The Odd One Out Task: Toward an Intelligence Test for Robots

Jivko Sinapov and Alexander Stoytchev
Developmental Robotics Laboratory
Iowa State University
{jsinapov, alexs}@iastate.edu

***Abstract*—Detecting the outlier in a set of objects is a fundamental task used in a wide variety of intelligence tests. This paper proposes a theoretical model that allows a robot to interactively estimate the pairwise similarity between everyday objects and use this knowledge to solve the odd one out task. That is, given a set of objects, the robot's task is to select the one object that does not belong in the group. In our experiments, the robot interacted with fifty different household objects (by applying five different exploratory behaviors on them) and perceived auditory and proprioceptive sensory feedback. Pairwise object similarity was estimated for different behavior and modality contexts. In a series of subsequent tests, three objects from a given category (e.g., three cups or three pop cans) along with one object from outside that category were selected and the robot's internal models were queried to pick the object that does not belong in the group. The object similarity relations learned by the robot were used to pick the most dissimilar object, with success rates varying from 45% to 100%, depending on the category. The results show that the learned similarity measures were sufficient to capture some of the common properties of human-defined object categories, such as cups, bottles, and pop cans.**

I. Introduction

Detecting an item that does not belong in a given set is a standard problem in modern Intelligence Quotient (IQ) tests. This is known as the *odd one out* task, which is formulated as follows: given a set of items, the participant is asked to decide which one among them is most dissimilar from the rest. Variants of this task have been used extensively in a wide variety of disciplines to test for brain abnormalities [1], learning disabilities [2], and categorization abilities [3]. It has also been used to probe the cultural and social foundations of cognition [4]. Typically, the presented items vary along one dimension (e.g., size, shape, color), which, if identified by the participant, could be used to pick the most dissimilar item. In more complex settings, however, picking the odd object requires comparing along multiple sensory dimensions [3]. This task has also been tried in the auditory domain with human participants [5], which indicates that the general principles used to pick the odd item are not necessarily tied to the visual sensory modality.

The ubiquity of the odd one out task makes it an attractive candidate for an intelligence test in developmental robotics. The task has been used extensively to study how humans estimate object similarity and form object categories. Therefore, it may also be a valuable tool for conducting experiments with robots. Recent work in robotics has focused on detecting object similarities and forming object categories [6], [7], [8], [9], [10], [11], [12], indicating that robots should, in principle, be capable of solving the odd one out task in a variety of settings.

We propose a framework that allows a robot to estimate the similarity between objects based on its prior interactive experience with them. A theoretical model is presented that uses the estimated object relations to solve the odd one out task by selecting the most dissimilar object from a given set. The experiments were conducted with an upper-torso humanoid robot, which interacted with fifty different objects by applying five types of exploratory behaviors (lift, shake, drop, crush, and push). Over the course of each interaction, the robot detected auditory and proprioceptive sensory feedback. The robot was able to estimate pairwise object similarity relations for each behavior-modality context, which were used to select the odd object in subsequent tests. The framework was repeatedly evaluated on six natural object categories (e.g., cups, bottles, pop cans, etc.). During each test, a group of three objects from the target category and one object from outside the category were presented. The robot's internal models were queried to pick the most dissimilar object. The results show that the estimated object relations were successful in capturing the properties of natural object categories, since the robot was able to solve the task with success rates substantially better than chance. This suggests that it may be possible to ground the semantic labels for many object categories in the robot's sensorimotor experience.

II. Related Work

A. Psychology and Cognitive Science

Asking participants to pick the odd item from a set is a task that can provide valuable insights into how humans categorize objects. One of the early experiments that used this task was performed by Luria, who studied how social and cultural upbringing affect development [4]. Uneducated Soviet peasants were shown images of four objects (e.g., hammer, saw, hatched, and wooden log) and asked to select the object that does not belong in the group. The goal of this test was to determine whether the participants grouped items together based on their semantic category (e.g., handheld tools) or not.

Other researchers have used the odd one out task to study how humans measure perceptual similarity. Stephens *et al.* [3] investigated how people establish similarity relations for three-dimensional models of animal-like objects called "greebles."

During each trial, the participants were asked to pick the odd one out from a set of three greebles. The data was used to generate a pairwise matrix that specified the similarity for each pair of greebles, as determined by the participants. The study presented in our paper solves the opposite problem: the robot first estimated pairwise measures of object similarity, and then used these measures to solve the odd one out task.

