Improving Robot Transparency: An Investigation With Mobile Augmented Reality

Alexandros Rotsidis[1], Andreas Theodorou[2], Joanna J. Bryson[3] and Robert H. Wortham[4]

***Abstract*— Autonomous robots can be difficult to understand by their developers, let alone by end users. Yet, as they become increasingly integral parts of our societies, the need for affordable easy to use tools to provide transparency grows. The rise of the smartphone and the improvements in mobile computing performance have gradually allowed Augmented Reality (AR) to become more mobile and affordable. In this paper we review relevant robot systems architecture and propose a new software tool to provide robot transparency through the use of AR technology. Our new tool, ABOD3-AR provides real-time graphical visualisation and debugging of a robot's goals and priorities as a means for both designers and end users to gain a better mental model of the internal state and decision making processes taking place within a robot. We also report on our on-going research programme and planned studies to further understand the effects of transparency to naive users and experts.**

I. INTRODUCTION

The relationship between transparency, trust, and utility is a complex one. By exposing the inner 'smoke and mirrors' of our agents, we risk of making them look less interesting. Moreover, the wide range of application domains for AI and of the different stakeholders interacting with intelligent systems should not be underestimated[1]. Therefore, What is effectively transparent varies by who the observer is, and what their goals and obligations are [2]. There is however a need for design guidelines on how to implement transparent systems, alongside with a 'bare minimum' standardised implementation [3]. In the end, the goal of transparency is should not be complete comprehension, that would severely limit the scope of human achievement. Instead, the goal of transparency is to provide sufficient information to ensure at least human accountability [4].

Still, the use real-time implementation can help users to *calibrate their trust* in the machine [5, and references therein]. Calibration refers to the correspondence between a person's trust in the system and the system's capabilities [6]. Calibrating of trust occurs when the end-user has a mental model of the system and relies on the system within the system's capabilities and is aware of its limitations. If we are to consider transparency as mechanism that exposes the decision-making of a system, then it can help users adjust their expectations and forecast certain actions from the system. This position about transparency is supported by [7], who conducted a study where the participants decide whether they trust a particular piece of pattern recognition software. The users were given only the percentage of how accurate the prediction of their probabilistic algorithm was in each image. Yet, by having access to this easy-to-implement transparency feature, they were able to calibrate their trust in real time.

Later work demonstrate how users of various demographic backgrounds had inaccurate mental models about a mobile robot [8]. The robot transmits a transparency feed to the real-time debugging software ABOD3 [9]. The transparency display is customised for a high-level end-user display of the robot's goals and process towards those goals. Participants without access to the transparency software ascribe unrealistic functionalities, potentially raising their expectations for its intelligence and safety. When the same robot is used with ABOD3, providing an end-user transparency visualisation, the users are able to calibrate their mental models, leading to more realistic expectations, but interestingly a higher respect for the system's intelligence.

Yet, despite its effectiveness, there is a major disadvantage with solutions like ABOD3: a computer and display is required to run the software. One solution might be to port ABOD3 to run directly on robots with built-in screens. Albeit that this is a technologically feasible and potentially interesting approach, it also requires that custom-made versions of ABOD3 will need to be made for each robotics system. Moreover, this is not a compatible solution for robots without a display.

Nowadays, most people carry a smartphone. Such mobile phones are equipped with powerful multi-core processors, capable of running complex computational-intensive applications, in a compact package. Modern phones also integrate high-resolution cameras, allowing them to capture and display a feed of the real world. That feed can be enhanced with the real-time superimposition of computer-generated graphics to provide Augmented Reality (AR) [10]. Unlike Virtual Reality that aims for complete immersion, AR focuses on providing additional information of and means of interaction with real-world object, locations, and even other agents.

In this paper we demonstrate new software, *ABOD3-AR*, which can run on mobile phones. ABOD3-AR, as its name suggests, uses a phone's camera to provide AR experience by superimposing the ABOD3's tree-like display of Instinct plans over a tracked robot. This allows real-time visualisation of a robot's priorities and plan execution for both end-user

[1]Alexandros Rotsidis is with Department of Computer Science, University of Bath, UK A.Rotsidis@bath.ac.uk

[2]Andreas Theodorou is with the Department of Computer Science, Umeå University, Sweden andreas.theodorou@umu.se

[3]Joanna J. Bryson is with the Department of Computer Science, University of Bath, UK J.J.Bryson@bath.ac.uk

[4]Robert H. Wortham is with the Department of Electronic & Electrical Engineering, University of Bath, UK r.h.wortham@bath.ac.uk

transparency and debugging purposes. In the next section we introduce ABOD3-AR and other relevant technologies and tools. Next, we present a user study conducted to investigate the effectiveness of our software. In the penultimate section, we discuss the results of our study. We conclude this paper by reviewing the work presented and discussing planned future work.

II. Tools and Technologies for Transparency

In this section we describe in some detail the tools and technologies used in our transparency experiments.

