Comparing Task-Based and Socially Intelligent Behaviour in a Robot Bartender

Manuel Giuliani[1] Ronald P.A. Petrick[2] Mary Ellen Foster[3]
Andre Gaschler[1] Amy Isard[2] Maria Pateraki[4] Markos Sigalas[4]
[1] fortiss GmbH, Munich, Germany [2] School of Informatics, University of Edinburgh, Edinburgh, UK
[3] School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, UK
[4] Foundation for Research and Technology – Hellas (FORTH), Heraklion, Crete, Greece
{giuliani,gaschler}@fortiss.org {rpetrick,amyi}@inf.ed.ac.uk m.e.foster@hw.ac.uk {pateraki,sigalas}@ics.forth.gr

ABSTRACT

We address the question of whether service robots that interact with humans in public spaces must express socially appropriate behaviour. To do so, we implemented a robot bartender which is able to take drink orders from humans and serve drinks to them. By using a high-level automated planner, we explore two different robot interaction styles: in the *task only* setting, the robot simply fulfils its goal of asking customers for drink orders and serving them drinks; in the *socially intelligent* setting, the robot additionally acts in a manner socially appropriate to the bartender scenario, based on the behaviour of humans observed in natural bar interactions. The results of a user study show that the interactions with the socially intelligent robot were somewhat more efficient, but the two implemented behaviour settings had only a small influence on the subjective ratings. However, there were objective factors that influenced participant ratings: the overall duration of the interaction had a positive influence on the ratings, while the number of system order requests had a negative influence. We also found a cultural difference: German participants gave the system higher pre-test ratings than participants who interacted in English, although the post-test scores were similar.

Categories and Subject Descriptors: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – Evaluation/methodology; I.2.9 [Artificial intelligence]: Robotics – Operator interfaces

Keywords: Social robotics; Multi-party interaction

1. INTRODUCTION

The market for service robotics is constantly growing and is projected to increase substantially in the next 20 years [21, 22]. Since the objective of many service robots is to interact with multiple humans, often in dynamic public spaces, it is essential that they possess the appropriate capabilities that are needed for fulfilling their assigned tasks: they need to be able to recognise human speech and non-verbal signals from audio and visual sensors, they need to be able to interpret these input signals to plan their actions, and they need to be able to execute these actions in a safe way to guarantee that their human interaction partners are not in any danger. But is that enough? Is it sufficient for a service robot to execute its assigned task as efficiently as possible, or should it also act in a socially appropriate manner with its human interaction partners?

ICMI'13, December 9–13, 2013, Sydney, Australia

ACM 978-1-4503-2129-7/13/12 ...$15.00.
http://dx.doi.org/10.1145/2522848.2522869.

Figure 1: Robot bartender

In this work, we address these questions using a robot bartender (Figure 1) which has all of the technical abilities listed above: the robot is able to recognise humans as customers, take their drink orders, and serve the ordered drinks, making use of multimodal input and output (Section 3). Moreover, by using a high-level automated planner to control the robot's behaviour (Section 4), we consider two different interaction styles in the bartender setting: on the one hand, we can set the robot to a *task only* setting, where the robot simply asks humans in front of the bar for their drink orders and then serves them their drinks; on the other hand, in the *socially intelligent* setting, the robot also exhibits social behaviour that is derived from the observation of natural bartender interactions. In a user study (Section 5), we tested how experiment participants respond to the two different versions of the robot bartender.

2. RELATED WORK

In recent years, robot bartenders have been used to demonstrate a number of aspects of human-robot interaction. For example, Wosch et al. [36] showed an efficient motion planner that guaranteed safe motions of a robot bartender that worked closely with humans; Masuda and Misaki [19] presented "T-Bartender," a robot that was able to serve green tea in a traditional Japanese way; Limbu et al. [17] implemented "FusionBot," a mobile robot bartender that served drinks in homes; while Grigore et al. [15] used a robot bartender as an

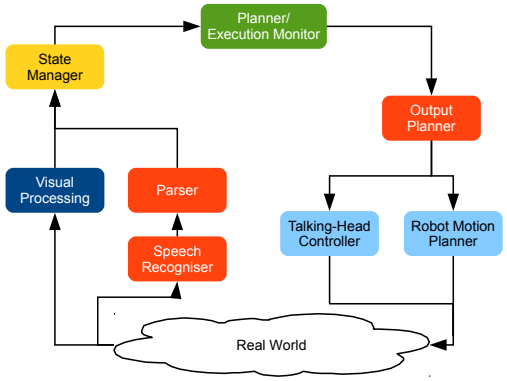

Figure 2: System architecture

example of a human-robot interaction scenario in which the safety of the human could be guaranteed by using verification techniques.

In contrast to the above robot bartenders, which concentrated mainly on the task-based aspects of the bartending scenario, our work fits into the active research area of *social robotics*, which Ge and Matarić [14] define as "the study of robots that interact and communicate with themselves, with humans, and with their environment, within the social and cultural structure attached to their roles." Most current social robots play the role of a companion, often in a long-term, one-on-one relationship with the user—e.g., the systems described by Breazeal [5], Castellano et al. [6], and Dautenhahn [7]. In this context, the primary goal of the robot is to build a relationship with the user through social interaction: the robot is primarily an interactive partner, and any task-based behaviour is secondary to this overall goal. Even for social robots that deal with multiple partners—e.g., those of Matsusaka et al. [20] and Mutlu et al. [23]—social interaction is still the primary goal.

In this work, we address a different style of interaction: the robot bartender—a service robot operating in a public space—interacts with multiple persons over short time spans, with social communication that takes place in the context of cooperative, physically grounded, task-based interaction. This is similar to the multimodal information kiosk of Bohus and Horvitz [4], but with the addition of physical (as opposed to virtual) embodiment, which has been shown to have a large effect on social interaction [e.g., 1, 18].

