The Impact of Stereotyping Errors on a Robot's Social Development

Alan R. Wagner, *Member, IEEE*

***Abstract*—Psychologists note that social cognition often involves the creation, refinement, and use of models of one's interactive partners. The influence of categorical thinking on interpersonal expectations is commonly referred to as a stereotype. Using an algorithm that we created for stereotype learning, we investigate problems that can occur when the robot acquires its first models of people and learns its first stereotypes—the robot's early social development. We examine if the errors related to the creation of these *initial* models have a disproportionate impact on the robot's developing social skills, perhaps even reflecting some of the same challenges faced by humans [1]. We hypothesized that errors in which the robot interacted with someone that did not represent the true nature of a category, an outlier, would impact the robot's performance on a social coordination task more if the error occurred earlier in the robot's social development rather than later. Results from simulation confirmed our hypothesis. The results of this work have potential implications for social robotics, autonomous agents, and possibly psychology.**

***Index Terms*—Autonomous mental development, Predictive models, Intelligent Robots, Service Robots.**

I. Introduction

Psychologists note that social cognition often involves the creation, refinement, and use of models of one's interactive partners [2, 3]. Further, over the course of social development an individual typically accumulates models of different interactive partners garnered from a variety of different social situations [4]. Humans generally organize this space of partner models by creating generalized partner models which represent individuals encountered from a specific perceptual or situational category [5]. These generalized partner models are commonly referred to as stereotypes.

Stereotypes act as a general source of information about a category or type of person [5]. Stereotypes are learned from interactions with the members of a category. During these interactions, models are developed which may include information about the person's beliefs, actions, desires, etc. This information is justly or unjustly related to the individual's perceptual features. For example, if all fire fighters that person encounters have beards, an individual may learn that the perceptual feature of having a beard is predictive of the actions and motives of a fire fighter.

Stereotypes impact social cognition in a variety of important ways [5]. The use of categories has been shown to simplify and speed up the process of person perception when encountering a new individual [6, 7]. Moreover, generalizations based on stereotypes allow individuals to predict traits that may or may not be related to the individual's perceptual features. Finally, the use of stereotypes may allow an individual to actively select which features to focus on during the performance of a social task. For instance, during an emergency a person will search for individuals adorned with perceptual features related to emergency response personnel such as uniforms rather than irrelevant perceptual features such as hair color.

In this article we argue that the robot's social development can roughly be equated to the initial stages of accumulating models of the robot's interactive partners and learning categories or stereotypes related to these models. At startup, the robot has no models or stereotypes of any interactive partners. Hence, it has no knowledge on which to base predictions about the behavior of others. Psychologists have shown that the accumulation of interactive experience is a key facet of social development [8, 9]. We investigate problems that can occur when the robot acquires its first models of people and learns its first stereotypes—the robot's early social development. We examine if the errors related to the creation of these *initial* models have a disproportionate impact on the robot's developing social skills, perhaps even reflecting some of the same challenges faced by humans [1].

This article uses an algorithm that we developed for stereotyping. As explained below, our algorithm creates stereotypes by clustering the partner models that a robot has previously learned. When encountering a new person a separate algorithm then matches the perceptual features of the new person to a stereotype model.

The remainder of this paper begins by discussing related work and introducing our framework for social action selection as well as our algorithm for stereotype creation and matching. Next we present an experiment which investigates different types of stereotyping errors and the timing of their occurrence. The paper concludes with a discussion of the results.

II. Related Work

Stereotypes and stereotyping has long been a topic of

Manuscript received June 29, 2012. This research has been support by the Georgia Tech Research Institute Social Robotics strategic initiative.

A. R. Wagner is with the Georgia Tech Research Institute, Atlanta, GA 30318 USA (404-407-6522; fax: 404-407-8081; e-mail: alan.wagner@gtri.gatech.edu).

investigation for psychologists [10]. Schneider provides a good review of the existing research [5]. 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 [11]. Smith and Zarate describe three general classes of stereotype models: attribute-based, schematic-based, and exemplars [3]. Psychologists have shown that stereotypes can have a strong influence on human social intelligence [12, 1]. Yamagishi et al., related social intelligence to the development of an accurate techniques for gauging trust which, in turn, required socialization methods for evaluating a partner's personal traits and intentions [1]. Moreover, Foddy et al., found that stereotyped evaluations related to whether or not a new partner was in-group or out-group plays a dominate role in evaluating these personal traits and intentions [13]. Developmental psychologists have long explored stereotyping [14, 15]. Stereotyping in children appears to be related to child's developing sense of a categorical self [16]. This process begins with a preference for one's own category followed by judgments of similarity to that group [5, 16]. Moreover, children tend to focus on the most salient and observable perceptual characteristics, such as hair color, when categorizing a person (e.g. [17]).