A. Behaviour Oriented Design

Behaviour Oriented Design is a cognitive architecture that provides an ontology of required knowledge and a convenient representation for expressing timely actions as the basis for modular decomposition for intelligent systems [11], [12]. It takes inspiration both from the well-established programming paradigm of object-oriented design (ODD) and its associated agile design [13], and an older but well-known AI systems-engineering strategy, *Behaviour-Based AI* (BBAI) [14].

BOD helps AI developers as it provides not only an ontology, addressing the challenge of 'how to link the different parts together', but also a development methodology; a solution to 'how do I start building this system'. It includes guidelines for modular decomposition, documentation, refactoring, and code reuse. BOD aims to enforce the good-coding practice 'Don't Repeat Yourself', by splitting the behaviour into multiple modules. Modularisation makes the development of intelligent agents easier, faster, reusable and cost efficient.

Behaviour modules also store their own memories, e.g. sensory experiences. Multiple modules grouped together form a *behaviour library*. This 'library' can be hosted on a separate machine, for example in the cloud.

The *planner* executing within the agent is responsible for exploiting a plan file; stored structures describing the agent's priorities and behaviour. This separation of responsibilities into two major components enforces further code reusability. The same planner, if coded with a generic API to connect to a behaviour library, can be deployed in multiple agents, regardless of their goals or embodiment. For example, the Instinct planner has been successfully used in both robots and agent-based modelling, while POSH-Sharp and UN-POSH have been deployed in a variety of computer games [9], [13].

B. POSH and Instinct

POSH planning is an action-selection system introduced by [11]. It is designed as a reactive planning derivative of BOD to be used in embodied agents. POSH combines faster response times, similar to reactive approaches for BBAI, with goal-directed plans. Its use of hierarchical fixed representations of priorities makes it easy to visualise in a human, non-expert directed graph and sequentially audit.

Instinct is a lightweight alternative to POSH, incorporating elements from the various variations and modifications of POSH released over the years [15]. The planner was first designed to run on low resources available on the ARDUINO micro-controller system, such as the one used by the R5 robot seen in Figure 2.

C. ABOD3

ABOD3 is a substantial revision and extension of ABODE (A BOD Environment), originally built by Steve Gray and Simon Jones. ABOD3 directly reads and visualises POSH, Instinct, and UN-POSH plans.

Moreover, it reads log files containing the real-time transparency data emanating from the Instinct Planner, in order to provide a real-time graphical display of plan execution. Plan elements are highlighted as they are called by the planner and glow based on the number of recent invocations of that element. Plan elements without recent invocations dim down over a user-defined interval, until they return to their initial state. This offers abstracted backtracking of the calls, and the debugging of a common problem in distributed systems: race conditions where two or more sub-components constantly trigger and interfere with or even cancel each other. ABOD3 is also able to display a video and synchronise it with the debug display. In this way it is possible to explore both runtime debugging and wider issues of AI Transparency.

The editor provides a user-customisable user interface (UI) in line with the good practices for transparency introduced by [2]. Plan elements, their sub-trees, and debugging-related information can be hidden, to allow different levels of abstraction and present only relevant information to the present development or debugging task. The application, as shown in Figure 3, allows the user to override its default layout by moving elements and zooming the display to suit the user's needs and preferences. Layout preferences can be stored in a separate file. We have successfully used ABOD3 in both [8].

D. ABOD3-AR

ABOD3-AR builds on the good practice and lessons learned through the extended use of ABOD3. It provides a mobile-friendly interface, facilitating transparency for both end users and experts. In this section, we not only present the final system, but also look at the technical challenges and design decisions faced during development.

1) Deployment Platform and Architecture: The Android Operating System (OS) is our chosen development platform due to its open-source nature and the a number of computer vision and augmented reality libraries existing for the platform. Moreover, no developer's license is required to prototype or release the final deliverable. Android applications are written in Java, like ABOD3, making it possible to reuse its back-end code. Unlike the original ABOD3, ABDO3-AR is aimed exclusively for embodied-agents transparency. At the time of writing, Instinct (see Section II-B) is the only supported action-selection system.

Our test configuration, as seen in Figure 1, includes the tried-and-tested R5 robot. In the R5 robot,the callbacks write textual data to a TCP/IP stream over a wireless (WiFi)

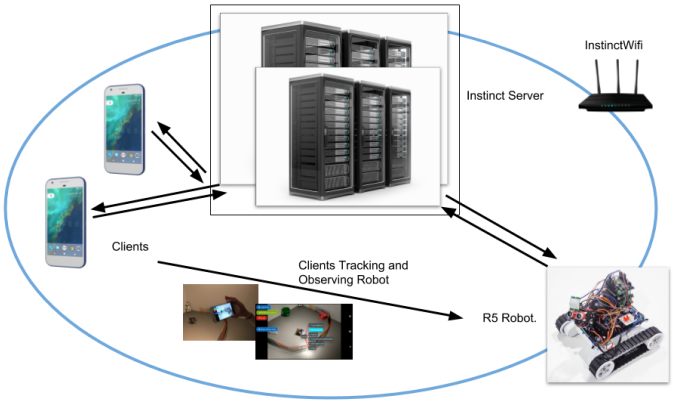

Fig. 1. R5 uses a WiFi connection to send the transparency feed to the Instinct Server for processing. Smartphones, running ABOD3-AR, can remotely connect to the server and receive the processed information.

link. A JAVA based Instinct Server receives this information, enriches it by replacing element IDs with element names and filters out low-level information, sending this information any mobile phones running ABOD3-AR. Clients do not necessarily need to be on the same network, but it is recommended to reduce latency. We decided to use this 'middle-man server' approach to allow multiple phones to be connected at the same time.