3. ROBOT HARDWARE AND SOFTWARE

In this work, we use the robot bartender shown in Figure 1. The robot hardware consists of two 6-degrees-of-freedom industrial manipulator arms with grippers, mounted to resemble human arms. Sitting on the main robot torso is an animatronic talking head capable of producing facial expressions, rigid head motion, and synthesised speech. The robot is equipped with two stereo cameras and two Kinect sensors: we use the depth sensor of one of the Kinects to support stereo camera-based vision processing, and the microphone array of the other Kinect for automatic speech recognition (ASR).

Figure 2 shows the software architecture of the robot. The robot senses events in its surroundings by a *speech recognition* and a *visual processing* module. The *parsing* component processes the output from speech recognition. The *state manager* takes the output from visual processing and parsing and transforms it into symbolic representations for the *planner / execution monitor* module. The planner then selects high-level actions for the robot, which are processed by the *output planner* for execution as concrete actions by the *talking-head controller* and the *robot motion planner*.

The initial version of the robot bartender [10] supported simple interactions with up to two customers where the only thing that a customer could do was order a drink. An initial study confirmed that the system performed successfully in this simple scenario. Since that initial version, the software components of the robot bartender have been extended in all areas, as described in the following sections.

3.1 Visual Processing

The vision system [24] is responsible for tracking multiple humans entering and leaving the bar area in front of the robot. To do so, we use a tracking and classification algorithm for hands and faces [2], to track multiple objects and people in a common framework and provide support for the detection of groups of humans in front of the robot. The algorithm periodically invokes a GPU implementation of an appearance-based face detector [33] in order to update the appearance models of existing hypotheses, making the vision system more robust against illumination changes in the environment.

A second task of the vision system is to detect the torso orientation and head pose of all human users in a scene, as this information forms the basis for determining a user's focus of attention. We detect torso orientation through a tracking approach, which uses depth information and extracts the full body pose in 3D. This method removes the need for an initialisation phase, while also exhibiting robustness in dynamic settings. In the previous version of the system [10], a Hidden Markov Model tracked the orientation of the human torso in the 1D space of all possible orientations [31]. Although adequate for simple cases, this method could not effectively cope with the problems of a real time scenario. Instead, we employ a data-driven model-based method for 3D torso pose estimation from RGB-D image sequences, using 3D geometric primitives to approximate the shoulder joints and the user's torso, and to derive the 3D body pose via a global optimisation scheme [24]. Shoulder joints are modelled as spheres, while the torso is modelled as an ellipsoid. The models exhibit features that guarantee robust adaptation on RGB-D data, unaffected by scale, pose, or body type variations across users.

The employed methodology consists of three major steps: (i) *agent segmentation*, where robust face identification triggers detection and segmentation of the human body silhouette, (ii) *shoulder joint approximation*, where sets of points on the RGB silhouette are selected, given the location of the face, delineating possible shoulder or armpit areas, and (iii) *body pose estimation*, where a set of 3D points, approximately along the user's upper-body, is selected and used for torso approximation via an ellipsoid, driven by the detection of the shoulder joints.

3.2 Linguistic Interaction

The linguistic interaction system consists of two parts: on the one hand, there are components to recognise and understand spoken utterances (the speech recogniser and parser); on the other hand, there is a dedicated module for generating embodied natural language (the output planner). All of the linguistic components are able to handle natural language in English and German, using a common core grammar to allow the rest of the system to operate in a language-independent manner. For speech recognition we use Microsoft Kinect and the associated Microsoft Speech API. The current system uses minimum confidence thresholds for speech to avoid processing incorrectly recognised speech; based on testing in the robot lab environment, the thresholds are currently set to 50% for English and 30% for German.

The output planner coordinates the movements of the robot's head and arms with aligned speech, using the animatronic head to generate embodied speech. For language parsing and language generation, we use a bi-directional, bilingual OpenCCG grammar [35]. On the input side, the parsed input is passed to the state manager along with an estimate of the source angle from the Kinect microphone array;

on the output side, the language generation component receives an XML representation from the planner which contains instructions for the multimodal output [28]. In addition to the basic task-based utterances, the system supports a number of additional utterances on the input and output sides to deal with the extra interactions necessary for social behaviour, such as group drink ordering.

3.3 State Management

The primary role of the state manager is to turn the continuous stream of messages produced by the low-level input and output components of the system into a discrete representation of the world, the robot, and all entities in the scene, integrating social, interaction-based, and task-based properties. The state is modelled as a set of *relations* such as `facePos(A)=(x,y,z)` or `closeToBar(A)` [27]. The state manager provides a query interface to allow other system components access to the relations stored in the state, and also publishes an updated state report every time there is a change which might require a response from the system (e.g., a customer appears, begins seeking attention, or makes a drink order).

In addition to storing all of the low-level sensor information, the state manager also infers additional relations that are not directly reported by the sensors. For example, it fuses information from vision and speech to determine which user should be assigned a recognised spoken contribution, and estimates which customers are in a group. Most importantly in the current social domain, the state manager also provides a constant estimate of whether each customer is currently seeking attention from the bartender (`seeksAttention(A)`). While the initial version of this estimator used a simple, hand-coded rule based on the observation of human behaviour in real bars [16], the current version [11] instead makes use of a supervised learning classifier trained on labelled recordings of humans interacting with the first version of the robot bartender.