Computer scientists have also explored techniques for stereotyping. Human Computer Interaction (HCI) researchers used categories and stereotypes of users to influence aspects of user interface design [18, 19]. The multi-agent systems community has also investigated the use of stereotypes. Ballim and Wilks use stereotypes to generate belief models of agents [20]. Denzinger and Hamdan develop a system by which an agent is tentatively stereotyped, then, after interacting with the target for a period of time, stereotype switching may occur [6]. Their results indicate that the system performs well regardless of the number and quality of stereotypes. Burnett et al. uses stereotypes to gauge an agent's trustworthiness [21].

Investigations of stereotyping by roboticists are comparatively scarce. Fong et al. employed predefined categories of users in conjunction with a human-robot collaboration task [22]. These categories influenced the robot's dialogue, actions, and the information presented and types of control afforded to the user. Duffy describes a framework for social embodiment in mobile autonomous systems that include methods for stereotyping [23]. 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. The sections that follow detail the framework on which our algorithm for stereotyping is based.

III. Partner Modeling

Our framework for social action selection uses the outcome matrix (fig. 1) as a computational representation of interaction. The term interaction, in turn, describes a discrete event in which two or more individuals select interactive behaviors as part of a social situation or social environment. The term individual is used to indicate a human, a social robot, or an agent. We focus on interaction involving two individuals—dyadic interaction.

Outcome matrices are computational representations; hence, it is possible to describe them formally. A representation of interaction consists of 1) a finite set N of interacting individuals; 2) for each individual $i \in N$ a nonempty set A^i of actions; 3) the utility obtained by each individual for each combination of actions that could have been selected. The action $a_j^i \in A^i$ represents an arbitrary action j from individual i's set of actions. If $\left(a_j^i, \ldots, a_k^N\right)$ denotes a combination of actions, one for each individual, and if u^i denotes individual i's utility function then $u^i\left(a_j^i, \ldots, a_k^N\right) \rightarrow \Re$ is the utility received by individual i when the action combination $\left(a_j^i, \ldots, a_k^N\right)$ is selected. We have shown in prior research that a robot can construct an outcome matrix representing an arbitrary interaction if it has 1) a model of its interactive partner and 2) a model of itself [7].

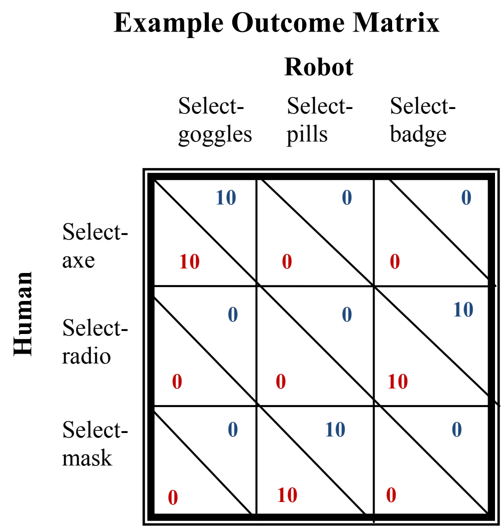

Fig. 1. An example outcome matrix is depicted above. This outcome matrix represents a coordination game in which the robot and the human only receive positive outcome if they select complimentary objects.

Norman used the term mental model to describe a person's concept of how something in the world works [24]. We use the term partner model (denoted m^{-i}) to describe a robot's mental model of its interactive human partner. The term self model (denoted m^i) is used to describe the robot's mental model of itself. The superscript *-i* is used to express individual i's partner [25].