2) Robot tracking: Developing an AR application for a mobile phone presents two major technical challenges: (1) managing the limited computational resources available to achieve sufficient tracking and rendering of the superimposed graphics, and (2) to successfully identify and continuously track the object(s) of interest.

3) Region of Interest: A simple common solution to both challenges is to focus object tracking only within a region of the video feed, referred to as the Region of Interest (ROI), captured by the phone's camera. It is faster and easier to extract features for classification and sequentially track within a limited area rather than over the full frame. The user registers an area as the ROI, by expanding a yellow rectangle over the robot. Once selected, the yellow rectangle is replaced by a single pivot located at the centre of the ROI.

4) Tracker: Various solutions were considered; from the built-in black-box tracking of *ARCore* [1] to building and using our own tracker. To speed-up development, we decided to use an existing library *BoofCV* [2], a widely-spread Java library for image processing and object tracking. BoofCV was selected due to its compatibility with Android and the range of trackers available for prototyping.

BoofCV receives a real-time feed of camera frames, processes them, and then returns required information to the Android application. A number of trackers, or *processors* as they are referred to in BoofCV, are available. We narrowed down the choice to the *Circulant Matrices* tracker [16] and *Track-Learning-Detect* (TLD) tracker (TLD) [17].

The Track-Learning-Detect tracker follows an object from frame to frame by localising all appearances that have been observed so far and corrects the tracker if necessary. The learning estimates the detector's errors and updates it to avoid such errors, using a learning process. The learning process is modelled as a discrete dynamical system and the conditions under which the learning guarantees improvement are found. However, the TLD is computationally intensive. In our testing we found that when TLD was used the application would crash in older phones, due to the high memory usage.

The Circulant Matrices tracker is fast local moving-objects tracker. It uses the theory of Circulant matrices, Discrete Fourier Transform (DCF), and linear classifiers to track a target and learn its changes in appearance. The target is assumed to be rectangular with a fixed size. A dense local search, using DCF, is performed around the most recent target location. Texture information is used for feature extraction and object description. However, as only one description of the target is saved, the tracker has a low computational cost and memory footprint. Our informal in-lab testing shown that the Circulant tracker provides robust tracking.

The default implementation of the Circulant Matrices tracker in BoofCV does not work with coloured frames. Our solution first converts the video feed, one frame at a time, to greyscale using a simple RGB averaging function. The tracker returns back only the coordinates of the centre of the ROI, while the original coloured frame is rendered to the screen. Finally, to increase tracking performance, the camera is set to record at a constant resolution of 640 by 480 pixels.

5) User Interface: ABOD3-AR renders the plan directly next to the robot, as seen in Figure 2. A pivot connects the plan to the centre of the user-selected ROI. The PC-targeted version of ABOD3 offers abstraction of information; the full plan is visible by default, but the user has the ability to hide information. This approach works on the large screens that laptops and desktops have. Contrary, at time of writing, phones rarely sport a screen larger than 15cm. Thus, to accommodate the smaller screen estate available on a phone, ABOD3-AR displays only high-level elements by default. Drives get their priority number annotated next to their name and are listed in ascending order. ABOD3-AR shares the same real-time transparency methodology as ABOD3; plan

