

Visualizing Causality in Mixed Reality for Manual Task Learning: A Study

Rahul Jain*, Jingyu Shi*, Andrew Benton, Moiz Rasheed, Hyungjun Doh, Subramanian Chidambaram, and Karthik Ramani

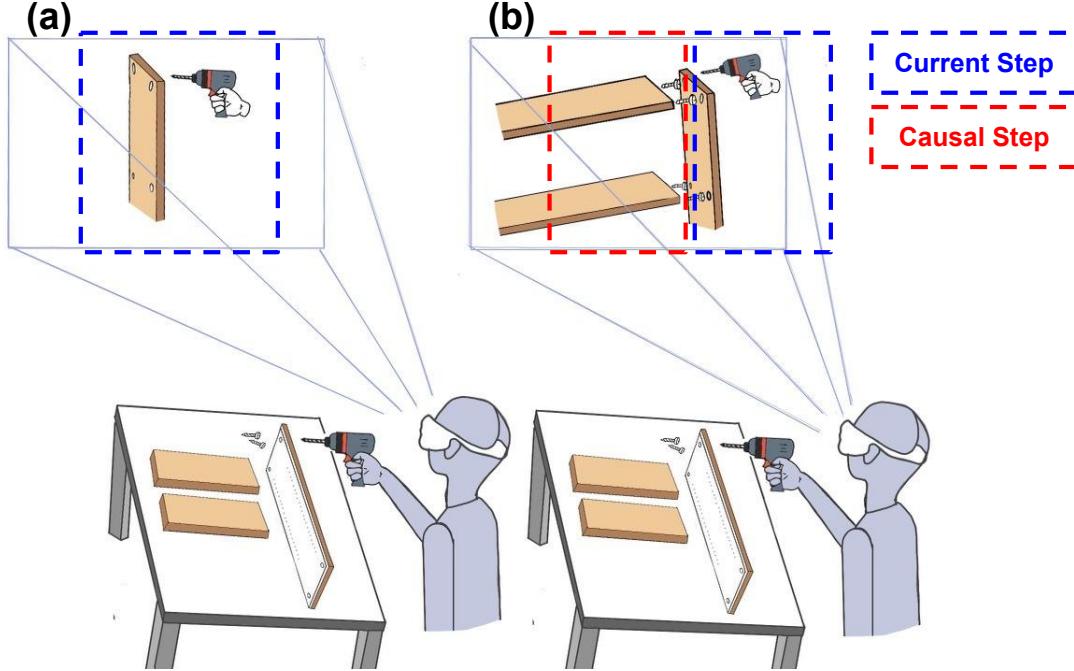


Fig. 1: An Illustration of the concept of visualizing causality. Consider an assembly task of a cabinet. Learners can learn the task by perceiving visualization of the steps from the MR device. In (a) learners are only shown the MR instructions of the current step (drill the screws into the plank), while in (b) learners are shown not only the current steps but also the future steps that are causally related (drill the screws into the plank so that it can fit into the other two planks). Our study aims to investigate the effect of showing the causal step in MR manual task learning.

Abstract—Mixed Reality (MR) is gaining prominence in manual task skill learning due to its in-situ, embodied, and immersive experience. To teach manual tasks, current methodologies break the task into hierarchies (tasks into subtasks) and visualize not only the current subtasks but also the future ones that are causally related. We investigate the impact of visualizing causality within an MR framework on manual task skill learning. We conducted a user study with 48 participants, experimenting with how presenting tasks in hierarchical causality levels (no causality, event-level, interaction-level, and gesture-level causality) affects user comprehension and performance in a complex assembly task. The research finds that displaying all causality levels enhances user understanding and task execution, with a compromise of learning time. Based on the results, we further provide design recommendations and in-depth discussions for future manual task learning systems.

Index Terms—Mixed Reality, Skill Learning, Causality, Visualization

I. INTRODUCTION

MIXED Reality (MR) is a technology that combines the physical and digital worlds and has been increasingly utilized for manual task learning across various domains such as assembly [1]–[5], machine tasks [6]–[8], and medical training [9]–[11]. Manual task learning involves the acquisition of skills necessary to perform activities that require hand-eye coordination and physical manipulation. It has thus been significantly enhanced by the introduction of MR applications because of their immersive and realistic scenarios that allow users to practice in a safe and realistic setting with various modalities of instructions.

Current methodologies for manual task learning in MR predominantly focus on guiding the users through the process by visualizing the current steps necessary to perform a particular task [3], [7], [12]–[14]. These approaches are designed to support the user's learning process by accurately and quickly guiding users through step-by-step instructions of a task.

Derived from these prior works, research pointed out that learners can anticipate their future steps [15], [16], connected

* These authors contributed equally to this research.

Email: (jain348|shi537|benton7|mrasheed|hdoh|schidamb|ramani)@purdue.edu

to their current steps. Similarly, discussion in the psychology community holds that human learns tasks by understanding the causality, i.e., the cause and effect relations among the steps that lie within the tasks [17]–[21]. Enlightened by these findings, prior works have also explored the visualization of causality in various methods such as pre-cueing [15], [22]–[24] or visualizing past or future events [25], [26] in MR. For example, Liu et al. [15] suggest pre-cueing for manual tasks, providing cues before the actual event occurs, which can prepare the user for the upcoming task. Similarly, showing past and future events in the task can help humans understand the causality within the task [25]. It is also manifested how understanding a task’s cause and effect helps prevent errors in future steps [22].

Prior works have focused on diverse applications of visualizing causality in MR and, therefore, enlightened our discussion over the effect of visualization of causality in MR on the human learning process, particularly in our scope, the manual task learning process. Based on our literature reviews, we categorize the MR visualization of causality in manual tasks into three levels: gesture, interaction, and event. Gesture-level causal visualization reveals the details of hand-object interaction that are required for future steps, e.g. angles, amount of translation, directions, etc. An interaction defines a set of dynamic gestures with an object, and interaction-level causal visualization displays the causal relationships between the interactions. An event defines a set of interactions with different objects, and event-level causal visualization reveals the causal relationships among the events. Although different levels of causal visualization have been implemented among the prior works, there exists no systematic study to reveal the effect of visualizing causality in MR on the learning gain of manual tasks. Moreover, there exists no discussion over the potential differences among different methodologies of causal visualization in MR.

To this end, we are motivated to contribute to this topic, driven by the belief that a deeper understanding of causality is crucial for more effective manual task learning. We aim to conduct a study to discover whether and how the visualization of causality in MR affects the understanding of causality in the task and consequently the manual task learning gain. This study aims to answer the following research questions to reveal future research directions for manual task learning in MR: **(1) Whether and how does visualizing causality help users learn manual tasks better in MR? (2) How different methods of causal visualization in MR helps users in learning manual tasks effectively?**

To answer these questions, we built a three-level hierarchy for causality representation in manual tasks, deriving from the existing psychology research on causality and human intention [27], [28], namely *event-level*, *interaction-level*, and *gesture-level*, corresponding to three levels of human-intention and causality. Based on the derived hierarchy, our embodied demonstrations followed four causal visualization options: ***no causality*, *task-level causality*, *interaction-level causality*, and *gesture-level causality***, where the users are shown the corresponding demonstrations of manual tasks that are causally connected to their current step.

To investigate the effects of these four options on the learning gain in manual task learning in MR, we conducted two sessions of studies ($N = 48$), where we embodied the step-by-step demonstration of the hand-object avatars in a manual assembly task in users’ MR view while they were learning the task with the physical components in situ. The contributions of our papers are as follows:

- **Study Design Rationale and Implementation** of a manual task learning scenario to compare four levels of causality visualization, where *no causality*, *event-level causality*, *interaction-level causality*, and *gesture-level causality* are shown in the demonstration of the manual tasks during users’ learning process.
- **Quantitative and Qualitative Results** showing users’ objective/subjective responses and learning gains from the sessions of manual tasks while following different visualizations of causality.
- **Design Recommendation and In-depth Discussion** summarized from the results and further discussion of the study, suggesting promising directions that deepen the research in manual task learning in MR.

II. RELATED WORK

A. Manual Task Learning in MR

Mixed reality (MR) integrates physical and digital worlds, creating an environment where physical and digital objects coexist and interact in real-time. MR is defined as a union of Virtual Reality (VR) and Augmented Reality (AR) by [29]. MR is gaining interest as a skill acquisition technology due to its uniquely immersive method of delivering educational content [30]. When used as a learning platform, MR has shown improved learner achievement, motivation, and attitude towards the learning materials [31]–[33]. Furthermore, embodied interaction, immersion, and situated learning in MR are beneficial in pedagogical contexts [30], [34]–[36] by enhancing learners’ spatial abilities [37], enhancing high-level critical thinking [38], and reducing the time and cognitive load [39].

These benefits have motivated researchers to investigate and develop MR-based manual-task learning tools and methods. MR naturally supports spatially-aware instructions for interacting with the physical environment [6]. Therefore, many task learning methods in MR provide the learner with various designs of spatially-aware, step-by-step instructions by text [40]–[42], numerical values [43], [44], 3D symbolic visual annotations [45]–[48], and 3D animation of the interactive objects [49]–[52].

Step-by-step instructions by text, annotation, or 3D animation merely explain what to do to the learners, failing to emphasize how to do it. One method for teaching learners how is by revealing temporal information of position, orientation, and the affordances of the objects along with temporal motion and gesture of the human hands [3], [7], [12]–[14]. Showing concurrent demonstration as guidance in MR to the manual task learners is the most common methodology in existing literature, across diverse domains such as assembly [1]–[5], machine task [6]–[8], instrument [53], physical tasks [54],

[55], beyond all. This method focuses on demonstrating the current steps to the users to achieve step-by-step manual task learning.

Based on this prior work, research has noticed the ability of users to anticipate the future steps of the tasks [15], and studied the effect of this comprehensive predicting ability [56]. The discussion was open over the effect of demonstration in MR of future and past steps in manual task learning. We categorize the prior works that involve visualizing cause and effects in MR into three levels: gesture-level, interaction-level, and event-level. Gesture-level causal visualization reveals the details of hand-object interaction that are required for future steps in a manual task. Such details involve the location and orientation of an object [15], [57], the hand pose or location [22], or the hand pose and object pose in a hand-object interaction holistically [58]. Interaction-level causal visualization explains a set of multiple gestures and object movements regarding one object [15], [57], [58]. Interactions depict the dynamic movements of the hands and objects, and, therefore, the temporal similarity among interactions affects the performance of the manual task [26]. An event is a set of interactions with different objects. Many prior works have investigated the effect of replaying causally related past events in MR [25], [59], [60], where sequences of past interactions have been captured and visualized to the users to assist in performing the current event.

Despite the coverage of prior research utilizing visualization of causality at different levels in MR, there exists yet no systematic research on the differences of visualizing causality at these levels, nor the discussions over their effects on manual task learning in MR. To this end, we aim to conduct a study, modeling the causal understanding in manual tasks at different levels, and provide insight into the effect of causal understanding in manual task learning in MR to maximize the learning gain.