3.4 High-level Planning and Monitoring

The high-level automated planner is responsible for managing interactions with customers, tracking multiple drink orders, and gathering additional information as needed with follow-up questions [27]. To do this, the planner takes state reports from the state manager and selects actions to be executed on the robot. Plans are generated using PKS (Planning with Knowledge and Sensing) [25, 26], a conditional planner that works with incomplete information and sensing actions. Unlike many general-purpose planners, PKS operates at the *knowledge level* and reasons about how its knowledge, rather than the world, changes due to action. PKS's knowledge state is represented symbolically by a set of five databases, each of which models a particular type of information, interpreted in a modal logic of knowledge. To ensure efficient reasoning, PKS restricts the knowledge it can represent while ensuring it is expressive enough to model many types of information that arise in common planning scenarios.

Actions in PKS are described by a set of *preconditions* which define the conditions that must be true for an action to be applied, and a set of *effects* that capture the changes an action makes to the planner's knowledge state. Preconditions ask simple questions about the planner's knowledge state, while effects modify the planner's databases in a STRIPS-like [9] manner, through additions and deletions which correspond to changes to the knowledge state. PKS constructs plans by reasoning about actions using a forward-chaining search process, and can build contingent plans by considering the potential outcomes arising from certain kinds of knowledge.

PKS is also aided by an execution monitor which controls replanning. The monitor takes as input a PKS plan and a description of the sensed state provided by the state manager. The task of the monitor is to assess how close an expected, planned state is to a sensed state in order to determine whether the current plan should continue to be executed. To do this, it tries to verify that the current state still permits the next action (or set of actions) in the plan to be executed. In the case such a test fails, the planner is directed to build a new plan, using the sensed state as its initial state.

3.5 Robot Behaviours

The robot motion planner communicates with the output planner to control the manipulation of bottles. In particular, an internal grasp planner works with arbitrary locations at run time, allowing drinks to be served to each customer's location. To calculate possible grasp configurations at run time, a numerical inverse kinematics algorithm is used, that includes task-space constraints, as implemented in the Robotics Library [29]. This allows us to define a degree-of-freedom around the axis of rotation of a bottle to obtain the full null space of solutions. In this null space, the solution with maximum distance from robot joint boundaries is found in order to achieve a good posture far away from the workspace boundaries. The trajectory to this grasping pose is then generated by quintic polynomial interpolation in joint space to generate very smooth robot motions.

A real-time collision detection layer works with the grasp planner to support dynamic locations at run time. A convex decomposition procedure is applied to store the robot geometry and its environment model as a small set of convex polyhedra, including a small offset for safety [13]. Using this representation, run-time collision checking only needs to process convex-convex checks, which are of linear complexity in the number of vertices. In our system, all collision checks can be run in less than a millisecond. Furthermore, all robot trajectories are verified within the hardware real-time control loop and a safe distance to all static obstacles is maintained, even if the robot is commanded to move to an invalid position.

In response to messages from the output planner, the animatronic robot head also generates synthesised embodied speech together with facial expressions and gestures such as nods. In addition, the robot head is able to look directly at defined positions, and can be turned towards customers while the robot is talking to them; the coordinates are derived from the locations reported by the vision system via the state manager.

4. PLANNING DOMAINS

As mentioned in Section 3.4, the high-level behaviour of the robot bartender is controlled by the PKS planner, which is guided by a symbolic domain model that includes a specification of the physical, sensory, and linguistic actions available to the planner. To allow us to assess the impact of social behaviour on the robot bartender's interactions, we have created two different PKS planning domains: one that implements a purely task-based view of the bartending scenario, and one that also incorporates a range of social behaviours as observed in data collected from a study of real bartenders.

4.1 A Task-Based Bartender

At a purely task-based level, the behaviour of a bartender can in general be summarised by the following two rules:

- If the bartender does not know a customer's drink order, ask them what they want.
- If a customer with a known drink order has not been served, serve them their drink.

At the planning level, we can model similar behaviour for the robot bartender by using the following PKS actions, included in the task-based version of the domain:

ask-drink(?a)	Ask customer ?a for a drink order,
serve(?a,?d)	Serve drink ?d to customer ?a.

Using these actions, the planner could construct the following plan for serving three customers (A1, A2, A3) in a bar:

ask-drink(A1),	[Ask A1 for drink order]
ask-drink(A2),	[Ask A2 for drink order]
serve(A2,request(A2)),	[Give the drink to A2]
ask-drink(A3),	[Ask A3 for drink order]
serve(A3,request(A3)),	[Give the drink to A3]
serve(A1,request(A1)).	[Give the drink to A1]

In the above plan, each customer's drink order is taken and the drink is served. (The term request(A) acts as a placeholder for the actual drink ordered by customer A.) However, the domain places no constraints on the order in which customer transactions take place. In particular, a customer whose drink order is taken early in the plan (e.g., A1) may not necessarily be served until much later, and possibly after all other customers have been served.

4.2 A Socially Intelligent Bartender

While the above domain captures the core task-based aspects of bartending, it fails to include many of the behaviours real bartenders exhibit in a natural context. In particular, observations of real bartenders interacting with customers [16] indicate that they also made use of a range of social behaviours, including the following:

- Only customers who are seeking to engage with the bartender are addressed.
- Customers are dealt with in the order that they arrive.
- The bartender acknowledges all drink orders as soon as they are given.
- If a group of customers approaches the bar, the bartender takes all of their drink orders in sequence and then serves all of the requested drinks.
- If a new customer appears while the bartender is engaged in a transaction, the customer is acknowledged with a nod, and then served after the current transaction is completed.

These additional social behaviours have been formalised in a second PKS planning domain that includes the following actions:

greet(?a,?g)	Greet customer ?a in group ?g,
ask-drink(?a,?g)	Ask ?a in group ?g for a drink order,
serve(?a,?d,?g)	Serve drink ?d to ?a in group ?g,
bye(?a,?g)	End an interaction with ?a in group ?g,
wait(?a,?g)	Tell ?a in group ?g to wait (e.g., nod),
ack-order(?a,?g)	Acknowledge ?a's order in group ?g,
ack-wait(?a,?g)	Thank ?a in group ?g for waiting.