In another notable experiment, Roberson *et al.* [2] used the odd one out task to study the relationship between perceptual similarity and object categorization. By examining a patient's performance on this task, they concluded that the mapping between a perceptual representation (e.g., color) and the corresponding category label (e.g., the name of the color) is not as transparent as previously thought. In relation to developmental robotics, this study suggests that the odd one out task may indeed be useful as a testbed for studying how well the robot's perceptual experience with an object matches the object's human-defined category label.

B. Robotics

Robots that can estimate the similarity between objects and form meaningful object categories would be more useful in dynamic and unstructured environments. Related work in robotics has demonstrated that, through active interaction, robots can derive a measure of perceptual as well as functional object similarity [6], [8], [9], [12]. For example, Natale *et al.* [7] used a Self-Organizing Map to illustrate the haptic similarity between objects, obtained as their robot repeatedly grasped them and recorded tactile sensations. Sinapov *et al.* [10] demonstrated that a robot can estimate the similarity between objects based on the sounds that the objects generate when different behaviors are performed on them.

Other related research has focused on categorizing objects in terms of their functional properties. The simulated robot in [13] was able to establish how similar two tools are based on what the tools allow the robot to do. Modayil *et al.* [14] introduced a general framework that allows a robot to discover classes of objects, based on their detected percepts over the course of an interaction. Griffith *et al.* [11] demonstrated that a robot can form the functional category of "containers" by repeatedly observing visual movement patterns of objects dropped in, or near, the container. Along with other published research, these results give a strong indication that, in the right setting, robots should be able to solve the odd one out task.

III. Experimental Setup

The experimental setup and the dataset used in this study are identical to the ones used by Bergquist *et al.* [15]. Due to space limitations, they are only briefly summarized here. The robot was an upper-torso humanoid robot with two 7-DOF Barrett WAMs for arms and two 3-finger Barrett Hands as end effectors. The robot's head was equipped with a Audio-Technica U853AW cardioid microphone.

The robot performed five exploratory behaviors on each object: *lift*, *shake*, *drop*, *crush*, and *push* (shown in Fig. 1). The behaviors were encoded with the Barrett WAM API and performed with the robot's left arm. The raw proprioceptive

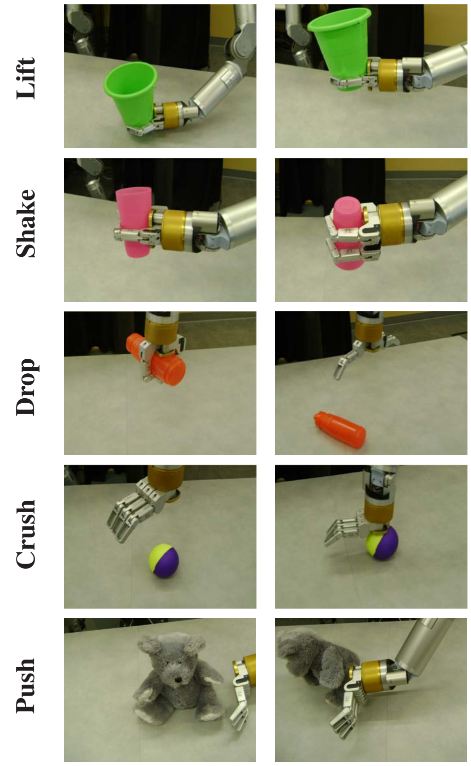


Fig. 1. *Before* and *after* snapshots of the five behaviors used by the robot.

data (i.e., joint torques of the left arm) and the raw audio were recorded for the duration of each behavior (start to end).

The proprioceptive and auditory feedback for each behavioral interaction were represented as discrete sequences, where each sequence element corresponded to the most highly activated state in an 6-by-6 Self-Organizing Map (SOM) [16]. One SOM was trained for each modality, as described in [15] and [10]. For example, given a specific joint-torque configuration (i.e., a vector in $\mathbb{R}^7$), the data point is fed as input to the proprioceptive SOM and the index of the most highly activated state in the map is appended as the next token in the proprioceptive sequence for that behavioral interaction. Similarly, given a Discrete Fourier Transform (DFT) of a recorded sound, each column vector of the DFT is given as input to the auditory SOM and the index of the most highly activated state is added as the next token in the auditory feedback sequence. The proprioceptive SOM was trained with sample joint-torque configurations experienced by the robot, while the auditory SOM was trained with a set of column vectors extracted from the recorded DFTs. This procedure is described in much more detail by Bergquist *et al.* [15] for proprioception and by Sinapov *et al.* [10] for audio.

