate these findings with a larger and more diverse group of participants.

Additionally, task characteristics may differ across different trades in construction, e.g., masonry construction involves different amounts of repetition and size of parts. This further highlights the challenge we noted above, where different task characteristics can influence the role of user control.

Lastly, our sessions lasted around one hour per user. This meant that our results focused only on immediate task performance and psychological needs perception. The effects of long-term use may reveal different trade-offs between high- and low-agency systems, e.g., considering skill retention, learning, and motivation. Bennett et al. highlighted similar limitations in current research on autonomy and agency in HCI, where studies conducted over single sessions cannot address the relationship of agency and users' long-term well-being [2]. This is an important challenge that needs to be addressed in future work.

# 7 Conclusions

This paper examined the differences between high- and low-agency interaction modes with AR support in construction assembly. We present insights on user perception and behaviour in situations where direct agency and control are reduced. We tailored the study to the needs of timber construction and interviewed carpenters to extract qualitative insights on the task validity, AR system evaluation, and their perspectives of user control when using AR systems. Our results point out opportunities for automation to mitigate drawbacks in interactive systems, e.g., higher mental workloads and focus on short-term results. These insights are summarised into design considerations to support future explorations of human-centred, adaptive AR systems for construction applications.

# Acknowledgments

This research was supported by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) under Germany's Excellence Strategy – EXC 2120/1 – 390831618 and BBSR as Zukunft Bau Project "MRK II Projektnr. 10.08.18.7-22.22". We would like to thank müllerblaustein HolzBauWerke for their generous support of this research.

# References

[1] Behzad Behzadnia, Paul J.C. Adachi, Edward L. Deci, and Hasan Mohammadzadeh. 2018. Associations between students' perceptions of physical education teachers' interpersonal styles and students' wellness, knowledge, performance, and intentions to persist at physical activity: A self-determination theory approach. *Psychology of Sport and Exercise* 39 (Nov. 2018), 10–19. doi:10.1016/j.psychsport.2018.07.003

[2] Dan Bennett, Oussama Metatla, Anne Roudaut, and Elisa D. Mekler. 2023. How does HCI Understand Human Agency and Autonomy?. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 375, 18 pages. doi:10.1145/3544548.3580651

[3] Bruno Berberian, Jean-Christophe Sarrazin, Patrick Le Blaye, and Patrick Haggard. 2012. Automation Technology and Sense of Control: A Window on Human Agency. *PLOS ONE* 7, 3 (March 2012), 1–6. doi:10.1371/journal.pone.0034075

[4] Joanna Bergström, Jarrod Knibbe, Henning Pohl, and Kasper Hornbæk. 2022. Sense of Agency and User Experience: Is There a Link? *ACM Trans. Comput.-Hum. Interact.* 29, 4, Article 28 (March 2022), 22 pages. doi:10.1145/3490493

[5] John Brooke et al. 1996. SUS: A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[6] Francesco Chiossi, Johannes Zagermann, Jakob Karolus, Nils Rodrigues, Priscilla Balestrucci, Daniel Weiskopf, Benedikt Ehinger, Tiare Feuchtner, Harald Reiterer, Lewis L. Chuang, Marc Ernst, Andreas Bulling, Sven Mayer, and Albrecht Schmidt. 2022. Adapting visualizations and interfaces to the user. *it - Information Technology* 64, 4-5 (2022), 133–143. doi:10.1515/itit-2022-0035

[7] Gabriele Cimolino, Renee (Xinyu) Chen, Carl Gutwin, and T.C. Nicholas Graham. 2023. Automation Confusion: A Grounded Theory of Non-Gamers' Confusion in Partially Automated Action Games. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 569, 19 pages. doi:10.1145/3544548.3581116

[8] J. Douglas Coatsworth and David E. Conroy. 2009. The effects of autonomy-supportive coaching, need satisfaction, and self-perceptions on initiative and identity in youth swimmers. *Developmental Psychology* 45, 2 (March 2009), 320–328. doi:10.1037/a0014027

[9] Steward T. Coffin. 1990. *The Six-Piece Burr.* Oxford University Press, Oxford, UK, Chapter 5, 204. https://www.johnrausch.com/PuzzlingWorld/chap05.htm

