Projecting Robot Intentions Through Visual Cues:
Static vs. Dynamic Signaling

Shubham Sonawani, Yifan Zhou and Heni Ben Amor

Abstract— Augmented and mixed-reality techniques harbor a great potential for improving human-robot collaboration. Visual signals and cues may be projected to a human partner in order to explicitly communicate robot intentions and goals. However, it is unclear what type of signals support such a process and whether signals can be combined without adding additional cognitive stress to the partner. This paper focuses on identifying the effective types of visual signals and quantify their impact through empirical evaluations. In particular, the study compares static and dynamic visual signals within a collaborative object sorting task and assesses their ability to shape human behavior. Furthermore, an information-theoretic analysis is performed to numerically quantify the degree of information transfer between visual signals and human behavior. The results of a human subject experiment show that there are significant advantages to combining multiple visual signals within a single task, i.e., increased task efficiency and reduced cognitive load.

I. INTRODUCTION

Among the many roles future robots are envisioned to assume, one particularly challenging role is that of a human teammate. In such collaborative scenarios, robots have to provide continuous assistance to a human partner, while also ensuring a mutual understanding of the cooperative task and its individual elements. Consequently, there has been significant research interest in generating *interpretable robot behavior* that allows a human partner to better anticipate the goals, intentions, and future actions of a robot for the purpose of fluent teaming. For instance, the Roadmap for U.S. Robotics report highlights that "humans must be able to read and recognize robot activities in order to interpret the robot's understanding" [1].

To achieve such a shared mental model, several approaches use *implicit cognitive cues* to communicate robot intentions, e.g., by adjusting robot motion to elicit a specific interpretation from a human observer [2]–[4]. Alternatively, other approaches use *explicit cognitive cues*, e.g., visual, haptic or auditory signals to improve human understanding of robot intentions [5]–[8]. To this end, the use of augmented and mixed-reality techniques has gained considerable attention in recent years [9]–[14]. The work in [15] presented a robot system which projects information about a collaborative task directly into the shared workspace – a mixed reality approach. For example, by projecting a warning sign onto a particular object in the scene, a robot may identify a goal object it intends to manipulate. As a result, the environment becomes a canvas for the display of perceptual messages

S. Sonawani, Y. Zhou, and H. Ben Amor are with the School of Computing and Augmented Intelligence, Arizona State University {sdsonawa, yzhou298, hbenamor}@asu.edu
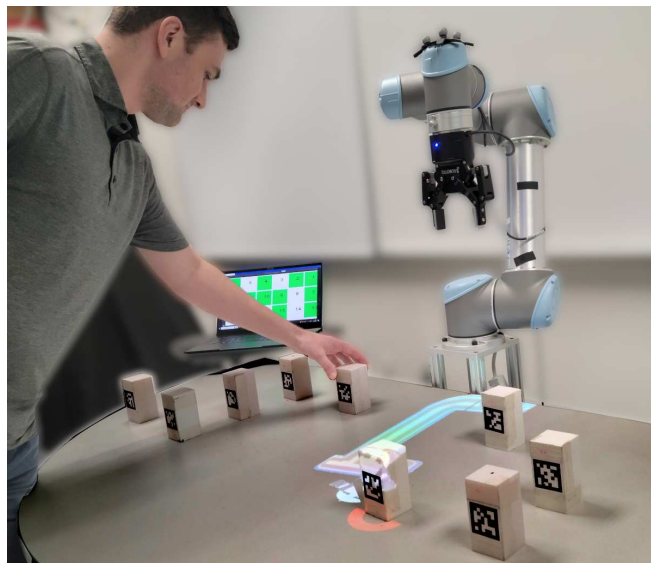
Fig. 1: The setup of our experiments: a human subject is tasked with sorting all cubes indicated by green squares, while a robot sorts all the cubes shown by white squares on the laptop screen. Here, a static cue (red semi-circular disc), and a dynamic cue (digital twin of the robot) are projected onto the physical environment showing the current goal of the robot.

that can rapidly be processed by the human visual cortex. Another way to provide visual cues is discussed in [16]–[21], e.g., using virtual reality glasses or head-mounted displays which augment the environment around humans. Various works have provided ample evidence that such visual projection of intent improves critical dimensions of human-robot collaboration tasks, e.g., efficiency, fluency, and trust.

However, there are still critical questions and challenges that are not well understood. In particular, it is unclear how to choose and design visual signals so as to achieve the desired transfer of information between the robot and the human partner. To date no objective measures of information transfer have been established that could answer such questions. In a similar vein, it is unclear whether static visual signals (e.g. signs) are preferable to dynamic ones (e.g. and an animation of a moving object or robot). Insights from related fields such as Semiotics [22], [23] and Human-Computer Interaction (HCI) [24], [25] can partially be transferred to this scenario but do not fully address the relationship between human, robot and physical environment.

