

# Towards Explainable Shared Control using Augmented Reality

Mark Zolotas and Yiannis Demiris

**Abstract**—Shared control plays a pivotal role in establishing effective human-robot interactions. Traditional control-sharing methods strive to complement a human’s capabilities at safely completing a task, and thereby rely on users forming a mental model of the expected robot behaviour. However, these methods can often bewilder or frustrate users whenever their actions do not elicit the intended system response, forming a misalignment between the respective internal models of the robot and human. To resolve this model misalignment, we introduce Explainable Shared Control as a paradigm in which assistance and information feedback are jointly considered. Augmented reality is presented as an integral component of this paradigm, by visually unveiling the robot’s inner workings to human operators. Explainable Shared Control is instantiated and tested for assistive navigation in a setup involving a robotic wheelchair and a Microsoft HoloLens with add-on eye tracking. Experimental results indicate that the introduced paradigm facilitates transparent assistance by improving recovery times from adverse events associated with model misalignment.

## I. INTRODUCTION

Shared control is an auspicious trend in the domain of human-robot interaction, especially for applications in assistive robotics. By establishing joint collaboration between a human operator and an automated controller in exerting control over a system, a robot can take into account the operator’s initiatives and administer “conditional assistance” [1]. For instance, robotic support may be supplied on an as-needed basis, such that a user’s authority is respected and their proficiency at completing a task continues to develop.

Although typical features of shared control, such as obstacle avoidance, are often included as a form of robotic support, they may instead obstruct and frustrate an individual whenever their actions do not elicit the intended system response [2]. This misalignment between a person’s mental model of the expected system behaviour and the robot’s internal model can lead to people rejecting the assistance. A further lack of consensus about the guidelines for developing control-sharing systems only exacerbates the issue [3].

To resolve the complications of model misalignment, we introduce Explainable Shared Control (XSC). Traditional processes of shared control, namely intention estimation and arbitration [4], [5], are commonly evaluated based on how well the output control commands align with the user’s actual task intent. In the proposed paradigm, we examine these processes from an additional perspective: *transparency*,

M. Zolotas and Y. Demiris are with the Personal Robotics Lab, Dept. of Electrical and Electronic Engineering, Imperial College London, SW7 2BT, UK; Email: {mark.zolotas12, y.demiris}@imperial.ac.uk

This research was supported in part by an EPSRC Doctoral Training Award to MZ, and a Royal Academy of Engineering Chair in Emerging Technologies to YD.

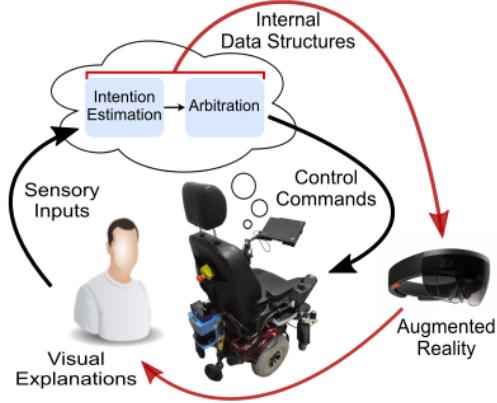


Fig. 1. Overview of Explainable Shared Control. Standard shared control will involve an estimation of the user’s task intent based on sensory inputs, followed by an arbitration of control to output assistive commands. In the explainable paradigm, the data structures of internal mechanisms are represented in augmented reality and communicated back to users as visual explanations about any resulting robot behaviours (indicated by red arrows).

where the objective is to best represent the underlying robot reasoning and feed it back to the user.

Coinciding with this objective, a suitable medium of communication is required to relay back information to the user. In many instances of shared control, the haptic channel is the selected modality of sensory feedback [5], [6]. However, augmented reality (AR) and head-mounted displays (HMDs) are an increasingly prevalent bridge of communication in human-robot interaction [7], [8]. These technologies offer a range of capabilities suited for creating embodied interfaces and granting “explanation as visualisation” [9].

By proposing a means of achieving XSC via AR, we aspire to settle the model mismatch between interacting humans and robots (overview shown in Fig. 1). Therefore, the main contributions of this paper are: 1) to outline XSC and then instantiate it with an AR-HMD interface for assistive robot navigation; 2) to discern the paradigm’s benefits in a setup involving a robotic wheelchair and a Microsoft HoloLens supplemented with eye tracking capabilities. This work extends our earlier AR-HMD wheelchair platform [10] by adopting a new shared control and AR system, as well as by conducting a user study tailored to model misalignment.

## II. RELATED WORK

Model misalignment or “model reconciliation” [11] is a well-known problem that has been explored in a related

field of research, known as explainable planning [12]. In this domain, the model reconciliation problem refers to the use of “explanations” or model updates [11] to resolve differences in a human’s expectations of a robot’s plan. As a result, the aim behind generating explanations is to modify the human’s model of the world to agree with the robot’s model.