[10] David Coyle, James Moore, Per Ola Kristensson, Paul Fletcher, and Alan Blackwell. 2012. I did that! Measuring users' experience of agency in their own actions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2025–2034. doi:10.1145/2207676.2208350

[11] Edward L. Deci, Richard M. Ryan, Marylène Gagné, Dean R. Leone, Julian Usunov, and Boyanka P. Kornazheva. 2001. Basic Psychological Need Satisfaction at Work Scale. doi:10.1037/t71065-000

[12] Hitesh Dhiman, Yutaro Nemoto, Michael Fellmann, and Carsten Röcker. 2023. Beyond productivity and efficiency: design tools, methods and frameworks for psychological well-being, and (un)exploited potentials to assist employees at work. *Frontiers in Computer Science* 5 (Aug. 2023), 19 pages. doi:10.3389/fcomp.2023.1129087

[13] Hitesh Dhiman, Gustavo Alberto Rovelo Ruiz, Raf Ramakers, Danny Leen, and Carsten Röcker. 2024. Designing Instructions using Self-Determination Theory to Improve Motivation and Engagement for Learning Craft. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 673, 16 pages. doi:10.1145/3613904.3642136

[14] Sarah Diefenbach, Daniel Ullrich, Tim Lindermayer, and Kaja-Lena Isaksen. 2024. Responsible Automation: Exploring Potentials and Losses through Automation in Human–Computer Interaction from a Psychological Perspective. *Information* 15, 8 (2024), 31 pages. doi:10.3390/info15080460

[15] Markus Funk, Andreas Bächler, Liane Bächler, Thomas Kosch, Thomas Heidenreich, and Albrecht Schmidt. 2017. Working with Augmented Reality?: A Long-Term Analysis of In-Situ Instructions at the Assembly Workplace. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '17)*. ACM, New York, NY, USA, 222–229. doi:10.1145/3056540.3056548

[16] Junhao Geng, Xinyu Song, Yuntao Pan, Jianjun Tang, Yu Liu, Dongping Zhao, and Yongsheng Ma. 2020. A systematic design method of adaptive augmented reality work instruction for complex industrial operations. *Computers in Industry* 119 (2020), 103229. doi:10.1016/j.compind.2020.103229

[17] C. Gerdenitsch, A. Sackl, and P. Hold. 2022. Augmented reality assisted assembly: An action regulation theory perspective on performance and user experience. *International Journal of Industrial Ergonomics* 92 (Nov. 2022), 103384. doi:10.1016/j.ergon.2022.103384

[18] Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. 2017. Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 23, 6 (2017), 1706–1724. doi:10.1109/TVCG.2016.2543720

[19] Jiajing Guo, Vikram Mohanty, Jorge H Piazentin Ono, Hongtao Hao, Liang Gou, and Liu Ren. 2024. Investigating Interaction Modes and User Agency in Human-LLM Collaboration for Domain-Specific Data Analysis. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 203, 9 pages. doi:10.1145/3613905.3651042

[20] S. G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in psychology* 52 (1988), 139–183. https://api.semanticscholar.org/CorpusID:15252590

[21] Derek Holton and David Clarke. 2006. Scaffolding and metacognition. *International Journal of Mathematical Education in Science and Technology* 37, 2 (March 2006), 127–143. doi:10.1080/00207390500285818

[22] Gaoping Huang, Xun Qian, Tianyi Wang, Fagun Patel, Maitreya Sreeram, Yuanzhi Cao, Karthik Ramani, and Alexander J. Quinn. 2021. AdapTutAR: An Adaptive Tutoring System for Machine Tasks in Augmented Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 417, 15 pages. doi:10.1145/3411764.3445283

[23] David Iacuone. 2005. "Real Men are Tough Guys": Hegemonic Masculinity and Safety in the Construction Industry. *The Journal of Men's Studies* 13, 2 (March 2005), 247–266. doi:10.3149/jms.1302.247

[24] Sunwook Kim, Maury A. Nussbaum, and Joseph L. Gabbard. 2019. Influences of augmented reality head-worn display type and user interface design on performance and usability in simulated warehouse order picking. *Applied Ergonomics* 74 (Jan. 2019), 186–193. doi:10.1016/j.apergo.2018.08.026

[25] Susanne Kirchner, Jessica Schroeder, James Fogarty, and Sean A. Munson. 2021. "They don't always think about that": Translational Needs in the Design of Personal Health Informatics Applications. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 210, 16 pages. doi:10.1145/3411764.3445587

