

Human-in-the-loop testing of the explainability of robot navigation algorithms in extended reality

Jérôme Guzzi¹, Alessandro Giusti¹

¹Dalle Molle Institute for Artificial Intelligence (IDSIA), USI-SUPSI, Lugano, Switzerland

Abstract

Developing robots to be deployed in spaces shared with people requires testing with humans-in-the-loop. In fact, only humans co-located with the robots are in a suitable position to judge the robots' behaviors. This is especially important when people participate in the same task as the robots, like when navigating a shared environment: the actions of people and robots influence each other. To test the resulting dynamic system, we need to let real people experience the interaction, which, when done in simulation, translates to immersing real users using Virtual or Mixed Reality. We implement and demonstrate this solution where users experience variable legibility, predictability, and explainability of the robot navigation algorithms, depending on if and how the robots explicitly communicate their intentions.

Keywords

Social robotic navigation, Human-in-the loop, Virtual Reality

1. Introduction

Mobile robots are increasingly deployed in spaces shared with people: from small vacuum cleaners to large autonomous vehicles for moving goods, their behavior impacts people surrounding them [1]. In particular, when people and robots move in the same space, they take each other states and actions into account to adjust their movements, with the shared goal of avoiding collisions and approaching targets [2]. For people and animals, occupying space is an implicit form of communication called *proxemics* [3]: humans are aware of personal spaces, respect social rules when approaching others, and infer the intentions of neighbors from current and past actions. During navigation, humans predict and account for future trajectories of neighbors: an ability that relies on the *legibility* (i.e., infer goals from actions) and *predictability* (i.e., infer actions from goals) of their neighbors' actions [4].

To promote acceptance by surrounding people, it is therefore important that the navigation behavior of robots is – in addition to safe, efficient, and respectful of basic social rules – legible and predictable [5]. How much robots should rely on proxemics or use more explicit forms of communication, e.g. through their gaze[6], is an open research question, whose answer depends on the context. Explicit communication is also common among humans when navigating or driving, like nodding or waving at each other at a crossing, or signaling intentions by activating the turn signals.

Late-breaking work, Demos and Doctoral Consortium, colocated with The 2nd World Conference on eXplainable Artificial Intelligence: July 17–19, 2024, Valletta, Malta

✉ jerome@idsia.ch (J. Guzzi); alessandrogiusti@idsia.ch (A. Giusti)

🆔 0000-0002-1263-4110 (J. Guzzi); 0000-0002-1263-4110 (A. Giusti)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



Even when the individual robot behavior is based on transparent and explainable bio-inspired rules that mimic pedestrians [7], trajectories result from a complex multi-agent dynamical system and are less predictable, in particular in densely packed spaces. Nonetheless, co-located humans need to understand the behavior of the robots, at least in terms of legibility and predictability, which in this context are primary factors for *explainability* [8, 9]

The development of navigation algorithms for social robots requires testing with humans-in-the-loop. In fact, only people moving among the robots are in a suitable position to judge the robots' behaviors: people can experience the influence of the surrounding robots, and experience if and how the robots' presence affects their psycho-physical effort. This brings the question of how to gather human feedback during early development phases when it is not yet possible, or convenient, to test with real robots but simulations are available. The answer we explore in this contribution is to use *Virtual Reality* to immerse subjects in simulations, where they experience interacting with the robots with enough realism to not impact significantly their feedback. In fact, Virtual Reality is well suited to experiment with cognitive Human-Robot Interaction, like when using pointing gestures to select objects or destinations [10] or to build a wall together [11]. Simulation speeds up development, for instance, by isolating control algorithms from perception; Virtual Reality extends it to include *real* humans in a human-centred development cycle.