After each behavioral interaction is performed, the robot records two sequences, $X_{prop} = p_1 p_2 \ldots p_k$ and $X_{audio} = a_1 a_2 \ldots a_l$, as shown in Fig.2. The two sequences are not necessarily of the same length, since proprioception and audio are sampled at different frequencies. Finally, the robot needs a metric that can establish the similarity between two sequences from the same sensory modality. As described in [15] and [10], the global alignment similarity function was used, which is a

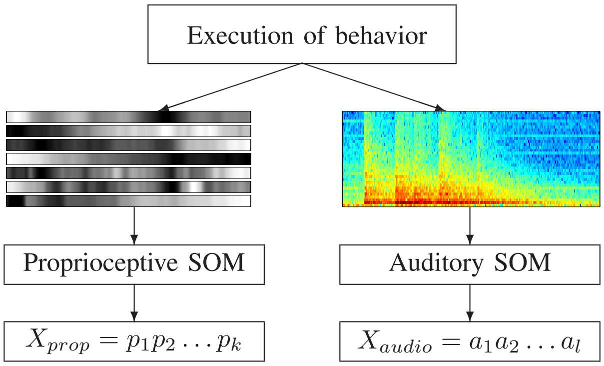


Fig. 2. The procedure used to turn the high-dimensional proprioceptive and auditory signals into low dimensional discrete sequences. First, the robot performs a behavior on the object and records the joint-torque data and the Discrete Fourier Transform of the audio signal. Each 7-dimensional configuration of the joint-torque data is fed as input to the proprioceptive SOM and the index of the SOM state with the highest activation value is added as a token in the resulting sequence, X_{prop}. Similarly, each DFT column vector of the recorded spectrogram is mapped to a state in the auditory SOM, resulting in the auditory feedback sequence X_{audio}.

common choice for comparing discrete sequences.

The robot interacted with a set of objects, $\mathcal{O}$, consisting of 50 common household objects, including cups, bottles, and toys (see Fig. 3). The figure also shows several natural object categories formed by the objects, which were used in the evaluation of the robot's performance on the odd one out task. During each test, 3 objects from a given category (e.g., pop cans) and 1 from outside the category were chosen. The robot's model was queried to select the object that, according to the robot's internal representation, does not belong in the group. The next section describes the theoretical model used by the robot to solve this task.

IV. Methodology

This section describes the three stages used to solve the odd one out task. First, the robot interacts with the set of all objects, $\mathcal{O}$, by performing each of its exploratory behaviors on every object while recording the detected proprioceptive and auditory feedback. Second, after the interaction stage is over, the robot estimates a pairwise $|\mathcal{O}| \times |\mathcal{O}|$ object similarity matrix, $\mathbf{W}$, such that W_{ij} denotes the similarity between objects i and j. Finally, when presented with 4 (or in the general case, K) different objects, the robot uses the similarity matrix $\mathbf{W}$ to select the one object that does not belong. The next subsections describe these three stages in more detail.

A. Interacting with Objects

Let $\mathcal{B} = \{$*lift*, *shake*, *drop*, *crush*, *push*$\}$ be the set of N exploratory behaviors of the robot and let M be the number of sensory modalities (in our case, $N = 5$, and $M = 2$). Each behavior-modality combination (e.g., *lift-proprioception*) determines a context, which we will denote by $c \in \mathcal{C}$, where $\mathcal{C}$ is the set of all contexts. In our experiments, the size of $\mathcal{C}$ was $|\mathcal{C}| = N \times M = 5 \times 2 = 10$.

Given a context $c \in \mathcal{C}$, and an object $i \in \mathcal{O}$, let $\mathcal{X}_c^i = [X_1, \ldots, X_D]$ be the set of sensory feedback sequences

Fig. 3. The six object categories, along with the remaining 25 objects, used in this study. An object may belong to more than one category - e.g., the three pop cans also belong to the set of *metal objects*. One of the pop bottles was full during the experiments and is not included in the *empty bottles* set.

detected while interacting with object i in context c. Each behavior was performed 10 times on each object, hence $|\mathcal{X}_c^i| = 10$. As described below, the robot estimates the similarity between objects using the sets $\mathcal{X}_c^i$ for all modality-behavior contexts $c \in \mathcal{C}$ and all objects $i \in \mathcal{O}$.