This paper extends prior work on intention projection by investigating the effectiveness of different types of signals in
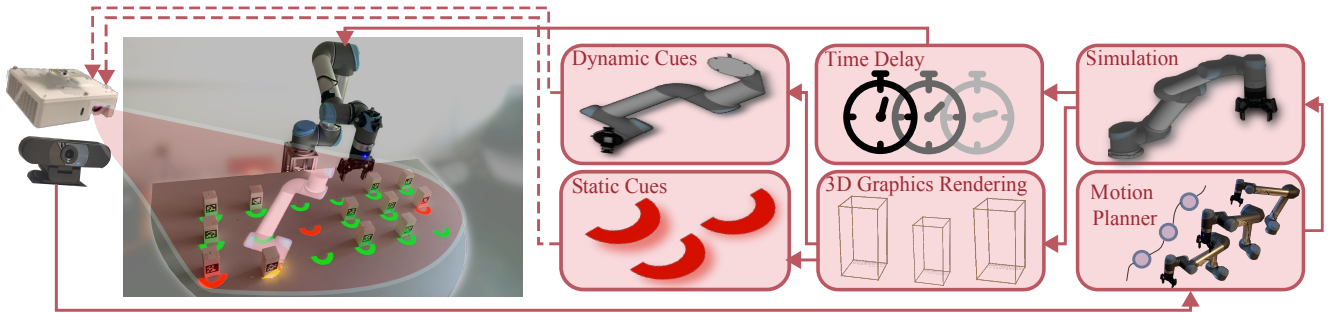
Fig. 2: Overview of the system architecture and experiment setup. Information about the current state of the real world environment is captured using a camera. In turn, a robot motion planner generates the intended following actions. The result is used to produce 3D visualizations of either static visual cues or dynamic motion cues. Finally, generated visual signals are projected into the world using a projection device. In this work, we use the projected signals only for the robot actions. However, the designed mixed reality system can also show other signals such as human goals (green) shown in the left image.

human-robot collaboration tasks. Specifically, we compare static and dynamic cues, as well as combinations thereof, to assess their impact on user performance. To this end, we discuss a collaborative sorting task in which a human user has to anticipate the robot's motion to avoid potential collisions or conflicting sub-goals, e.g., reaching for the same object as the robot. We discuss a dynamic signal in which a simulated digital twin of the robot (projected using mixed reality) performs future actions ahead of the real, physical one. Accordingly, the human user can visually anticipate the upcoming motion of the robot. We contrast this mode with a static signaling type, in which the target object is highlighted using a stationary visual cue. Rather than focusing on the actions of the robot, this mode focuses on the underlying goal object only. By comparing the performance of users in dynamic and static conditions, we aim to understand whether different types of signals affect the human user in different ways. The contributions of this paper can be summarized as follows:

- A mixed reality system for static and dynamic signaling of robot intention. The system features a novel mode for dynamic signaling that leverages a projected digital twin of the robot to preview upcoming actions.
- A human subject experiment focusing on multiple projection modes, along with an extensive analysis of subjective and objective metrics. In addition, information-theoretic approaches are used to numerically quantify the amount of information transferred to the user through visual cues.
- An open-source release of the proposed system along with all necessary components to reproduce the described experiments or investigate other visual cues. link: https://github.com/ir-lab/IntPro.git

II. RELATED WORK

In human-robot collaboration, efficiently communicating a robot's intentions to a human co-worker is a well-known challenge [26]. Inherently, humans are excellent at understanding and communicating to each other through nonverbal cues. However, this ability does not apply when the human tries to predict a robot's motion or trajectories in a collaborative setting. To date, robots lack the skill, physical subtlety, and human-like appearance to provide such nonverbal cues effectively. Thus, substantial research has been devoted to different modalities such as gesture, gaze, and haptic feedback to overcome this communication gap [27]–[30]. All of these modalities have shown promising results in improving human-robot collaboration. However, humans are visual creatures and can often process explicit visual cues faster than implicit or indirect cues. In addition, visual cues can be co-located with the environment or context they refer to. To convey the presence of a dangerous object, for example, we can display a hazard sign in close proximity or on top of it. Virtual or mixed-reality frameworks provide an excellent technological platform for visualizing such cues in an engaging and interactive fashion. For example, a survey and overview of different modes of visualization that can be used in industrial applications can be found in [11]. The work in [31] uses a head-mounted display to show augmented reality signals indicating an indoor drone's path and navigation points. In a similar vein, [32] uses head-mounted displays to depict robot workspace and trajectory information. A technical requirement to achieve this effect is virtual reality headsets or see-through displays. However, as a side effect of this requirement, users may develop fatigue or nausea during the operation of the task. Similarly, such setups may make involving multiple humans in the interaction scenario difficult since one headset per participant is needed. Alternatively, a mixed-reality setup can be used. Specifically, a camera-projection stereo system can be used to accurately project information on 3D surfaces without needing external hardware such headset. The concept of providing the robot intentions via mixed-reality projections framework called intention projection was previously explored in [33]. This work compared different interfaces, such as textual descriptions, monitor displays, and projections in human-robot collaboration tasks. Results show that users find the projection interface to be more reliable and effective. Similarly, [15] discusses how projected patterns can form a rich visual language used in a specific context or domain,

e.g., collaborative assembly. Independently of the specific technical implementation, all of the above papers build upon the same principle – the use of virtual, augmented, or mixed-reality to visually communicate information to a human interaction partner [19], [34], [35]. Unlike previous work, the main focus of our paper is on the types of signals and their effectiveness in transferring information between agents rather than on the technical details of intention projection. The aim is to gain insights into various signaling strategies that can be used in any system. By providing a methodology for studying the effectiveness of different types of visual cues, the paper provides a framework for designing and improving intention projection systems.