In addition to ask-drink and serve from the task-based domain, this domain includes actions for controlling particular aspects of a transaction (e.g., greet, wait), as well as actions for acknowledging certain customer behaviours (e.g., ack-order, ack-wait). Most notably, all actions are modelled around the idea that customers may be part of groups, which affects how individual customers are served. For instance, consider the situation where there are three customers in the bar: A1 and A2 are part of a group denoted by G1, and A3 is in a singleton group G2. In this case, the planner might build the following plan for serving drinks to all customers:

wait(A3,G2),	[Tell G2 to wait]
greet(A1,G1),	[Greet group G1]
ask-drink(A1,G1),	[Ask A1 for drink order]
ack-order(A1,G1),	[Acknowledge A1's order]
ask-drink(A2,G1),	[Ask A2 for drink order]
ack-order(A2,G1),	[Acknowledge A2's order]
serve(A1,request(A1),G1),	[Give the drink to A1]
serve(A2,request(A2),G1),	[Give the drink to A2]
bye(A2,G1),	[End G1's transaction]
ack-wait(A3,G2),	[Acknowledge G2's waiting]
ask-drink(A3,G2),	[Ask A3 for drink order]
ack-order(A3,G2),	[Acknowledge A3's order]
serve(A3,request(A3),G2),	[Give the drink to A3]
bye(A3,G2).	[End G2's transaction]

The plan first directs the robot to tell group G2 to wait before transacting with group G1. The robot then collects drink orders from all customers in G1 before serving their drinks and completing the transaction. After that, the robot thanks group G2 (i.e., customer A3) for waiting before taking the final drink order and serving the drink.

5. USER EVALUATION

The two planning domains described above can result in very different interactive behaviour from the robot bartender in practice. For example, Figure 3 shows how each version of the bartender deals with three customers. At the end of both interactions, all three customers have received their desired drink. However, unlike in Interaction 1, the customers in Interaction 2 are served in the same sequence that they arrive at the bar; each customer's drink order is acknowledged after it is given; and the bartender takes into account that A1 and A2 are together by serving them as a group. The robot in Interaction 2 also deals with customer A3 more appropriately: it acknowledges A3's arrival with a nod, and then completes the current transaction before thanking A3 for waiting and serving A3.

To assess the practical impact of the two planning domains, we carried out an experiment in which human experiment participants enacted the scenario shown in Figure 3. Before and after the study, the participants answered a questionnaire regarding their experience of interacting with the robot bartender. Half of the participants interacted with the purely task-based bartender, which served drinks without social actions as in Interaction 1, while the other half interacted with the socially intelligent bartender, which included extra actions as in Interaction 2. In addition to the questionnaire, we also gathered a range of other measures assessing the quality of the interaction based on data gathered from system log files.

Interaction 1 (Task-based)	**Interaction 2 (Socially intelligent)**
Two people, A1 and A2, approach the bar together	
Robot (to A1): How can I help you?	Robot (to A1): How can I help you?
A1: A pint of cider, please.	A1: A pint of cider, please.
	Robot (to A1): Okay
A third person, A3, approaches the bar	
Robot (to A3): How can I help you?	Robot: (Nods at A3)
A3: I'd like a pint of beer.	Robot: (to A2) And you?
Robot: (Serves A3)	A2: A glass of white wine.
Robot: (To A2) What would you like?	Robot (to A2): Okay.
A2: A glass of white wine.	Robot: (Serves A1 and A2)
Robot: (Serves A2)	Robot (to A3): Thanks for waiting.
Robot: (Serves A1)	How can I help you?
	A3: I'd like a pint of beer.
	Robot (to A3): Okay.
	Robot: (Serves A3)

Figure 3: Example interactions in a bar setting

5.1 Participants

40 participants (28 male), drawn from university departments outside the robotics group involved in developing the bartender, took part in this experiment. The mean age of the participants was 27.9 (range 16–50), and their mean self-rating of experience with human-robot interaction systems was 2.50 on a scale of 1–5. Each participant could choose whether to interact with the robot in English or German; in total, 26 chose German, while 14 chose English.

5.2 Procedure

Before the experiment, we told participants that their task would be to go to the robot bartender to order a drink. Without showing the robot to the participants, we then asked them to rate their expectations in the interaction on a computer-based questionnaire. After filling out the questionnaire, we introduced the participants to the robot. The only instructions that we gave to the participants were that they should order a drink from the robot and that they should use specific names when referring to the drinks the robot had to offer: Coke, blue lemonade, or green lemonade (or the German equivalents). After the instructions, each participant carried out the three-person drink-ordering transaction shown in Figure 3, along with two confederates: the participant acted as customer A1, while customers A2 and A3 were played by the confederates. After the experiment, the participants completed the same questionnaire as at the start of the experiment.

5.3 Independent measures

We manipulated one feature of the robot bartender during this study: half of the participants interacted with a bartender that reacted at a purely task-based level, and half with a system making use of all of the social behaviours described in Section 4. We used a between-participants design in this experiment, which means that each participant interacted with one version of the robot. Participants were assigned alternately to the two versions of the system: in total, 20 participants ordered drinks from the task-based robot, and 20 participants interacted with the socially intelligent robot.

5.4 Dependent measures

We gathered two classes of dependent measures: objective measures derived from the system logs, as well as subjective measures computed from the pre- and post-experiment questionnaires.