B. Estimating the Similarity between Objects

Next, the robot estimates an $|\mathcal{O}| \times |\mathcal{O}|$ pairwise object similarity matrix $\mathbf{W}$ such that each entry W_{ij} denotes how similar objects i and j are. The similarity matrix is calculated in two steps: 1) for each of the 10 contexts $c \in \mathcal{C}$, estimate an object similarity matrix $\mathbf{W}^c$; and 2) combine the 10 estimated similarity matrices $\mathbf{W}^c$ into a single consensus similarity matrix $\mathbf{W}$.

Let $\mathcal{X}_c^i$ and $\mathcal{X}_c^j$ be two sets containing the sensory feedback sequences detected in context c with objects i and j, respectively. In our experiments, each set contained 10 such sequences, recorded while performing the same behavior ten times with each object. Let $sim(X_a, X_b)$ be the global alignment similarity function that measures the similarity between two sequences $X_a \in \mathcal{X}_c^i$ and $X_b \in \mathcal{X}_c^j$. For context $c \in \mathcal{C}$, the similarity between two objects i and j can be defined as the expected pairwise similarity of two sequences X_a and X_b:

$$W_{ij}^c = \mathbf{E}[sim(X_a, X_b) | X_a \in \mathcal{X}_c^i, X_b \in \mathcal{X}_c^j]$$

The expected value is estimated as follows:

$$\frac{1}{|\mathcal{X}_c^i| \times |\mathcal{X}_c^j|} \sum_{X_a \in \mathcal{X}_c^i} \sum_{X_b \in \mathcal{X}_c^j} sim(X_a, X_b)$$

In other words, the entry W_{ij}^c is estimated by calculating the average similarity of all possible pairs of sensory feedback sequences in the two sets $\mathcal{X}_c^i$ and $\mathcal{X}_c^j$. Let $\mathbf{W}^c \in \mathbb{R}^{|\mathcal{O}| \times |\mathcal{O}|}$ be the resulting pairwise object similarity matrix for behavior-modality combination c. The matrices $\mathbf{W}^c$ for all contexts are used to construct a single consensus similarity matrix, $\mathbf{W}$, using a weighted combination:

$$W_{ij} = \sum_{c \in \mathcal{C}} \alpha_c \times W_{ij}^c$$

where α_c is the weight assigned to context c (i.e., the consensus object similarity matrix $\mathbf{W}$ is a linear combination of the similarity matrices $\mathbf{W}^c$ for all contexts $c \in \mathcal{C}$).

Two different weighting schemes were used to calculate $\mathbf{W}$. In the first, the weights are uniform, i.e., $\alpha_c = \frac{1}{|\mathcal{C}|}$. In the second, it is assumed that the robot can estimate how useful each behavior-modality context is for the task of object recognition. A context that enables the robot to better distinguish between objects is deemed more useful and assigned a higher weight. Let a_c be the object recognition accuracy achieved in context c, estimated by performing 10-fold cross validation on all data recorded in that context and evaluating a classifier that estimates the object identity given the sensory feedback sequence as input. Once these accuracies are estimated, the weights α_c are computed such that $\alpha_c \propto a_c$ and $\sum_{c \in \mathcal{C}} \alpha_c = 1.0$. The classifier used in this stage was the k-Nearest Neighbor classifier with k set to 3, using the global alignment similarity function to rank neighbors. The classifier, the similarity function $sim(X_a, X_b)$, and the cross-validation setup were identical to the ones used by Sinapov *et al.* [10] and Bergquist *et al.* [15].

C. Detecting the Odd Object

Given an object similarity matrix (either a context-specific matrix $\mathbf{W}^c$ or a consensus matrix $\mathbf{W}$), the robot's model is queried to select the most dissimilar object from a test set $\mathcal{T}$ of K objects, where $\mathcal{T} \subset \mathcal{O}$. For example, if presented with three pop cans and a cowboy hat, we expect the hat to be selected as the object that does not belong in that group. The robot's model selects the odd object i such that the pairwise object similarity within the remaining group of $K-1$ objects is maximized, while the similarity between the selected object i and the remaining $K-1$ objects is minimized.