[26] Thomas Kosch, Markus Funk, Albrecht Schmidt, and Lewis L. Chuang. 2018. Identifying Cognitive Assistance with Mobile Electroencephalography: A Case Study with In-Situ Projections for Manual Assembly. *Proceedings of the ACM on Human-Computer Interaction* 2, EICS (June 2018), 1–20. doi:10.1145/3229093

[27] Christian Krauter, Katrin Angerbauer, Aimée Sousa Calepso, Alexander Achberger, Sven Mayer, and Michael Sedlmair. 2024. Sitting Posture Recognition and Feedback: A Literature Review. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. ACM, New York, NY, USA, Article 943, 20 pages. doi:10.1145/3613904.3642657

[28] Evie Lee. 2021. Manual Handling Weight Limits. https://cpdonline.co.uk/knowledge-base/health-and-safety/manual-handling-weight-limits/ Available at https://cpdonline.co.uk/knowledge-base/health-and-safety/manual-handling-weight-limits/, Last accessed: May 2, 2024.

[29] James R. Lewis and Jeff Sauro. 2009. The Factor Structure of the System Usability Scale. In *Human Centered Design*, Masaaki Kurosu (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 94–103.

[30] Xiao Li, Wen Yi, Hung-Lin Chi, Xiangyu Wang, and Albert P.C. Chan. 2018. A critical review of virtual and augmented reality (VR/AR) applications in construction safety. *Automation in Construction* 86 (2018), 150–162. doi:10.1016/j.autcon.2017.11.003

[31] Chunlin Liao, Yiheng Zhang, and Rongrong Chen. 2024. The Sense of Agency Improves Visual Search and Holds Up Against External Salient Distractions in Virtual Reality. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 319, 7 pages. doi:10.1145/3613905.3652006

[32] Hannah Limerick, David Coyle, and James W. Moore. 2014. The experience of agency in human-computer interactions: a review. *Frontiers in Human Neuroscience* 8 (Aug. 2014), 10 pages. doi:10.3389/fnhum.2014.00643

[33] Martin Löwstedt and Rikard Sandberg. 2020. Standardizing the free and independent professional: The case of construction site managers in Sweden. *Engineering, Construction and Architectural Management* 27, 6 (Jan. 2020), 1337–1355. doi:10.1108/ecam-11-2018-0489

[34] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J. Vera Liao, James Choi, Kaiyue Fan, Sean A. Munson, and Alexis Hiniker. 2021. How the Design of YouTube Influences User Sense of Agency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 368, 17 pages. doi:10.1145/3411764.3445467

[35] Tiffany Luong, Nicolas Martin, Anaïs Raison, Ferran Argelaguet, Jean-Marc Diverrez, and Anatole Lécuyer. 2020. Towards Real-Time Recognition of Users Mental Workload Using Integrated Physiological Sensors Into a VR HMD. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Porto de Galinhas, Brazil, 425–437. doi:10.1109/ISMAR50242.2020.00068

[36] Geneviève A Mageau and Robert J Vallerand. 2003. The coach–athlete relationship: a motivational model. *Journal of Sports Sciences* 21, 11 (Nov. 2003), 883–904. doi:10.1080/0264041031000140374

[37] Kane Meissel and Esther Yao. 2024. Using Cliff's Delta as a Non-Parametric Effect Size Measure: An Accessible Web App and R Tutorial. *Practical Assessment, Research and Evaluation* 29, 2 (2024), 11 pages. doi:10.7275/PARE.1977

[38] Grégoire Mompeu, Florence Danglade, Frédéric Mérienne, and Christophe Guillet. 2024. Methodology for augmented reality-based adaptive assistance in industry. *Computers in Industry* 154 (Jan. 2024), 104021. doi:10.1016/j.compind.2023.104021

[39] Alberto Monge Roffarello and Luigi De Russis. 2019. The Race Towards Digital Wellbeing: Issues and Opportunities. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3290605.3300616

[40] Florian 'Floyd' Mueller, Darren Edge, Frank Vetere, Martin R. Gibbs, Stefan Agamanolis, Bert Bongers, and Jennifer G. Sheridan. 2011. Designing sports: a framework for exertion games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 2651–2660. doi:10.1145/1978942.1979330

[41] Mojtaba Noghabaei, Arsalan Heydarian, Vahid Balali, and Kevin Han. 2020. Trend Analysis on Adoption of Virtual and Augmented Reality in the Architecture, Engineering, and Construction Industry. *Data* 5, 1 (March 2020), 26. doi:10.3390/data5010026

[42] U.S. Bureau of Labor Statistics. 2022. Census of Fatal Occupational Injuries Summary, 2022. https://www.bls.gov/news.release/pdf/cfoi.pdf last accessed: 2025.02.01.