B. Causality and Human Intention in Task Learning

1) *Definition of Causality*: There's a rich history of philosophical debate on the essence of causality. Definitions predominantly fall into two types: (1) the abstract of the progress of the world [61], [62], and (2) the relationship between human actions and attempts [63]–[67]. Similarly, debates are heated over the definition of intention. Some philosophers [68]–[72] reduce intention to beliefs and desires to act, while others [69], [72] see intention as a belief that taking a particular action leads to positively evaluated results [69], [72]–[74].

Since multiple definitions of causality and intention exist, their applicability varies across domains and scenarios. To ground our work with psychology and philosophy theories that better explain our motivations and framework, we follow the definition of causality and human intention associated with human actions. By this definition, actions are events that are intentional [72], [75], i.e., they are caused by human intention. The events take place necessarily in planned ways in a task [76]–[78]. The human's ability to understand and replicate the planned occurrence is the key to learning a

task, and we define this planned causality as **intention-driven causality**.

Although the definition above can be debated, such as the absence of unintentional actions [79], [80] and awareness of unintended side-effects (i.e., oblique intentions) [81], we argue that in the task learning process, learners will only be taught intended tasks. Thus the definitions provided serve as a sufficiently working model.

2) *Causality in Task Learning*: Research in cognitive and developmental psychology indicates the significance of understanding causality and intention in task learning. Humans have been found to parse and understand actions from intentions instead of motion from infancy [82]–[84] and early childhood [85]. In addition, Heyes et al. [86] state that humans can interpret the intentions of others and generalize this knowledge to predict their actions in novel contexts.

Causality can be inferred via probabilistic learning [17], [18] or through one's trials of the task [19]–[21]. However, learning causality by observation has shown to be more effective [87], [88]. Learners comprehend the cause and effect of each action by observing the demonstrations of instructors/teachers [89]–[91]. Such observational causal learning is guided by intention. When learners apply the task they have learned, with causality in mind, they bind, extend, and generalize the causes and effects of their actions [92].

III. DESIGN CONSIDERATION

A. Hierarchy of Causality in Task

Within the domains of Cognitive Science, Psychology, and Neuroscience, scholars have put forth a postulation that humans possess a tendency to segment complex tasks into distinct groups [93], [94]. Additionally, people understand those ongoing tasks in partonomic hierarchies [95]. Similarly, we break the hierarchical task structure into numerous discrete events, which can be further parsed into interactions and subsequently into gestures. These entities — events, interactions, and gestures —are henceforth denoted as elements within this hierarchical framework. The occurrence of these elements is decided by the intentions at different levels which are P-Intentions (Events), D-Intentions(Interactions), and M-Intentions(gestures) [27].

Moreover, intentions not only decide the occurrences but also decide the order and pattern of the elements in the inter-layer(i.e., they are only connected in the same layer). These interlayer connections between elements are related by cause and effect(causality) [94]. This implies that the intentions themselves drive the causality observed at each level of the hierarchy, and this **intention-driven causality** Figure 2. includes 1) Event level causality 2) Interaction level causality and 3) gesture level causality.

Event Level Causality: This refers to causal links between events. Events are connected in a temporal sequence, often involving cause-and-effect relationships. The causality between events is driven by intention, particularly the intention to complete a task.

Interaction Level Causality: This type of causality focuses on the relationships between interactions within an event.

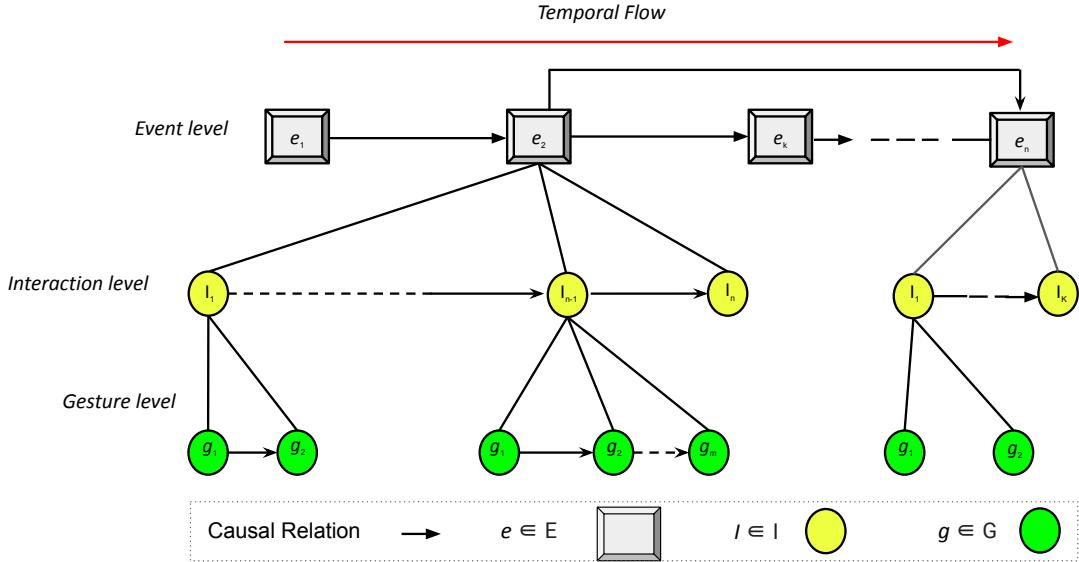


Fig. 2: A three-level hierarchy for causality in a manual task. The three levels are event, interaction, and gesture, respectively. Horizontally, each node is connected by causal relations, explaining the cause and effect among them. E.g., the occurrence of event 1 leads to the occurrence of event 2. A node from a higher level consists of one or multiple children at the lower level. We aim to differentiate different levels of causality and study their effects in manual task learning.

Interactions could be more specific actions, behaviors, or steps that contribute to the completion of an event. The causal relations between interactions are established due to the intention set at the event level. The overall aim here is to accomplish the event successfully.

Gesture Level Causality: Gesture-level causality deals with the temporal links between poses (referring to physical postures, configurations, or conditions) within an event. These causal relations are influenced by the intentions set at the interaction level. In this context, poses represent more granular components of interactions, and their causality is aimed at fulfilling the requirements of the interactions.

B. Visualization

Effective task learning often involves the emulation of an expert practitioner, where learners observe and replicate demonstrations provided by a proficient master [6]. These demonstrations are typically recorded and delivered to the learners. Prior research has emphasized the importance of visualizing task hierarchies in a Mixed Reality environment [8], [54], [55]. We follow similar approach for our study in which we present the visualization of the task by showing the hierarchies using a graph - Events, Interactions, and Gestures in a Mixed reality environment. The graph structure involves three nodes for events, interactions, and gestures as shown in Figure 3

1) *Causality Visualization:* There are other tutoring systems that provide information on the current task as well as the future task in AR [15]. In a similar way instead of showing the next task, we show the causal task (effect of the current task) to the learner. Other prior works have used textual and visuals [25] narrative with graphs to help users understand temporality in the task. We prefer textual and visual

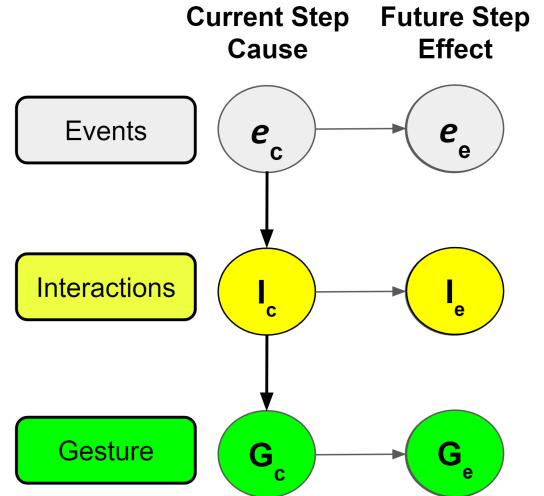


Fig. 3: Our methodology to visualize the causality in the manual task at three levels. In the Cause column, we will show the demonstration of (event, interaction, or gesture) of the current step, which is the cause. In the Effect column, we will show the corresponding level of demonstration of the future step that is causally related to the current step.

information to show causal graphs because visuals stick to memory more effectively. For our study, We show current event, interactions, and gesture and their effect (the future) together in the mixed reality in the form of a graph as shown in Figure 3.

2) *Demonstration of Content:* When a user is trying to learn, the most effective way is to observe and follow the demonstration of an experienced master. The expert records

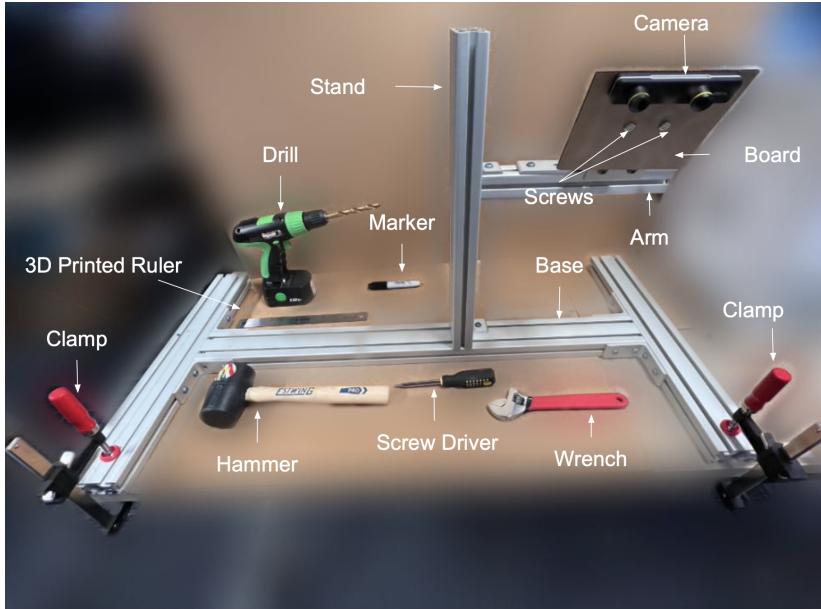


Fig. 4: The camera setup to be assembled and the tools to be used in our test bed.