Our research targets navigation algorithms for autonomous wheelchairs in the context of the Horizon Europe project REXASI-PRO (REliable & eXplAinable Swarm Intelligence for People with Reduced mObility)¹ that is developing trustworthy-by-design tools to help people with reduced mobility. We are designing social-compliant behaviors that are comfortable for the user sitting in the wheelchair, as well as perceived as friendly by people sharing the same spaces. In this work, we present a *demonstration* of the setup we are using to test navigation algorithms with humans-in-the-loop: subjects, wearing a wireless VR headset, immerse themselves in a Virtual Reality simulation populated with virtual wheelchairs and virtual pedestrians, are free to move around and experience how the virtual agents behave and react to their actions. Subjects experience navigation scenarios in one of two modalities: in Virtual Reality, they perceive virtual agents in a *virtual environment*; in Mixed Reality, they perceive virtual agents in the *real* environment instead. In the scenarios, robots are equipped with different behaviors, some using explicit communication (e.g., switching on/off LEDs) to signal their intentions. We believe that this demonstration offers an interesting and unique opportunity for people to experience interaction with AI agents in a simple but ubiquitous task that displays the interplay between legibility, predictability, explainability, and communication.

2. Extended reality testbed for social robotics navigation

2.1. Infrastructure

To interact with virtual robots, users wear a VR headset. The headset runs an application that connects as a client to a robotic simulator. The simulator regularly broadcasts the state of the simulated world, which the client uses to render the scene from the point of view of their user.

¹<https://rexasi-pro.spindoxlabs.com>

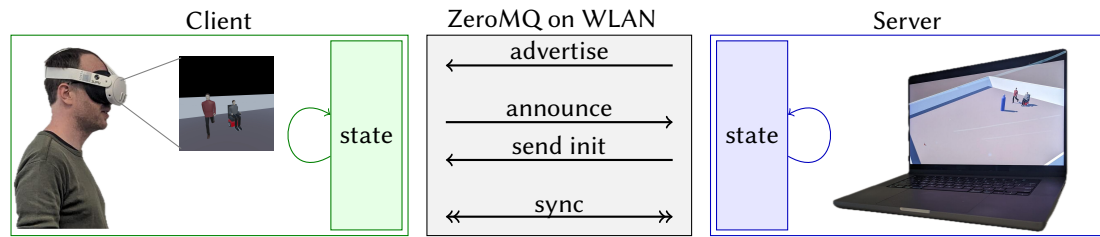


Figure 1: Synchronization of the simulation state between the client running on the VR headset (left) and the server running on a laptop (right). The state update loops on the client and on the server are decoupled and run at different rates. They are synchronized via ZeroMQ pub/sub.

In the following, we describe the components and their integration illustrated in Figure 1.

Navground Navground² is a collection of navigation algorithms and tools to simulate/integrate/test them. It’s a cross-platform open-source project that provides benchmarks and reference implementations for robotic navigation. Users can extend from C++ or Python the navigation algorithms, and other components like kinematics models, and test them against baselines in different navigation scenarios. It is integrated with ROS2 and with machine learning libraries, like Gymnasium,³ to support ML-based navigation policies.

Simulator *CoppeliaSim* [12] is one of the most used simulators for robotics research. It is very flexible, supporting different ways – from embedded LUA scripts to C++ plugins and ROS interfaces – to program robots. It offers an intuitive WYSIWYG interface to edit simulation scenes. Navground supports CoppeliaSim by providing LUA bindings to configure simulations where robots or other agents (e.g., pedestrians) use Navground’s algorithms to navigate.

Headset The *Quest 3* from Meta is a relatively inexpensive VR headset that supports color passthrough for Mixed Reality applications. We develop an application, using the official Unity API,⁴ that renders on the headset a scene simulated by CoppeliaSim on a *separated* computer (a laptop, in our setup): through the application, the headset acts as a client to a server running CoppeliaSim. The scene is rendered at up to 120 Hz on the headset (to avoid motion sickness), interpolating and extrapolating the state updated at a lower rate in CoppeliaSim (typically at 20 Hz).

Communication between headset and simulator We use ZeroMQ to synchronize the state of the simulation between the server and the clients, as depicted in Figure 1. CoppeliaSim instances are automatically discovered by (wireless) clients in the same local network, which announce themselves to the server. The server sends to each client a description of the current scene containing the pose and shape of all visible objects. At each simulation step, the server publishes poses and velocities of objects that have moved, which clients use to update their