5.4.1 Objective measures

Using the interaction logs, we gathered a set of general objective measures about the recorded interactions, which were based on the dimensions proposed by the PARADISE dialogue evaluation framework [34]. **Task success** was assessed by counting how many drinks were served by the system; **dialogue quality** was measured by counting how many of the user's attempted contributions fell below the confidence threshold of the speech-recognition system, how many times a time-out expired while waiting for a user response, and how many times the robot had to ask for a customer's drink order; while for **dialogue efficiency**, we computed the mean time between a customer's initial appearance and the time that the bartender acknowledged that customer (verbally or non-verbally), the time taken to serve the first drink in a trial, as well as the total duration of the trial as measured both in seconds and in system turns.

In addition, we counted the number of times that the fully social bartender employed two specific behaviours only included in that domain: dealing with customers as a group (as opposed to individually), and explicitly acknowledging a customer that arrived while the bartender was already serving another. As shown in Figure 3, the target scenario should have included both of these behaviours.

Measure	Mean	Median	Min	Max
Drinks served	2.63	3	1	3
Low ASR turns	6.25	6	1	14
Timeouts	1.05	0	0	12
Order requests	7.10	6	1	22
Response time	14.7	12.9	1.8	52.8
Time to first drink	51.0	46.1	33.0	110.1
Total time (s)	111.6	109.0	41.7	214.5
Total system turns	13.2	13	6	35
Group orders	0.75	1	0	2
Acknowledgements	2.55	2	1	7

Table 1: Summary objective results

5.4.2 Subjective measures

We measured participants' subjective experiences via a questionnaire based on the GODSPEED questionnaire series [3]. The GODSPEED questionnaire is designed to be a standard user measurement tool for human-robot interaction, and includes items asking the participant to assess the robot on five scales: anthropomorphism (five items), animacy (six items), likeability (six items), perceived intelligence (five items), and perceived safety (three items) All responses were given on a six-point semantic differential scale, with lower scores corresponding in each case to a more negative assessment of the robot or the interaction; for each question, the participant could also choose "no answer" if they could not or did not want to respond.

We administered the GODSPEED questionnaire before the experiment to measure user expectations, and then again after the experiment to test any changes in user opinions. That is, before the study, one of the survey items was as follows:

I expect the robot to be:

human-like 1 2 3 4 5 6 machine-like

After the study, the corresponding question was posed as follows:

The robot was:

human-like 1 2 3 4 5 6 machine-like

This pre-test/post-test strategy was chosen to allow us to assess the impact of the two interaction styles more directly, controlling for user expectations [8]. The GODSPEED questionnaire has successfully been used in a pre-test/post-test context in other studies, e.g., [32].

5.5 Results

5.5.1 Objective results

Table 1 shows the overall results on the general objective measures. In general, the task success was reasonably high, with nearly all customers receiving a drink (a mean value of 2.63 out of a maximum possible value of 3). However, dialogue quality was affected by the number of attempted user turns that fell below the ASR confidence threshold, and the system often had to repeat its request for a drink order several times. Regarding efficiency, users were acknowledged on average about 15 seconds after they first became visible, the first drink was generally served about 50 seconds after the interaction began (which included approximately 20 seconds for the robot arm to physically grasp and hand over the drink), and the whole interaction took an average of about two minutes, or 13 system turns. The interactions with a very large number of ASR failures, timeouts, and system turns generally indicate either that the speech recogniser had particular difficulty with the speech of a participant, or—in some cases—that the customers stood in such a way that the vision system had difficulty tracking them. The bottom two rows of the table show the counts for the two additional

Measure	Task-only (sd)	Full social (sd)
Time to first drink	55.6 (17.0)	46.3 (15.5)
Total system turns	11.0 (4.8)	15.5 (5.8)

Measure	Male (sd)	Female (sd)
Order requests	6.4 (3.8)	9.2 (6.4)
Time to first drink	48.4 (12.3)	58.5 (24.2)
Total system turns	11.6 (4.6)	17.2 (6.9)

Table 2: Influence of planning domain and gender

social actions. In general, about three-quarters of the interactions with the fully social system involved the system treating customers as a group, and the system acknowledged newly-arrived customers an average of twice in an interaction. The trials with a very high number of acknowledgements (e.g., 7) indicate an issue with the vision system tracking one of the customers, who therefore would have been repeatedly treated as a new customer. This was only an issue in three of the 20 social trials.

To assess the impact on these results of the experimental manipulation (task-based vs. socially intelligent) and the four demographic factors of the participants (age, gender, HRI experience, interaction language), we carried out a multiple regression analysis. This analysis found that two factors had a significant impact on the objective results: the interaction domain used by the robot, and the participants' gender. In particular, when the robot used the task-based planning domain, the robot required significantly fewer system turns overall ($\beta = -0.33, p < 0.05$ in the regression model) and took significantly longer to serve the first drink ($\beta = 0.37, p < 0.05$). On the other hand, when the participant was male, the robot had to make significantly fewer drink-order requests ($\beta = -0.36, p < 0.05$), took fewer system turns overall ($\beta = -0.42, p < 0.01$), and served the first drink more quickly ($\beta = -0.35, p < 0.05$). Table 2 shows the per-group means for these objective measures. We carried out an ANOVA analysis to check for interactions between the factors, and found no significant interaction: $F(1, 35) = 0.14, p = 0.71$ for system turns, and $F(1, 35) = 1.36, p = 0.25$ for the first drink time.

5.5.2 *Subjective results*

In order to analyse the questionnaire results, we first processed the responses to replace any non-answers with the mean response for that item. For each of the five groups on the GODSPEED questionnaire (anthropomorphism, animacy, likeability, perceived intelligence, perceived safety), we then computed the mean pre-test and post-test scores for each participant, and also then computed the mean score change for each category. These summary results are shown in bold in Table 3: in general, the scores decreased in all categories from the pre-test to the post-test to varying degrees.