Intelligent systems within explainable planning tackle model reconciliation by adopting three qualities: trust, interaction and transparency [12], all of which can be mapped to shared control. For example, *trust* has been identified in human-robot interaction as a property mostly influenced by the robot’s performance [13]. Likewise, *interaction* and *transparency* are comparable to how shared control systems continuously operate in a perception-action cycle and communicate their internal mechanisms, respectively [3]. Improving any of these qualities in the control-sharing can thus hope to mitigate model misalignment.

Nevertheless, transparency continues to elude most existing shared control systems [2]. One potential reason for this is the growing use of deep learning and other “black-box” approaches to intention estimation [14] and arbitration [15], which are far from explainable to end-users [12]. Another cause for concern in attaining transparency is the choice of sensory feedback and its capacity to convey information. Despite the pervasiveness of haptic technologies in shared control [3], [5], [6], we advocate for AR as a more promising means of revealing the latent robot decision-making processes via *visualisation* [9].

Dating back to the earliest applications of AR in robotics, there have been reports on its efficacy at guiding the control of human operators [16]–[18]. In many human-robot collaborations that require a human to teleoperate a robot’s end-effectors, AR serves as a communication channel between the two agents [19]. By overlaying the robot’s perspective onto the operator’s view and displaying the predicted effects of interacting with the surroundings, the user is capable of executing accurate remote control [17], [19].

In spite of the potential of AR as a medium of exchange, explicating the hidden rationale of a robot is not a trivial task. Motion inference suffers when the appearance of a robot lacks anthropomorphic features and attempts to transmit its intention back to a user [7]. Furthermore, the model misalignment problem can be worsened by ineffective visualisations, such as those that do not account for the influence of depth perception [20] or utilise any conformal graphics [21]. We seek to resolve these challenges with a pertinent AR-HMD interface that guides users into building a more precise mental model of the robot rationale.

### III. EXPLAINABLE SHARED CONTROL

In this section, we outline XSC and how a head-mounted AR system is required to demystify the underlying shared control by closing the feedback loop. The paradigm is then situated in an assistive robot navigation setting as a concrete example of one of its many possible instantiations (final platform is shown in Fig. 2).

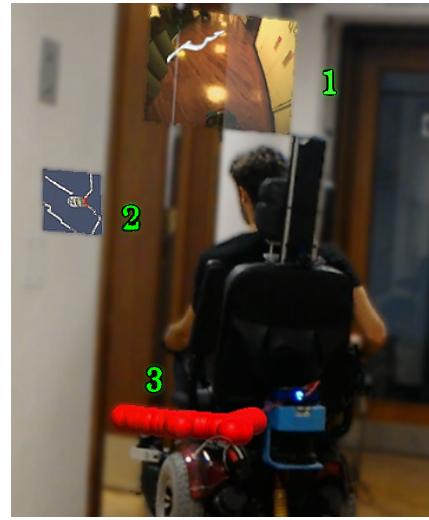


Fig. 2. Composite image of the visualisations rendered during assistive robot navigation. The rear view display (1) overlays virtual objects onto an image taken from a camera situated on the back of the seat. The mini-map panel (2) depicts a bird’s eye view of the wheelchair configuration and its surroundings. Finally, red spheres (3) are placed atop of real-world referents to highlight collisions.

#### A. Transparency in Shared Control

XSC refers to a novel paradigm where the shared control simultaneously complements the abilities of a human operator and facilitates rich information exchange. There is a plethora of research on how to complement an operator’s abilities and instill trust in the interaction [13]. Nonetheless, the literature on making processes of shared control interpretable is sparse. In turn, the focus of XSC is to expose any latent mechanisms in AR using “explanation as visualisation” [9] and comply with standard axioms of shared control [3]. The following guidelines on how to realise XSC are from the perspective of developing internal models, as well as constructing a head-mounted AR interface.

Firstly, minimising conflict in the human’s understanding of robot behaviours (Axiom 1 [3]) should require the shared control to exhibit *causality* [12]. In other words, the internal mechanisms must be able to draw connections between inputs of the world (*e.g.* robot state, user commands, sensor readings) and the individual stages involved in generating the final output commands. Model-based intention estimation [4] and arbitration algorithms make this an easier endeavour [12]. Effectively, this helps answer user questions, such as: *why* and *how* did the robot perform that action?

Answering the *why* and *how*, an AR interface guideline of XSC is to represent causality by designing *contextual* visualisations. Contextual visual aids are those that achieve state summarisation of the robot’s perceived environment and any immediate action-effect relationships (*e.g.* an occupancy grid constructed from LiDAR data, or end effector kinematics resulting from actuation). However, the active role of a participant in shared control – as opposed to a passive observer – compels their continuous engagement with the task-at-hand. For such scenarios, we advocate for the visual

context to be represented using “embodied” cues [7]. These cues are generated atop of the robot morphology as virtual extensions, *e.g.* a radar attached to a mobile base that depicts the constructed occupancy grid, or arrow vectors originating from an arm’s end effector to illustrate planned motion [7].