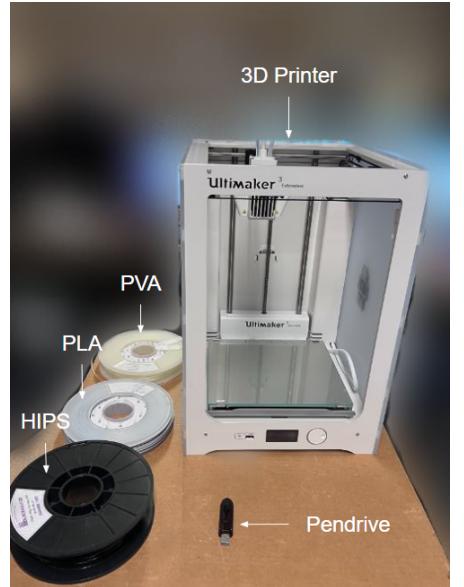


Fig. 5: The printer setup in our test bed.

the content and delivers it to the user. Similar to prior work [3] which allows both video and demonstration. We further provide the users with the functions to see the 3D hand-object-interaction demonstrations of the manual tasks in MR. This allows users to see instructions as well as demonstrations in MR.

IV. TEST BED

We implemented an MR scene to evaluate the effect of causal visualization in learning tasks. The user interface and demonstrations of the content were developed in Unity software.

A. Hardware Implementation

We developed an MR system by attaching the ZED Mini stereo camera to the Oculus Quest 2. The system uses the ZED Mini to capture both the physical scene and depth information of the environment while the Quest 2 gathers hand-pose information and displays the visualization and demonstration in the MR scene. The creation of virtual content and hardware integration was done using the Unity [41] game engine.

B. Test scene

In our study, we used the task of assembling a camera setup Figure 4 as a test scene. The task was chosen because of its procedural nature and complexity (i.e. it is hard to perform the task without looking at instructions). The task supports all the elements in the hierarchy which are events, interactions, and gestures. The task consists of five events: (1) Move the base onto a table, fix the base with two clamps, and check that it's set by striking it with the hammer. (2) Attach materials to a 3D printer, insert a flash drive, and press the start button to print a ruler. (3) Attach a vertical beam to a designated location on the

base and fasten it using a screw and screwdriver. (4) Attach an arm to a designated location on the vertical beam and fasten it using a screw and screwdriver. (5) Attach a camera mount to the arm and fasten it with a screwdriver. Then, drill two holes in a wooden board and use two bolts to secure the board to the mount. The detailed procedure of the task is present in Table I.

C. Authoring Content

We developed a system that converts live demonstrations to 3D content for generating tutorials for the learner. To create 3D tutorials of the camera setup, CAD models of the objects were created through scanning by an Intel RealSense 435i camera. The camera was moved 360 degrees around the object to capture RGB-D frames which were used by BADSLAM [96], an RGB-D SLAM method to generate a 3D mesh. The 3D models for all objects involved in the camera setup were imported to Unity. Then one of the authors wears an Oculus 2 with a zed camera enabling passthrough to create the content. In the beginning, the scanned mesh models are aligned manually with the respective physical object. This alignment provides the initial six degrees of freedom (6DoF) pose of the objects. After that author performs the entire procedure. To create 3D content, hand tracking, and object 6 Dof tracking are needed. During the demonstration, Hand-tracking data and RGB-D frames were collected along with RGB frames from the Zed Camera and Oculus 2. The collected RGB and RGB-D frames were used to track object 6DoF using Megapose [97]. Both 3D hand tracking data and object 6Dof data were used for creating 3D content for the entire camera setup tutorial.

To find the elements of the camera assembly task, we manually segment the entire 3D content into five events. To segment interactions in each event, we temporally segment

TABLE I: Events and interactions of the manual assembly tasks with our test bed.

Event	No.	Interactions	Object involved in interaction
Base Installation	1	Move a base panel up onto a table	Base, desk
	2	fix the base panel with clamp A	Base, desk and clamp A
	3	fix the base panel with clamp B	Base, desk and clamp B
	4	Strike hammer to check the stability	Hammer, base
Printing Ruler from a 3D printer	5	Rotate the printer	Printer
	6	Attach PVA to printer	PVA and Printer
	7	Attach PLA to printer	PLA and Printer
	8	Attach HIPS to printer	HIPS and Printer
	9	Insert hard drive to load CAD model	Hard drive and printer
	10	Press Start button	Printer
	11	Pick ruler from printer	Ruler and printer
	12	Mark 12 inch from the end on the base with ruler	Ruler, base and base
Stand Installation	13	Move the bolt to the location	Screwdriver and bolt
	14	Align the stand to the bolt	Stand and bolt
	15	Fix the stand with screwdriver	Screwdriver, stand and base
	16	Mark 12 inches on stand from the top with ruler	Ruler, stand and base
Arm Installation	17	Slide the arm to the location	Arm and stand
	18	Fix the arm to the stand	Screwdriver, stand and arm
	19	Slide the mount into the arm	Mount and arm
Mount Installation	20	Fix the mount with screwdriver	Screwdriver mount and stand
	21	Align board with mount	Board and mount
	22	Drill two holes on the board	Drill gun and board
	23	Screw the bolts	Wrench, bolts and boards
Camera Attachment	24	Mark 12 inch from the end on the base with ruler	Ruler, base and base

the events into interactions. Each interaction consists of four stages: (1) hand approaching an object, (2) grasping the object, (3) manipulating the object, and (4) releasing the object. For finding pose, we only consider gesture used in stage 2 when the hand is grasping the object. Causal links between the elements are also manually established using the same method used by Fender and Holz, (2022) [25]. These causal links and the order of operations generate our learning graph for the task of authoring content.

V. USER STUDY

We conducted a two-session user study with our test bed. In the first session, we studied the impact of the existence of causality visualization on the learning process of a manual task. In the second session, we studied the impact of different levels of causality visualization on the learning process.

A. Hypotheses

Based on our research questions and literature review, we formulated the following hypotheses for Session 1:

- **H1.** Participants will spend less time learning the tasks if shown the demonstrations of steps of the task that are causally related to their current step than if shown only the demonstration of their current task.

- **H2.** Participants will perform faster in the task after they learn the task by understanding the causality in the task.
- **H3.** Participants will make fewer mistakes in the task after they learn the task by understanding the causality in the task.

For Session 2:

- **H4.** Among three levels of causality, participants will spend the least time learning if shown the gesture-level causality, and the most time if shown the event-level causality.
- **H5.** Among three levels of causality, participants will perform fastest after they learn the gesture-level causality of the task and slowest after they learn the event-level causality.
- **H6.** Among three levels of causality, participants will make the least mistakes in the task after they learn the gesture-level causality of the task and the most mistakes after they learn the event-level causality.

B. Study Design

- 1) **Participants:** We recruited 48 users (25 self-identified as males and 17 self-identified as females, aged 18 to 55) for our user study. Among the user group, 14 have used either VR or AR devices before, 12 have used both before, and 20 have never used either one. Thirty-six users have a background

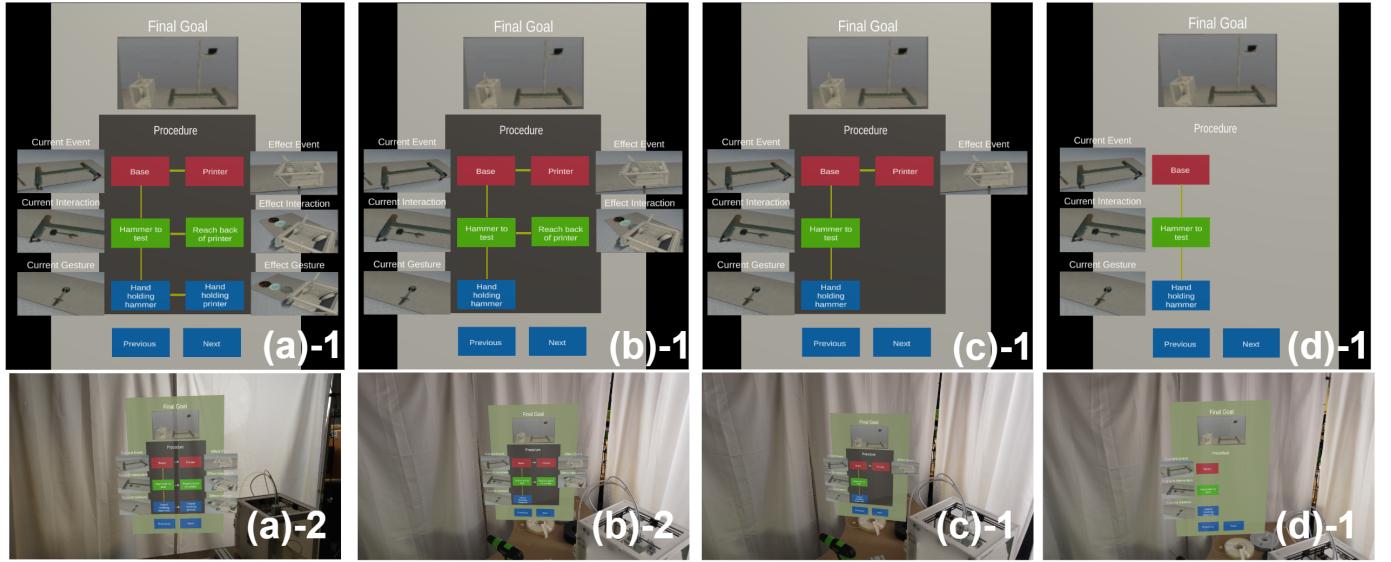


Fig. 6: UIs that were shown to 4 different groups and AR animation. a) shows the event, interaction, and affordance/gesture, b) shows the event and interaction level relation, c) shows the event level causal relation and finally d) shows the graph and no causal relation.

in engineering, eight in science, one in education, and three in others. Each participant was compensated with a \$15 gift card. All participants completed the entire study. We divided our users into four groups, each provided with different levels of information from the learning graph.

- 1) The Control Group **GC** were shown 3D demonstrations of the hand gestures, object affordances, interactions, and events of the current step Figure 6 (d).
- 2) The First Group **G1** were shown 3D demonstrations of the hand gestures, object affordances, interactions, and events of the current step. Meanwhile, they are shown demonstrations of the steps that are causally related at the event level Figure 6 (c).
- 3) The Second Group **G2** were shown 3D demonstrations of the hand gestures, object affordances, interactions, and events of the current step. Meanwhile, they are shown demonstrations of the steps that are causally related at the interaction and event levels Figure 6 (b).
- 4) The First Group **G3** was shown 3D demonstrations of the hand gestures, object affordances, interactions, and events of the current step. Meanwhile, they are also shown demonstrations of the steps that are causally related at all three levels Figure 6 (a).

We evaluated the learning performance by modeling the causality for Groups (**G1**, **G2**, and **G3**), while the Control group (**GC**) was provided no information on causality. This approach enabled comparison of the learning performance between participants exposed to different levels of our taxonomy for modeling causality. In our study, we would like to clarify that the understanding of causality has a cliff-like learning effect, i.e., "know if a thing will happen or not," is an approximate binary variable. We design our user study with a between-subject setup.