Table 3 also breaks down the GODSPEED scores by interaction language. In general, the German participants gave systematically lower pre-test scores ($M = 3.35, SD = 0.73$) than did the English participants ($M = 3.96, SD = 0.60$), while the post-test scores are more similar (German $M = 2.94, SD = 0.64$; English $M = 2.74, SD = 0.72$). When we assessed this difference with a T test, the difference in pre-test scores was indeed found to be significant, ($t(31.5) = 2.83, p < 0.001$), as was the difference in the overall score change ($t(28.9) = -3.67, p < 0.001$), while the post-test means did not differ significantly ($t(23.9) = -0.87, p = 0.40$).

As with the objective factors, we again used a multiple regression analysis to assess the impact of the experimental manipulation and the demographic features. Based on the above analysis, we used the post-test scores as our response variable, as the variance of those scores was less than that of the pre-test/post-test change. In the initial regression analysis, neither the planning domain nor any of the demographic factors had any significant effect on any of the GODSPEED categories. We then carried out an ANOVA analysis to test for interactions between the language and the planning domain on each of the GODSPEED categories. This analysis found that neither the planning domain nor the language had a main effect on the responses in any category. However, a significant interaction was found on the animacy scores ($F(1, 36) = 4.19, p < 0.05$), with marginal effects for liking ($F(1, 36) = 4.06, p \approx 0.05$) and perceived intelligence ($F(1, 36) = 3.00, p \approx 0.09$): in all cases, the effect was that the scores for the German participants for the task-only planning domain were lower than those for the fully social domain.

Category	Pre (sd)	Post (sd)	Change (sd)
Anthropomorphism	**2.74 (0.95)**	**2.02 (0.64)**	**-0.73 (0.90)**
German	2.34 (0.78)	1.90 (0.60)	-0.43 (0.86)
English	3.51 (0.76)	2.24 (0.65)	-1.27 (0.71)
Animacy	**3.04 (0.90)**	**2.30 (0.70)**	**-0.73 (0.96)**
German	2.83 (0.97)	2.25 (0.66)	-0.59 (1.01)
English	3.43 (0.59)	2.42 (0.78)	-1.01 (0.82)
Likeability	**4.43 (1.01)**	**3.70 (1.17)**	**-0.74 (1.34)**
German	4.27 (1.03)	4.05 (0.96)	-0.22 (1.01)
English	4.74 (0.92)	3.04 (1.27)	-1.70 (1.38)
Perc. Intelligence	**3.77 (0.80)**	**2.87 (0.83)**	**-0.90 (0.86)**
German	3.52 (0.78)	2.92 (0.90)	-0.60 (0.81)
English	4.22 (0.65)	2.76 (0.70)	-1.46 (0.64)
Perc. Safety	**4.15 (1.10)**	**3.97 (1.12)**	**-0.18 (1.20)**
German	4.12 (1.11)	4.17 (1.08)	+0.05 (0.82)
English	4.22 (1.12)	3.62 (1.14)	-0.60 (1.65)

Table 3: Summary of GODSPEED scores

5.5.3 *Comparing objective and subjective measures*

We considered a number of objective and subjective measures, all of which varied widely across participants and across trials. We therefore investigated which of the objective measures had the largest effect on users' subjective judgements, using stepwise multiple linear regression as suggested by the PARADISE evaluation framework [34]. This process produces coefficients describing the relative contribution of each objective predictor to subjective user satisfaction. If a predictor does not contribute significantly, its coefficient is zero after the stepwise process, so only significant predictors remain.

Table 4 shows the results of the PARADISE procedure on the data from this study. Each column corresponds to a possible predictor, while the cells indicate how each factor contributes to the post-test score in each of the GODSPEED categories. The sign (+ or −) indicates the direction of influence, while the number of symbols indicates the strength of the influence: three for $p < 0.001$, two for $p < 0.01$, and one for $p < 0.05$. For example, for likeability, the number of order requests made by the system had an extremely negative impact ($p < 0.001$), while the duration had a moderately positive impact ($p < 0.01$). The R^2 column indicates the percentage of variance explained by each predictor function.

5.6 Discussion

The overall objective results of this study indicate that the robot bartender system was generally successful at its core task of serving drinks. Despite the minimal instructions given to the participants, the objective success rate was very high across the conditions; however, interactions with the system were affected by the hard-coded threshold of the speech recogniser, which led to attempted user contributions being discarded. In the trials with the fully social bartender, the first drink was served more quickly on average: this is

Category	NumDrinks	OrderReq	FirstDrink	Duration	SysTurn	GroupO	Ack	R^2
Anthropomorphism			–				–	0.10
Animacy								*na*
Likeability		– – –		++				0.24
Perc. Intelligence				+	−−	++		0.24
Perc. Safety	–	−−		+	–			0.34

NumDrinks: number of drinks served; *OrderReq*: number of system order requests; *FirstDrink*: time to serve first drink; *Duration*: total length in seconds; *SysTurn*: total length in system turns; *GroupO*: number of group order requests; *Ack*: number of customer acknowledgements

Table 4: Significant objective predictors of GODSPEED post-test scores

likely a reflection of the stronger ordering imposed on order-taking and serving by the additional social constraints. Also, while those fully social trials took more system turns to complete, there was no significant difference in the overall interaction time, indicating that the additional system turns required to implement the social behaviour did not affect the overall efficiency of the dialogue.

The objective dialogue efficiency was also significantly affected by the gender of the participants: in the interactions involving male participants, the first drink was served more quickly, and the interaction was also shorter overall and involved fewer system turns. There was no difference in the number of timeouts or low-ASR turns, indicating that the problem was probably not to do with speech recognition. Instead, this finding suggests that the issue with the female participants may have been at least partly due to the performance of the vision system, which was trained primarily on males: if vision has difficulty detecting or tracking a customer, then the interaction will proceed less smoothly.