The partner and self models used in this research contain three types of information: 1) a set of partner features $\left(f_1^{-i}, \ldots, f_n^{-i}\right)$; 2) an action model, A^{-i}; and 3) a utility function u^{-i}. Partner features are perceptual features used for partner recognition. These features allow the robot to recognize the person in subsequent interactions. The action model contains a list of actions available to that individual. The utility function includes information about the outcomes obtained by that individual when the robot and the human select a pair of actions. Information about the partner's beliefs, knowledge, personality, etc. could also conceivably be included in these models but were not included in the research described here.

But how does a robot learn a partner model? One simple method for learning a model of a partner is to just interact with the person, observe their features and action selections, and store this information in the partner model. In previous work we showed that a robot could eventually learn a model of its interactive partner, assuming that the partner's model was static [6].

IV. Stereotyped Partner Models

Edwards defines a stereotype as a perceptual stimulus which arouses standardized preconceptions that, in turn, influence one's response to the stimulus [11]. With respect to this framework, a stereotype is a type of generalized partner model used to represent a collection or category of individual partner models. We have developed algorithms for creating stereotypes from a collection of partner models and for matching of a new interactive partner's perceptual features to an existing stereotype [26]. Stereotype creation is a two phase process. First, partner models are clustered with the centroids of the clusters becoming the partner model stereotype. Next, using the cluster centroids as data, a mapping from partner features to the stereotypes is learned. The following section describes the stereotype creation process in detail.

A. The Create Stereotypes Algorithm

The **create stereotypes algorithm** (fig. 2 top) takes as input a new partner model. Individual partner models are learned by successively interacting with an individual and updating a model with the results from the interaction.

Initially the robot has no partner models at all in its model space. As the robot gains experience socializing with new people, it adds models to its model space. Initially however, the robot must seed its model space with a model of itself, its self model. The self model influences the robot's predictions about others in a manner that appears to be similar to the way children use the categorical self to determine similarity [16]. We argue that this process of adding partner models to the model space and learning categories over this space may relate to the experience gained over the course of human social development. We discuss this proposition in greater detail in the next section.

Once a model of a new partner has been learned, the first step of the algorithm adds the new model to the model space. Next, in lines 2 and 3, each model in the space is assigned to a unique cluster. Lines 4 and 5 perform agglomerative clustering, iterating through each cluster and, if the clusters meet a predetermined distance threshold, merging them. The Jaccard distance, a measure used for determining the similarity of sets, is used to determine the distance between two clusters [27]. The cluster centroids that remain after step four are stereotypes, denoted $s_1, \ldots, s_n$.

In the next phase, the C4.5 decision tree algorithm is used to create a mapping, denoted ψ, from the partner's perceptual features to stereotypes. Data is created by pairing each model's perceptual features to an associated stereotype and that data is used to train the decision tree classifier.

Create Stereotypes Algorithm

Input: Partner model m^{-i}.
Output: Classifier ψ mapping m^{-i}(*features*) to a stereotype.

<u>Cluster phase</u>
1. **Add** m^{-i} to partner model space
2. **for** all models in model space
3. make a cluster
4. **while** centroid-distance $(c_1, c_2) < k$
5. merge-clusters(c_1, c_2)

<u>Function learning phase</u>
6. **for** all models *n* in model space
7. **set** data$_j$←make-pair(m_j(*features*), centroid$_j$)
8 ψ ←train-classifier(data)
9. **return** ψ

Match to Stereotype Algorithm

Input:Partner features $f_1^{-i}, \ldots, f_n^{-i}$.
Output: Partner model m^{-i}.

1. **convert** $f_1^{-i}, \ldots, f_n^{-i}$ to instance of classifier data
2. result ← ψ(classify(instance))
3. m^{-i}← StereotypeList(result)
4. **return** m^{-i}

Fig. 2. Algorithms for creating stereotypes and for matching newly perceived individuals to existing stereotypes. The **create stereotypes algorithm** operates by clustering partner models and then constructing a classifier mapping a partner's perceptual features to a stereotype. The **match to stereotype algorithm** uses the classifier to match a new partner's perceptual features to the closest stereotype.

B. Match To Stereotype Algorithm

When perceiving a new interactive partner, the robot matches the new person's perceptual features to an existing stereotype. This process begins by converting the partner's features into an instance of data for the classifier and then using the classifier to select the correct model (fig. 2 bottom).