²<https://github.com/idsia-robotics/navground>

³<https://gymnasium.farama.org>

⁴<https://developer.oculus.com/documentation/unity>

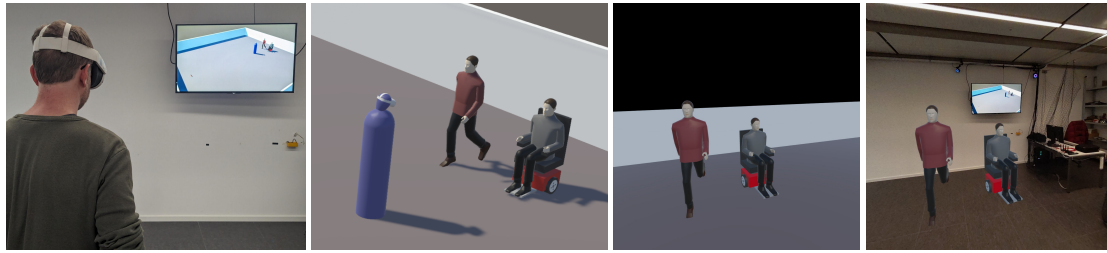


Figure 2: A participant in the demonstration observes a virtual pedestrian and a virtual user of an autonomous wheelchair navigating using a bio-inspired algorithm. Left: the participant, wearing the VR headset, looks at virtual agents located in front of him; a screen displays the simulation from an external point of view. Center: a magnification of the view shown on the screen with the participant's avatar in blue. Right: participant's first person view, as rendered by the headset, in Virtual (center-right) and Mixed Reality (right).

rendering. At the same time, clients stream the position of the headset (tracked by onboard sensors) to the server. CoppeliaSim passes the location, shape, and velocity of each user to the navigation algorithms controlling the simulated agents, which take them into account to avoid collisions.

2.2. Usage

With the infrastructure described in Section 2.1, we can immerse users in any scene that we can simulate in CoppeliaSim: “what you simulate, you can interact with”. This is very powerful because it extends the domain of a robotic simulator like CoppeliaSim to any human-robot interaction that does not require physical contact between robots and people.

Freedom of movement Users wear a *wireless* headset and are free to move in the space to observe (and interact with) robots from different places/viewpoints: this is required to experience proximity-dependant interactions.

Virtual and Mixed Reality Users can interact with the robots in two possible modes (see Figure 2). In Virtual Reality, they can immerse themselves completely in a world that could be different and possibly larger their physical environment, and focus on the interaction without distractions. In Mixed Reality, users may feel safer as they are aware of the physical environment, perceive other people or real robots more realistically, and perform interactions situated in the real environment, like when interacting with a virtual robot placed on a real desk or observing a virtual smart wheelchair passing through a real door. However, Mixed Reality is constrained by the environment.

Multiple users The infrastructure supports multiple users immersed in the same simulation at the same time. In Virtual Reality, users see the avatars of other users, while in Mixed Reality they directly see other users. Moreover, in Mixed Reality, users have to physically share the space, while in Virtual Reality they may actually be in separate spaces that only virtually overlap.

In both cases, the only requirement to support multiple users is that all headsets are connected to the same wireless network and that users agree on a common reference frame, which is done through a simple procedure that only needs to be performed once in each environment.

Displacing objects Using their hands or VR controllers, users can move virtual objects. When an object is in the user's hand, CoppeliaSim stops simulating its dynamics but resumes as soon as the user drops the object. The ability to move objects (or even robots or parts of robots) increases the range of interactions that we can simulate. For instance, we could interact with a robotic arm by moving a target point for its end-effector, we could move way-points used by navigating robots, or we can modify obstacles, like when opening or closing doors.

Multiple scenes The same client/server applications described in Section 2.1 work with any scene that the user selects from CoppeliaSim or from the headset: changes are propagated to all connected clients. No software update is required when adding new scenes from CoppeliaSim: one VR application can therefore support different scenes that developers are free to add. For instance, researchers working with a particular robot only need to integrate it in CoppeliaSim to be able to render it in VR.

2.3. Testing social navigation algorithms

We use the infrastructure described in Section 2.1 to develop, test, and validate algorithms for social navigation. After developing an algorithm with Navground, we design one or more simulation scenes in CoppeliaSim where some of the agents are configured to use this algorithm. Scenes reproduce many different situations, ranging from empty spaces (e.g. a plaza or a large hall), to narrow corridors, to unstructured environments with obstacles at random positions.