2) *Procedure:* Each participant was warmly welcomed by the researcher and provided with a concise introduction to

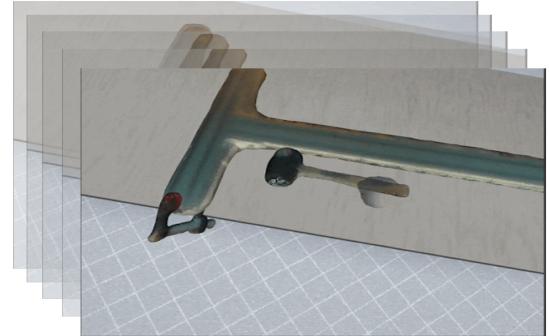


Fig. 7: The VR content shown to the user in the visualization graph.

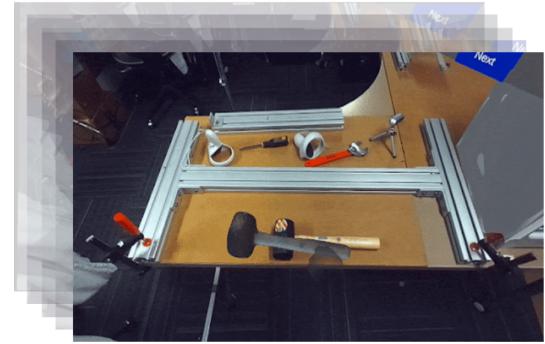


Fig. 8: AR demonstration of the current interaction.

both the study and the task. Upon signing the consent form, participants received a brief overview of the task's details. Then, they were asked to wear headsets with pass-through capabilities to get familiar with our system in the MR environment. The study includes two phases: the learning phase and the testing phase. In the learning phase, each user was

provided with sufficient time (1 hour) to learn the content using our system. Every participant finished learning within one hour. Then, participants performed the entire assembly task without any help in the testing phase. This phase has a 15-minute time limit to complete the task. After finishing the study, users provided feedback about the study through a semi-structured interview and questionnaires.

3) Data Collection: Throughout the study, we recorded participants' performances from the third-person camera as well as a screen capture of the AR view during the session. We then manually computed the length of time of each session, event, and interaction alongside task/sub-task accuracy in post-processing. We visually compared the procedural consistency between participants' demonstrations of tutorials at all three levels. We analyzed the number of times each participant from **Gc** and **G3** revisited the instruction, i.e., they referred to the instruction again for more information. We did not include participants from **G1** and **G2** since the coverage in the length of the instructions is different. Therefore, it is of no need or rigor to compare with them.

After the study, participants provided feedback on our system through the NASA TLX [98] form and a semi-structured interview. Participants also completed a 5-point Likert scale questionnaire. The questions from the survey were designed to determine the participant's memory of the task, self-rating of their memory from learning, motivation to learn, and comprehension of the causality and intention within the tasks.

VI. RESULTS AND DATA ANALYSIS

This section evaluates our workflow by analyzing participants' performance, subjective feedback, and other interesting findings.

The independent variable manipulated between our four groups is the level of causality being visualized in the learning system. As described in subsection V-B, **Gc** saw no intent and causality information while **G1, G2**, and **G3** were exposed to various levels of our taxonomy. Note that the level of causality (being visualized) is defined by the cause-and-effect pair of the level, hence both should be visualized accordingly. The dependent variables for each of the hypotheses in our experiments are time spent in learning (**H1, H4**), time spent completing the task after learning (**H2, H5**), and the mistakes made while completing the task (**H3, H6**).

Thus, we make four comparisons (**Gc** vs **G3**, **G1** vs **G2**, **G1** vs **G3**, and **G2** vs **G3**). It is shown by the Shapiro-Wilk normality test ($p < 0.005$) that the data from **neither** our quantitative nor qualitative evaluation follows a normal distribution. Therefore, we conducted a Friedman test for **G1** vs **G2** vs **G3** and a Wilcoxon signed-rank test for all four comparisons individually. To prove the hypothesis of session 1, we compare (**Gc** vs **G3**) and for session 2 the other three comparisons.

A. Objective Performance

1) Learning Performance: We first demonstrate the overall learning performance by comparing the total time spent in the

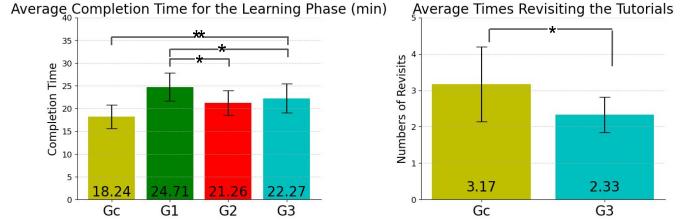


Fig. 9: Objective Results from Learning phase. The comparison of **average learning time** is on the left-hand side. The comparison of the **average number of revisits in the instructions** is on the right-hand side. (** = $p < .005$, * = $p < .05$, error bars represent standard deviations)

learning phase, including viewing the content and practicing. The results are shown in Figure 9.

The Friedman test shows significant differences among the groups in learning time($\chi^2(3) = 18.725, p = 0.003$). The average learning time shows that the participants spend the longest amount of time (all in minutes) when provided event-level causal demonstration in **G1** ($M = 24.71, SD = 3.08$). To evaluate **H1**, we compare the **Gc** and **G3**. The average learning time of participants from **G3** ($M = 22.27, SD = 3.23$) spend more time than those from **Gc** ($M = 18.24, SD = 2.60, p = 0.001, Z = -2.981$). This shows that **G3** which includes causal information increases the learning (training) time as compared to **Gc** which shows only the current task. Hence **H1** is failed. For **H4**, we compare among **G1, G2** and **G3**. Participants from **G2** ($M = 21.26, SD = 2.71$) spend insignificantly more time ($p = 0.140, Z = -1.098$) than those from **G3**. Meanwhile, participants from **G2** spend significantly less time in learning than those from **G1** ($p = 0.008, Z = -2.401$). Participants from **G1** also spent significantly more time learning than those from **G3** ($p = 0.050, Z = -1.647$). We conclude that **H4** is partly proved as the most time was taken by **G1** and least by **G2**.

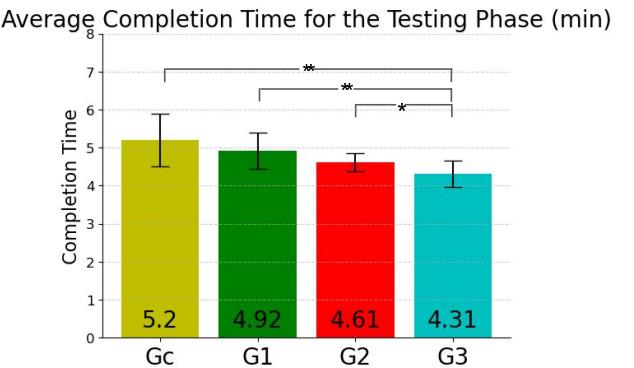


Fig. 10: **Average time (in minutes) spent in the testing phase** by groups. (** = $p < .005$, * = $p < .05$, error bars represent standard deviations)

2) Testing Performance: We evaluate the remaining four hypotheses **H2, H3, H5**, and **H6** in this section. The participant's performance during the two testing phases with respect to the average total task time of the first phase is shown in Figure 10 (**H2** and **H5**). The number of mistakes of three

different levels made during both phases is demonstrated in Figure 11 (**H3** and **H6**).

The Friedman test shows significant differences among the groups in testing time ($\chi^2(3) = 15.9, p = 0.001$) The average testing time of the participants from **G3** ($M = 4.31, SD = 0.35$) is significantly lower than all three other groups, namely **Gc** ($M = 5.20, SD = 0.69, p = 0.002, Z = -2.824$), **G1** ($M = 4.92, SD = 0.47, p = 0.001, Z = -2.981$), and **G2** ($M = 4.61, SD = 0.24, p = 0.036, Z = -1.804$). This proves **H2 and H5**. The average testing time of the participants from **G1** is insignificantly lower than those from **G2** ($p = 0.058, Z = -1.5689$).

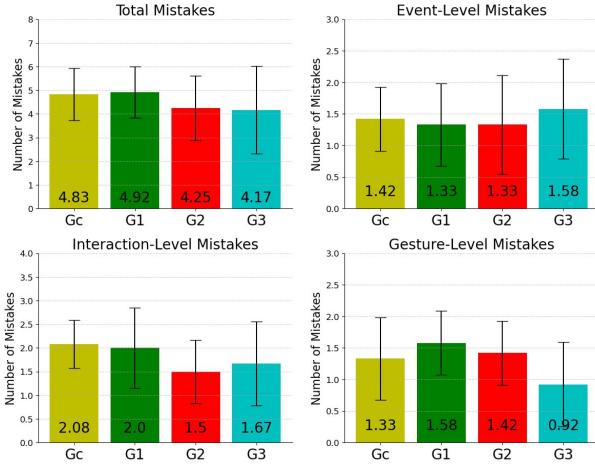


Fig. 11: **Average Mistakes made per participant** during the testing phase in different levels. (error bars represent standard deviations)

B. Subjective Feedback

This also proves **H5** We also evaluated the number of mistakes made by the participants during the testing phase. The Friedman tests show no significant difference among the groups in total mistake numbers ($\chi^2(3) = 3.3, p = 0.348$), event-level mistake numbers ($\chi^2(3) = 0.5, p = 0.919$), event-level mistake numbers ($\chi^2(3) = 5.825, p = 0.120$), and event-level mistake numbers ($\chi^2(3) = 4.525, p = 0.210$). This shows that both **H3 and H6** fails.

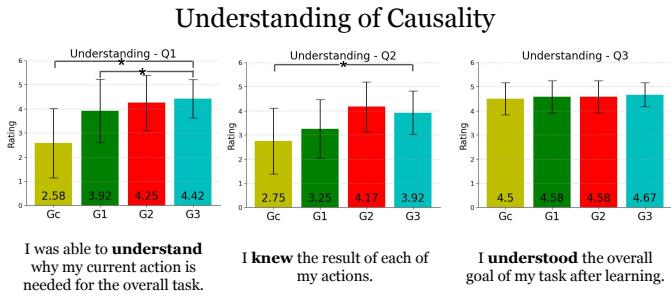


Fig. 12: Subjective User Rating on their **Understanding of the task**, overall or in detail. The results are average scores by group from a 5-point Likert-scale questionnaire. $** = p < .005$, $* = p < .05$, error bars represent standard deviations.

Self-confidence of the Learning Gain

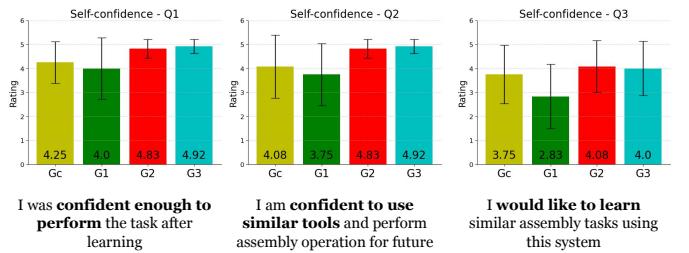


Fig. 13: Subjective User Rating on their **Self-confidence in the learning gain**. The results are average scores by group from a 5-point Likert-scale questionnaire. No significant difference was found. Error bars represent standard deviations.