In previous work we found that use of this algorithm for stereotyping required 5.75 fewer interactions (out of 20) to obtain an 80 percent rate of correct partner action prediction when compared to relearning a new model for each individual [28].

V. False Stereotypes and Social Development

Even if stereotyped partner models can bootstrap the process of learning about a new person, various types of error can effect stereotype creation and retrieval. Recently we explored two types of error that significantly impact the use of stereotypes: modeling error and outlier error. Modeling error occurs when the robot misperceives the information contained within the partner model, such as the action being performed. For example, if the robot incorrectly perceives the action of putting out a fire as the action of making an arrest, this is a perceptual modeling error. In this case, the robot's action model for that particular firefighter would indicate that the firefighter makes arrests. Modeling errors could potentially cause the robot's stereotype to be inaccurate with respect to the true category.

Stereotyping error can also result from statistical anomalies. Ideally, the stereotype model is created from individuals that are representative of the category. It is possible, however, that the individuals from which the stereotype is created are actually outliers with respect to the overall category. Here the stereotype created is not representative of the category. Consider, for example, a robot that creates a stereotype based on models it has learned from interactions with firefighters. Rather than put out fires and rescue victims, these particular firefighters act as police officers, making arrests and writing tickets. Because the robot's stereotype has been created from outliers of the overall category, predictions based on the stereotype will be incorrect when the robot interacts with a non-outlier member of the category. Outliers differ from modeling errors in that an outlier will consistently select actions that are inaccurate with respect to the true nature of the category. Modeling errors, on the other hand, are random with respect to the individuals.

As mentioned in the previous section, initially the robot's model space is devoid of partner models. The robot must seed its model space with its self model. After interacting with a new partner the robot learns a model of the partner which is then added to the model space. Influenced by work from social psychologists, principally Yamagishi et al., we wondered what would happen if one of the initial models added to the robot's model space was an outlier [1]. In other words, if one of the robot's earliest models does not reflect the true type, how does this impact performance with later individuals. We hypothesized that encountering an outlier early in the robot's social development would impact the robot's performance longer and to a greater extent than encountering an outlier later in development. We believed that, because these models act as the foundation for a new category, early errors would strongly influence the development of the category. On the other hand, we further hypothesized that timing (early or late) would not affect modeling errors.

VI. Experiment

A coordination game was used to test our hypothesis. A coordination game is a game-theoretic social situation in which both individuals receive maximal reward only if they select coordinating actions [25]. Figure 1 depicts an example of an outcome matrix representing a coordination game. In this example, both individuals receive an outcome of 10 if they select action pairs (*select-goggle*, *select-axe*), (*select-badge*, *select-radio*), or (*select-pills*, *select-mask*) and 0 outcome if any other action pair is selected. The notional scenario for the experiment is a situation in which a robot acts as a cooperative assistant to a human. In this scenario, the robot must select the best tool to assist its human partner. The robot, however, does not initially know, and must learn, the tool preferences for each type of partner. Table I lists all of the tools used in these experiments and the groupings of matching tools. In order to receive maximal outcome the robot needs to predict the tool that the person is going to select and to then select the tool that matches.

We conducted a numerical simulation to evaluate impact of modeling error and outlier error. A numerical simulation of interaction focuses on the quantitative results of the algorithms and processes under examination and does not attempt to simulate aspects of the robot, the human, or the environment. As such, this technique offers advantages and disadvantages as a means for discovery. One advantage of a numerical simulation experiment is that a proposed algorithm can potentially be tested on thousands of outcome matrices representing thousands of social situations. This allows one to evaluate the statistical significance of the results. One disadvantage is that, because it is not tied to a particular robot, robot's actions, human, human's actions, or environment, the results, while extremely general, have not been shown to be true for any existent social situation, robot, or human. We have, however, conducted several experiments involving a real robot and human using this paradigm to test other hypotheses.