[1]https://developers.google.com/ar/
[2]https://boofcv.org/

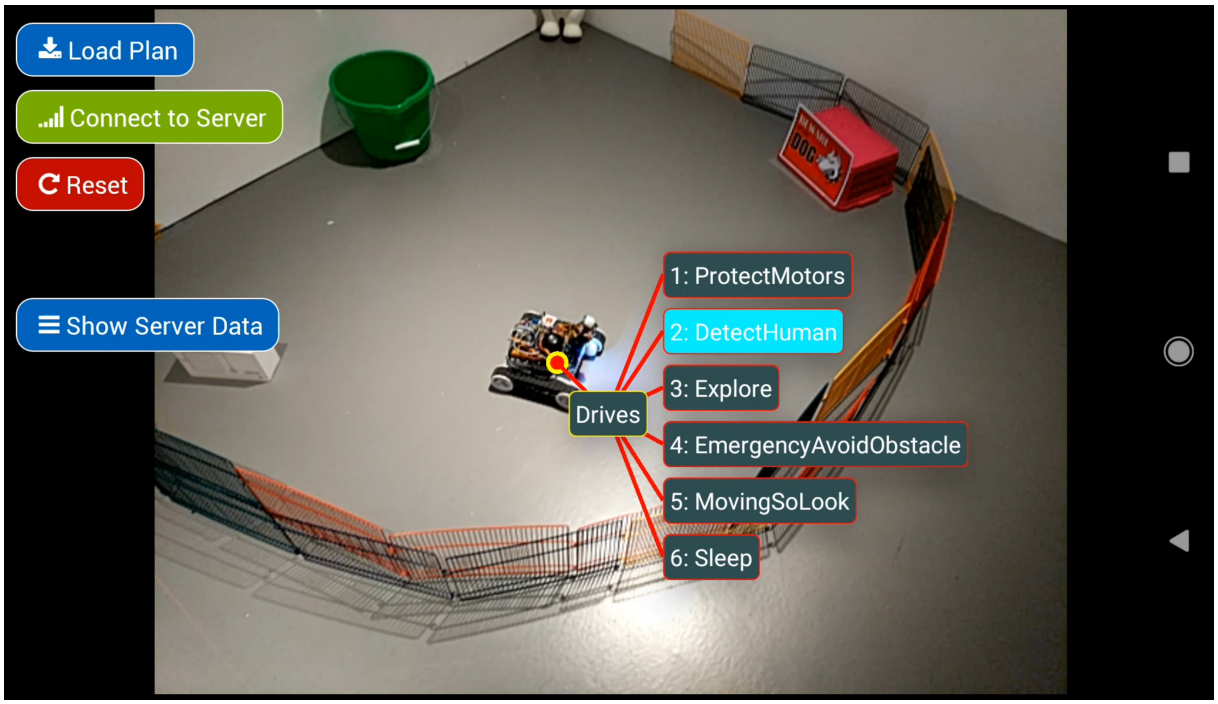

Fig. 2. Screenshot of ABOD3-AR demonstrating its real-time debugging functionality. The plan is rendered next to the robot with the drives shown in a hierarchical order based on their priority. The robot here is executing one of its predefined behaviours —detecting for humans and lighting up its LEDs.

Fig. 3. The ABOD3 Graphical Transparency Tool displaying a POSH plan in debugging mode. The highlighted elements are the ones recently called by the planner. The intensity of the glow indicates the number of recent calls. ABOD3 (used as an IDE for the entire hierarchical cycle) show everything ABOD3-AR uses parts only (2 levels only).

elements light up as they are used, with an opposing thread dimming them down over time.

Like its 'sibling' application, ABOD3-AR is aimed to be used by both end users and expert roboticists. A study conducted by [18] demonstrates how users of AR applications aimed at developers that provide transparency-related information require an AR interface that visualizes additional technical content compared to naive users. These results are in-line with good practices [2] on how different users require different levels of abstraction and overall amount of information. Still, we took these results into consideration by allowing low-level technical data to be displayed in ABOD3-AR upon user request. A user can tap on elements to expand their substree. In order to avoid overcrowding the screen, plan elements not part of the subtree 'zoomed in' become invisible (see Figure 4). [18] shows that technical users in an AR application prefer to have low-level details. Hence, we added an option to enable display of the Server data, in string format, as received by ABOD3-AR.

III. User Study

A. Experimental Design

A user study was carried out to investigate the effectiveness of ABOD3-AR. The study ran over five days. The principle hypothesis of this experiment is that observers of a robot with access to ABOD3-AR will be able to create more accurate mental models. In this section, we present our results, and discuss how ABOD3-AR provides an effective alternative to ABOD3 as a means to provide robot transparency. Moreover, we argue that our results demonstrate that the implementation of transparency with ABOD3-AR increases not only the trust towards the system, but also its likeability.

The R5 robot is placed in a small pen with a selection of objects, e.g. a plastic duck. The participants are asked to observe the robot and then answer our questionnaires. The participants are split in two groups; Group 1 used the AR app and Group 2 did not use the app. Participants are asked to observe the robot for at least three minutes. A total of 45 participants took part in the experiment ($N = 45$). The majority of users were aged 36 to 45. Each group had same number of females and males. Although they worked