Memory of the Task

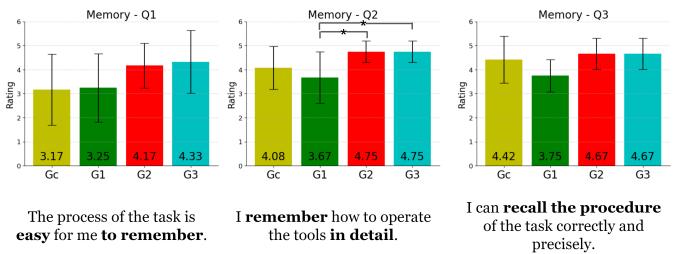


Fig. 14: Subjective User Rating on their **Memory of the learning content**. The results are average scores by group from a 5-point Likert-scale questionnaire. $** = p < .005$, $* = p < .05$, error bars represent standard deviations.

1) Likert-Scale Questionnaire: We qualitatively evaluate user feedback using a 5-point Likert Scale Questionnaire consisting of 3 major categories Understanding, Motivation, and Memory. We performed the Friedman test to compare all the questions in the respective categories. The results are presented below Figure 12.

- Understanding.** The Friedman test shows a significant difference between all four groups for Q1 ($\chi^2(3) = 8.525, p = 0.036$) and Q2 ($\chi^2(3) = 8.05, p = 0.044$) but not Q3 ($\chi^2(3) = 0.175, p = 0.98$). We performed pairwise comparisons among the groups for Q1 and Q2. **G3** ($M = 4.42, SD = 0.75$) understands the importance of the task better as compared to **G1** ($M = 3.92, SD = 1.21, p = 0.020, Z = -2.038$) and **Gc** ($M = 2.58, SD = 1.24, p = 0.008, Z = -2.400$). **G3** ($M = 3.92, SD = 0.9$) also knows better the outcome of their actions beforehand as compared to **Gc** ($M = 2.75, SD = 1.35, p = 0.038, Z = -1.765$) Figure 12.

- Self Confidence.** The Friedman test does not reveal any significant difference for all the questions Q1 ($\chi^2(3) = 5.725, p = 0.125$), Q2 ($\chi^2(3) = 6.775, p = 0.079$), and Q3 ($\chi^2(3) = 4.675, p = 0.197$) in self-confidence Figure 13.
- Memory.** The Friedman test shows significant variation in Q2 ($\chi^2(3) = 7.041, p = 0.029$) but not in Q3 ($\chi^2(3) = 5.975, p = 0.112$) and Q1 ($\chi^2(3) = 5.875, p = 0.117$).

G3 ($M = 4.75$, $SD = 0.45$) users can remember to operate the tools better than **G1** ($M = 3.67$, $SD = 1.07$, $p = 0.0007$, $Z = -2.310$). Also, **G2** can better remember the tool use as compared to **G1** ($M = 3.67$, $SD = 1.07$, $p = 0.007$, $Z = -2.446$).

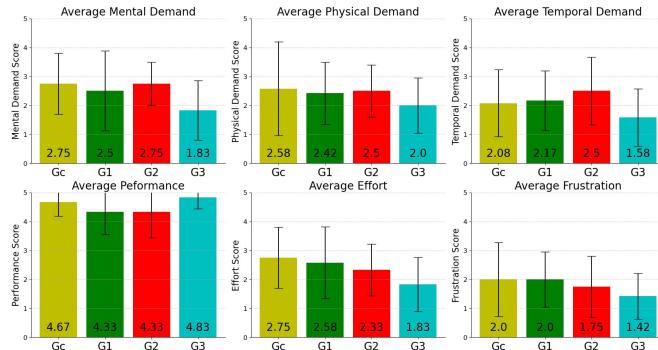


Fig. 15: **NASA TLX results.** Error bars indicate the standard deviations. We found no significant difference among the groups.

2) *NASA TLX*: We used NASA TLX's responses to evaluate the perceived subjective mental workload experience for the sessions. The findings are present in Figure 15. After confirming the non-normality with a Shapiro-Wilk test, a Friedman test was conducted on the results obtained from an unweighted NASA TLX to compare all the sessions. None of the metrics showed significance: Mental Demand ($\chi^2(3) = 3.875$, $p = 0.275$), Physical Demand ($\chi^2(3) = 1.8$, $p = 0.614$), Temporal Demand ($\chi^2(3) = 4.225$, $p = 0.238$), performance ($\chi^2(3) = 2.625$, $p = 0.453$), Effort ($\chi^2(3) = 3.525$, $p = 0.317$) and frustrations ($\chi^2(3) = 2.125$, $p = 0.546$). We further do not perform any adjustments in p-values for multiple comparisons, as there is no significant result as shown in Figure 15.

C. Research Question Analysis

- **RQ1:** We observed that **H1** and **H3** did not yield successful outcomes, whereas **H2** was confirmed to be accurate. **H1** posited that the learning time in **G3** would surpass that of **Gc**. After failing of two hypotheses we further analyzed more results. To explore performance improvements during the learning phase, we incorporated an additional metric based on the number of tutorial revisits [6]. Interestingly, the average revisits for **Gc** ($M = 3.17$, $SD = 0.72$) were significantly higher compared to **G3** ($M = 2.33$, $SD = 0.49$, $p = 0.018$, $Z = -2.090$), as depicted in Figure 9. This shows the difficulty in learning the content. Regarding **H3**, it was inconclusive due to the absence of a significant difference. We conclude that the absence of significance in the number of mistakes is due to the similar mistakes made during testing, which led to close counts of mistakes across the groups at all three levels.

Moreover, the NASA TLX workload assessment results did not yield significance across all four groups. Through a combination of our original hypotheses and additional analyses, we can deduce that the inclusion of causal

information improves the quality of the learning gain and thus enhances the performance of the participants in the same task. However, additional causal information induces more content to learn, resulting in a longer learning time. Yet, it is worth noticing that despite the extra content to learn, participants from **G3** revisited the tutorials less than those from **Gc**, implying a better immediate learning gain with an understanding of the causal information in the tasks.

- **RQ2:** We can see from the results that **H4** is rejected although **G1** did take the most time which corresponds to event-level causality. **H5** was proven valid whereas **H6** is not supported. From further analyzing the data we found that the reason behind the failure of **H6** is the same as discussed for **H3** in the above paragraph for the mistakes made in the task. Overall, we can conclude that the complete hierarchy of causality with all three levels improves the performance of the task.

VII. DISCUSSION

In this section, we summarize key results and provide insight drawn from our analysis and post-study interviews.

A. Better Learning Results with Intention-driven Causality

The average learning time of **Gc** is significantly less than that of **G3** indicating that participants are spending more time learning additional information about the intention-driven causality of the assembly task. However, the significantly high number of revisits from **Gc** indicates that with the knowledge of causality, the learners understand the tasks with fewer times referring to the instructions because they get insights from the additional causality information. In terms of the testing performance, participants from **G3** perform significantly faster than participants from **Gc**, with fewer total mistakes made during their demonstrations. The ratings from the post-study questionnaire **Understanding - Q1** and **Understanding - Q2** show that participants from **G3** give significantly higher ratings than those from **Gc** regarding their understanding of the consequences of and intention of their actions in question, which leads to better performance of the task "*I knew I needed to clamp the base tight enough when I saw my next step was to hit it with a hammer (G2, P4)*". This finding aligns with the conclusion from the prior work [15], [92], [99], [100] that in observational causal learning, the understanding of the causality is key to the learning performance and learning outcome.

B. Active Inference of Causality

Although we have discovered the gap of understanding in intention-driven causality of their actions between **Gc** and **G3** (**Understanding - Q1, Understanding - Q2**), it is interestingly worth noticing that both groups show insignificant difference in their comprehension of the overall goal (**Understanding - Q3**). As suggested in [89]–[91], learners can infer the causality and the intention of the demonstrator while observing. We anticipate that during our study, learners will

be also able to actively infer the causality within the task, in addition to passively learning causality from the causal information provided by the system. Between **G2** and **G1**, there exist significant gaps in the understanding of causal relation, indicating that participants from **G1** (who were not provided with interaction level causality information) cannot infer interaction level causality as much as that we provide to the participants from **G2**. We summarize that learners are not able to fully acknowledge the causality connecting their current interaction, although they can infer the event-level causality by observational learning. Moreover, it remains an interesting open research question to what level and how much of the causality the learners can infer by observation. We give details of future work in this direction in subsubsection VIII-A3.

C. Levels of Details: the Deeper, the Better

From our cross-comparisons among **G1**, **G2**, and **G3**, we notice that both **G2** and **G3** have significantly less learning time and less testing time than **G1**. The higher learning time of **G1** comes from the fact that showing only event-level instruction makes it hard for the learners to memorize the procedure and results in the learners more frequently referring to the instructions. "*The instruction is too long for me to remember, and I have to rewind to the part I am doing (G1, P3).*" Meanwhile, both **G2** and **G3** have less numbers of mistakes during their demonstrations. The ratings from the question **Q1** show that **G1** do not possess as much causal understanding as **G3**, which indicates that showing only the event level intention-driven causality does not provide a deeper knowledge of the causal relations within the tasks. The feedback from **Memory - Q2** also indicates that without learning the interaction and gesture/affordance levels of intention-driven causality, **G1** has a poorer understanding of how to operate the tool. A similar conclusion can be drawn from **Memory - Q3** where **G1** gives lower ratings on their performance than **G2** and **G3** regarding their precision. Although the Friedman test revealed insignificant differences among the group in responses to **Memory - Q3**, we conducted pair-wise Wilcoxon signed rank tests for **G2** vs **G1** ($p = 0.018$, $Z = -2.073$) and **G3** vs **G1** ($p = 0.023$, $Z = -1.988$) just to validate the similar conclusion from **Memory - Q2**.

Participants from **G2** spend insignificantly less time in the learning phase than those from **G3**, since **G2** are not learning the additional information about the gesture/affordance level causality. However, **G3** have a significant reduction in testing time and fewer total mistakes during testing.

The results above indicate that learners' efficiency and understanding are positively correlated with more detailed intention-driven causality information, although compromises are made in terms of the insignificant amount of extra time spent on learning. Prior works [8], [101], [102] have shown that the more details provided to the users of the MR system, the higher cognitive load is forced on the users, resulting in decreasing learning performance. Yet, with details added in our study, we do not observe a significant increase in learning time or a significant difference in users' subjective confidence about

the task they have learned (**Self-confidence - Q1,2,3**). To sum up, we conclude that including causal information in manual task learning in MR provides the learners with more and better learning gains, however, might not be the best practice when efficiency in learning is the priority.