Externalising the “brain” of shared control is another step towards transparency, which demands that robot reasoning about the task and system design is made explicit (Axiom 2 [3]) through high-level *abstraction* [9]. Simply representing raw data streams of the system’s inputs or outputs will not suffice, as it risks overloading users with information that is already observable. Conversely, a trace or trajectory that highlights semantic task-specific characteristics and shows the provenance of information captured in the environment is better for visual portrayal [8]. A critical question answered here is: *when* does the robot decide to intervene?

Hence, another AR guideline of XSC is to develop *predictive* visualisations that capture the reasoning behind *when* shared control intervenes. Predictive visualisations are those that possess a temporal element and inform the high-level planning of users. A limitation of the aforementioned contextual visual aids is that they display only a snapshot of the current state, which is unlikely to explain when the robot might intervene and can cause even greater misunderstandings [10]. This is particularly problematic for shared control settings, where the active involvement of the user calls for advance planning. By presenting visual traces of the historic or future world states (*e.g.* the evolution of a virtual robot’s arm motion trajectory), users will be equipped with the necessary information to act preemptively.

Viewing “explanation” as a process of visualisation [11], immersive HMDs are an integral component of the paradigm. Whilst most shared control systems achieve information exchange via force feedback, haptic interfaces are limited at explainability in complex task settings unless combined with a visual modality for a multimodal approach [5]. Given the rich visual feedback requirements of XSC, we stipulate that embodied interfaces in AR offer a superior medium of unveiling the robot’s inner workings. In particular, AR-HMDs have the potential to better circumvent model mismatch during an active collaboration like shared control, as they provide users with an increase sense of presence and engagement over monitors or projectors [22], [23].

Acknowledging that the presented outlook on XSC can be tackled in a multitude of ways, a concrete instance of its application in assistive robot navigation is now detailed.

### B. Assistive Robot Navigation

Inspired by the motivation to ensure that severely disabled individuals can exercise mobility in a transparent manner [2], we employ XSC for assistive robot navigation, specifically robotic wheelchairs. There are two principal functions of the navigational assistance: trajectory generation and an arbitration phase to output safe control commands.

The trajectory generation process initially takes place to estimate user navigational intent. Given that the wheelchair’s mobile base follows differential-drive kinematics [24], its

motion is constrained by:

$$-\dot{x} \sin \theta + \dot{y} \cos \theta = 0 \quad (1)$$

Where  $(x, y, \theta)$  represents the robot location  $(x, y)$  and orientation  $\theta$ . The forward simulated state is thus:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} \cos \theta & 0 \\ \sin \theta & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v \\ \omega \end{bmatrix} \quad (2)$$

With linear  $v$  and angular  $\omega$  velocities. Projecting trajectories fits XSC’s *causality* guideline, as it represents a trace of recognised intent that can also be used to visually explain the wheelchair dynamics when supplied with input commands.

The ensuing arbitration phase is composed of two independent stages, for which the first involves collision-checking. Each edge of the rectangular base is tested along the simulated trajectory for intersections with obstacle points constructed from rangefinder data. Assuming the trajectories can be approximated by piecewise circular arcs [25], either a line-line or arc-line intersection check occurs depending on whether motion is purely translational or not, respectively. Performing this navigability check yields a vector of potential collision points. This geometric method of collision-checking coincides with XSC on *abstraction* by providing a high-level representation of threats based on sensor data.

The arbitration’s second stage is to adjust user input commands whenever they are deemed unsafe. A gap-based navigation method is chosen for this stage [26], where “gaps” are extracted from rangefinder data and a strategy of traversing the closest gap is applied. The concept of identifying gaps in the surroundings is particularly compelling for XSC, as it adds another semantic layer of *abstraction* around sensory state information.

The next phase of XSC is to convey this navigational assistance using a head-mounted AR interface, which is described in the following section.

### C. Augmented Reality System

With the interface design guidelines of XSC in mind, three visualisations are developed and categorised as either *environmental* or *embodied*. Any graphical cues that directly overlay the real-world surroundings are considered *environmental*, whilst *embodied* visualisations are fixed to either the robot or headset’s orientation and motion [7], [21]. First-person perspectives are demonstrated in Fig. 3.

The first virtual aid is a collision sphere paired with a directional arrow. Highlighting collision referents in the physical *environment* with salient red spheres (see Fig. 3-(b)) augments the *contextual* awareness of users, enabling them to identify why their actions might lead to unexpected behaviour. Directional arrows are also employed as headset-*embodied* cues that signal where these imminent collisions are situated from the operator’s perspective (shown in Fig. 3-(c) & (d)). These arrows are constrained to always appear within the headset’s field of view (FoV).