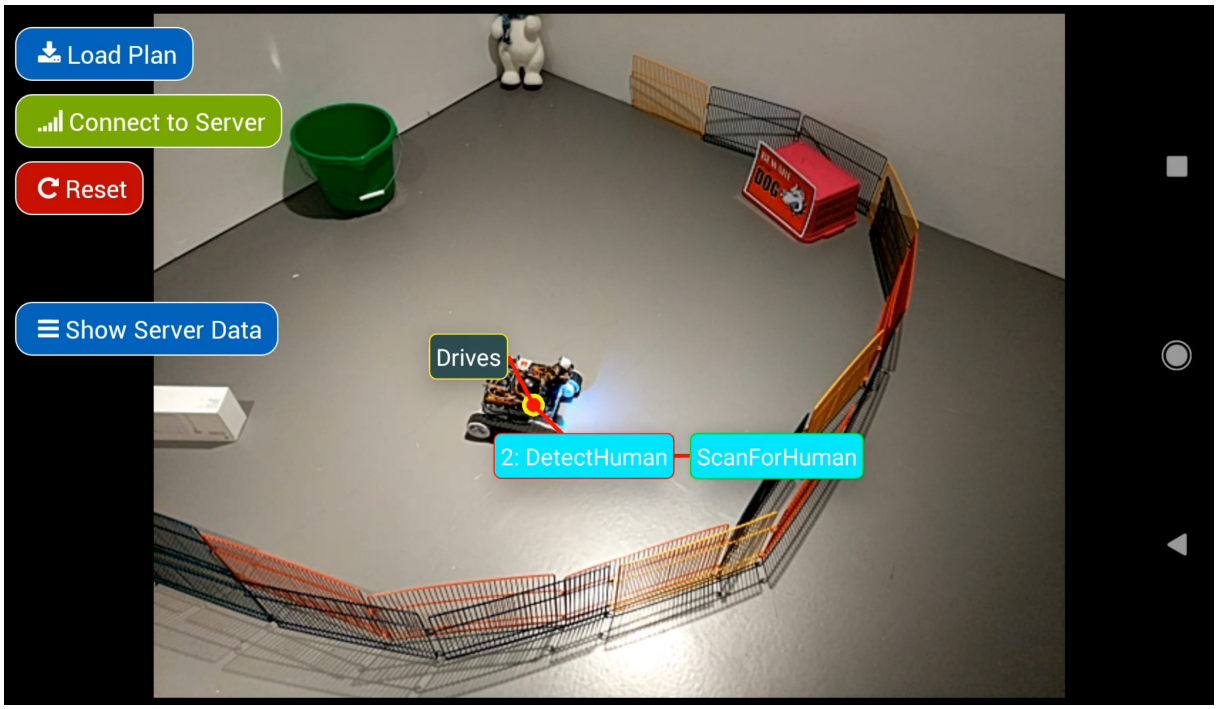

Fig. 4. Screenshot from ABOD3-AR showing how a user can access additional information for a plan element by clicking on it. Other plan elements of its same level, seen in Figure 2, become hidden to increase available screen estate.

regularly with computers, most of them did not have a STEM background — This was the main difference with participants in previous research [19].

The *Godspeed questionnaire* by [20] is used to measure the perception of an artificial embodied agent with and without access to transparency-related information. These are standard questions often used in research regarding Human Robot Interaction (HRI) projects, and also used in similar research [21]. We used a Likert scale of 1 to 5. In addition to the standard Godspeed questionnaire, participants were asked to answer the questions shown in Table I. The first question was added, as a follow up to our previous work, to test difference in perceiving the robot as *thinking* between the two groups. Questions 3 and 4 were added to investigate if transparency increases trust. Finally, the last questions are included to gather additional empirical evidence on the effectiveness of ABOD3-AR as a means to provide real-time transparency.

TABLE I

ABOD3-AR EXPERIMENT: ADDITIONAL QUESTIONS GIVEN TO ALL PARTICIPANTS.

Ref. No.	Question	Response
1.	Is the robot thinking?	Yes/No
2.	Would you feel safe to interact with the robot (for example putting your hand in front of it))?	Yes/No
3.	Would you trust a robot like this in your home?	Yes/No
4.	In your own words, what do you think the robot is doing?	Free Text

B. Results

Individuals who had access to ABOD3-AR were more likely to perceive the robot as *alive* (M = 3.27, SD = 1.202) compare the ones without access to the app; $t(43)$ = –*0.692* and p = 0.01. Moreover, participants in the no-transparency condition described the robot as more *stagnant* (M = 3.30, SD = 0.926) compare to the ones in Group 2 (M = 414, SD = 0.710) who described the robot as *Lively*; $t(43)$ = –3.371, p = 02. Finally, in the ABOD3-AR condition, participants perceived the robot to be *friendlier* (M = 3.17, SE = 1.029) than participants in Group 1 (M = 3.77, SE = 0.869); $t(43)$ = –2.104, p = 041. No other significant results were reported. These results are shown in Table II.

Only 41 from our participants answered the question *Is the robot thinking?*. To test the null hypothesis that access to ABOD3-AR does not increase the perception of *thinking*, we run a Chi-square test in the contingency table shown in Table III. ABOD3-AR does not increase the perception of thinking; χ^2 = 0.0232, p = 0.878828, and DF = 1.

Table IV shows the results gathered for the question "Would you feel safe to interact with the robot (for example putting your hand in front of it)?" There is no significant interaction between the two groups; p = 1. Unfortunately, just 20 participants per group answered the question "Would you trust a robot like this in your home?" We run a Chi-square test on our results (Table V) which returned back χ^2 = 4.2857, p = 0.038434, DF = 1, demonstrating that the results are significant. Access to ABOD3-AR helps users increase their trust to the machine.