III. METHODOLOGY

In this section, we will first describe the task used throughout the paper. Thereafter, we will provide details regarding our intention projection system and two types of visual signals that are representative of a broader class of signals. Furthermore, details about the transfer entropy and its use for finding a causal relationship between the robot and human is explained.

A. Collaborative Task

In order to investigate the effects of visual signals, an object sorting task was designed, see Fig. 1. The task involves sorting eighteen 3D-printed cubes placed on a tabletop surface. Human participants are asked to sort twelve cubes while an autonomous robot (Universal Robot UR5) is assigned six cubes. Sorting involves picking assigned cubes, one at a time, and placing them in designated areas. For participants, this designated area is a second table placed on their left-hand side. By contrast, the robot has to place cubes into boxes to its left and right. Visual signals are used throughout the task to visualize the target objects or the motion of the robot.

B. Intention Projection System

Following the rationale of this paper, visual cues of robot intention are projected into the joint workspace. Fig. 2 depicts these cues and provides a detailed overview of our intention projection system. First, a webcam observes the current scene and tracks the location of physical objects on a table. Tracking is performed using simple fiducial markers [36]. The resulting scene information (object positions and orientations) is sent to a motion planner (developed using [37]) in order to generate valid robot motions to reach the intended next goal. In turn, the plan is simulated, and the resulting data is used to generate visualizations of future states of the system (robot or object). Finally, the generated visual cues are projected into the physical environment using a projection device. For calibration purposes, we use the method explained in [38]: the camera-projector system can be treated as a stereo system consisting of a monocular camera and a projector (inverse camera) by projecting multiple binary patterns on a checkerboard and, in turn, computing intrinsic and extrinsic parameters of the projector in relation to the

Fig. 3: Dynamic signal: A virtual robot (bottom) moves ahead of the real robot resulting in a brief window for visualization of future motions.

camera via homographies. To render different projections, we leveraged the combination of the Unity game engine and OpenCV. Furthermore, to provide seamless communication between the robot and the simulation, we used the Robot Operating System (ROS) Noetic version.

C. Visual Signal Types

As mentioned before, we are interested in contrasting two types of signals, i.e., Static and Dynamic visual cues of robot intent. Static visual signals, depicted as red semi-circular disks, highlight the target cube to be picked next by the robot. As Dynamic visual signals, the projection system displays a continuous animation of the intended robot motion. It is important to note that robot motions are shown *ahead of time* – the user sees a preview of upcoming actions. An example of this Dynamic mode is shown in Fig. 3. A virtual twin of the physical robot is projected onto the table. This virtual robot moves ahead of time and provides a window into the future motion of the real robot. This information allows the user to avoid areas of the shared workspace that are soon to be inhabited by the robot. In addition, this information provides an early indication of the object (or group of objects) that will likely be the target. The temporal offset, or delay, between the virtual and real robots, is an adjustable parameter. These two visual signal modes allow four distinct mode combinations:

1) **No-Projection Mode**: This mode provides no visual cues and merely displays information about assigned cubes to the human subject on a laptop screen in the form of a grid with green colored squares. The six cubes assigned to the robot are shown in white.
2) **Static Mode**: A visual cue showing a semi-circular disc is projected for one second to indicate the next target cube (out of six possible cubes). This visual information is intended to provide the human partner with information about the robot's next objective.
3) **Dynamic Mode**: A real-time rendered animation of a virtual robot is projected onto the table. This rendered digital twin previews the motion of the physical robot before they occur. A time delay of one second between the virtual and real robot arm is used.
4) **Dual Mode**: A hybrid mode combining visual cues from both Static and Dynamic Mode.

In our study, we utilized a 3×6 grid displayed on a screen to depict the location of 18 cubes on a table, as shown in Fig. 1. The grid comprises twelve green squares and six white squares that are randomly distributed across the grid. Users are tasked with picking up cubes located on the green squares, while the robot is assigned to collect cubes located on the white squares. The Static, Dynamic, or Dual Mode may provide users with visual cues about the robot's target objects.

D. Measuring Information Transfer and Causality

A critical question for identifying efficient signals for intention projection is how to measure their influence on human behavior. How much information transfer is there between the robot projecting its intent using a specific visual signal and the interacting human partner? How can this be quantified in an objective manner?

To this end, we employ a formal, information-theoretic approach to describe information transfer. More specifically, we calculate the Transfer Entropy (TE) [39] between the sender (robot) and receiver (human). In this formulation, the visual signals form a communication channel between the two partners. TE measures the directed transfer of information between two processes and is widely used for inferring causal relationships between observed processes [40]. We can calculate it as:

$$\text{TE}_{X \to Y} = H(Y_t | Y_{t-1:t-K}) - H(Y_t | Y_{t-1:t-K}, X_{t-1:t-K})$$

where X and Y are the source and target time series. In our specific case, X corresponds to information about the robot projections (when did the robot project), whereas Y corresponds to observations of the human behavior (when did the human pick an object). H indicates the Shannon entropy while K is a history length or past observations of the source time series. For best practices on how to set an optimal value of K we refer the reader to [41]. In this case, we set the parameter to the average time window between two human pick events, i.e., $K = 9$. For small sample sizes, the transfer entropy estimates are known to be biased [42]. To correct for bias, we use a specific variant of TE called Effective Transfer Entropy ($\widehat{\text{TE}}$) [42]:

$$\widehat{\text{TE}}_{X \to Y} = \text{TE}_{X \to Y} - \frac{1}{Z} \sum_{1}^{Z} \text{TE}_{X' \to Y}$$

where $Z = 100$ as proposed in [41] and X' is the shuffled source time series. TE can be calculated in a data-driven fashion, i.e., by performing an experiment and collecting data about the timing of when certain visual signals were projected as well as data about when the human performed a certain action (e.g. a pick or lifting action).

