Who, how, where: Using Exemplars to Learn Social Concepts

Alan R. Wagner[1], Jigar Doshi[2]

[1]Georgia Institute of Technology Research Institute, Atlanta GA
[2]Georgia Insitute of Technology, Dept. of Electrical & Computer Engineering, Atlanta, GA

alan.wagner@gtri.gatech.edu, jdoshi8@gatech.edu

Abstract. This article introduces exemplars as a method of stereotype learning by social robot. An exemplar is a highly specific representation of an interaction with a particular person. Methods that allow a robot to create prototypes from its set of exemplars are presented. Using these techniques, we develop the situation prototype, a representation that captures information about an environment, the people typically found in an environment, and their actions. We show that this representation can be used by a robot to reason about several important social questions.

Keywords: exemplars, stereotype, prototype, social learning, user modeling

1 Introduction

If robots are to ever act and interact in a wide variety of social contexts and with a wide variety of people, then these robots will need computational techniques that allow them to adapt and learn from their social experiences. Social robots will need to reason about who will be found at a particular location, where a particular type of person can be found, and how people will act in a given context. Consider, as a motivating example, a robot tasked with searching for a teenager in a crowded mall. Having a model that includes the interests and disinterests of a typical teenager will allow the robot to rule out large portions of the search space and make the problem more tractable by focusing on areas likely to result in locating the missing youth. As a further example, consider a robot looking for a human partner with a particular skill set, such as those of an EMT. Can a robot use knowledge gained by interacting with different people to predict a likely place to find people with a specific set of skills? In this paper we develop techniques that allow a robot to reason and predict facets of social interaction such as these.

Our research has focused on the development of a principled, formal framework for social action selection that allows a robot to represent and reason about its interactions with a human partner [1]. This framework operates by modeling the robot's interactive partner and using its model of the partner to predict the partner's preferences and guide its social interaction. The research presented here contributes a key extension of our framework into the domains of concept learning and mental

simulation. In previous work we developed a method for creating stereotypes based on a prototype model of concept learning [2]. This work indicated that prototype-based stereotypes could be used to bootstrap the process of learning about a new person, used to preselect perceptual features important for recognition, and predict a person's appearance from their behavior.

Nevertheless, prototype-based stereotypes have important limitations such as their lack of context and details. These deficits have motivated our exploration of exemplars as an alternative model of social concept learning. The research presented here focuses on the development of methods designed to create exemplars from a robot's sensor data, create prototypes from a set of exemplars, and produce exemplars and prototypes which are context specific. We demonstrate that our approach can be used by a simulated robot to make predictions about several important aspects of a social environment, such as what types of people the robot will find in a given context and how people will act.

This article makes two key contributions. First, it begins to explore the development of categories related to context and the person's actions from a robot's interactive experience. This contribution has the potential to allow a social robot to reason about where certain categories of people can be found and the types of behavior these people exhibit within a social environment. Second, we begin to develop the representational and theoretical underpinnings that will allow a robot to utilize the exemplar model of stereotyping. Exemplars are an important potential representation because they act as an extensive source for generating predictions and inferences about the human partner's interactive behavior. For this research, exemplars are used to predict who is interacting, how they will act, where an interaction will occur. Further, the system we present is not given category labels or even the number of categories a priori. Rather the system learns these characteristics from its experience.

This article begins with a review of related work. Next, exemplar stereotypes are introduced followed by our techniques for creating representations that include context specific information. Experiments using this system as well as results are presented next. We conclude with a discussion of the results, their limitations, and provide directions for ongoing and future work.

2 Related Work

Stereotypes and stereotyping have long been a topic of investigation for psychologists. Schneider provides a good review of the existing work [3]. Numerous definitions of the term stereotype exist. Edwards defines a stereotype as a perceptual stimulus which arouses standardized preconceptions that, in turn, influence one's response to the stimulus [4]. Smith and Zarate describe three general classes of stereotype models: attribute-based, schematic-based, and exemplars [5].

With respect to computer science, the inclusion of techniques for stereotyping is not new. Human Computer Interaction (HCI) researchers have long used categories and stereotypes of users to influence aspects of user interface design [6]. The multi-

agent systems community has also explored the use of stereotypes. Ballim and Wilks use stereotypes to generate belief models of agents [7].