During user-tests, subjects wearing the VR headset are immersed in the scene simulated in CoppeliaSim, and are free to move in an area of the physical environment that is free of obstacles. We instruct subjects to perform a navigation task (like "walk along the corridor", "follow the yellow arrow", or "reach the yellow target"). During the tests, we record their poses using the VR headset's native tracking capabilities. At the end of the tests, we ask subjects to provide feedback about the difficulty of the task and the behavior of the agents.

We use objective and subjective feedback from subjects to assess and improve the agent's navigation behavior, which we then test again. This development cycle is an effective way to incorporate users' feedback. Compared with real-world tests, it is easier and faster to deploy/test algorithms in simulation and we avoid any safety issues from potential collisions between robots and people. In contrast with tests where subjects passively look at a simulation/video to judge the robots' behaviors, we can gather more realistic feedback because being part of the navigation scenario significantly changes its perception. For instance, from the outside, virtual robots generally appear slower to passive observers than to observers who are actively avoiding them in the VR simulation. Our goal is that robots are *perceived* as friendly by *nearby* people, therefore we should test their behavior in similar settings.

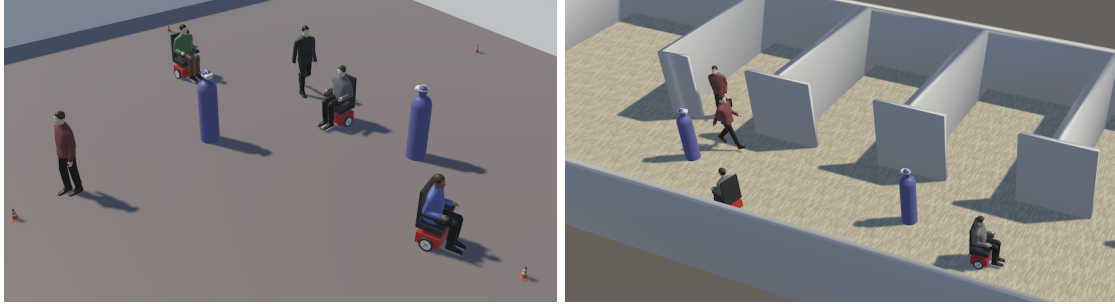


Figure 3: Two users (blue avatars) immersed in two navigation scenarios: left moving between the cones, crossing in the middle; right travelling along an corridor.

3. The demonstration

3.1. Communication and explainability in social robotics navigation

We propose a demonstration that focuses on the relationship between communication and legibility for robotic navigation in spaces shared with people. As a form of implicit communication (*proxemics*), people and robots partially expose their intentions by *moving* towards a target: if we see that an autonomous wheelchair, traveling along a corridor, turns towards a door, we can be quite sure that it would like to pass that door and enter a room. In more formal terms, we can *read* the intention from the trajectory, use this knowledge to *predict* the future trajectory, and take actions, like moving out of its way. In a more complex scenario, like when there are many robots navigating in the same space, their behaviors lose legibility – additional factors influence the trajectories behind the goal to enter the room – which in turn makes humans less able to predict and adapt to the robots’ behavior. From the point of view of the robots, very similar dynamics happen with respect to surroundings robots and people. This results in a dynamical system where, although the algorithms covering the single (non-human) agents may be deterministic and explainable, the collective behavior is chaotic and more opaque.

To improve legibility, and predictability, and let co-located humans better understand the behavior of robots, we can make robots expose their intentions more explicitly. In the proposed demonstration, we test, in simulation, simple communication strategies for ground robots, like: 1) switching on LEDs when about to enter/exit a room or pass through a narrow door; 2) projecting the goal trajectory or direction on the floor using a laser; 3) hinting the target goal on a front-facing screen (like the Astro robot by Amazon directs its virtual eyes). How much do these strategies impact legibility, predictability, and explainability? Which is one do surrounding people find more intuitive and comfortable? We demonstrate and investigate these questions in a simulation with humans-in-the-loop using the Extended Reality test-bed described in Section 2.

3.2. User experience during the demonstration

For each user, the demonstration starts by explaining the general functioning and goals. Then, they move to the middle of an area free of obstacles and wear the VR headset, which immerses

them in a sequence of navigation scenes they are guided through, as illustrated in Figure 2.