IV. EXPERIMENTS AND RESULTS

To compare the efficiency and impact of different visual cues, an Institutional Review Board (IRB)[1]-approved human

[1]This study was approved by Arizona State University (#STUDY00016445)

subject study was conducted with 22 subjects between the ages of 18-28. All participants voluntarily agreed to participate in the experiment, which was advertised as a sorting game with a robot in a flyer. They were not given any form of compensation for their participation. Additionally, participants were not provided with any information regarding the technical aspects or analysis of the experiment, except for the logistics of how the experiment would be conducted. Participants were asked to engage in the cube sorting task described in Section III-A. Throughout the task, the human-robot team has to sort all cubes simultaneously. Hence, the coordination of actions is critical for safety and efficiency. Since the study is conducted in a within-subjects (or repeated measures) manner, the order in which the four modes were introduced to each participant was randomized, making sure any potential bias or influence due to familiarity with the task can be minimized, allowing for a more accurate assessment of the effects of each mode on task execution. In addition, the robot speed was set to variable safe speeds sampled from a Gaussian distribution with $\mu = 0.44$, $\sigma = 0.35$ (m/s).

An overall underlying question in our experiments is whether the modes introduced above for visual signaling provide different and distinct degrees of improvement in the specific human-robot collaboration task discussed here. More specifically, the following hypotheses are investigated in this experiment:

- **H1**: At least one projection mode enhances task efficiency compared to the No-projection mode.
- **H2**: Cognitive load indices are substantially lower in projection modes compared to the No-projection mode.

To provide evidence for or against the above hypotheses, we combined both subjective and objective analyses. Participants were asked to fill in the NASA-TLX (Task Load Index) questionnaire [43] after experimenting with each mode, which comprises six sub-scales (*Mental Demand, Physical Demand, Temporal Demand, Performance, Effort,* and *Frustration*) rated on a 0-20-point scale. However, for visualization purposes these scores were normalized between 0-100. To evaluate data distribution, subjective and objective measures were analyzed with the Friedman test, as described in [44]. In particular, the Friedman test was used to determine if there were any significant differences in the distribution of data across the four modes. Subsequently, a Wilcoxon signed rank test, as described in [45], was conducted to identify any statistically significant differences between the four modes. The estimated p values were adjusted with a Bonferroni correction to prevent Type-I error. These analyses are discussed in detail respectively in Sec. IV-A and Sec. IV-B.

A. Hypothesis 1: Task Efficiency

To investigate task efficiency, we analyze the relative finish time of the robot with respect to the human interaction partner across all users and modes. Relative finish time was used to gauge human-robot collaboration effectiveness. This decision was influenced by its capability to account for varying task conditions like robot speed fluctuations and visual

TABLE I: Results on objective (Relative Finish Time) and subjective (NASA-TLX) measures using Wilcoxon signed rank test on pairs of proposed modes as their p value. Here, the (+) and (-) notation results from whether the difference between the mean values of the compared modes (e.g. for Du vs St difference is calculated as μ(Du)-μ(St) for objective or subjective measures) is positive or negative. [Du = Dual, St = Static, Dy = Dynamic and No = No-Projection], [Statistically Significant, Not Statistically Significant]

| | **Objective Measure** (p value) | **Subjective Measures** (p value) | | | | | |
	Relative Finish Time	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration
Du vs St	(+) 0.1907	(-) 0.0062	(-) 0.4380	(-) 0.0721	(-) 0.1154	(-) 0.2314	(-) 0.1157
Du vs Dy	(+) 0.2760	(-) 0.6163	(-) 0.4669	(-) 0.1441	(-) 0.3740	(-) 0.3232	(-) 0.2467
Du vs No	(-) 0.0101	(-) 0.0007	(-) 0.1350	(-) 0.0039	(-) 0.1081	(-) 0.0135	(-) 0.0191
St vs Dy	(+) 0.4120	(+) 0.3591	(+) 0.4577	(+) 0.5701	(+) 0.8885	(+) 0.4308	(+) 0.6182
St vs No	(-) 0.0481	(-) 0.0551	(-) 0.3104	(-) 0.3464	(-) 0.3130	(-) 0.3307	(-) 0.0770
Dy vs No	(+) 0.2029	(-) 0.0754	(-) 0.3217	(-) 0.2044	(-) 0.9629	(-) 0.0104	(-) 0.0254

signal changes. It allowed to quantitatively determine which visual signals significantly enhanced human task efficiency in the collaborative scenario. We calculate the relative finish time $\Delta t = (t_r - t_h)$ where t_r is the robot finish time, and t_h is the human finish time. A positive value for Δt indicates that the participant finished the sorting task before the robot. A negative value of Δt indicates that the human was slower than the robot at sorting. Fig. 4 shows the distributions of relative finish times across all modes, and after passing the Friedman test ($p < 0.05$) over all four modes, the corresponding p values of paired Wilcoxon signed ranked tests with regards to Δt were calculated and shown in Table I. Based on the results presented in Fig. 4, it can be observed that the mean Δt is higher for the Dual mode (-1.2 sec.) when compared to the No-projection mode (-7.1 sec.). This result has been statistically validated in the significance test presented in Table I ($p < 0.05$). Conversely, the difference in Δt between the Static and Dynamic modes was found to be insignificant ($p > 0.05$), despite the mean value of Δt being higher in the Static mode than in the Dynamic mode. Additionally, the Static mode exhibited a significantly faster Δt on average (56 %) compared to the No-projection mode ($p < 0.05$).