The second visualisation is a mini-map panel that portrays a birds-eye view of the mobile base and its forward state.

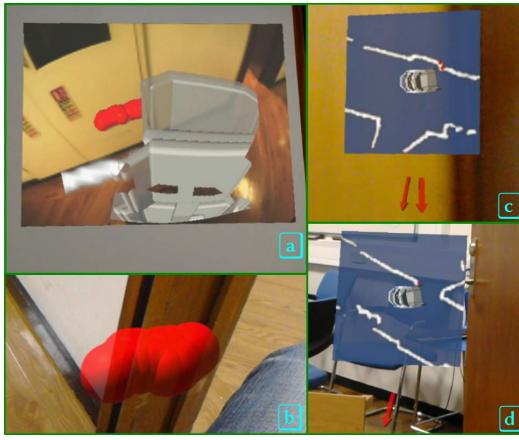


Fig. 3. First-person perspectives on the visualisations. (a) illustrates the rear-view display during reverse motion, where a wheelchair trajectory is projected outwards. (b) shows the red collision spheres overlaid in the environment. (c) & (d) display the mini-map utility and arrow indicators. The mini-map reveals the wheelchair's future trajectory and can hint at upcoming obstacles e.g. in (d) where the heading angle is too tight for the doorway and requires adjusting.

In order to explicate the navigational assistance discussed in Section III-B, the mini-map is annotated with laser scan readings and forecasts of the robot's estimated poses after applying input commands. For instance, Fig. 3-(d) illustrates a scenario in which the operator has selected a command that leads to a tight angle for door traversal. The *predictive* visual feedback of both the red collision marker and projected robot trajectory suggests that the user should adjust their heading before attempting the doorway.

Lastly, the rear-view display incorporated in our previous AR-HMD proposal for wheelchair navigation [10] is persisted. We extend the display to supplement the rear-view with a virtual wheelchair avatar during backwards motion (translucent grey avatar shown in Fig. 3-(a)). This aid conforms to XSC by supplying users with a wider *contextual* perspective of the robot's navigational reasoning, and by rendering the *predicted* effects of issuing reversal commands.

Certain design considerations are taken to avoid distracting users throughout operation, which can pose a problem for HMDs [23]. The red obstacle spheres are instantiated at the physical targets detected by laser sensors attached to the mobile base, thus providing an accurate depiction of the robot's collision-checking process. Whenever these spheres fall outside of the egocentric FoV, small arrows appear in the navigator's periphery as a non-obtrusive indicator. Both the panel and rear-view are robot-*embodied* cues that are rigidly attached to the mobile base. In order to not clutter an operator's FoV, these two virtual objects are positioned outside the natural viewing angle of wheelchair navigation.

#### IV. EXPERIMENT

We conducted an indoor navigation study to investigate the influence of using XSC with an AR-HMD whilst operating a robotic wheelchair. The mobile platform has a rectangular shape ( $0.90\text{m} \times 0.65\text{m}$ ) and is equipped with two laser

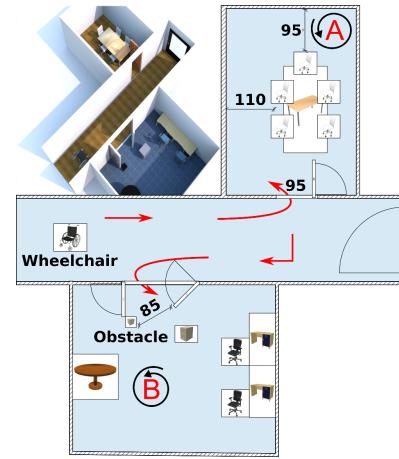


Fig. 4. A floor plan of the navigation route used for a single experiment trial, with a 3D overhead view in the top-right corner. Each participant is requested to manoeuvre from the wheelchair position, to goal A, and then to B, before finally returning back to the starting wheelchair position. The small numeric labels denote the centimetre widths of spaces along the route.

scanners at the front and one at the back for a full FoV. All navigation processes are implemented atop of ROS [27] and run using an on-board laptop. The Unity 3D game engine was utilised for AR application development and deployed on a HoloLens ( $30^\circ \times 17.5^\circ$  FoV), which has been supplemented with an add-on for eye tracking [28]. Readers interested in how to integrate a HoloLens onto a robotic wheelchair can refer to our prior work [10].

#### A. Experimental Setup & Protocol

We invited a total of 18 able-bodied volunteers (4 female, 14 male) aged 22-65 (median: 25) to take part in the experiment. Participants reported their familiarity with powered mobility, robotic wheelchairs and AR, with the most common grouping reporting no prior experience in any.

A 2D floor plan of the experimental route is illustrated in Fig. 4. Each subject was requested to complete this route twice, once with and without the aid of the proposed visualisations. To counterbalance the effects of trial order, even-numbered participants performed their first trial with visual feedback before proceeding onto their second trial without, and vice versa for the odd-numbered. Participants wore the headset across both trials for the purpose of data collection and fairness in comfort.