Investigations of stereotyping with respect to robots are comparatively scarce. Fong et al. used predefined categories of users in conjunction with a human-robot collaboration task [8]. These categories influenced the robot's dialogue, actions, the information presented to the human, and the types of control afforded to the user. Duffy presents a framework for social embodiment in mobile autonomous systems [9]. His framework includes methods for representing and reasoning about stereotypes. He notes that stereotypes serve the purpose of bootstrapping the evaluation of another agent and that the perceptual features of the agent being stereotyped are an important representational consideration.

3 Exemplar Stereotypes

The exemplar approach to stereotyping emphasizes concrete rather than abstract representations [5]. Exemplars were developed as a means for reflecting the specificity and detail often associated with a person's cognitive representations. Exemplar researchers tend to stress that these memories are stored as highly detailed mental images and that the inclusion of these details serves as an additional source of information for making inferences. For instance, if one is asked to think of a child the mental image that results tends to be a specific memory trace related to a child that the person has recently or often interacted with. Exemplar categorization is achieved by comparing new instances of the perception of a person to stored representations of previously encountered individuals. Different dimensions of comparison can be weighted more heavily to reflect the robot's social motivation or objective [3].

Table 1. Features and values used in experiments.

Partner Feature	Possible Values
Gender	Male, female
Age	Baby, child, teen, middle-aged, senior
Skin color	Pale white, white, dark white, brown, black
Skin texture	Smooth, wrinkled
Height	short, medium, tall
Hair style	bald, short, medium, long, bald
Hair color	brown, black, white, gray, red, blonde
Weight	thin, medium, heavy
Shirt style	collared t-shirt, t-shirt, long sleeve, collared long-sleeve, sweater, coat, no sleeves
Shirt/pant color	black, white, green, blue, red, yellow, brown, gray, multi-color
Glasses/facial hair	yes, no
Pant style	long, shorts, jeans, skirt, pajamas
Shoe style	tennis, sandal, high-heels, flip-flop, boot, slipper, no-shoes; dress
Actions	Eating, sitting, talking, drinking, walking, listening, watching, playing, writing, reading, coloring, standing, building, answering, running, singing, kicking, sleeping, teaching, hitting
Context Features	Chalkboard, clock, TV, book, wheelchair, food, fork, spoon, desk, chair, computer, pencil, wipe board, paper, toy, window, game, plants, table, art, plate, toilet, sand, cart, bicycle, car, ladder, bed, fence, glass

For a social robot, exemplars should ideally be represented as a collection of images or video segments capturing the robot's interaction with a person. In order for newly created exemplars to be compared to previously encountered exemplars, a process must exist during which the robot uses perceptual recognition and classification techniques to capture feature/value pairs related to the exemplar. These feature/value pairs are stored as a set and serve to discretely represent the exemplar. We term this intermediate representation an attributized exemplar. Table 1 shows our list of feature/value pair used during this experiment.

The creation of the attributized exemplar allows for the use of set distance measures, such as the Jaccard index, to be used with the representation. The Jaccard index,

$$j\left(m^A, m^B\right) = \frac{\left|m^A \cup m^B\right| - \left|m^A \cap m^B\right|}{\left|m^A \cup m^B\right|} \qquad (1)$$

gauges the distance from one exemplar to another. The Jaccard index is used in conjunction with agglomerative or "bottom-up" clustering to determine if two exemplars are sufficiently similar to warrant their merging. Clustering affords a means for creating prototypes from exemplars. Agglomerative clustering operates by initially assuming each exemplar as a unique cluster and then merging exemplars which are sufficiently close to one another. The centroid that results from merging these clusters serves as a prototype representing a general category of individuals. As the distance required to merge decreases from 1 to 0, the generality of the category and concept being represented increases (Fig. 1).