D. Neglecting Details

An interesting finding from the numbers is that all four groups give similarly high ratings in **Understanding - Q3**, indicating that regardless of the methodology applied, all learners show a good understanding of the overall goal of the task. Likewise, all groups report high ratings of their confidence in performing the task in **Self-confidence - Q1**. Moreover, **G2** and **G3** show similar confidence in their tool operation (**Self-confidence - Q2**) and task precision (**Memory - Q3**). Yet their performance in time and the number of mistakes made at the gesture level are significantly different (Despite the insignificance among the groups shown by the Friedman test, we ran an individual comparison test for gesture-level mistakes ($p = 0.034$, $Z = -2.0226$)). This contradiction can be best explained by a lack of understanding. Learners exposed to more detailed causality and intent information had a deeper understanding of the process, enabling a more complete picture of how tools and actions contribute to the goal. By seeing the future that is causally connected to their present, they can obtain more knowledge about their present. With this knowledge, comes the ability to ascertain (1) which parts of the details of the instructions are causally linked, (2) which parts of the details can be safely ignored, and (3) which parts of the details are most important for efficient, accurate task completion "*I saw that I was about to attach a camera, then I just guessed that I needed to make the hole big enough for the bolts to go through (G3, P9).*"

However, we further argue that even in **G3**, where all intention-driven causality information is presented, learners still neglect important details due to the static nature of our implementation. In other words, the causality is presented to the learners by merely showing the ideal future state instead of showing how the future is being affected by the present. "*I knew I would install the arm later onto the stand, but I didn't know I had put the stand backward (which was wrong) until I was about to install the arm (G3, P1).*" To address this problem, in subsubsection VIII-A1 we propose a method for learners to immediately understand their errors by predicting the action's influence on the intention.

VIII. FUTURE WORK AND LIMITATION

A. Future Work

Given the insights from the results, we would like to provide recommendations for further integration of causality and intention in future MR manual task learning systems.

1) *Dynamic Causality*: As described in subsection VII-D, showing merely one future causally connected to the present results in neglected details is not optimal for learning. Therefore, we propose enhancing the representation of causality and human intention in manual task learning by **dynamically modeling causality**. In other words, the learner sees how the

causally connected futures are affected by their actions in the present. Furthermore, by showing the future dynamically, the learners can visualize and understand a broader range of causality, providing a complete understanding of the task.

Another motivation for ***dynamically modeling causality*** is the suggestion from some psychology research [103], [104] that humans can learn not only from correct instruction but also from the adverse effects of their actions. By showing the possible future failure caused by their current mistake, learners may re-evaluate their reasoning and gain a deeper understanding of the task's principles. We anticipate promising results from incorporating a dynamic representation of the future in MR manual task learning systems.

2) *Visualization of Causality*: One of the less satisfying results from our study is the compromise in learning time learners have to make when learning additional information about causality and intention of the task, as mentioned in subsection VII-C. There is always a trade-off between the content's volume and the time spent learning. For example, "*I could understand that my action would reflect the task, but it took me a while to realize how (G4, P11.)*" Therefore, we suggest that natural, user-friendly, and dedicated visualization techniques can be applied to causality and intention in MR. Take the scenario of making an omelet for example. Should the learners set the heat too high when preheating the pan, a 3D animation of an omelet being overcooked can be overlaid on the pan in MR. Such dedicated visualizations can intuitively acknowledge the user's causality within the tasks without significantly increasing the time and cognitive cost. Yet, such visualizations' design, formulation, and generalization remain challenging. One promising direction is to utilize human attention during an action. State-of-the-art artificial intelligence algorithms can predict human attention based on their actions [105]–[107]. By predicting the engagement based on learners' interaction with the objects and better positioning the visualization, it is possible that the learner can naturally perceive knowledge without a high cognitive cost in transitioning.

3) *Adaptive Causality*: Another workaround to reduce the time and cognitive cost of learning is applying adaptive Levels of Detail in the causality and intention of the learners. As suggested in [89]–[91], learners can infer the demonstrator's causality and intention while observing. Our findings, to some degree, suggest that even without explicitly presenting the causality and intention, the learners can infer them (**Understanding - Q1,2,3**). We identify this ability as active inference in subsection VII-B. Based on our work, we anticipate future research on two major research questions:

How much of the causality can learners actively infer? In our study, we conclude that learners cannot actively infer all interaction-level causality. Yet, how much of the causality can learners infer? And how do individual intelligence and cognition contribute to the inference?

What is the difference between the preserved and the inferred causality? A promising research prospect is to detect the learners' inference of causality and to quantify and measure the distance between the preserved causality from authors and the inferred causality from learners. We envisage that future research will shed light on quantifying the understanding of

causality in MR manual task learning.

With knowledge of the inferred causality of the learners, adaptive learning methods can be applied. Should more causal information be provided to learners who infer less and vice versa, and intended causality supplied to learners who infer differently, we expect more efficiency and less cognitive load in MR manual task learning methods.

B. Limitation

Our workflow for generating 3D instructions and segmenting tasks relies on the tracking 6 DoF of the objects and 3D gestures of the hands. Hence, we suffer from the same limitations state-of-the-art tracking algorithms face, such as occlusion, blurry images/feed, small-size tracking, and computation efficiency. Therefore, relying entirely on camera-based systems and generalizing the workflow remains challenging to obtain high-quality 3D content. However, compromises can be made, such as using 6 DoF sensors, multiple cameras, or markers-based computer vision to enhance the accuracy. One may also be concerned about tracking non-rigid objects, such as flexible wires and cables. However, with advances in computer graphics, state-of-the-art vision-based algorithms can also obtain a 3D model of non-rigid objects [108], which is encouraging. Finally, we would like to state that the methodology applied here for tracking does not relegate the comprehension of causality and intention matter in manual task learning.

To test our hypothesis, we have implemented a systematic instance of the workflow that is capable of preserving the causal relation and intention and delivering them to the learners. Along with this implementation, we designed a use case for an assembly task that involves interaction with multiple objects and causally related events. Even though the use case is designed based on real-world hand-object interactions and assembly tasks, it is yet a test bed. We want to acknowledge that this study aims to evaluate the implementation of our theory, and the results are analyzed to justify our approach and to bring insights into the potential contribution the understanding of causality and intention can make to manual task learning in the MR community. We also identified a blemish in our experiment setup, as the task in the user study might have been simplified, the mistake numbers across the groups did not show significant differences, which failed to support **H3** and **H6**. We anticipate a more rigorous setup and metrics to evaluate the learning gain of understanding causality in terms of mistakes made during the tasks.

Finally, we propose our work as an elicitation for future MR manual task learning systems.

IX. CONCLUSION

In this paper, we studied the effect of understanding causality and intention in manual task learning in MR. We studied the importance of causal understanding in the learning process and present the concept of ***Intentional-driven Causality for manual task learning in MR***, consisting of three layers of causality information in a manual task, namely, event level, interaction level, and gesture level. We then built a test bed to

study the effect of different layers of causality on the learning gain of an assembly task. We conducted a thorough user study with 48 users, aiming to answer the research questions of whether and how visualizing causality helps manual task learning in MR, as well as the effect of different levels of causality in manual task learning. With the quantitative and qualitative results from the study, we discuss and attribute the improvement in learning shown in the analysis, as well as all other phenomena and insights, worthy of more research. We conclude that participants perform the assembly task better after learning the causality information and gesture-level causal information grants the best performance among all three levels. Besides, we also draw other conclusions on more findings from the study. Finally, we envision the future work from our study and the concept of ***Intentional-driven Causality for manual task learning in MR*** to help the community create natural, efficient, and learner-friendly manual task learning methodologies with MR.

ACKNOWLEDGEMENT

This work is partially supported by the NSF under the Future of Work at the Human-Technology Frontier (FW-HTF) 1839971. We also acknowledge the Feddersen Distinguished Professorship Funds and a gift from Thomas J. Malott. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the funding agency.