These findings support hypothesis **H1**: both the Static and Dual modes show a significant difference when compared to the No-projection mode. Significant improvements in task efficiency can be observed – especially in the case of the Dual mode. Compared to that, the Dynamic mode shows no significant difference from the No-projection mode.

B. Hypothesis 2: Cognitive Load

To address the second hypothesis, we focus on the subjective feedback provided by users in the form of NASA TLX scores. Fig. 5 shows a summary of mean scores across all modes and workload indices (lower values correspond to better subjective responses). Looking at Fig. 5, we notice that all projection modes consistently produce better scores when compared to the No-projection mode. The Dual mode (combining visual cues) substantially decreases all indices, e.g., frustration is reduced by 48%, the mental demand is reduced by 37% and effort sees a 39% reduction compared to No-projection mode. Furthermore, all cognitive load indices across all four modes show significant differences in the
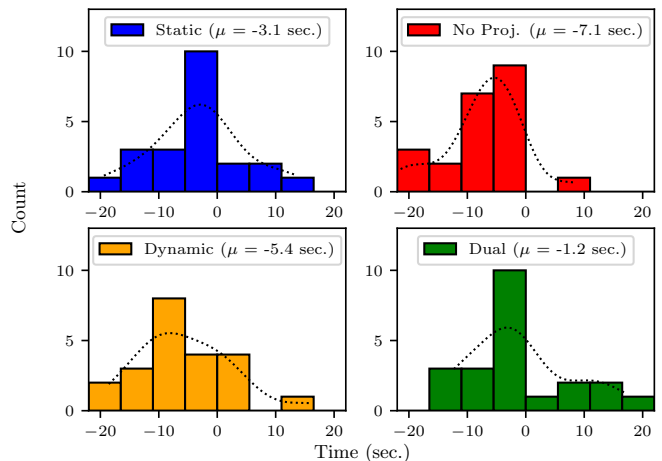

Fig. 4: Histogram of Δt between robot vs human finish times. A positive Δt indicates that the subject finished earlier than the robot. In other words, the efficiency was higher. From this figure, there is an explicit lean toward the right in Dual mode, indicating a higher efficiency than other modes.

subjective score as verified by the Friedman test ($p < 0.05$). Table I provides additional details regarding the pairwise significance of the four modes across six different cognitive load indices. With regards to *Mental Demand*, we find that the Dual mode results in significantly better scores (highlighted in green) when compared to the Static projection and No-projection modes ($p < 0.05$). However, when looking at the difference between the mean values of compared modes, we note that both the Dual and Dynamic modes performed better than the Static and No-projection modes, as indicated in Table I. Interestingly, the Dynamic mode showed better scores than the Static mode. Contrary to what we observed for relative finish times in Sec. IV-A, the Dynamic mode seems to provide marginal improvements in reducing cognitive load.

On the other hand, cognitive load indices such as *Physical Demand* and *Performance* did not see any statistically significant change in scores ($p > 0.05$). However, given that the sorting task remained the same across all four modes, with the only variation being the information feedback in the form of visual signals to human subjects, it is reasonable

to assume that their *Performance* and *Physical Demand* would remain similar across all four modes. Regarding *Effort* and *Frustration*, we find that both the Dynamic and Dual modes have statistically significant scores (highlighted in green) when compared to the No-projection mode ($p < 0.05$). Furthermore, although the Static mode did not demonstrate a significant improvement in terms of frustration and effort when compared to the No-projection mode, the differences in their mean scores indicate that human subjects still preferred the Static mode over the No-projection mode. Finally, *Temporal Demand* shows that Dual mode has a statistically significant (lower) score when compared to No-projection mode, which means that human subjects were not hurried or rushed by the robot's actions and were able to finish tasks with ease. Nonetheless, the Dual mode is again slightly better than the Static mode, i.e., differences in mean and close to significant p value in the first row of *Temporal Demand* column in Table I.