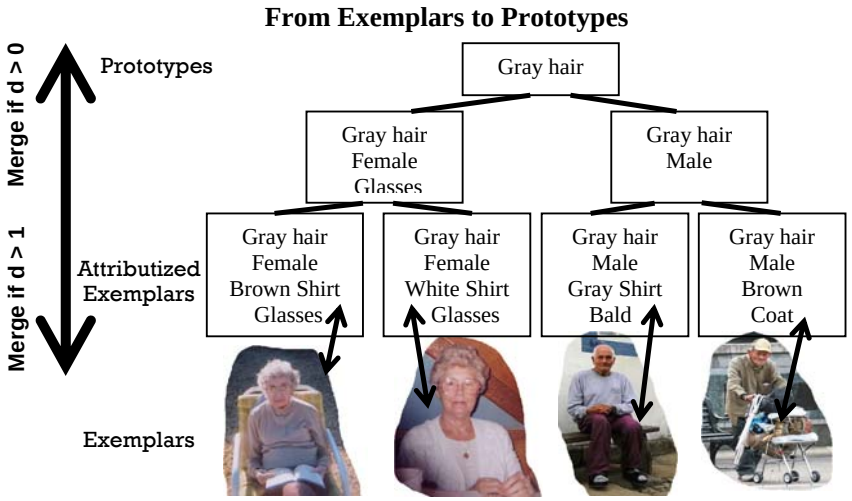

Fig. 1. The figure above depicts the transformation from exemplar to attributized exemplar to prototype. Exemplars are converted to attributized exemplars through a series of feature detectors. An attributized exemplar represents an exemplar as a set of feature value pairs. Agglomerative clustering results in increasingly general prototypes depending on the distance required to merge individual clusters.

3.1 Situation Prototype Representation

Consider, as a motivating example, a robot entering an elementary school classroom (Fig. 2). The people that the robot would likely encounter would typically be children. Perceptually these children would be shorter than adults, tend to have smooth skin, little facial hair, and wearing tennis shoes. Further, the people in this environment would utilize a specific set of actions such as raising their hand before speaking. Finally, the environment itself would include specific objects such as desks, bulletin boards, and books. A representation, which we call a situation prototype, consisting of a set of context features and a prototype of the type of person found in the environment captures the type of people in an environment, but also their behaviors and the features describing the context itself.. Context features relate to either salient objects or characteristics of the environment. In a classroom, for example, desks, chairs, books, bulletin boards, etc. could serve as context features.

To create a situation prototype, first exemplars and attributized exemplars are

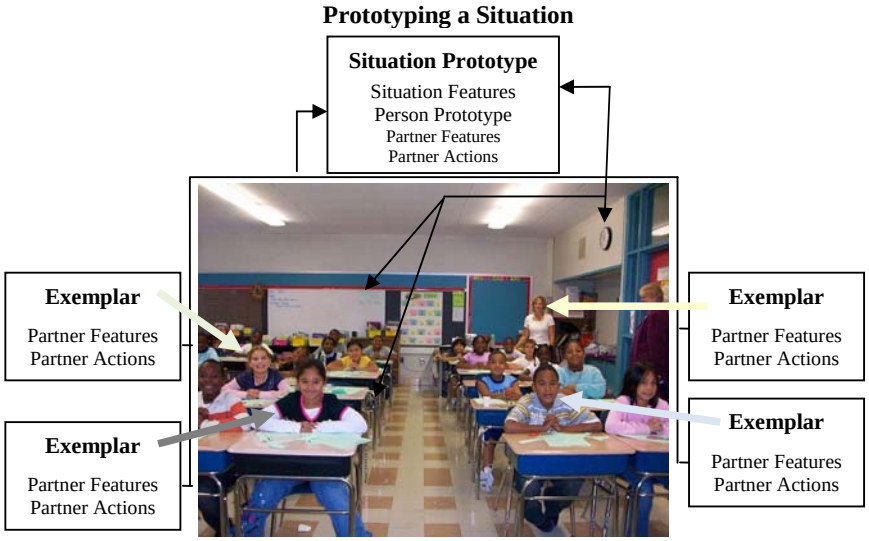

Fig. 2. The figure depicts the creation of a situation prototype from image information. Context features in the form of objects, such as clocks, desks, etc. are captured from the environment. Exemplars are created from the people in the environment. The exemplars are then used to create a person prototype consisting of prototypical partner features and actions. The lower portion of the figure depicts the process used to answer a question about the social environment. For example, the question "where do people look like this?" is examined by creating a list of partner features representing the person in question, for each context cluster the distance is calculated to determine the most similar target, and finally the cluster's context features are returned.

created from several or all of the people found in the environment. The exemplars that are created capture both the person's perceptual features (feature/value pairs related to what the person looks like) and the action or actions they are performing in the environment (see Table 1). Next, these individual exemplars are used to create a prototype representing the typical person found in the environment. The process of creating a prototype from a set of exemplars within the scene involves determining