[43] Jeesun Oh, Wooseok Kim, Sungbae Kim, Hyeonjeong Im, and Sangsu Lee. 2024. Better to Ask Than Assume: Proactive Voice Assistants' Communication Strategies That Respect User Agency in a Smart Home Environment. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 846, 17 pages. doi:10.1145/3613904.3642193

[44] Jordan Herrema Rafael Radkowski and James Oliver. 2015. Augmented Reality-Based Manual Assembly Support With Visual Features for Different Degrees of Difficulty. *International Journal of Human–Computer Interaction* 31, 5 (2015), 337–349. doi:10.1080/10447318.2014.994194

[45] Richard M. Ryan and Edward L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55, 1 (2000), 68–78. doi:10.1037/0003-066x.55.1.68

[46] Jeff Sauro and Joseph S. Dumas. 2009. Comparison of three one-question, post-task usability questionnaires. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1599–1608. doi:10.1145/1518701.1518946

[47] Ben Shneiderman, Maxine Cohen, Steven Jacobs, Catherine Plaisant, Nicholas Diakopoulos, and Niklas Elmqvist. 2017. *Designing the User Interface Strategies for Effective Human-Computer Interaction*. Pearson Deutschland, München, Deutschland. 624 pages.

[48] Ana Stanescu, Peter Mohr, Dieter Schmalstieg, and Denis Kalkofen. 2022. Model-Free Authoring by Demonstration of Assembly Instructions in Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (2022), 3821–3831. doi:10.1109/TVCG.2022.3203104

[49] Low Sui Pheng. 2001. Construction of dwellings and structures in ancient China. *Structural Survey* 19, 5 (Dec. 2001), 262–274. doi:10.1108/02630800110412859

[50] Yuko Suzuki, Fridolin Wild, and Eileen Scanlon. 2023. Measuring cognitive load in augmented reality with physiological methods: A systematic review. *Journal of Computer Assisted Learning* 40 (2023), 375–393. doi:10.1111/jcal.12882 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/jcal.12882

[51] Yi Tan, Wenyu Xu, Shenghan Li, and Keyu Chen. 2022. Augmented and Virtual Reality (AR/VR) for Education and Training in the AEC Industry: A Systematic Review of Research and Applications. *Buildings* 12, 10 (2022), 1–25. doi:10.3390/buildings12101529

[52] Huayuan Tian, Gun A. Lee, Huidong Bai, and Mark Billinghurst. 2023. Using Virtual Replicas to Improve Mixed Reality Remote Collaboration. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (May 2023), 2785–2795. doi:10.1109/tvcg.2023.3247113

[53] Markus Wallmyr, Taufik Akbar Sitompul, Tobias Holstein, and Rikard Lindell. 2019. Evaluating Mixed Reality Notifications to Support Excavator Operator Awareness. In *Human-Computer Interaction – INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part I* (Paphos, Cyprus). Springer-Verlag, Berlin, Heidelberg, 743–762. doi:10.1007/978-3-030-29381-9_44

[54] Xiliu Yang, Felix Amtsberg, Lior Skoury, Hans Jakob Wagner, and Achim Menges. 2022. Vizor, Facilitating Cyber-physical Workflows in Prefabrication through Augmented Reality. In *Proceedings of the 27th Conference on Computer Aided Architectural Design Research in Asia (CAADRIA) [Volume 2] (CAADRIA 2022)*. CAADRIA, Hong Kong, 141–150. doi:10.52842/conf.caadria.2022.2.141

[55] Beste F. Yuksel, Kurt B. Oleson, Lane Harrison, Evan M. Peck, Daniel Afergan, Remco Chang, and Robert JK Jacob. 2016. Learn Piano with BACh: An Adaptive Learning Interface that Adjusts Task Difficulty Based on Brain State. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5372–5384. doi:10.1145/2858036.2858388