An entire experiment run took approximately 30 minutes and consisted of the following five phases: (1) preliminaries, (2) training if necessary, (3) eye gaze calibration, (4) navigation task and (5) post-experiment questionnaire. Phase (1) asked volunteers to fill out an introductory questionnaire, sign a consent form and watch video demonstrations of the AR assistance. Phase (2) was an optional 5-minute training of wheelchair control, especially for those with no prior experience. Step (3) involved fitting the HoloLens on the subject and calibrating the eye tracker using the plugin for HMDs [28]. Phase (4) consisted of the two navigation trials, followed by a post-experiment questionnaire in stage (5).



Fig. 5. Illustrates two participants navigating the office doorway entrance (refer to map in Fig. 4) mid-trial. The volunteer on the left had to regularly attend to the obstacle box without any graphical aid. On the other hand, the volunteer on the right could manoeuvre around the box with little need to perform downward glances due to the graphical assistance.

There are two noteworthy locations along the navigation route that test the benefits of XSC guidelines on AR interface design. Firstly, the passageway to goal ‘A’ includes a chair in the top-left corner of the room, which leads to a tight bend around the table that is challenging to manoeuvre (*contextual*). Secondly, the obstacle box located in the office is small and requires advance notice for smooth circumvention without regular downward glances (*predictive* – see Fig. 5).

#### B. Evaluation Metrics

A variety of task-specific metrics are examined to evaluate the XSC and AR system. Aside from standard performance measures for mobile robotics, such as time-to-completion, the focus of evaluation is also directed towards human factors associated with model misalignment in shared control.

One notable human factor that has previously been investigated for its impact on shared control is cognitive workload [29]. Self-reported questionnaires are often utilised to assess this metric, but eye gaze is also known to be correlated with heightened workload or difficulty in manoeuvring a wheelchair [29], [30]. As a result, we report eye gaze patterns as a physiological measure on the mental models of participants.

Additionally, we record the time to traverse specific navigation events that are relevant to XSC and any task load incurred due to model mismatch. Bypassing doorways and avoiding incidents where the wheelchair gets stuck are both prominent issues for powered mobility [30]. Given the role of transparent assistance in these situations, we identified events where participants encountered doors and “stucks”, so as to record the time it takes to overcome such events. Door positions are set by their midpoints and tracked as events whenever they fall within the wheelchair’s footprint. Likewise, “stucks” occur whenever the wheelchair does not escape its own clearance for a duration of 10 seconds.

Lastly, a post-experiment survey was handed out to volunteers for subjective feedback. The survey asked users to rate

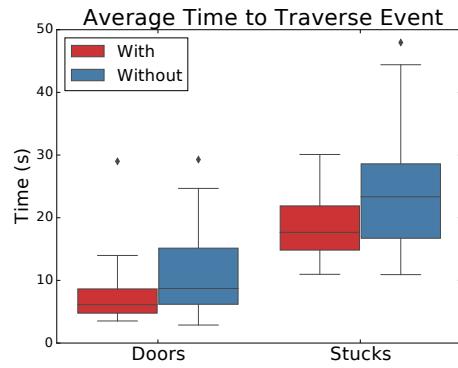


Fig. 6. Average timing results across commonly occurring events that can result from model misalignment in shared control for robot-assisted navigation. For both doorways and “stucks”, participants overcame these circumstances faster in their visualisation-aided trial, highlighting the benefit of predictive cueing at inciting quicker recovery times.

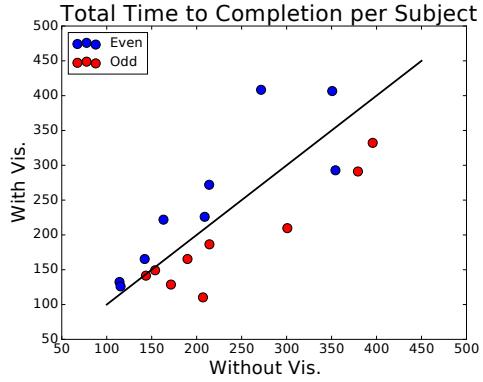


Fig. 7. Time-to-completion per participant, with and without visualisations. Even-numbered subjects had graphical aid for their first trial, and odd-numbered subjects on their second. Points below the main diagonal show subjects that attained lower times with visual aid in comparison to without.

the benefit of the different visualisations and their general perceptions of the overall system (5-point Likert scale).

#### C. Empirical Findings

Focusing on the evaluation of transparent assistance attributed to XSC, Fig. 6 demonstrates the traversal times for the events described in Section IV-B. The results indicate that volunteers recovered faster from “stucks” ( $p = 0.03$ ) when guided with AR (median: 18.61s, IQR = 22.7 – 15.17s) than without (median: 26.23s, IQR = 44.41 – 16.88s). Similarly, doorways had quicker traversal times ( $p = 0.093$ ) on visualisation trials (median: 6.13s, IQR = 8.64 – 4.77s) in comparison to without (median: 8.71s, IQR = 15.14 – 6.19s). These findings support the fact that subjects overcame hazardous and otherwise jarring incidents more effectively.