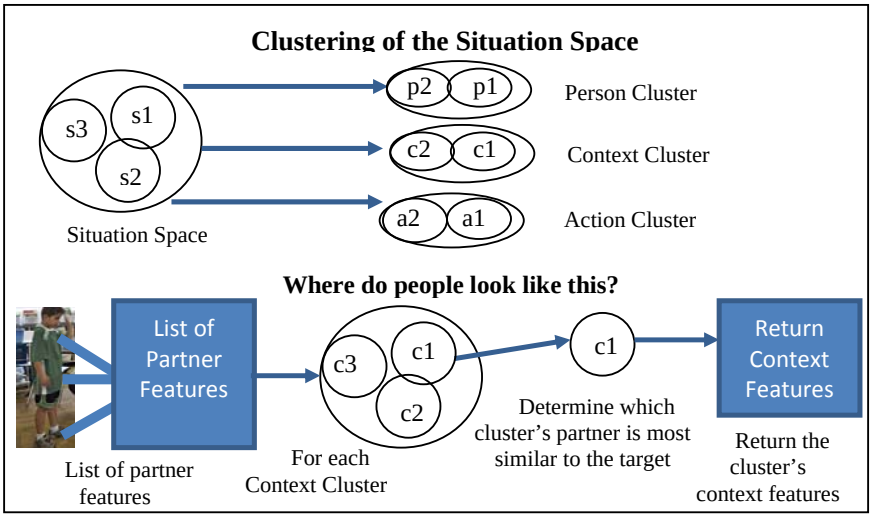

Fig. 3. The top portion of the figure depicts the situation space being clustered with respect to people, contexts and actions. The bottom portion then depicts the use of these clusters to reason about the six questions described in section 3.2.

which partner features and/or actions should be included in the prototype. A feature/action voting scheme in which individual features or actions receive votes from each exemplar in the scene was used to create the person prototype for a situation. Finally, context features are collected and added to a set. The result is a representation of the situation which includes information about what a typical member of the environment looks like, acts like, and the objects or characteristics of the environment itself (Fig. 2).

3.2 Six Questions

Given a robot with a collection of situation prototypes representing the individuals, contexts, and actions encountered in specific situations, we hypothesized that one could cluster over this situation space to generate centroids reflecting the concepts of who, how, and where people, actions, and environments could be found. More specifically, clustering over the situation space with respect to the person prototype portion of the representation generates centroids reflecting the types of people encountered irrespective of the environment. Clustering with respect to the context features, on the other hand generates concepts specific to particular environments. Finally clustering with respect to the action sets reflects the behaviors that occur independent of the environment. The top portion of Fig. 3 depicts this clustering of the situation space.

The process described thus far, (1) creating situation prototypes that reflect a social context (Fig. 2), and (2) clustering the resulting situation space with respect to people, context and actions (Fig. 3 top), results in concepts describing categories of people, places and behavior. As a final step, one can measure the distance from a

particular exemplar to each context cluster (for example), select the closest cluster, and use this information to determine which context features are most likely to be associated with this type of person (e.g. Fig. 3 bottom). In other words, a robot can use the system to answer questions like, "where do people look like this?" where "this" refers to an exemplar indicating a person's appearance. Overall, we hypothesized that the system could be used by a robot to reason about the following six questions:

1. **Where do people look like this?** Given a set of partner features representing a person predict a set of context features detailing the type of environment in which people with this appearance are likely to be found.
2. **Where do people act like this?** Given a set of actions, predict a set of context features describing the environment in which these actions occur.
3. **What do people look like that act like this?** Given a set of actions predict a set of partner features describing the appearance of people that perform these actions.
4. **What do people look like here?** Given a set of context features predict a set of partner features describing the appearance of people in this environment.
5. **How do people act here?** Given a set of context features predict a set of actions likely to be performed in this context.
6. **How do people that look like this act?** Given a set of partner features predict a set of actions likely to be performed by people that look like this.

A procedure that allows a robot to generate predictions related to these six questions could afford significant advantages. For instance, such a system could afford a robot operating in a search and rescue mission to use information about the victims being found to predict its current location. For example, if many of the victims are elderly then the search location may be a nursing home. Alternatively, if the robot knows its current context than it can use this procedure to predict how to search based on predictions about the type of people found in that context. In order to test the success of this procedure an empirical evaluation was performed.