At first, in a simple scene where a virtual autonomous wheelchair and a virtual pedestrian go back and forth between two points, users get comfortable with switching between Mixed and Virtual Reality and experience how the virtual agents react to their movements. Then, they progress through more complex scenes where, for instance, agents have to pass an area with many crossings, or move in an indoor space where only one agent at a time can pass through doors (see Figure 3). Users are encouraged to participate in the navigation task (e.g., enter/exit the rooms) but may also just passively observe the scene and the behavior of the virtual agents.

When users are immersed in the simulation, spectators can observe the current scene on a computer screen – where the environment, virtual agents, and real users are rendered from a fixed external point of view – to understand what is going on (see Figure 2).

For each user, a demonstration lasts about 5 minutes. The demonstration is very similar to the procedure we use in our lab to test and validate the navigation algorithm with humans-in-the-loop, with the main difference that, as a demonstration, we do not follow a protocol, or record users' actions and feedback.

4. Conclusions

We presented an interactive demonstration where participants move among virtual autonomous wheelchairs in Extended Reality. On one hand, participants experience how their behavior influences the surrounding robots' navigation behavior. On the other hand, participants also observe how different navigation algorithms impact the legibility and predictability of the robots, which in turn influences the participants' actions. In particular, they experience how explicitly exposing robots' intentions impacts the intuitive explainability of the individual and collective behaviors of the robots. The demonstration is part of a research effort to improve the ability of robots to navigate spaces shared with people in a safe, efficient, and friendly way.

Acknowledgments

This work was supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract no. 22.00291 (REXASI-PRO project). The project has been selected within the European Union's Horizon Europe research and innovation programme under grant agreement ID: 101070028 (call HORIZON-CL4-2021-HUMAN-01-01). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the funding agencies, which cannot be held responsible for them.

References

- [1] D. S. Syrdal, K. Dautenhahn, K. L. Koay, M. L. Walters, W. C. Ho, Sharing spaces, sharing lives—the impact of robot mobility on user perception of a home companion robot, in: *Int. Conf. on Social Robotics*, Springer, 2013, pp. 321–330.

- [2] C. Mavrogiannis, F. Baldini, A. Wang, D. Zhao, P. Trautman, A. Steinfeld, J. Oh, Core challenges of social robot navigation: A survey, *ACM Transactions on Human-Robot Interaction* 12 (2023) 1–39.
- [3] E. T. Hall, et al., Proxemics, *Current anthropology* 9 (1968) 83–108.
- [4] A. D. Dragan, K. C. Lee, S. S. Srinivasa, Legibility and predictability of robot motion, in: *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI)*, 2013, pp. 301–308.
- [5] C. Lichtenthäler, A. Kirsch, Towards legible robot navigation-how to increase the intend expressiveness of robot navigation behavior, in: *Int. Conf. on Social Robotics*, 2013.
- [6] M. M. Neggers, P. A. Ruijten, R. H. Cuijpers, W. A. IJsselsteijn, Effect of robot gazing behavior on human comfort and robot predictability in navigation, in: *IEEE Int. Conf. on Advanced Robotics and Its Social Impacts (ARSO)*, 2022, pp. 1–6.
- [7] J. Guzzi, A. Giusti, L. M. Gambardella, G. Theraulaz, G. A. Di Caro, Human-friendly robot navigation in dynamic environments, in: *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2013, pp. 423–430.
- [8] A. Abdul, J. Vermeulen, D. Wang, B. Y. Lim, M. Kankanhalli, Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda, in: *Conf. on Human Factors in Computing Systems (CHI)*, 2018, pp. 1–18.
- [9] S. Wallkötter, S. Tulli, G. Castellano, A. Paiva, M. Chetouani, Explainable embodied agents through social cues: a review, *ACM Transactions on Human-Robot Interaction (THRI)* 10 (2021) 1–24.
- [10] J. Guzzi, G. Abbate, A. Paolillo, A. Giusti, Interacting with a conveyor belt in virtual reality using pointing gestures, in: *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI)*, 2022, pp. 1194–1195.
- [11] S. Shayesteh, H. Jebelli, Toward human-in-the-loop construction robotics: Understanding workers' response through trust measurement during human-robot collaboration, in: *Construction research congress*, 2022, pp. 631–639.
- [12] E. Rohmer, S. P. N. Singh, M. Freese, V-rep: A versatile and scalable robot simulation framework, in: *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2013, pp. 1321–1326.