REFERENCES

- [1] M. Whitlock, G. Fitzmaurice, T. Grossman, and J. Matejka, “Authar: Concurrent authoring of tutorials for ar assembly guidance,” in *Proceedings of Graphics Interface 2020*, ser. GI 2020. Canadian Human-Computer Communications Society, 2020, pp. 431 – 439.
- [2] M. Yamaguchi, S. Mori, P. Mohr, M. Tatzgern, A. Stanescu, H. Saito, and D. Kalkofen, “Video-annotated augmented reality assembly tutorials,” in *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, ser. UIST ’20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1010–1022. [Online]. Available: <https://doi.org/10.1145/3379337.3415819>
- [3] S. Chidambaram, H. Huang, F. He, X. Qian, A. M. Villanueva, T. S. Redick, W. Stuerzlinger, and K. Ramani, “Processar: An augmented reality-based tool to create in-situ procedural 2d/3d ar instructions,” in *Designing Interactive Systems Conference 2021*, ser. DIS ’21. New York, NY, USA: Association for Computing Machinery, 2021, p. 234–249. [Online]. Available: <https://doi.org/10.1145/3461778.3462126>
- [4] S. Chidambaram, S. S. Reddy, M. Rumple, A. Ipsita, A. Villanueva, T. Redick, W. Stuerzlinger, and K. Ramani, “Editar: A digital twin authoring environment for creation of ar/vr and video instructions from a single demonstration,” in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2022, pp. 326–335.
- [5] L.-C. Wu, I.-C. Lin, and M.-H. Tsai, “Augmented reality instruction for object assembly based on markerless tracking,” in *Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, 2016, pp. 95–102.
- [6] Y. Cao, X. Qian, T. Wang, R. Lee, K. Huo, and K. Ramani, “An exploratory study of augmented reality presence for tutoring machine tasks,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–13. [Online]. Available: <https://doi.org/10.1145/3313831.3376688>
- [7] J. Kong, D. Sabha, J. P. Bigham, A. Pavel, and A. Guo, “Tutoriallens: Authoring interactive augmented reality tutorials through narration and demonstration,” in *Symposium on Spatial User Interaction*, ser. SUI ’21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3485279.3485289>
- [8] G. Huang, X. Qian, T. Wang, F. Patel, M. Sreeram, Y. Cao, K. Ramani, and A. J. Quinn, “Adaptutar: An adaptive tutoring system for machine tasks in augmented reality,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3411764.3445283>
- [9] D. Eckhoff, C. Sandor, D. Kalkoten, U. Eck, C. Lins, and A. Hein, “Tutor: Semi-automatic generation of augmented reality tutorials for medical education,” in *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. Piscataway, NJ, USA: IEEE, 2018, pp. 430–431.
- [10] R. De Ponti, J. Marazzato, A. M. Maresca, F. Rovera, G. Carcano, and M. M. Ferrario, “Pre-graduation medical training including virtual reality during covid-19 pandemic: a report on students’ perception,” *BMC medical education*, vol. 20, pp. 1–7, 2020.
- [11] M. H. Knudsen, N. Breindahl, T.-S. Dalsgaard, D. Isbye, A. G. Mølbak, G. Tiwald, M. B. S. Svendsen, L. Konge, J. Bergström, and T. Todsen, “Using virtual reality head-mounted displays to assess skills in emergency medicine: Validity study,” *Journal of Medical Internet Research*, vol. 25, p. e45210, 2023.
- [12] J. Blattgerste, K. Luksch, C. Lewa, and T. Pfeiffer, “Train ar: A scalable interaction concept and didactic framework for procedural trainings using handheld augmented reality,” *Multimodal Technologies and Interaction*, vol. 5, no. 7, p. 30, 2021.
- [13] P. Ajidarma and S. Y. Nof, “Skill and knowledge sharing in cyber-augmented collaborative physical work systems with hub-ci,” 2022.
- [14] A. Stanescu, P. Mohr, D. Schmalstieg, and D. Kalkofen, “Model-free authoring by demonstration of assembly instructions in augmented reality,” *IEEE Transactions on Visualization and Computer Graphics*, vol. earlyaccess, no. earlyaccess, pp. 1–11, 2022.
- [15] J.-S. Liu, B. Tversky, and S. Feiner, “Precueing object placement and orientation for manual tasks in augmented reality,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 11, pp. 3799–3809, 2022.
- [16] D. A. Rosenbaum, R. G. Cohen, S. A. Jax, D. J. Weiss, and R. van der Wel, “The problem of serial order in behavior: Lashley’s legacy,” *Human Movement Science*, vol. 26, no. 4, pp. 525–554, 2007, european Workshop on Movement Science 2007. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167945707000280>
- [17] A. Gopnik, D. M. Sobel, L. E. Schulz, and C. Glymour, “Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation,” *Developmental psychology*, vol. 37, no. 5, p. 620, 2001.
- [18] T. Kushnir and A. Gopnik, “Young children infer causal strength from probabilities and interventions,” *Psychological science*, vol. 16, no. 9, pp. 678–683, 2005.
- [19] A. Gopnik, L. Schulz, and L. E. Schulz, *Causal learning: Psychology, philosophy, and computation*. Oxford, UK: Oxford University Press, 2007.
- [20] J. Woodward, *Making things happen: A theory of causal explanation*. Oxford, UK: Oxford university press, 2005.
- [21] ——, “19 Interventionist Theories of Causation in Psychological Perspective,” in *Causal Learning: Psychology, Philosophy, and Computation*. Oxford, UK: Oxford University Press, 04 2007. [Online]. Available: <https://doi.org/10.1093/acprof:oso/9780195176803.003.0002>
- [22] J. Wolf, Q. Lohmeyer, C. Holz, and M. Meboldt, “Gaze comes in handy: Predicting and preventing erroneous hand actions in ar-supported manual tasks,” in *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2021, pp. 166–175.
- [23] J.-S. Liu, C. Elvezio, B. Tversky, and S. Feiner, “Using multi-level precueing to improve performance in path-following tasks in virtual reality,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 11, pp. 4311–4320, 2021.
- [24] J.-S. Liu, P. Wang, B. Tversky, and S. Feiner, “Adaptive visual cues for guiding a bimanual unordered task in virtual reality,” in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2022, pp. 431–440.
- [25] A. R. Fender and C. Holz, “Causality-preserving asynchronous reality,” in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: <https://doi.org/10.1145/3491102.3501836>
- [26] J. Illing, P. Klinke, M. Pfingsthorn, and W. Heuten, “Less is more! support of parallel and time-critical assembly tasks with augmented reality,” in *Proceedings of Mensch Und Computer 2021*, ser. MuC ’21. New York, NY, USA: Association for

- Computing Machinery, 2021, p. 215–226. [Online]. Available: <https://doi.org/10.1145/3473856.3473861>
- [27] G. Riva, J. A. Waterworth, E. L. Waterworth, and F. Mantovani, “From intention to action: The role of presence,” *New Ideas in Psychology*, vol. 29, no. 1, pp. 24–37, 2011. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0732118X09000725>
- [28] E. Pacherie, “Towards a dynamic theory of intentions,” *Does consciousness cause behavior*, pp. 145–167, 2006.
- [29] M. Speicher, B. D. Hall, and M. Nebeling, “What is mixed reality?” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–15. [Online]. Available: <https://doi.org/10.1145/3290605.3300767>
- [30] B. Dalgarno and M. J. Lee, “What are the learning affordances of 3-d virtual environments?” *British Journal of Educational Technology*, vol. 41, no. 1, pp. 10–32, 2010.
- [31] A. Villanueva, Z. Liu, Y. Kitaguchi, Z. Zhu, K. Peppler, T. Redick, and K. Ramani, “Towards modeling of human skilling for electrical circuitry using augmented reality applications,” *International Journal of Educational Technology in Higher Education*, vol. 18, no. 1, pp. 1–23, 2021.
- [32] A. M. Kamarainen, S. Metcalf, T. Grotzer, A. Browne, D. Mazzuca, M. S. Tutwiler, and C. Dede, “Ecomobile: Integrating augmented reality and probeware with environmental education field trips,” *Computers & Education*, vol. 68, pp. 545–556, 2013.
- [33] S.-J. Lu and Y.-C. Liu, “Integrating augmented reality technology to enhance children’s learning in marine education,” *Environmental Education Research*, vol. 21, no. 4, pp. 525–541, 2015.
- [34] S. Martirosov and P. Kopecek, “Virtual reality and its influence on training and education-literature review.” *Annals of DAAAM & Proceedings*, vol. 28, pp. 0708–0717, 2017.
- [35] S. Van Der Land, A. P. Schouten, F. Feldberg, B. van Den Hooff, and M. Huysman, “Lost in space? cognitive fit and cognitive load in 3d virtual environments,” *Computers in Human Behavior*, vol. 29, no. 3, pp. 1054–1064, 2013.
- [36] A. Ipsita, L. Erickson, Y. Dong, J. Huang, A. K. Bushinski, S. Saradhi, A. M. Villanueva, K. A. Peppler, T. S. Redick, and K. Ramani, “Towards modeling of virtual reality welding simulators to promote accessible and scalable training,” in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: <https://doi.org/10.1145/3491102.3517696>
- [37] H.-C. K. Lin, M.-C. Chen, and C.-K. Chang, “Assessing the effectiveness of learning solid geometry by using an augmented reality-assisted learning system,” *Interactive Learning Environments*, vol. 23, no. 6, pp. 799–810, 2015.
- [38] F. Saltan and Ö. Arslan, “The use of augmented reality in formal education: A scoping review,” *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 2, pp. 503–520, 2016.
- [39] M. Yuan, S. Ong, and A. Nee, “Augmented reality for assembly guidance using a virtual interactive tool,” *International journal of production research*, vol. 46, no. 7, pp. 1745–1767, 2008.
- [40] T. Wójcicki, “Supporting the diagnostics and the maintenance of technical devices with augmented reality,” *Diagnostyka*, vol. 15, no. 1, pp. 43–47, 2014.
- [41] J. Zhu, S.-K. Ong, and A. Y. Nee, “A context-aware augmented reality assisted maintenance system,” *International Journal of Computer Integrated Manufacturing*, vol. 28, no. 2, pp. 213–225, 2015.
- [42] A. Peniche, C. Diaz, H. Trefftz, and G. Paramo, “Combining virtual and augmented reality to improve the mechanical assembly training process in manufacturing,” in *Proceedings of the 6th WSEAS International Conference on Computer Engineering and Applications, and Proceedings of the 2012 American Conference on Applied Mathematics*, ser. AMERICAN-MATH’12/CEA’12. Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS), 2012, p. 292–297.
- [43] J.-R. Chardonnet, G. Fromentin, and J. Outeiro, “Augmented reality as an aid for the use of machine tools,” *Res. & Sci. Today*, vol. 13, p. 25, 2017.
- [44] E. Schoop, M. Nguyen, D. Lim, V. Savage, S. Follmer, and B. Hartmann, “Drill sergeant: Supporting physical construction projects through an ecosystem of augmented tools,” in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA ’16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1607–1614. [Online]. Available: <https://doi.org/10.1145/2851581.2892429>
- [45] M. Funk, “Augmented reality at the workplace: a context-aware assistive system using in-situ projection,” Ph.D. dissertation, University of Stuttgart, 2016.
- [46] G.-S. Jo, K.-J. oh, I. Ha, K.-S. Lee, M.-D. Hong, U. Neumann, and S. You, “A unified framework for augmented reality and knowledge-based systems in maintaining aircraft,” in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, ser. AAAI’14. Palo Alto, California, US: AAAI Press, 2014, p. 2990–2997.
- [47] S. Kim, G. Lee, W. Huang, H. Kim, W. Woo, and M. Billinghamurst, “Evaluating the combination of visual communication cues for hmd-based mixed reality remote collaboration,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–13. [Online]. Available: <https://doi.org/10.1145/3290605.3300403>
- [48] G. Westerfield, A. Mitrovic, and M. Billinghamurst, “Intelligent augmented reality training for assembly tasks,” in *Artificial Intelligence in Education*, H. C. Lane, K. Yacef, J. Mostow, and P. Pavlik, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 542–551.
- [49] A. Monroy Reyes, O. O. Vergara Villegas, E. Miranda Bojorquez, V. G. Cruz Sanchez, and M. Nandayapa, “A mobile augmented reality system to support machinery operations in scholar environments,” *Computer Applications in Engineering Education*, vol. 24, no. 6, pp. 967–981, 2016.
- [50] S. Ong and Z. Wang, “Augmented assembly technologies based on 3d bare-hand interaction,” *CIRP annals*, vol. 60, no. 1, pp. 1–4, 2011.
- [51] S. Webel, U. Bockholt, T. Engelke, N. Gavish, M. Olbrich, and C. Preusche, “An augmented reality training platform for assembly and maintenance skills,” *Robotics and autonomous systems*, vol. 61, no. 4, pp. 398–403, 2013.
- [52] Z. Zhu, V. Branzoi, M. Wolverton, G. Murray, N. Vitovitch, L. Yarnall, G. Acharya, S. Samarasekera, and R. Kumar, “Ar-mentor: Augmented reality based mentoring system,” in *2014 IEEE international symposium on mixed and augmented reality (ISMAR)*. Piscataway, NJ, USA: IEEE, 2014, pp. 17–22.
- [53] Z. Liu, Z. Zhu, E. Jiang, F. Huang, A. M. Villanueva, X. Qian, T. Wang, and K. Ramani, “Instrumentar: Auto-generation of augmented reality tutorials for operating digital instruments through recording embodied demonstration,” in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–17.
- [54] J. Blattgerste, P. Renner, and T. Pfeiffer, “Authorable augmented reality instructions for assistance and training in work environments,” in *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, 2019, pp. 1–11.
- [55] B. Thoravi Kumaravel, F. Anderson, G. Fitzmaurice, B. Hartmann, and T. Grossman, “Loki: Facilitating remote instruction of physical tasks using bi-directional mixed-reality telepresence,” in *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, ser. UIST ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 161–174. [Online]. Available: <https://doi.org/10.1145/3332165.3347872>
- [56] B. Volmer, J. Baumeister, S. Von Itzstein, I. Bornkessel-Schlesewsky, M. Schlesewsky, M. Billinghamurst, and B. H. Thomas, “A comparison of predictive spatial augmented reality cues for procedural tasks,” *IEEE transactions on visualization and computer graphics*, vol. 24, no. 11, pp. 2846–2856, 2018.
- [57] J.-S. Liu, B. Tversky, and S. Feiner, “Precueing sequential rotation tasks in augmented reality,” in *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*, ser. VRST ’22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: <https://doi.org/10.1145/3562939.3565641>
- [58] —, “A testbed for exploring multi-level precueing in augmented reality,” in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2022, pp. 540–541.
- [59] H. Cho, M. L. Komar, and D. Lindlbauer, “Realityreplay: Detecting and replaying temporal changes in situ using mixed reality,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 7, no. 3, pp. 1–25, 2023.
- [60] E. Bonnail, W.-J. Tseng, M. McGill, E. Lecolinet, S. Huron, and J. Gugenheimer, “Memory manipulations in extended reality,” in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–20.
- [61] J. L. Mackie, *The cement of the universe: A study of causation*. Walton St. Oxford OX2 6DP United Kingdom: Clarendon Press, 1980.
- [62] A. N. Whitehead, *Process and reality*. NY, US: Simon and Schuster, 2010.