4 Empirical Evaluation

Our empirical evaluation examined whether and to what extent situation prototypes could be created and used to answer the preceding six questions. Three different contexts were used: nursing homes, elementary school classrooms, and restaurants. We believed that clustering situation prototypes with respect to the context features based on several classrooms, nursing homes, and restaurants would create a centroid representing a classroom that is distinct from a nursing home and restaurant. The resulting centroid would itself be a situation prototype and as such contain prototype information about the appearance of the person typically found in this environment and their behavior.

Situation prototypes were generated from images. Thirty photos from each environment were obtained by searching Google images for the key words "classroom", "restaurant" and "nursing home". Images that depicted the environment

in its natural state were selected (e.g. images with people posing or obvious mismatches were rejected).

Amazon.com's Mechanical Turks were used create attributized exemplars from the images. The Amazon Mechanical Turk is a crowdsourcing Internet marketplace where the *Requesters* pay a pre-determined amount to the *Workers* for satisfactorily completing the Human Intelligence Task (HIT). Our study used only workers who were categorization masters and had at least a 90% acceptance rate. Images were prepared by marking a person in the image that the worker answered questions about. We marked people in the image whose faces were obviously visible and could have been easily detected by simple face recognition algorithms. To ensure consistency throughout the dataset, only 4 faces per image were selected. Workers were paid 50 cents for successfully completing a survey describing the person's features, actions, and context features.

The HITs that workers were asked to complete included of an image taken from a classroom, a nursing home, or a restaurant. In the scene an arrow pointed to a person about whom information is sought. For each marked individual in the image, questions were asked about the person's physical appearance and attire. Workers were also asked to select between 1-3 actions from a pre-defined set of 20 actions that the person was likely performing in the scene. Finally, questions were asked about the context such as the floor type, lighting and about the objects present in the room. From a pre-defined list of 30 objects, workers selected a minimum of three and a maximum of five objects. For each marked individual we collected independent surveys from three different workers. Table 1 lists the set of partner features, actions, and context features collected. All surveys that 1) had any answers that matched more than one possible value or did not match any possible values or 2) had a missing value for any feature (i.e. no response) were rejected. Any answers that were clearly typographical errors, had mismatched cases, and missing or additional special characters were accepted. Actions and context objects were selected from a dropdown menu. All entries that did not match the minimum quantity requirement (one for actions and three for context) were rejected.

Duplicate surveys were used to identify erroneous or random responses by the workers. The data relating to the perceptual features of a person depicted in an image, their actions, and characteristics of the context was aggregated by assigning votes to each of the HITs. Feature values with more than 50% agreement were accepted.

5 Experiments

Experiments were conducted in simulation and used the data generated by the workers. For each question from section 3.2, sixty people were randomly selected from the data. Depending on the question, either the features describing the person's appearance, the person's set of actions, or the features describing the context were used as input. Next, the distance from this input to each of the clusters was calculated. Finally the cluster with the smallest distance was used to predict the features in question. For example, the question, "where do people look like this?" takes as input a set of partner features describing a person's appearance (see Fig. 3 bottom). These

features are compared to the person prototypes in each context cluster. The context features from the cluster with the smallest distance were used to predict the environment in which the person would likely be found. This prediction is then compared with the actual data to determine the correctness of the prediction.

Two types of controls were used to for comparison. A mode control used the most common person prototype, context, and action set to predict the answers to the questions posed in section 3.2. A random control used a partner prototype created from random features, a random action set, and random context features.