Finally, randomly-picked answers of the free-text question

Question	Group 1 ($N = 23$)	Group 2 ($N = 22$)	p-value
Dead - Alive	2.39 (σ=0.988)	3.27 (σ=1.202)	**0.01**
Stagnant - Lively	3.30 (σ=0.926)	4.14 (σ=0.710)	**0.02**
Mechanical - Organic	1.91 (σ =1.276)	1.45 (σ =0.8)	0.158
Artificial - Lifelike	1.96 (σ =1.065)	1.95 (σ =1.214)	0.995
Inert - Interactive	3.26 (σ =1.176)	3.68 (σ =1.041)	0.211
Dislike - Like	3.57 (σ =0.728)	3.77 (σ =1.02)	0.435
Unfriendly - Friendly	3.17 (σ=1.029)	3.77 (σ0.869)	**0.041**
Unpleasant - Pleasant	3.43 (σ0.788)	3.77 (σ1.066)	0.232
Unintelligent - Intelligent	3.17 (σ=0.937)	3.14 (σ=1.153)	0.922
Bored - Interested	3.80 (σ=0.834)	4.19 (σ=0.680)	0.110
Anxious - Relaxed	4.15 (σ=0.933)	3.81 (σ=1.167)	0.308

TABLE II

ABOD3-AR EXPERIMENT: MEANS (SD) OF THE RATINGS GIVEN BY EACH GROUP AT VARIOUS QUESTIONS. THE RESULTS SHOW THAT PARTICIPANTS IN GROUP 2 PERCEIVE THE ROBOT AS SIGNIFICANTLY MORE *alive* IF THEY HAD USED ABOD3-AR COMPARE TO PARTICIPANTS IN GROUP 1. MOREOVER, PARTICIPANTS IN THE NO-APP CONDITION DESCRIBED THE ROBOT AS MORE *stagnant* COMPARED TO THE ONES IN GROUP 2. FINALLY, IN THE ABOD3-AR CONDITION, PARTICIPANTS PERCEIVED THE ROBOT TO BE *friendlier* THAN PARTICIPANTS IN GROUP 1.

TABLE III

ABOD3-AR EXPERIMENT: THE CONTINGENCY TABLE FOR THEN ANSWERS GIVEN TO THE BINARY QUESTION "IS THE ROBOT THINKING?" (N = 41). THERE IS NO SIGNIFICANT DIFFERENCE BETWEEN THE TWO GROUPS WITH $\chi^2 = 0.0232$, p = 0.878828, AND TEXTITDF = 1. IN CURVY BRACKETS THE EXPECTED CELL TOTALS AND IN SQUARE BRACKETS THE CHI-SQUARE STATISTIC FOR EACH CELL.

Result	Group 1 ($N = 21$)	Group 2 ($N = 20$)
Yes	11 (10.76) [0.01]	10 (10.24)[0.01]
No	10(10.24)[0.01]	10 (9.76)[0.01]

TABLE IV

ABOD3-AR EXPERIMENT: THE CONTINGENCY TABLE FOR THEN ANSWERS GIVEN TO THE BINARY QUESTION *Do you think the robot is performing the way it should be?* (N = 45). THERE IS NO SIGNIFICANT DIFFERENCE BETWEEN THE TWO GROUPS; p = 0.6078.

Result	Group 1 ($N = 20$)	Group 2 ($N = 21$)
Yes	19	20
No	1	1

"In your own words, what do you think the robot is doing?" are included bellow. Note, multiple participants in Group 1 referred to the robot as a 'he', while none of the Group 2 participants did. Group 1:

- [the robot is] Trying to build a memory of the distance between itself and the objects to judge its own location in space.
- [the robot is] Processing Data.
- [the robot is] Random.
- [the robot] is actively looking for something specific. At some points he believes he has found it (flashes a light) but then continues on to look.
- [the robot is] Taking pictures of the objects.
- [the robot is] Occasionally taking pictures.
- He is looking for something.

Group 2:

TABLE V

ABOD3-AR EXPERIMENT: THE CONTINGENCY TABLE FOR THEN ANSWERS GIVEN TO THE BINARY QUESTION *Would you trust a robot like this in your home?* (N = 41). THERE IS NO SIGNIFICANT DIFFERENCE BETWEEN THE TWO GROUPS WITH $\chi^2 = 4.2857$, p = 0.038434, DF = 1 IN CURVY BRACKETS THE EXPECTED CELL TOTALS AND IN SQUARE BRACKETS THE CHI-SQUARE STATISTIC FOR EACH CELL.

Result	Group 1 ($N = 21$)	Group 2 ($N = 20$)
Yes	11 (14)[0.64]	17 (14)[0.64]
No	9 (6)[1.5]	3 (6)[1.5]

- [the robot is] Exploring its surroundings and trying to detect humans.
- [the robot is] Roaming detecting objects and movement through sensors.
- [the robot is] The robot likes to scan for obstacles, humans and find new paths to follow it can understand animals and obstacles.
- [the robot is] imitating commands, responding to stimuli.
- [the robot is] registering programmed behaviours and connecting it to it surroundings.
- [the robot 's] movement looks random I would say it is using sensors to avoid the obstacles.
- [the robot is] Occasionally taking pictures.