TABLE I. GROUPINGS OF TOOL TYPES

Type	Tools				
1	Extinguisher	Axe	Flashlight	Helmet	Goggles
2	Antiseptic	Mask	Neckbrace	Pills	Bandage
3	Binoculars	Radio	Handcuffs	Badge	Batton

TABLE II. PARTNER FEATURES AND POSSIBLE VALUES

Feature Name	Values
Badge	yes, no
Uniform color	brown, green, blue
Head Gear	yes, no
Head Gear Color	black, green, blue
Hair Color	black, blonde, red
Beard	yes, no
Facial Symmetry	highly, symmetric, asymmetric
Facial Length	very wide, square, long, very long
Skin Color	light, dark
Glasses	yes, no
Age	young, old, medium
Body Type	thin, heavy, medium
Height	tall, small, medium
Gender	male, female

The human was simulated by providing the robot with a list of perceptual features from Table II representing a nominal person. The robot used this information in conjunction with the algorithms for creating and matching stereotypes (fig 2.) to obtain a stereotyped partner model of the person. This partner model was then used to predict the tool that would be selected by the simulated person. Using this prediction the robot made its own tool selection. Finally, the simulated human selected their tool in accordance with the experimental condition and a numerical outcome value was awarded.

The dependent variable in this experiment was the mean number of correct coordinations performed by the simulated human and the robot. In this paradigm, correct coordinations represent a measure of task success. The independent variables were the type of error introduced (outlier or misrecognition) and whether the error occurs early or late. Hence four different conditions were examined. In the modeling error condition error was added at a rate of 50 percent. Modeling error was introduced by giving the robot a 50 percent chance of incorrectly perceiving the human's action selection. In the outlier error conditions two partners were assigned the role of outlier. Again outliers, although perceptually similar to a firefighter, consistently selected tools not associated with firefighting. For both types of error in the early conditions these errors occurred while interacting with the 2^{nd} and 3^{rd} partners. In the late conditions these errors occurred while interacting with the 12^{th} and 13^{th} partners.

Thirty trials of the experiment were run in order to obtain statistical significance. A single trial consisted of 15 interactions in the game with 15 different individuals. Hence, a score of 15 correct coordinations is the best possible score for interactions with a particular partner. The build stereotype algorithm was used to create new stereotypes after interacting with each partner. The stereotypes that resulted were then used to predict the partner's action selection.

The results of this experiment are presented in Fig. 3. When an early outlier is introduced (the red line) the robot's performance in the coordination task is impacted for the remainder of the experiment. Although the robot's performance gradually improves going from 5.10 to 9.53 over the course of partners 4 thru 15, in this condition the performance remains significantly $(p < 0.01)$ below the performance obtained before the introduction of the error (14.93). In contrast, when the robot encounters the same type of error later in the experiment (partners 12 and 13), its performance rebounds after only two partners to 11.97. Put another way, when the outlier error occurs late it only takes one partner for the performance to rebound from 3.8 to 11.43. When the same error occurs early, after 11 additional partners the performance still has not fully recovered.

The results do not depict the same trend for modeling errors. Modeling error (purple and blue lines) only appears to impact performance while these errors are occurring. Moreover, the sum of the performance over all 15 partner is 207.13, 213.20, 125.20, and 186.70 for the early modeling error, late modeling error, early outlier error, and late outlier error respectively. Thus earlier outliers impact performance to a greater extent than any other type of error.

Hence, we can conclude that the experiments support our hypothesis that earlier outliers impact a robot's performance longer and to a greater extent than late outliers. In addition, we have shown that outlier errors but not modeling errors, affect performance in this manner.

VII. Conclusion

This article has used an algorithm for stereotype creation and matching to explore the possibility that early outlier errors impact performance more than late outlier errors. Simulation results confirmed our hypothesis. Namely we found that outlier errors which occur early in the robot's social development impact social task performance to a significantly greater extent than when the same errors occur later during the robots social develop. In other words, the first examples corresponding to a specific type of person are critical for learning that category of individual. Moreover, we found that this critical period impacts one type of error (outliers) but does not impact another (modeling errors).