The scores on the subjective questionnaire generally decreased from the pre-test to the post-test. The score decrease was generally larger for the participants who interacted in English: the pre-test scores were significantly higher for the English participants, while the post-test scores were more similar. It is worth noting that on a recent study involving participants from 142 countries worldwide [12], Germans had the fourth-lowest level of optimism as measured by the expected difference between current and future quality of life: so the difference in pre-test scores may well be related to cultural differences. Also, this experiment was carried out in Germany, so the participants who chose to use German were primarily native Germans, while the participants who chose English were mainly international students whose first language was neither German nor English. This mixture of native and non-native speakers may also have affected the experimental results; for example, [30] also found differences between native speakers and non-native speakers on both subjective and objective measures in their dialogue system evaluation. Although the interaction style had no main effect on the subjective scores, there was an interaction between language and interaction style: the post-test scores from the German participants on animacy, liking, and perceived intelligence were generally lower for the task-based system than for the fully social system.

In a PARADISE-style stepwise regression analysis, we found that the objective factors that had the greatest impact on the subjective responses were the number of system order requests and the overall duration of the trials: other factors that had some impact included the number of system turns and the presence of social behaviours such as group ordering and asking customers to wait. Some predictors have an unexpected or even contradictory effect—e.g., the scores for perceived safety were actually negatively correlated with objective task success—and the R^2 values for the PARADISE predictor functions, while in line with those from similar studies [e.g., 10, 30], were generally quite low. This suggests that the users' subjective judgements were also affected by factors other than the log-based objective measures considered here.

Note that all of the objective measures are currently based only on the data from the log files, along with some underlying assumptions about user behaviour based on the scenario given to the participants (Figure 3): for example, we assume that all customers were seeking to engage with the bartender, and that customers *A1* and *A2* were in a group together while *A3* was isolated. We do not yet have ground-truth data as to the actual verbal and non-verbal behaviour of the customers in the scene, such as their actual spoken utterances or their true attention-seeking and group behaviour. For this reason, we are currently annotating the videos resulting from this study to add that information, which should allow a more detailed objective analysis of the interactions: for example, we can assess the performance of the vision system, and also compute additional measures such as the word error rate from the speech synthesiser. Such additional measures will allow the interactions to be compared more closely, and may also shed more light on the influence of the demographic factors on the results; we also expect that adding them to the PARADISE analysis will increase the R^2 values of the predictor functions.

6. CONCLUSION

Our guiding research question in this work was whether a service robot needs to show social behaviour when it interacts with humans in a task-based context. To address this question, we used a high-level automated planner to implement two different behaviours on a robot bartender that could serve drinks to human customers. In a user study in which the robot served drinks to multiple humans, the robot interacted in one of two ways: either it executed only task-relevant actions, or it also executed socially intelligent actions in addition to the actions necessary to complete the task. The socially intelligent system served the first drink more quickly than it did in the purely task-based system; and while the socially intelligent version used more system turns to complete an interaction, this did not affect the overall time taken. Interactions involving a male participant were also found to be somewhat more efficient: we hypothesise that this was due at least in part to the vision system having more difficulty tracking the female participants. The biggest effect on the subjective questionnaire was the interaction language: participants who did the experiment in German gave significantly lower scores on the GODSPEED pre-test, although their scores on the post-test were similar to those of the English participants. An analysis of the post-test scores found an interaction between language and interaction style: the German participants tended to give the task-only system lower scores on animacy, liking, and perceived intelligence. We were also able to show that certain objective experiment measurements had an influence on participant ratings: interactions that required more system turns and more order requests generally resulted in lower satisfaction, while longer interactions and those that included specific social behaviours tended to be rated more highly.

It was surprising that the different robot behaviours had such a subtle influence on the ratings of the experiment participants: intuitively, comparing the two interactions from Figure 3, one would

expect that a socially intelligent bartender would be perceived as a clearly better interaction partner. Therefore, we plan to refine our methodology in the future to analyse this question in more depth. First, we will annotate the videos of the human-robot interactions from our experiments in order to have ground truth data about the participants' actions. This data cannot be obtained from the robot log files, and will allow the experiment to be analysed in more detail; it should also shed more light on the demographic factors that emerged from the analysis. Second, we will conduct an experiment in which we compare an even *less* social domain to the socially intelligent domain. The robot could, for example, just act as a soda machine, and only serve a drink to a predefined position when explicitly told to do so, without any reference to specific customers. For subsequent studies, we will recruit a demographically balanced selection of participants, particularly with respect to gender and native language, to ensure that the effects of any individual differences are minimised. Finally, we will enhance all components of the robot to improve dialogue quality. In particular, we will fine-tune the speech recognition thresholds, and we will also implement an improved clarification strategy to deal with lower-confidence recognised utterances.

7. ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 270435, JAMES: Joint Action for Multimodal Embodied Social Systems (`james-project.eu`). Thanks to Ingmar Kessler for helping run the experiment, and to Jan Peter de Ruiter for advice on experiment design and statistics.

8. REFERENCES

[1] W. Bainbridge, J. Hart, E. Kim, and B. Scassellati. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics*, 3:41–52, 2011.

[2] H. Baltzakis, M. Pateraki, and P. Trahanias. Visual tracking of hands, faces and facial features of multiple persons. *Machine Vision and Applications*, 23(6):1141–1157, 2012.

[3] 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*, 1:71–81, 2009.

[4] D. Bohus and E. Horvitz. Dialog in the open world: platform and applications. In *Proceedings of ICMI-MLMI 2009*, pages 31–38, 2009.

[5] C. Breazeal. Socially intelligent robots. *interactions*, 12(2):19–22, 2005.