In the post-survey responses, participants made various comments that reinforce the quantitative findings on event traversal times. Some claimed that the aids “helped with understanding *where* the problem was”, as well as “explaining *why* the safety algorithm was changing the way the wheelchair behaved”. These observations echo the XSC

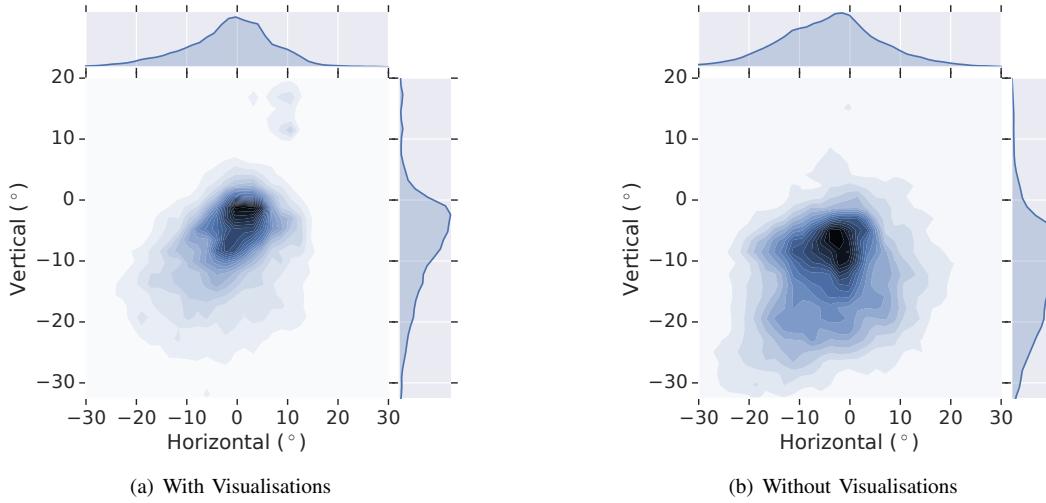


Fig. 8. Joint angular gaze distribution across all participants along the horizontal and vertical axes. When operating the wheelchair with visualisations (a), volunteers maintain a more centrally-oriented gaze angle and perform less downward glances in comparison to without (b).

guidelines discussed in Section III-A and fortify the value of explicating shared control, such that human operators can quickly recover from model misalignment.

Time-to-completion per participant is shown in Fig. 7. We report average relative improvement in timings, with median values of 16.05% (IQR = 24.91 – 12.86%) when switching to visualisations, and 13.75% (IQR = 21.36 – 8.49%) for no visualisations on the second trial. Despite the positive trend in improvements, the experimental scenario only involves two trials and thus cannot discount the possibility of a larger effect due to natural learning rates between successive trials.

Joint angular gaze distribution plots of eye tracking results are shown in Fig. 8. Adhering to the manufacturer's eye tracker guidelines, we filtered all gaze points below a specified confidence threshold before generating these plots. For the angular coordinates with visualisations, Fig. 8-(a) presents mean angles of  $-1.74^\circ$  (SD  $9.63^\circ$ ) and  $-7.57^\circ$  (SD  $9.88^\circ$ ) in the horizontal and vertical directions, respectively. Conversely for non-visualisation trials, Fig. 8-(b) presents mean angles of  $-3.97^\circ$  (SD  $11.39^\circ$ ) and  $-12.54^\circ$  (SD  $9.97^\circ$ ) in the horizontal and vertical directions, respectively.

Although no strong claims can yet be made for the relationship between these gaze patterns and mental models, there are still a few noteworthy observations. First, subjects occupied the negative vertical region less frequently in the visually-aided trial, signifying that they could successfully complete the task without repeatedly glancing downwards for obstacles (exemplified in Fig. 5). This implies that the AR-HMD interface provided an easily accessible source of contextual information regarding the shared control. Furthermore, volunteers maintained a more centrally-oriented gaze when operating the wheelchair with visualisations, than without. A reduction in variability could be indicative of lower mental workload during the prescribed navigation task [29]. Lastly, greater care might need to be taken to not divert user attention through visual explanations, as there are more instances of looking upwards in visualisation trials.

TABLE I  
 SUMMARY OF RESPONSES TO: "RATE THE BENEFIT OF THE FOLLOWING VISUALISATION FROM 1-5 (1 = VERY POOR, 5 = VERY GOOD)".