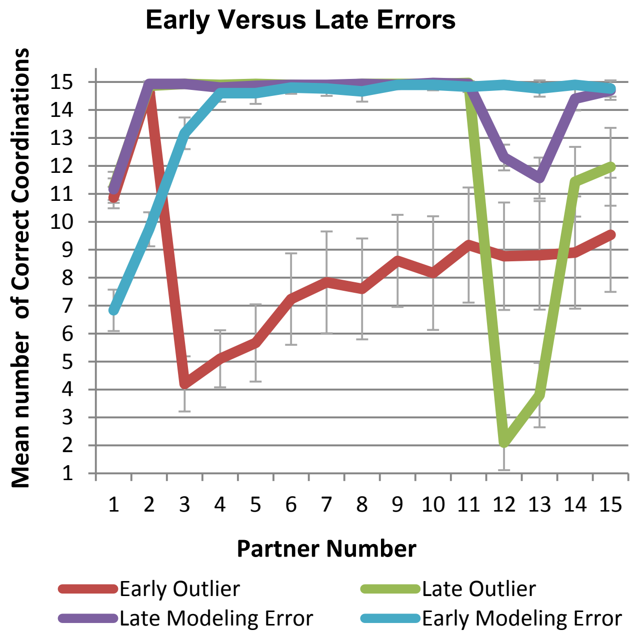

Fig. 3. The figure above depicts results from an experiment investigating the difference between early and late errors. The red line depicts the coordination performance when an outlier occurs early. In this case, performance rebounds slowly over the course of the remaining partners. The same behavior is not observed in the late outlier (green), early modeling error (blue), or late modeling error conditions (purple).

These results are important for several reasons. First, they indicate not just that different types of errors impact social category learning in different ways, but also that the timing of these errors is critical. Hence the creation of a robot that learns and uses stereotypes would be aided by ensuring that the robot's first interactive partners are not outliers. To take this point a bit further, because the data highlights the importance of a robot's first interactive partners, it could be argued that the work foreshadows the potential importance of a type of robot parentage. This parentage would operate by providing correct, stable initial models from which the robot could contrast the behavior of later individuals that it meets.

The results also indicate the resiliency of the stereotyping algorithm when presented with modeling errors. Even at a rate of 50% modeling error, this type of error has little impact on the algorithm's performance regardless of whether these errors occur early or late in the robot's social development. Because modeling errors are perceptual in nature, this result might serve as evidence that strong perceptual ability by a robot is not critical for early social development. Similar results for children have also been reported in the developmental psychology literature [29].

Finally, and most importantly, although the data is unquestionably a reflection of the learning algorithms used, the results do appear to resemble some of the macroscopic social phenomena that is witnessed by social psychologists [1]. Notably, Yamagishi, Kikuchi, and Kosugi have demonstrated that one's initial partner models have an important influence on one's trust and gullibility later in life. Clearly the experiments conducted for this paper are not sufficient to claim that the use of these algorithms results in the same or a similar social development as that which occurs in human beings. Nevertheless, the data may be an indicator that some of the same computational process that serves as the foundation for our stereotyping algorithm also underlies portions of human social development.

The purpose of this research is not simply to design socially optimized robots but rather to create computational algorithms that allow a robot to develop socially. The algorithm and the data that result from the use of the algorithm rest on a very basic set of assumptions, namely that the robot and its interactive partners have reward functions and that the robot learns and stores information about the individuals with which it has interacted. We therefore find the possibility that the results may potentially hint at the cause of macroscopic social phenomena often witnessed during normal human social interaction intriguing.

Still, the preliminary results presented in this article are meant to serve more as a starting point then a conclusion. This article is limited in that it only explores one type of stereotype (prototypes), over a limited number or partners, and during a limited number of interactions. Moreover, the models that the robot created of each individual were rather narrowly tailored to a well-defined task. Future work should examine whether there are limitations to the technique's scalability and, if so, whether these limitations influence the results presented above.

This research is a small portion of a larger effort. To date we have conducted several experiments related to robot stereotype creation and usage. These experiments have been performed both in simulation and with real robots. In addition to firefighters, we have created stereotyped models of police officers and EMTs. Some of our ongoing experiments examine the use of stereotype models as a source of inference about a new person and the inclusion of situation specific characteristics determining the appropriateness of a stereotype. Future work may also explore potential applications of this research in the areas of home healthcare and security.

REFERENCES

[1] T. Yamagishi, M. Kikuchi, and M. Kosugi, "Trust, gullibility, and social intelligence," *Asian Journal of Social Psychology*, no. 2, pp. 145-161, 1999.

[2] R. D. Ashmore, "Sex stereotypes and implicit personality theory," in *Cognitive processes in stereotyping and intergroup behavior* . Hillsdale, NJ: Erlbaum, 1981, pp. 37-81.

[3] E. R. Smith and M. A. Zarate, "Exemplar-based model of social judgement," *Psychological Review*, pp. 3-21, 1992.