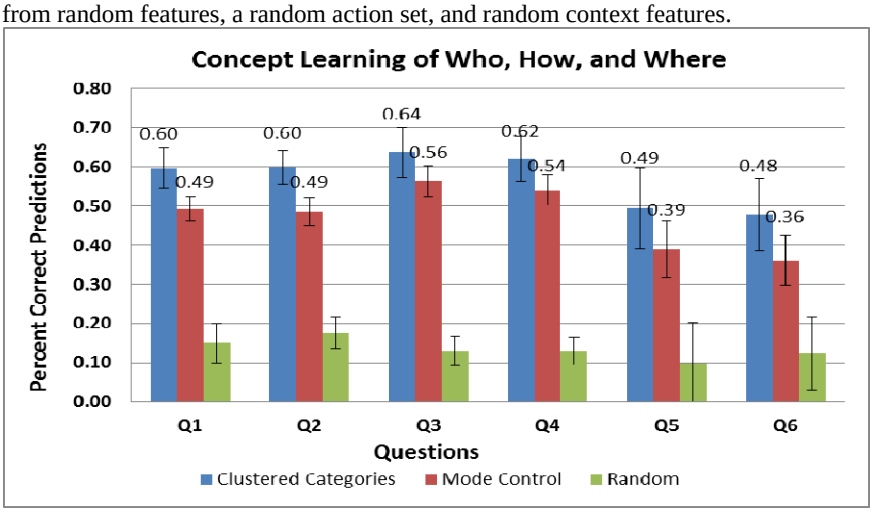

Fig. 4. The results from the experiment are depicted above. The bar to the left (blue) depicts the experimental condition. The center and rightmost bars depict the control conditions. The graph depicts results for each of the six questions from section 3.2.

The results from this experiment are depicted in Fig. 4. The experimental procedure is depicted by the bars on the left in blue. The middle bars depict the mode control and the random control is on the right. The experimental procedure performed significantly better than the both controls on Q1-Q4 (p<0.04 for all), yet did not perform better than the mode control on Q5 and Q6. Question 5 examines the type of actions performed in a context and question 6 examines the actions performed by a particular type of person. The lack of significance for these questions likely stems from the fact that the use of single images made distinguishing different actions from one another very difficult for the workers. Because of this, clustering the situation prototypes with respect to the action set lead to only a single cluster. We strongly believe that the use of movies or actual robots capable of detecting motion would result in better predictions of the partner's actions.

6 Conclusions

We have presented a system that uses exemplars in conjunction with prototypes to create a representation that captures the typical person and actions in an environment.

A technique that clusters this situation prototype representation and matches a person's appearance, actions, and context to a cluster centroid was described as a method affording a means for reasoning about important social questions.

The results presented make assumptions that would need to be addressed in order to use the system on a robot. First, clearly the use of Mechanical Turk workers would need to be replaced with the robot's own perceptual system. We have developed several classifiers toward this goal. Further, the procedure learns the category of people in context if such a category exists. Restaurants, for example, do not have a natural category of people that they serve. In this case, the system does not learn to associate any particular partner features with the restaurant context. Still, the actions and contextual features are predictive and would be of use to a robot reasoning about the social interactions likely to occur at this locale. Finally, the system was not given any a priori knowledge related to category labels or even the number of categories. Rather these characteristics were learned from experience.

We are currently implementing this system on a real robot. We intend to apply some of these techniques to the challenge of turn-taking during social interaction. We admit that the ideas and techniques are simplistic. Nevertheless, as we have shown, these simple ideas could afford a social robot with a powerful means for reasoning about its social environment. Our goal is to develop techniques that will allow a robot to act and interact in a wide variety of social contexts and with a wide variety of people, the methods presented here are a step towards allowing them to adapt and learn from their social experiences.

Acknowledgements

This work was funded by award #N00141210484 from the Office of Naval Research.

References

1. Wagner, A. R.: Creating and Using Matrix Representations of Social Interaction. In: The 4th ACM/IEEE International Conference on Human-Robot Interaction, San Diego CA (2009).
2. Wagner, A. R.: Using Cluster-based Stereotyping to Foster Human-Robot Cooperation. Proceedings of IEEE International Conference on Intelligent Robots and Systems (IROS 2012. Villamura, Portugal (2012).
3. Schneider, D. J.: The Psychology of Stereotyping. The Guilford Press. New York (2004)
4. Edwards, A.: Studies of Stereotypes: I. The directionality and uniformity of responses to stereotypes. J. of Soc. Psy, 12, 357-366 (1940).
5. Smith, E. R., Zarate, M. A.: Exemplar-Based Model of Social Judgement. Psy.Rev, (1992)
6. Rich, E.: User Modeling via Stereotypes. Cognitive Science, 3(197), 329-354 (1979)
7. Ballim, A., Wilks, Y.: Beliefs, stereotypes, and dynamic agent modeling. User Modeling and User-adapted Interaction, 1(1), 33-65 (1991).
8. Fong, T., Thorpe, C., Baur, C.: Collaboration, dialogue, and human–robot interaction. Proceedings of the International Symposium on Robotics Research. (2001).
9. Duffy, B.: Robots Social Embodiment in Autonomous Mobile Robotics. International Journal of Advanced Robotic Systems, 1(3) (2004).