Visualisation	Mean User Rating	Standard Deviation
Red Collision Indicators	4.39	0.92
Mini-map Display	3.22	1.17
Rear-view Display	4.06	0.94

TABLE II  
 SUMMARY OF SUBJECTIVE RESPONSES TO GENERAL PERCEPTIONS FROM 1-5 (1 = STRONGLY DISAGREE, 5 = STRONGLY AGREE).

Question	Mean Rating $\pm$ Std. Dev.
I felt <i>clarity</i> on the visualisations	$4.11 \pm 0.68$
I found the visualisations <i>effective</i>	$3.67 \pm 0.84$
I felt <i>distracted</i> by the visualisations	$2.17 \pm 0.99$

#### D. Survey Results

An uplifting result from the survey responses was the positive user appraisal of all AR visualisations. Table I lists the average scores from 1-5 (5 being most positive) for the graphical aids. Most notably, the red collision indicators garnered nearly universal approval ( $4.39 \pm 0.92$ ) and even surpassed the popularity of the rear-view display ( $4.06 \pm 0.94$ ). Although the mini-map was less appreciated ( $3.22 \pm 1.17$ ), the result was still in favour of its inclusion.

Table II presents general opinions on the overall system, which promote the efficacy of conforming to XSC through AR-HMDs. The general disagreement with the notion of feeling *distracted* ( $2.17 \pm 0.99$ ) asserts that our visualisations did not have any inherently misleading effects. In line with the aim to administer interpretable assistance, subjects responded with high levels of *clarity* on the purpose of each virtual aid ( $4.11 \pm 0.68$ ). Finally, the positive tendency for users to find the visualisations *effective* at predicting the robot's behaviour ( $3.67 \pm 0.84$ ) bolsters the prospects of predictive cueing in AR for XSC.

## V. CONCLUSIONS

In this paper, XSC was introduced to resolve the model misalignment problem that frequents many human-robot interactions. An instantiation of XSC for assistive navigation was presented, where an AR headset played the integral role of visually demystifying the shared control. Experimental results on the effectiveness of XSC demonstrated quicker user recovery times from adverse situations commonly encountered during model misalignment.

There are many ways of instantiating XSC, however the guidelines applied in this work set a precedent for multiple application domains. In medical applications, such as surgical navigation, the weight of impeding performance through inappropriately placed virtual objects can have life-threatening consequences [31]. *Predictive* aids that augment the surgeon's trajectory planning and *contextual* awareness are an exemplar use-case of XSC. Likewise, we anticipate that designing shared control mechanisms to exhibit *causality* and *abstraction* could help disambiguate robot intentions in other HMD-based human-robot collaborations, such as aerial navigation [7] and shared workspace manipulation [23].

Future work will explore how to generalise the XSC paradigm to be robot-agnostic and less task-dependent. The resounding appreciation for the red collision indicators suggests that visualisations augmenting the *environment* are perhaps more tightly coupled with XSC, warranting further investigation in other task domains. Another worthwhile avenue of research is to formalise the XSC guidelines on visualisations for “explanation” [9] and identify how exactly they reconcile model misalignment. One way of testing this could be to purposefully inject incorrect robot behaviours into the shared control and observe how user misconceptions are corrected with different visualisations.

