Using Stereotypes to Understand One's Interactive Partner (Extended Abstract)

Alan R. Wagner
Georgia Institute of Technology
85 Fifth Street, Room S27
Atlanta, GA
1.404.894.9311
alan.wagner@gatech.edu

ABSTRACT
Psychologists note that humans regularly use categories to simplify and speed the process of person perception [1]. The influence of categorical thinking on interpersonal expectations is commonly referred to as a stereotype. This research explores the construction and use of stereotypes in human-robot interaction. We present a novel algorithm that creates generalized models of a robot's interactive partner. The results of this work have potential implications for social robotics, autonomous agents, and possibly psychology.

Categories and Subject Descriptors
I.2.9 [**Artificial Intelligence**]: Robotics – *autonomous vehicles, operator interfaces*

General Terms
Algorithms, Human Factors.

Keywords
Mental model, interaction, interdependence theory, game theory.

1. INTRODUCTION
Macrae and Bodenhausen suggest that categorical thinking influences a human's evaluations, impressions, and recollections of the target. The influence of categorical thinking on interpersonal expectations is commonly referred to as a stereotype. For better or for worse, stereotypes have a profound impact on interpersonal interaction [2]. Information processing models of human cognition suggest that the formation and use of stereotypes may be critical for quick assessment of new interactive partners [3]. From the perspective of a roboticist the question then becomes, can the use of stereotypes similarly speedup the process of partner modeling for a robot?

This question is potentially critical for robots operating in complex, dynamic social environments, such as search and rescue. In environments such as these the robot may not have time to learn a model of its interactive partner through successive interactions. Rather, the robot will likely need to bootstrap its modeling of the partner with information from prior, similar partners. We argue that stereotypes can serve this purpose.

This paper presents an algorithm for creating and using stereotyped partner models to hasten learning about a robot's interactive partner. Our techniques are not tied to specific social environments or paradigms. Moreover, the algorithm contributed is not just limited to robots per se, but rather constitute a general investigation of the use of stereotypes by robots, agents, or interactive control software. This extended briefly outlines our algorithm for building and using stereotype partner models.

Cite as: Using Stereotypes to Understand One's Interactive Partner (Extended Abstract), Alan R. Wagner, *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. XXX-XXX.

2. STEREOTYPE PARTNER MODELS
We use the term partner model (denoted m^{-i}) to describe a robot's mental model of its interactive human partner. The superscript -*i* is used to express individual *i*'s partner. Our partner model contains 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}. We use the notation $m^{-i}.A^{-i}$ and $m^{-i}.u^{-i}$ to denote the action model and utility function within a partner model. Partner features are used for partner recognition. Partner features allow the robot to recognize the partner in subsequent interactions. The partner's action model contains a list of actions available to that individual. The partner's utility function includes information about the outcomes obtained by the partner when the robot and the partner select a pair of actions.

Sears, Peplau and Taylor define a stereotype as an interpersonal schema relating perceptual features to distinctive clusters of traits [4]. Hence a stereotype is a type of generalized partner model used to represent a collection or category of individual partner models. Thus, the creation of stereotypes requires the creation of these generalized partner models. Moreover, to be useful, techniques capable of matching a new interactive partner's perceptual features to an exiting stereotype must exist. Stereotype building will therefore be a two phase process. First, we cluster partner models with the centriods of the clusters becoming the partner model stereotype. Next, using the cluster centroids as data, we learn a mapping from partner features to the stereotypes.

2.1 Building Stereotypes
The building stereotypes algorithm (Figure 1) takes as input a new partner model. The first step of the algorithm adds the new model to the model space. Next each model in the space is assigned to a unique cluster. The third and fourth steps perform agglomerative clustering, iterating through each cluster and, if the clusters meet a predetermined distance threshold, merging them. Equations (1) and (2) from section 2.2 (below) for partner model distance are

used to determine if the clusters meet the predetermined distance threshold for merging. The cluster centroids that remain after step four are the stereotypes, denoted $S_1, \ldots, S_n$. A list of stereotype models is kept by the robot.

Building Stereotypes Algorithm

Input: Partner Model m^{-i}.

Output: Classifier ψ mapping m^{-i}.features to a stereotype.