[4] F. E. Aboud, "Interest in ethnic information: A cross-cultural developmental study," *Canadian Journal of Behavioral Science*, no. 9, pp. 134-146, 1977.

[5] D. J. Schneider, *The Psychology of Stereotyping*. New York: The Guilford Press, 2005.

[6] J. Denzinger and J. Hamdan, "Improving observation-based modeling of other agents using tentative stereotyping and compactification through kd- tree structuring," *Web Intelligence and Agent Systems: An International Journal*, vol. 4, no. 3, pp. 255-270, 2006.

[7] A. R. Wagner, "Creating and using matrix representations of social interaction," in *Human-Robot Interaction (HRI)*, San Diego, CA, 2009.

[8] L. Festinger, "A theoretical interpretation of shifts in level of aspiration," *Psychological Review*, vol. 49, pp. 235-250, 1942.

[9] L. A. Hirschfeld, *Race in the making: Cognition, culture, and the child's construction of human kinds*. Cambridge, MA: MIT Press, 1996.

[10] L. Terveen, "An overview of human–computer collaboration," *Knowledge-Based Systems*, vol. 8 , no. (2–3), 1994.

[11] A. L. Edwards, "Studies of stereotypes: I. The directionality and uniformity of responses to stereotypes," *Journal of Social Psychology*, vol. 12, pp. 357-366, 1940.

[12] T. Yamagishi, "Trust as a form of social intelligence," in *Trust in Society*. New York, NY: Russell Sage Foundation, 2001, pp. 121-147.

[13] M. Foddy, M. J. Platow, and T. Yamagishi, "Group-based trust in strangers," *Psychological Science*, vol. 20, no. 4, pp. 419-421, 2009.

[14] B. Lasker, *Races attitudes in children*. New York: Holt, 1929.

[15] M. E. Goodman, *Race awareness in young children*. Cambridge, MA: Addison-Wesley, 1952.

[16] F. E. Aboud, "Social and cognitive bases of ethnic identity constancy," *Journal of Genetic Psychology*, vol. 145, pp. 217-230, 1984.

[17] P. G. Ramsey, "Young children's thinking about ethnic differences," in *Children's ethnic socialization*. Beverly Hills, CA: Sage, 1987, pp. 52-72.

[18] E. Rich, "User modeling via stereotypes," *Cognitive Science*, vol. 3, no. 197, pp. 329-354, 1979.

[19] R. Kass and T. Finin, "General user modelingD: A facility to support intelligent interaction," in *Intelligent User Interfaces*. ACM Press, 1991.

[20] A. Ballim and Y. Wilks, "Beliefs, stereotypes, and dynamic agent modeling," *User Modeling and User-adapted Interaction*, vol. 1, no. 1, pp. 33-65, 1991.

[21] C. Burnett, T. J. Norman, and K. Sycara, "Bootstrapping trust evaluations through stereotypes," in *International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, Toronto, CA, 2010.

[22] T. Fong, C. Thorpe, and C. Baur, "Collaboration, dialogue, and human–robot interaction," in *Proceedings of the International Symposium on Robotics Research*, 2001.

[23] B. Duffy, "Robots social embodiment in autonomous mobile robotics," *International Journal of Advanced Robotic Systems*, vol. 1, no. 3, 2004.

[24] D. Norman, *Some Observations on Mental Models*, D. Genter and A. Stevens, Eds. Hillsdale, New Jersey: Erlbaum Associates, 1983.

[25] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*. Cambridge, MA: MIT Press, 1994.

[26] A. R. Wagner, "Using cluster-based stereotyping to foster human-robot cooperation," in *Proceedings of the International Conference on Intelligent Robots and Systems* (IROS), Villamoura, Portugal, 2012.

[27] P. Jaccard, "Étude comparative de la distribution florale dans une portion des Alpes et des Jura," *Bulletin de la Société Vaudoise des Sciences Naturelles*, vol. 37, p. 547–579, 1901.

[28] A. R. Wagner, *The Role of Trust and Relationships in Human-Robot Social Interaction*. Ph.D. diss., School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, 2009.

[29] L. E. Berk, *Child Development*, 3rd ed. Needham Heights, MA: Allyn and Bacon, 1994.