- [63] J. C. Bishop, J. Bishop *et al.*, *Natural agency: An essay on the causal theory of action*. Cambridge, UK: Cambridge University Press, 1989.
- [64] N. McCain, H. Turner *et al.*, “Causal theories of action and change,” in *AAAI/IAAI*. Princeton, New Jersey: Citeseer, 1997, pp. 460–465.
- [65] W. A. Davis, “The causal theory of action,” *A Companion to the Philosophy of Action*, vol. 1, pp. 32–39, 2010.
- [66] R. Clarke, “Skilled activity and the causal theory of action,” *Philosophy and Phenomenological Research*, vol. 80, no. 3, pp. 523–550, 2010.
- [67] J. H. Aguilar and A. A. Buckareff, *Causing Human Actions: New Perspectives on the Causal Theory of Action*. add: The MIT Press, 2010. [Online]. Available: <http://www.jstor.org/stable/j.ctt5hhjnkw>
- [68] T. Honderich, *The Oxford companion to philosophy*. Oxford, UK: OUP Oxford, 2005.
- [69] K. Setiya, “Intention,” in *The Stanford Encyclopedia of Philosophy*, Fall 2022 ed., E. N. Zalta and U. Nodelman, Eds. Stanford, CA, US: Metaphysics Research Lab, Stanford University, 2022.
- [70] A. Mele, *Intention and Intentional Action*. Oxford, UK: Oxford University Press, 2009, ch. 0, p. 0.
- [71] D. M. Borchert, *The Encyclopedia of Philosophy 2nd Edition. Vol. 3*. Farmington Hills, MI, US: Thomson Gale, 2006.
- [72] E. Craig *et al.*, *Routledge encyclopedia of philosophy: Index*. Milton Park, Oxfordshire: Taylor & Francis, 1998, vol. 8.
- [73] A. R. Mele, “Are intentions self-referential?” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, vol. 52, no. 3, pp. 309–329, 1987.
- [74] A. S. Roth, “The self-referentiality of intentions,” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, vol. 97, no. 1, pp. 11–52, 2000.
- [75] J. Stuchlik, “From voluntionalism to the dual aspect theory of action,” *Philosophia*, vol. 41, no. 3, pp. 867–886, 2013.
- [76] B. McLaughlin, A. Beckermann, S. Walter *et al.*, *The Oxford handbook of philosophy of mind*. Oxford, UK: Oxford University Press, 2009.
- [77] R. Audi, *The Cambridge Dictionary of Philosophy*, 3rd ed. Cambridge: Cambridge University Press, 2015.
- [78] G. Wilson, “Action,” in *Stanford Encyclopedia of Philosophy*. Standford, CA, US: Metaphysics Research Lab Philosophy Department Stanforad University Stanford, 2008.
- [79] R. Vernon, “Unintended consequences,” *Political theory*, vol. 7, no. 1, pp. 57–73, 1979.
- [80] R. Boudon, *The unintended consequences of social action*. New York, NY, USA: Springer, 2016.
- [81] S. Guglielmo and B. F. Malle, “Can unintended side effects be intentional? resolving a controversy over intentionality and morality,” *Personality and social psychology bulletin*, vol. 36, no. 12, pp. 1635–1647, 2010.
- [82] A. N. Meltzoff and R. Brooks, “Like me” as a building block for understanding other minds: Bodily acts, attention, and intention,” *Intentions and intentionality: Foundations of social cognition*, vol. 171191, p. 171–191, 2001.
- [83] A. N. Meltzoff, “Understanding the intentions of others: re-enactment of intended acts by 18-month-old children.” *Developmental psychology*, vol. 31, no. 5, p. 838, 1995.
- [84] C. George and J. Solomon, “The caregiving system: A behavioral systems approach to parenting.” In *J. Cassidy & P. R. Shaver (Eds.), Handbook of attachment: Theory, research, and clinical applications*, vol. 1, p. 833–856, 2008.
- [85] J. W. Astington, “The paradox of intention: Assessing children’s metarepresentational understanding,” *Intentions and intentionality: Foundations of social cognition*, vol. 4, pp. 85–103, 2001.
- [86] C. M. Heyes and C. D. Frith, “The cultural evolution of mind reading,” *Science*, vol. 344, no. 6190, p. 1243091, 2014.
- [87] A. N. Meltzoff, A. Waismeyer, and A. Gopnik, “Learning about causes from people: observational causal learning in 24-month-old infants.” *Developmental psychology*, vol. 48, no. 5, p. 1215, 2012.
- [88] E. B. Bonawitz, D. Ferranti, R. Saxe, A. Gopnik, A. N. Meltzoff, J. Woodward, and L. E. Schulz, “Just do it? investigating the gap between prediction and action in toddlers’ causal inferences,” *Cognition*, vol. 115, no. 1, pp. 104–117, 2010.
- [89] A. Bandura, *Observational Learning*. Hoboken, New Jersey: John Wiley & Sons, Ltd, 2008. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781405186407.wbieco004>
- [90] R. Douglas Greer, J. Dudek-Singer, and G. Gautreaux, “Observational learning,” *International journal of psychology*, vol. 41, no. 6, pp. 486–499, 2006.
- [91] D. W. Teo and D. Ong, “Learning from agentic actions: Modelling causal inference from intention,” in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 43. CA, US: eScholarship, 2021, p. 0.
- [92] M. J. Buehner and G. R. Humphreys, “Causal binding of actions to their effects,” *Psychological Science*, vol. 20, no. 10, pp. 1221–1228, 2009, PMID: 19732384. [Online]. Available: <https://doi.org/10.1111/j.1467-9280.2009.02435.x>
- [93] R. G. Barker and H. F. Wright, “One boy’s day; a specimen record of behavior.” 1951.
- [94] A. Michotte, *The perception of causality*. Routledge, 2017, vol. 21.
- [95] C. A. Kurby and J. M. Zacks, “Segmentation in the perception and memory of events,” *Trends in cognitive sciences*, vol. 12, no. 2, pp. 72–79, 2008.
- [96] T. Schöps, T. Sattler, and M. Pollefeys, “Bad slam: Bundle adjusted direct rgb-d slam,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 134–144.
- [97] Y. Labb  , L. Manuelli, A. Mousavian, S. Tyree, S. Birchfield, J. Tremblay, J. Carpentier, M. Aubry, D. Fox, and J. Sivic, “Megapose: 6d pose estimation of novel objects via render & compare,” *arXiv preprint arXiv:2212.06870*, 2022.
- [98] N. NASA, “Task load index (tlx) v. 1.0 manual,” NASA, NASA-Ames Research Center Moffett Field, 1986.
- [99] A. K. Gardiner, M. L. Greif, and D. F. Bjorklund, “Guided by intention: Preschoolers’ imitation reflects inferences of causation,” *Journal of Cognition and Development*, vol. 12, no. 3, pp. 355–373, 2011.
- [100] E. Luchkina, J. A. Sommerville, and D. M. Sobel, “More than just making it go: Toddlers effectively integrate causal efficacy and intentionality in selecting an appropriate causal intervention,” *Cognitive Development*, vol. 45, pp. 48–56, 2018.
- [101] D. Lindlbauer, A. M. Feit, and O. Hilliges, “Context-aware online adaptation of mixed reality interfaces,” in *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, ser. UIST ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 147–160. [Online]. Available: <https://doi.org/10.1145/3332165.3347945>
- [102] M. Tatzgern, V. Orso, D. Kalkofen, G. Jacucci, L. Gamberini, and D. Schmalstieg, “Adaptive information density for augmented reality displays,” in *2016 IEEE Virtual Reality (VR)*, 2016, pp. 83–92.
- [103] K. A. Chillarege, C. R. Nordstrom, and K. B. Williams, “Learning from our mistakes: Error management training for mature learners,” *Journal of business and psychology*, vol. 17, no. 3, pp. 369–385, 2003.
- [104] K. A. A. Professor, “How to learn from our mistakes,” *Philosophical Explorations*, vol. 7, no. 3, pp. 247–263, 2004. [Online]. Available: <https://doi.org/10.1080/1386979042000258349>
- [105] Z. Yang, L. Huang, Y. Chen, Z. Wei, S. Ahn, G. Zelinsky, D. Samaras, and M. Hoai, “Predicting goal-directed human attention using inverse reinforcement learning,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [106] S. Bae, E. Pakdamanian, I. Kim, L. Feng, V. Ordonez, and L. Barnes, “Medirl: Predicting the visual attention of drivers via maximum entropy deep inverse reinforcement learning,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 13 158–13 168.
- [107] G. J. Zelinsky, Y. Chen, S. Ahn, and H. Adeli, “Chapter eight - changing perspectives on goal-directed attention control: The past, present, and future of modeling fixations during visual search,” in *Gazing Toward the Future: Advances in Eye Movement Theory and Applications*, ser. Psychology of Learning and Motivation, K. D. Federmeier and E. R. Schotter, Eds. Academic Press, 2020, vol. 73, pp. 231–286. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0079742120300268>
- [108] Z. Lakhili, A. El Alami, A. Mesbah, A. Berrahou, and H. Qjidaa, “Rigid and non-rigid 3d shape classification based on 3d hahn moments neural networks model,” *Multimedia Tools and Applications*, pp. 1–24, 2022.