[6] G. Castellano, I. Leite, A. Pereira, C. Martinho, A. Paiva, and P. W. McOwan. Affect recognition for interactive companions: challenges and design in real world scenarios. *Journal on Multimodal User Interfaces*, 3(1):89–98, 2010.

[7] K. Dautenhahn. Socially intelligent robots: dimensions of human-robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):679–704, 2007.

[8] D. M. Dimitrov and P. D. Rumrill, Jr. Pretest-posttest designs and measurement of change. *Work: A Journal of Prevention, Assessment and Rehabilitation*, 20(2):159–165, 2003.

[9] R. E. Fikes and N. J. Nilsson. STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2:189–208, 1971.

[10] M. E. Foster, A. Gaschler, M. Giuliani, A. Isard, M. Pateraki, and R. P. A. Petrick. Two people walk into a bar: Dynamic multi-party social interaction with a robot agent. In *Proceedings of ICMI 2012*, pages 3–10, 2012.

[11] M. E. Foster, A. Gaschler, and M. Giuliani. How can I help you? comparing engagement classification strategies for a robot bartender. In *Proceedings of ICMI 2013*, 2013.

[12] M. W. Gallagher, S. J. Lopez, and S. D. Pressman. Optimism is universal: Exploring the presence and benefits of optimism in a representative sample of the world. *Journal of Personality*, 2013.

[13] A. Gaschler, R. P. A. Petrick, M. Giuliani, M. Rickert, and A. Knoll. KVP: A knowledge of volumes approach to robot task planning. In *Proceedings of IROS 2013*, 2013.

[14] S. S. Ge and M. J. Matarić. Preface. *International Journal of Social Robotics*, 1(1):1–2, 2009.

[15] E. C. Grigore, K. Eder, A. Lenz, S. Skachek, A. G. Pipe, and C. Melhuish. Towards safe human-robot interaction. In *Towards Autonomous Robotic Systems*, pages 323–335. Springer, 2011.

[16] K. Huth, S. Loth, and J. De Ruiter. Insights from the bar: A model of interaction. In *Proceedings of Formal and Computational Approaches to Multimodal Communication*, August 2012.

[17] D. Limbu, Y. Tan, C. Wong, R. Jiang, H. Wu, L. Li, E. Kah, X. Yu, D. Li, and H. Li. Experiences with a barista robot, FusionBot. *Progress in Robotics*, pages 140–151, 2009.

[18] E. Márquez Segura, M. Kriegel, R. Aylett, A. Deshmukh, and H. Cramer. How do you like me in this: User embodiment preferences for companion agents. In *Proceedings of IVA 2012*, 2012.

[19] T. Masuda and D. Misaki. Development of Japanese green tea serving robot "T-Bartender". In *Proceedings of ICMA 2005*, volume 2, pages 1069–1074, 2005.

[20] Y. Matsusaka, T. Tojo, and T. Kobayashi. Conversation robot participating in group conversation. *IEICE Transactions on Information and Systems*, 86(1):26–36, 2003.

[21] Global Industry Analysts Inc. Service robotics – a global market report, 2010. `http://www.strategyr.com/Service_Robotics_Market_Report.asp`.

[22] International Federation of Robotics. Service robot statistics, 2012. `http://www.ifr.org/service-robots/statistics/`.

[23] B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, and N. Hagita. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Proceedings of HRI 2009*, pages 61–68, 2009.

[24] M. Pateraki, M. Sigalas, G. Chliveros, and P. Trahanias. Visual human-robot communication in social settings. In *Proceedings of ICRA Workshop on Semantics, Identification and Control of Robot-Human-Environment Interaction*, 2013.

[25] R. P. A. Petrick and F. Bacchus. A knowledge-based approach to planning with incomplete information and sensing. In *Proceedings of AIPS 2002*, pages 212–221, 2002.

[26] R. P. A. Petrick and F. Bacchus. Extending the knowledge-based approach to planning with incomplete information and sensing. In *Proceedings of ICAPS 2004*, pages 2–11, 2004.

[27] R. P. A. Petrick and M. E. Foster. Planning for social interaction in a robot bartender domain. In *Proceedings of ICAPS 2013, Special Track on Novel Applications*, pages 389–397, 2013.

[28] R. P. A. Petrick, M. E. Foster, and A. Isard. Social state recognition and knowledge-level planning for human-robot interaction in a bartender domain. In *Proceedings of the AAAI 2012 Workshop on Grounding Language for Physical Systems*, pages 32–38, 2012.

[29] M. Rickert. *Efficient Motion Planning for Intuitive Task Execution in Modular Manipulation Systems*. Dissertation, Technische Universität München, 2011.

[30] V. Rieser and O. Lemon. Comparing reinforcement and supervised learning of dialogue policies with real users. In *Reinforcement Learning for Adaptive Dialogue Systems*, pages 167–188. Springer, 2011.

[31] M. Sigalas, H. Baltzakis, and P. Trahanias. Visual tracking of independently moving body and arms. In *Proceedings of IROS 2009*, pages 3005–3010, 2009.

[32] E. Torta, J. Heumen, R. Cuijpers, and J. Juola. How can a robot attract the attention of its human partner? a comparative study over different modalities for attracting attention. In *Social Robotics*, pages 288–297. Springer, 2012.

[33] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[34] M. Walker, C. Kamm, and D. Litman. Towards developing general models of usability with PARADISE. *Natural Language Engineering*, 6(3&4):363–377, 2000.

[35] M. White. Efficient realization of coordinate structures in Combinatory Categorial Grammar. *Research on Language and Computation*, 4 (1):39–75, 2006.

[36] T. Wosch, W. Neubauer, G. Wichert, and Z. Kemény. Robot motion control for assistance tasks. In *Proceedings of IEEE RO-MAN 2002*, pages 524–529, 2002.