## REFERENCES

- [1] Y. Demiris, “Knowing when to assist: Developmental issues in lifelong assistive robotics,” in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 3357–3360.
- [2] P. D. Nisbet, “Who’s intelligent? wheelchair, driver or both?” in *Int. Conference on Control Applications*, vol. 2, 2002, pp. 760–765.
- [3] D. A. Abbink, T. Carlson, M. Mulder, J. C. F. de Winter, F. Aminravan, T. L. Gibo, and E. R. Boer, “A Topology of Shared Control Systems—Finding Common Ground in Diversity,” *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 5, pp. 509–525, 2018.
- [4] Y. Demiris, “Prediction of intent in robotics and multi-agent systems,” *Cognitive Processing*, vol. 8, no. 3, pp. 151–158, 2007.
- [5] D. P. Losey, C. G. McDonald, E. Battaglia, and M. K. O’Malley, “A review of intent detection, arbitration, and communication aspects of shared control for physical human-robot interaction,” *Applied Mechanics Reviews*, vol. 70, no. 1, pp. 010 804–010 804–19, 2018.
- [6] A. Kucukyilmaz and Y. Demiris, “Learning Shared Control by Demonstration for Personalized Wheelchair Assistance,” *IEEE Transactions on Haptics*, vol. 11, no. 3, pp. 431–442, 2018.
- [7] M. Walker, H. Hedayati, J. Lee, and D. Szafir, “Communicating Robot Motion Intent with Augmented Reality,” in *ACM/IEEE International Conference on Human-Robot Interaction*, 2018, pp. 316–324.
- [8] T. Chakraborti, S. Sreedharan, A. Kulkarni, and S. Kambhampati, “Projection-aware task planning and execution for human-in-the-loop operation of robots in a mixed-reality workspace,” in *IEEE/RSJ Int. Conference on Intelligent Robots and Systems*, 2018, pp. 4476–4482.
- [9] T. Chakraborti, K. P. Fadnis, K. Talamadupula, M. Dholakia, B. Srivastava, J. O. Kephart, and R. K. E. Bellamy, “Visualizations for an Explainable Planning Agent,” in *International Joint Conference on Artificial Intelligence*, 2018, pp. 5820–5822.
- [10] M. Zolotas, J. Elsdon, and Y. Demiris, “Head-Mounted Augmented Reality for Explainable Robotic Wheelchair Assistance,” in *IEEE/RSJ Int. Conference on Intelligent Robots & Systems*, 2018, pp. 1823–1829.
- [11] T. Chakraborti, S. Sreedharan, Y. Zhang, and S. Kambhampati, “Plan Explanations As Model Reconciliation: Moving Beyond Explanation As Soliloquy,” in *International Joint Conference on Artificial Intelligence*, 2017, pp. 156–163.
- [12] M. Fox, D. Long, and D. Magazzeni, “Explainable Planning,” in *International Joint Conference on Artificial Intelligence Workshop on Explainable Planning*, 2017.
- [13] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser, and R. Parasuraman, “A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction,” *Human Factors*, vol. 53, no. 5, pp. 517–527, 2011.
- [14] D. Nicolis, A. M. Zanchettin, and P. Rocco, “Human Intention Estimation based on Neural Networks for Enhanced Collaboration with Robots,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2018, pp. 1326–1333.
- [15] S. Reddy, A. Dragan, and S. Levine, “Shared Autonomy via Deep Reinforcement Learning,” in *Robotics: Science and Systems*, 2018.
- [16] P. Milgram, S. Zhai, D. Drascic, and J. Grodski, “Applications of augmented reality for human-robot communication,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3, 1993, pp. 1467–1472.
- [17] P. Milgram, A. Rastogi, and J. J. Grodski, “Telerobotic control using augmented reality,” in *IEEE International Workshop on Robot and Human Communication*, 1995, pp. 21–29.
- [18] R. T. Azuma, “A Survey of Augmented Reality,” *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, 1997.
- [19] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, “Recent advances in augmented reality,” *IEEE Computer Graphics and Applications*, vol. 21, no. 6, pp. 34–47, 2001.
- [20] C. Diaz, M. Walker, D. A. Szafir, and D. Szafir, “Designing for Depth Perceptions in Augmented Reality,” in *IEEE International Symposium on Mixed and Augmented Reality*, 2017, pp. 111–122.
- [21] H. Kim, J. L. Gabbard, A. M. Anon, and T. Misu, “Driver Behavior and Performance with Augmented Reality Pedestrian Collision Warning: An Outdoor User Study,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 4, pp. 1515–1524, 2018.
- [22] A. Alshaer, H. Regenbrecht, and D. O’Hare, “Immersion factors affecting perception and behaviour in a virtual reality power wheelchair simulator,” *Applied Ergonomics*, vol. 58, pp. 1–12, 2017.
- [23] E. Sibirtseva, D. Kontogiorgos, O. Nykvist, H. Karaoguz, I. Leite, J. Gustafson, and D. Krägic, “A comparison of visualisation methods for disambiguating verbal requests in human-robot interaction,” in *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2018, pp. 43–50.
- [24] J. Minguez, F. Lamiraux, and J.-P. Laumond, “Motion Planning and Obstacle Avoidance,” in *Springer Handbook of Robotics*, 2016, pp. 1177–1201.
- [25] D. Fox, W. Burgard, and S. Thrun, “The dynamic window approach to collision avoidance,” *IEEE Robotics and Automation Magazine*, vol. 4, no. 1, pp. 23–33, 1997.
- [26] M. Mujahed, D. Fischer, and B. Mertsching, “Admissible gap navigation: A new collision avoidance approach,” *Robotics and Autonomous Systems*, vol. 103, pp. 93–110, 2018.
- [27] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Y. Ng, “ROS: an open-source Robot Operating System,” in *IEEE ICRA Workshop on Open Source Software*, 2009.
- [28] M. Kassner, W. Patera, and A. Bulling, “Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction,” in *ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 2014, pp. 1151–1160.
- [29] T. Carlson and Y. Demiris, “Collaborative Control for a Robotic Wheelchair: Evaluation of Performance, Attention, and Workload,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 3, pp. 876–888, 2012.
- [30] R. C. Simpson, “How many people would benefit from a smart wheelchair?” *The Journal of Rehabilitation Research and Development*, vol. 45, no. 1, pp. 53–72, 2008.
- [31] B. J. Dixon, M. J. Daly, H. Chan, A. D. Vescan, I. J. Witterick, and J. C. Irish, “Surgeons blinded by enhanced navigation: the effect of augmented reality on attention,” *Surgical Endoscopy*, vol. 27, no. 2, pp. 454–461, 2013.