In summary, our hypothesis **H2** is partially supported by the evidence: both the Dual and Dynamic modes see significant or close to significant reductions in load indices with respect to *Mental Demand, Temporal Load, Effort*, and *Frustration*. With regard to these cognitive load indices, the Dynamic mode marginally outperforms the Static mode.
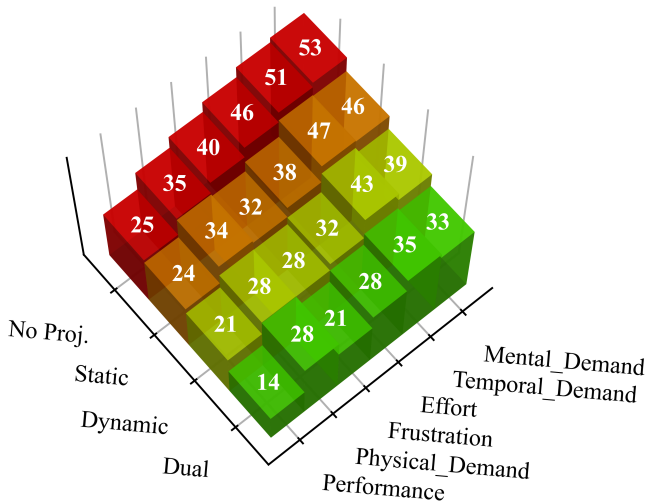

Fig. 5: Visualization of all subjective mental workload indices across all the modes.

C. Transfer Entropy Analysis

The insights and conclusions drawn above are based on objective metrics (relative finish time) and subjective human feedback (cognitive load indices). In this section, we investigate if similar insights can be drawn without any insight into the task, i.e., purely from the motion data of both humans and the robot.

More specifically, we use a (task-agnostic) information-theoretic metric to analyze recorded data during experiments and evaluate whether the findings corroborate the results observed in Sec. IV-A and Sec. IV-B. We conducted a posthoc analysis using Transfer Entropy [39], [42] and videos

collected during the experiment. We manually annotated time stamps in which one of the following events/actions occur: human pick (hp), robot stop (rs), virtual robot stop (vs), and Static goal (sg). Events (hp) indicate timestamps at which the human picked a cube. Similarly, (rs) indicates timestamps wherein the real robot stops moving to pick up an object. Events (vs) indicate timestamps in which the *virtual* robot stopped moving, while the event (sg) shows the time stamp at which the Static goal was projected. These four actions result in four separate time series τ_a with binary values of 0 or 1, depending on whether the respective action occurred or not. For Static and No-projection mode, (vs) will be a time series of zero values. We define each participant as P_n where $n \in \{1, 2, ..., N\}$. For each participant, we define time series for four actions (a) per mode:

$$\tau_a(m) = [v_1, v_2, ..., v_T] \text{ where } a \in \{\text{hp, rs, vs, sg}\},$$

with $m \in \{\text{No-Projection, Dynamic, Static, Dual}\}$ and T being the number of time steps. The variables $v_1, v_2, ..., v_t$ are binary variables indicating whether the action a occurs at the current timestamp (t).

TABLE II: Average Effective Transfer Entropy Across Different Modes (m) for annotated time series.

	No Proj.	Static	Dynamic	Dual
$\widehat{\text{TE}}_{[\text{vs}] \rightarrow [\text{hp}]}$	0.0	0.0	**0.00864755**	**0.0148222**
$\widehat{\text{TE}}_{[\text{sg}] \rightarrow [\text{hp}]}$	0.0	**0.01296397**	0.0	0.00418501
$\widehat{\text{TE}}_{[\text{rs}] \rightarrow [\text{hp}]}$	0.00968205	0.00919081	0.00484169	0.01176273

Based on the above time series we compute the effective Transfer Entropy for different modes as seen in Table II. Here, $\widehat{\text{TE}}_{[\text{vs}] \rightarrow [\text{hp}]}$ signifies the information transfer between the virtual robot stopping and the human picking an object, since this visual cue is only used in the Dynamic and Dual modes, we only observe an information transfer in either one of these modes. $\widehat{\text{TE}}_{[\text{sg}] \rightarrow [\text{hp}]}$ shows information transfer between the static goal (signal) and human picking action and final $\widehat{\text{TE}}_{[\text{rs}] \rightarrow [\text{hp}]}$ shows information transfer between physical robot and human picking. The $\widehat{\text{TE}}^2$ of about 0.0086 for the virtual robot in the Dynamic mode is higher than the real robot in the No-projection mode with $\widehat{\text{TE}} = 0.0091$. Moreover, an even higher value of 0.0129 can be achieved when projecting a static signal. The highest overall value for the effective Transfer Entropy of 0.0148 is observed in the Dual mode – which again emphasizes the power of projecting multiple visual cues in conjunction.

The $\widehat{\text{TE}}$ between the real robot and human picks never exceeds 0.011, i.e., robot actions influence human behavior to a lesser degree than the visual signals. Generally, it can be noticed that the information transfer is highest between visual signals and human actions. The overall trend identified via Transfer Entropy mirrors the similar trends found in the earlier analysis of relative finish times and subjective

---

[2]General notation for Effective Transfer Entropy. Includes all the different combinations of source and target time series show in Table II.

mental workload assessment: the Dual mode shows the best performance, followed by the individual projection modes (Static and Dynamic), and finally, the No-projection mode.

V. Discussion and Limitations

Contrary to previous assumptions, this paper found that projecting combination of visual cues during human-robot interactions significantly improves collaborative task performance. Objective, subjective, and information-theoretic metrics all support this conclusion. Moreover, carefully designing static and dynamic visual cues can enhance collaborative tasks. Nevertheless, further research is necessary to explore the boundaries of this finding. For instance, it remains to be seen whether there is an upper bound to the effective design and combination of these visual cues that, if exceeded, might hamper collaborative task efficiency. One potential avenue for future research is to investigate whether the visual signals convey different aspects of the task. For the current study, a possible rationalization about the design aspect of signals is that the static signal communicates the robot's intended destination. In contrast, the dynamic signal conveys how the robot will reach that location. Additionally, while the time delay between virtual and real robot trajectories was constant in this study, future investigations could explore the impact of variable time delays on human-robot collaboration.