Cluster phase

1. **Add** m^{-i} to partner model space
2. **for** all models in model space
3. make a cluster
4. **while** `centroid_distance`$\left(c_j, c_k\right) < k$,
5. `merge_clusters`$\left(c_j, c_k\right)$

Function learning phase

6. **for** all models *n* in model space
7. **set** data[*n*]←`make_pair(`m^{-i}.features, centroid)
8. ψ ←`build_classifier(` data)
9. **return** ψ

Figure 1. Algorithms for building stereotypes. The building stereotypes algorithm operates by clustering partner models and then constructing as classifier mapping a partner's perceptual features to a stereotype.

In the next phase we use the C4.5 algorithm to create decision trees, denoted ψ, mapping the partner's perceptual features to the stereotype. Line 7 from Figure 1 creates data for the C4.5 algorithm by pairing each model's perceptual features to a stereotype. In the final steps, this data is used to train a classifier mapping partner features to the stereotyped model.

The stereotype building algorithm makes two important assumptions. First, it assumes the existence of a distance function, $d\left(m_i^{-i}, m_j^{-i}\right)$, capable of measuring the difference between two partner models. We describe below our method for measuring partner model distance (see section 2.2). If, however, additional information (such as the partner's beliefs, motivations, goals, etc.) is added to the partner model, then creating a distance function may become difficult because this information may not naturally have a measure for determining distance. Second, the stereotype building algorithm assumes that partner models can be merged to create new partner models. In order to merge a partner model one must merge the components of the partner model. For this work that meant merging the action models and utility functions. Action models were merged by adding an individual action to the stereotype only if the action was included in half of the data that composed the merged model. Similarly, merged utility values were derived from the average utility value of the composition utility functions.

To use a stereotype the robot simply converts a newly encountered partner's perceptual features into an instance of data for the classifier and then uses the classifier to select the correct stereotype model. One important question is how the algorithm reacts to partners that conflict with its stereotypes. Briefly, if interaction with the new partner does not match what is predicted by the stereotype, then the model for the individual can be altered and add back to the partner model space resulting in a more generalized stereotype.

2.2 Determining Model Accuracy

But how do we measure the distance from one partner model to another? For example, given a particular human partner with action set $m^{-i}.A^{-i}$ and utility function $m^{-i}.u^{-i}$, how close is the robot's partner model m^{-i} to the actual model ${}^{*}m^{-i}$? We address this problem by viewing action models and utility functions as sets. The action model is a set of actions and a utility function is a set of triplets ($\left\langle a^i, a^{-i}, r \in \Re \right\rangle$) containing the action of each individual and a utility value. We can then do set comparisons to determine the accuracy of the robot's partner model m^{-i}.

Two types of error are possible. Type I error (false positive) occurs if an action or utility is added to the robot's partner model (m^{-i}) which is not in the actual model (${}^{*}m^{-i}$). Type II error (false negative) occurs if an action or utility in the actual model (${}^{*}m^{-i}$) is not included in robot's partner model (m^{-i}). The two types of errors are averaged in the equation,

$$d = 0.5\left(\frac{\left|m^{-i} - {}^{*}m^{-i}\right|}{\left|m^{-i}\right|}\right) + 0.5\left(1 - \frac{\left|{}^{*}m^{-i} \cap m^{-i}\right|}{\left|{}^{*}m^{-i}\right|}\right) \qquad (1)$$

to create *d*, an overall measure of model accuracy (or distance) for either an action model (d^a) or a utility function (d^u). To determine overall partner model accuracy we average the error from both components of the partner model,

$$d^{-i} = \frac{d^a + d^u}{2} \qquad (2)$$

3. REFERENCES

[1] C. N. Macrae and G. V. Bodenhausen, "Social Cognition: Thinking Categorically about Others " *Annual Review of Psychology*, vol. 51, pp. 93-120, 2000.

[2] J. A. Bargh, M. Chen, and L. Burrows, "Automaticity of social behavior: direct effects of trait construct and stereotype activation on action," *Journal of Personality and Social Psychology*, vol. 71, pp. 230-44, 1996.

[3] G. V. Bodenhausen, C. N. Macrae, and J. Garst, "Stereotypes in thought and deed: social-cognitive origins of intergroup discrimination," in *Intergroup Cognition and Intergroup Behavior*, C. Sedikides, J. Schopler, and C. A. Insko, Eds. Mahwah, NJ: Erlbaum, 1998, pp. 311–36.

[4] D. O. Sears, L. A. Peplau, and S. E. Taylor, *Social Psychology*. Englewood Cliffs, New Jersey: Prentice Hall, 1991.