IV. DISCUSSION

We found a statistically significant difference ($p < 0.05$) in three Godspeed questions: *Dead/Alive*, *Stagnant/Lively*, and *Unfriendly/ Friendly*. The R5 has connecting wires and various chipsets exposed. Yet, participants with access to ABOD3-AR were more likely to describe the robot as *alive*, *lively*, and *friendly*. All three dimensions had mean values over the 'neutral' score of 3. Although not significantly higher, there was an indicatively increased attribution of the descriptors *Interactive* and *Pleasant*; again both with values over the neutral score. At first glance, these results suggest an increase of anthropomorphic —or at least

biologic— characteristics. However, transparency decreased the perception of the robot being *Humanoid* and *Organic*; both characterizations having means below the neutral score.

Action selection takes place even when the robot is already performing a lengthy action, e.g. moving, or when it may appears 'stuck', e.g. it is in `Sleep` drive to save battery. The transparency display makes the constant selection and performance of actions visible to the users and, therefore, the robot to be appear as more 'lively'. These results also support that a sensible implementation of transparency, in line to the principles set by [2], can maintain or even improve the user experience and engagement. It is, however, worth noting that unlike our previous work, we found no statistical difference in the question "Is the robot thinking?".

An explanation for the high levels of *Interest* (3.8 mean for Group 1 and 4.19 mean for Group 2) is that embodied agents —unlike virtual agents— are not widely available. Participants in both groups may have been intrigued by the ideal of encountering a real robot. Nonetheless, our findings indicate that transparency does not necessary reduces the utility or 'likeability' of a system. Instead, the use of a transparency display can increase the utility and likeability of a system, as it could provide a more interactive experience.

Our results also suggest an increase of trust, when the user is in the transparency condition. There was a statistical significant difference between the number of people who answered *Yes* in the question "Would you trust a robot like this in your home?" between the two groups. Users with ABOD3-AR were more likely to have a robot like the R5 at home. Further work that includes a more detailed questionnaire is required to explore this. Our hypothesis is that some of their concerns were addressed; for example, subjects with ABOD3-AR could see that the robot does not have any audiovisual recording equipment that could compromise the privacy of its users.

On the contrary, there was no significant difference in the perception of safety between the two groups. Both groups overwhelming answered *Yes* in the question "Would you feel safe to interact with the robot (for example putting your hand in front of it) ?" Thus, some participants would feel safe to interact with the robot in a 'neutral' environment, but not feel comfortable having it at their homes. Still, this was expected as the R5 does not have any sharp edges or other threatening-looking characteristics. Moreover, the robot moves at slow speeds, something directly observables, alleviating any concerns for causing damage from an accidental impact. Furthermore, there is no significance difference between the two groups in questions *Anxious/Relaxed*, *Calm/Agitated*, and *Quiescent/Surprised* designed to measure the perceived Safety of the participant.

Finally, the answers found in the freetext question indicate that ABOD3-AR is an effective mean of producing a significantly better understanding of what a robot's functionality and capabilities are. This was expected, as previous work demonstrate that even naive users can develop more accurate mental models for a robot, when the artifact is accompanied by a transparency provision [22]. Interestingly, some of the participants in our control group, without access to ABOD3-AR, referred the robot as a 'he', while none of the participants in the transparency condition did. This indicates, in addition to our discussion above, a decrease of anthropomorphising the machine, albeit the higher rating of the robot in the questions *Dead/Alive* and *Stagnant/Lively*.

V. Conclusions and Future Work

In this paper we presented a new tool, ABOD3-AR, which runs on modern mobile phones to provide transparency-related information to end users. Our tool uses a purpose-made user interface with augmented-reality technologies to display the real-time status of any robot running the Instinct planner.

As far as we are aware this is the first use of mobile augmented reality focusing solely on increasing transparency in robots and users' trust towards them. Previous research regarding transparency in robots relied on screen and audio output or non real-time transparency. Building upon past research, we provide an affordable, compact solution, which makes use of augmented reality. There are several characteristics of augmented reality that makes it a promising platform to provide transparency information for both industrial and domestic robots. These include the affordability of AR enabled devices, its availability on multiple platforms such as mobile phones and tablets, the rapidly increasing progress in mobile processors and cameras, and the convenience of not requiring headsets or other paraphernalia unlike its competitor virtual reality.

The work presented in this paper is part of a research programme to investigate the effects of transparency on the perceived expectations, trust, and utility of a system. Initially this is being explored using the non-humanoid R5 robot and later we plan to expand the study using the *Pepper* humanoid robot manufactured by SoftBank Robotics. We argue that humanoid appearance will always be deceptive at the implicit level. Hence, we want see how explicit understanding of the robot's machine nature effects its perceived utility. Moreover, if transparency alters trust given to the machine by its users.

Planned future work also aims at improving the usability of the application further. Currently, the robot-tracking mechanism requires the user to manually select an area of ROI which contains the robot. Future versions of ABOD3 - AR would skip this part and replace it with a machine learning (ML) approach. This will enable the app to detect and recognize the robot by a number of features, such as colour and shape. The app will also be enhanced to be able to retrieve the robot type and plan of execution from a database of robots.

ACKNOWLEDGMENT

We would like to thank The Edge Arts Centre for hosting us during data collection. Thanks also to our helpful reviewers. We also acknowledge EPSRC grant [EP/L016540/1] for funding Rotsidis and Theodorou. Finally, we thank AXA Research Fund for part-funding Bryson.