The results of our experiment showed a clear distinction between Dual and No-projection mode. However, there was no significant distinction between dynamic and static signals. We suggest conducting a more careful investigation of the experiment dimension to better understand the extent of the role of visual signals in human-robot interactions. Specifically, further research should focus on studying the type and appearance of visual signals to confirm their vital role in affecting how human subjects interact with the robot system.

Finally, our investigation on Transfer Entropy indicated that information-theoretic measures are able to provide early indications regarding the amount of information transferred by a visual cue to human users. Nevertheless, more in-depth studies are required to confirm the effectiveness of Transfer Entropy on a broader setup in the future.

References

[1] H. I. Christensen, T. Batzinger, K. Bekris, K. Bohringer, J. Bordogna, G. Bradski, O. Brock, J. Burnstein, T. Fuhlbrigge, and R. Eastman, "A roadmap for us robotics: from internet to robotics," *Computing Community Consortium and Computing Research Association, Washington DC (US)*, 2009.

[2] A. D. Dragan, K. C. Lee, and S. S. Srinivasa, "Legibility and predictability of robot motion," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2013, pp. 301–308.

[3] Y. Zhang, S. Sreedharan, A. Kulkarni, T. Chakraborti, H. H. Zhuo, and S. Kambhampati, "Plan explicability and predictability for robot task planning," in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 1313–1320.

[4] Z. Han, E. Phillips, and H. A. Yanco, "The need for verbal robot explanations and how people would like a robot to explain itself," *J. Hum.-Robot Interact.*, vol. 10, no. 4, sep 2021.

[5] B. Mutlu, N. Roy, and S. Šabanović, "Cognitive human–robot interaction," *Springer handbook of robotics*, pp. 1907–1934, 2016.

[6] S. M. Fiore, T. J. Wiltshire, E. J. Lobato, F. G. Jentsch, W. H. Huang, and B. Axelrod, "Toward understanding social cues and signals in human–robot interaction: effects of robot gaze and proxemic behavior," *Frontiers in psychology*, vol. 4, p. 859, 2013.

[7] J. Dumora, F. Geffard, C. Bidard, T. Brouillet, and P. Fraisse, "Experimental study on haptic communication of a human in a shared human-robot collaborative task," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 5137–5144.

[8] S. Lackey, D. Barber, L. Reinerman, N. I. Badler, and I. Hudson, "Defining next-generation multi-modal communication in human robot interaction," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 55, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2011, pp. 461–464.

[9] G. C. Burdea and P. Coiffet, *Virtual reality technology*. John Wiley & Sons, 2003.

[10] M. J. Schuemie, P. Van Der Straaten, M. Krijn, and C. A. Van Der Mast, "Research on presence in virtual reality: A survey," *CyberPsychology & Behavior*, vol. 4, no. 2, pp. 183–201, 2001.

[11] G. d. M. Costa, M. R. Petry, and A. P. Moreira, "Augmented reality for human–robot collaboration and cooperation in industrial applications: A systematic literature review," *Sensors*, vol. 22, no. 7, p. 2725, 2022.

[12] E. Costanza, A. Kunz, and M. Fjeld, "Mixed reality: A survey," in *Human machine interaction*. Springer, 2009, pp. 47–68.

[13] S. Rokhsaritalemi, A. Sadeghi-Niaraki, and S.-M. Choi, "A review on mixed reality: Current trends, challenges and prospects," *Applied Sciences*, vol. 10, no. 2, p. 636, 2020.

[14] C. E. Hughes, C. B. Stapleton, D. E. Hughes, and E. M. Smith, "Mixed reality in education, entertainment, and training," *IEEE computer graphics and applications*, vol. 25, no. 6, pp. 24–30, 2005.

[15] R. K. Ganesan, Y. K. Rathore, H. M. Ross, and H. B. Amor, "Better teaming through visual cues: How projecting imagery in a workspace can improve human-robot collaboration," *IEEE Robotics and Automation Magazine*, vol. 25, pp. 59–71, 6 2018.

[16] M. Dianatfar, J. Latokartano, and M. Lanz, "Review on existing vr/ar solutions in human–robot collaboration," *Procedia CIRP*, vol. 97, pp. 407–411, 2021.

[17] E. Sibirtseva, D. Kontogiorgos, O. Nykvist, H. Karaoguz, I. Leite, J. Gustafson, and D. Kragic, "A comparison of visualisation methods for disambiguating verbal requests in human-robot interaction," in *2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 2018, pp. 43–50.

[18] M. Ostanin, S. Mikhel, A. Evlampiev, V. Skvortsova, and A. Klimchik, "Human-robot interaction for robotic manipulator programming in mixed reality," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2805–2811.

[19] T. Williams, D. Szafir, T. Chakraborti, and H. Ben Amor, "Virtual, augmented, and mixed reality for human-robot interaction," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 2018, pp. 403–404.

[20] R. T. Chadalavada, H. Andreasson, M. Schindler, R. Palm, and A. J. Lilienthal, "Bi-directional navigation intent communication using spatial augmented reality and eye-tracking glasses for improved safety in human–robot interaction," *Robotics and Computer-Integrated Manufacturing*, vol. 61, p. 101830, 2020.

[21] O. Liu, D. Rakita, B. Mutlu, and M. Gleicher, "Understanding human-robot interaction in virtual reality," in *2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 2017, pp. 751–757.

[22] T. Taniguchi, T. Nagai, T. Nakamura, N. Iwahashi, T. Ogata, and H. Asoh, "Symbol emergence in robotics: a survey," *Advanced Robotics*, vol. 30, no. 11-12, pp. 706–728, 2016.

[23] S. Coradeschi, A. Loutfi, and B. Wrede, "A short review of symbol grounding in robotic and intelligent systems," *KI-Künstliche Intelligenz*, vol. 27, no. 2, pp. 129–136, 2013.

[24] J. Katona, "A review of human–computer interaction and virtual reality research fields in cognitive infocommunications," *Applied Sciences*, vol. 11, no. 6, p. 2646, 2021.

[25] R. J. Holden, E. Abebe, J. R. Hill, J. Brown, A. Savoy, S. Voida, J. F. Jones, and A. Kulanthaivel, "Human factors engineering and human-computer interaction: supporting user performance and experience," *Clinical informatics study guide*, pp. 119–132, 2022.

[26] Z. Gong and Y. Zhang, "Behavior explanation as intention signaling in human-robot teaming," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2018, pp. 1005–1011.

[27] P. Neto, M. Simão, N. Mendes, and M. Safeea, "Gesture-based human-robot interaction for human assistance in manufacturing," *The International Journal of Advanced Manufacturing Technology*, vol. 101, no. 1, pp. 119–135, 2019.

[28] S. Waldherr, R. Romero, and S. Thrun, "A gesture based interface for human-robot interaction," *Autonomous Robots*, vol. 9, no. 2, pp. 151–173, 2000.

[29] M. A. Cabrera, J. Heredia, J. Tirado, V. Panov, F. Ragos, and D. Tsetserukou, "Cohaptics: Development of human-robot collaborative system with forearm-worn haptic display to increase safety in future factories," in *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2021, pp. 74–80.

[30] C. Brosque, E. G. Herrero, Y. Chen, R. Joshi, O. Khatib, and M. Fischer, "Collaborativewelding and joint sealing robots with haptic feedback," in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, vol. 38. IAARC Publications, 2021, pp. 1–8.

[31] M. Walker, H. Hedayati, J. Lee, and D. Szafir, "Communicating robot motion intent with augmented reality," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 2018, pp. 316–324.

[32] E. Rosen, D. Whitney, E. Phillips, G. Chien, J. Tompkin, G. Konidaris, and S. Tellex, "Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays," *The International Journal of Robotics Research*, vol. 38, no. 12-13, pp. 1513–1526, 2019.

[33] R. S. Andersen, O. Madsen, T. B. Moeslund, and H. B. Amor, "Projecting robot intentions into human environments," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2016, pp. 294–301.

[34] T. Williams, M. Bussing, S. Cabrol, E. Boyle, and N. Tran, "Mixed reality deictic gesture for multi-modal robot communication," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 191–201.

[35] J. Hamilton, T. Phung, N. Tran, and T. Williams, "What's the point? tradeoffs between effectiveness and social perception when using mixed reality to enhance gesturally limited robots," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 2021, pp. 177–186.

[36] J. Wang and E. Olson, "Apriltag 2: Efficient and robust fiducial detection," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 4193–4198.

[37] H. Bruyninckx, "Open robot control software: the orocos project," in *Proceedings 2001 ICRA. IEEE international conference on robotics and automation (Cat. No. 01CH37164)*, vol. 3. IEEE, 2001, pp. 2523–2528.

[38] D. Moreno and G. Taubin, "Simple, accurate, and robust projector-camera calibration," *Proceedings - 2nd Joint 3DIM/3DPVT Conference: 3D Imaging, Modeling, Processing, Visualization and Transmission, 3DIMPVT 2012*, pp. 464–471, 2012.

[39] T. Schreiber, "Measuring information transfer," *Physical review letters*, vol. 85, no. 2, p. 461, 2000.

[40] M. Staniek and K. Lehnertz, "Symbolic transfer entropy," *Physical review letters*, vol. 100, no. 15, p. 158101, 2008.

[41] T. Bossomaier, L. Barnett, M. Harré, and J. T. Lizier, "Transfer entropy," in *An introduction to transfer entropy*. Springer, 2016, pp. 65–95.

[42] R. Marschinski and H. Kantz, "Analysing the information flow between financial time series," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 30, no. 2, pp. 275–281, 2002.

[43] M. Feick, N. Kleer, A. Tang, and A. Krüger, "The virtual reality questionnaire toolkit," in *Adjunct Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, 2020, pp. 68–69.

[44] F. De Pace, G. Gorjup, H. Bai, A. Sanna, M. Liarokapis, and M. Billinghurst, "Leveraging enhanced virtual reality methods and environments for efficient, intuitive, and immersive teleoperation of robots," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 12 967–12 973.

[45] M. Hirschmanner, C. Tsiourti, T. Patten, and M. Vincze, "Virtual reality teleoperation of a humanoid robot using markerless human upper body pose imitation," in *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2019, pp. 259–265.