References

[1] R. H. Wortham and A. Theodorou, "Robot transparency, trust and utility," *Connection Science*, vol. 29, no. 3, pp. 242–248, jul 2017. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/09540091.2017.1313816

[2] A. Theodorou, R. H. Wortham, and J. J. Bryson, "Designing and implementing transparency for real time inspection of autonomous robots," *Connection Science*, vol. 29, no. 3, pp. 230–241, jul 2017. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/09540091.2017.1310182

[3] M. Boden, J. Bryson, D. Caldwell, K. Dautenhahn, L. Edwards, S. Kember, P. Newman, V. Parry, G. Pegman, T. Rodden, T. Sorrell, M. Wallis, B. Whitby, and A. Winfield, "Principles of robotics: regulating robots in the real world," *Connection Science*, vol. 29, no. 2, pp. 124–129, 2017. [Online]. Available: https://doi.org/10.1080/09540091.2016.1271400

[4] J. Bryson, Joanna and A. Theodorou, "How Society Can Maintain Human-Centric Artificial Intelligence," in *Human-centered digitalization and services*, M. Toivonen-Noro, E. Saari, H. Melkas, and M. Hasu, Eds., 2019.

[5] J. B. Lyons, "Being Transparent about Transparency : A Model for Human-Robot Interaction," *Trust and Autonomous Systems: Papers from the 2013 AAAI Spring Symposium*, pp. 48–53, 2013.

[6] J. D. Lee and N. Moray, "Trust, self-confidence, and operators' adaptation to automation," *International Journal of Human - Computer Studies*, vol. 40, no. 1, pp. 153–184, jan 1994.

[7] M. T. Dzindolet, S. A. Peterson, R. A. Pomranky, L. G. Pierce, and H. P. Beck, "The role of trust in automation reliance," *International Journal of Human Computer Studies*, vol. 58, no. 6, pp. 697–718, 2003.

[8] R. H. Wortham, A. Theodorou, and J. J. Bryson, "Robot transparency: Improving understanding of intelligent behaviour for designers and users," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10454 LNAI, pp. 274–289, 2017.

[9] A. Theodorou, "AI Governance Through a Transparency Lens," Ph.D. dissertation, University of Bath, 2019.

[10] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoper. Virtual Environ.*, vol. 6, no. 4, pp. 355–385, Aug. 1997. [Online]. Available: http://dx.doi.org/10.1162/pres.1997.6.4.355

[11] J. J. Bryson, "Intelligence by Design : Principles of Modularity and Coordination for Engineering Complex Adaptive Agents," Ph.D. dissertation, 2001.

[12] ——, "Action selection and individuation in agent based modelling," in *Proceedings of Agent 2003: Challenges in Social Simulation*, D. L. Sallach and C. Macal, Eds. Argonne, IL: Argonne National Laboratory, 2003, pp. 317–330.

[13] S. Gaudl, S. Davies, and J. J. Bryson, "Behaviour oriented design for real-time-strategy games: An approach on iterative development for STARCRAFT AI," *Foundations of Digital Games Conference*, pp. 198–205, 2013.

[14] R. A. Brooks, "New Approaches to Robotics," *Science*, vol. 253, no. 5025, pp. 1227–1232, 1991. [Online]. Available: http://people.csail.mit.edu/brooks/papers/new-approaches.pdf http://www.sciencemag.org/cgi/doi/10.1126/science.253.5025.1227

[15] R. H. Wortham, S. E. Gaudl, and J. J. Bryson, "Instinct: A biologically inspired reactive planner for intelligent embedded systems," *Cognitive Systems Research*, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389041717301912

[16] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7575 LNCS, no. PART 4, 2012, pp. 702–715. [Online]. Available: http://www.robots.ox.ac.uk/ joao/publications/henriques_eccv2012.pdf

[17] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-Learning-Detection." *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 1, pp. 1409–1422, 2011.

[18] E. K. Subin, A. Hameed, and A. P. Sudheer, "Android based augmented reality as a social interface for low cost social robots," in *Proceedings of the Advances in Robotics on - AIR '17*. New York, New York, USA: ACM Press, 2017, pp. 1–4.

[19] R. H. Wortham, A. Theodorou, and J. J. Bryson, "Improving robot transparency: Real-time visualisation of robot AI substantially improves understanding in naive observers," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, vol. 2017-Janua. IEEE, aug 2017, pp. 1424–1431. [Online]. Available: http://ieeexplore.ieee.org/document/8172491/

[20] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International Journal of Social Robotics*, vol. 1, no. 1, pp. 71–81, Jan 2009. [Online]. Available: https://doi.org/10.1007/s12369-008-0001-3

[21] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, "Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '15. New York, NY, USA: ACM, 2015, pp. 141–148. [Online]. Available: http://doi.acm.org/10.1145/2696454.2696497

[22] R. H. Wortham, "Using Other Minds: Transparency as a Fundamental Design Consideration for Artificial Intelligent Systems," Ph.D. dissertation, 2018.