Given a set of objects, $\mathcal{T}$, the odd object is selected as the object i that maximizes the following objective function:

$$q(\mathcal{T}, i) = \alpha_1 \sum_{j \in \mathcal{T}/i} \sum_{k \in \mathcal{T}/i} W_{jk} - \alpha_2 \sum_{j \in \mathcal{T}/i} W_{ij}$$

The first term captures the pairwise object similarity between the remaining objects in $\mathcal{T}$ (i.e., after i is removed from $\mathcal{T}$). The second term captures the similarity between the selected object i and the remaining $K-1$ objects in $\mathcal{T}$. It is worth noting that the objective function is based on the general normalized-cut criterion, which is commonly used in the machine learning community for clustering data points whose similarity is specified by an affinity matrix [17]. The constants α_1 and α_2 are normalizing weights, which ensure that the objective function is not biased towards either one of the two terms. In our case, the weights were set to:

$$\alpha_1 = \frac{1}{(|\mathcal{T}|-1) \times (|\mathcal{T}|-1)}, \; \alpha_2 = \frac{1}{|\mathcal{T}|-1}$$

D. Evaluation

The framework was evaluated as follows. Given a target category (e.g., metal objects), three objects from the category and one from outside the category were chosen at random. The robot's model was then queried to pick the odd object. If the selected object matched the object from outside the category, then the solution was deemed a success. This process was repeated for all possible combinations of three objects from the category and one object from outside the category. For example, consider the *metal objects* category, which has 5 objects (see Fig.3). There are $\binom{5}{3} = 10$ possible ways to select three objects out of five. For each of these, there are $50-5=45$ ways to select a fourth object from the dataset that does not belong to that category. Thus, a total of $10 \times 45 = 450$ odd one out tests were performed with that category. The extensive evaluations of these tests were performed off-line after the robot interacted on-line with all 50 objects.

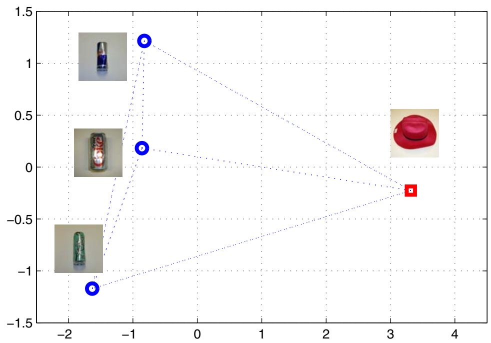
V. Results

A. Example

Figure 4 shows an example task in which the robot is presented with the three pop cans along with the cowboy hat, and is asked to select the object that does not belong in this group. Figure 4.a) shows images of the objects and the pairwise object similarity for these four objects (i.e., a sub-matrix of the uniformly-weighted consensus similarity matrix $\mathbf{W}$). As expected, the matrix shows that the three pop cans are far more similar to each other, than they are to the cowboy hat. To better visualize the similarity relationships between the four objects, the similarity matrix is embedded onto the 2D plane by first converting it into a distance matrix and then applying the ISOMAP method for dimensionality reduction [18]. Figure 4.b) shows the resulting graph. The distance between two nodes in the graph is an approximation of their distance specified in the input to the ISOMAP algorithm. The hat object, maximizes the objective function defined earlier.

Figure 5 shows three more example tasks, including one in which the robot's model makes a mistake. The object selected by the robot's model is denoted by a red square glyph, while the remaining objects are denoted by blue circle glyphs. Figure 5.a) shows an example task in which the robot's model is queried to pick the odd object out of three different types of bottles and a mug. The visualization shows that the mug is clearly the most different object. Figure 5.b) shows a test in which the four objects presented to the robot include three that have contents inside of them (a box of rice, a bottle with pills, and a box with screws) and one that does not (a PVC pipe).

a) Pairwise object similarity matrix for the four objects in this test.



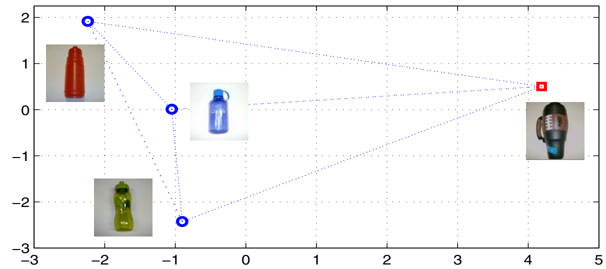
b) 2D embedding of the object similarity matrix shown in a).

Fig. 4. An example odd one out task. Four objects are presented: three pop cans and a cowboy hat. As expected, the hat is selected by the robot's model as the object that does not belong in that group. a) The pairwise object similarity matrix for the four objects (a sub-matrix of the unweighted consensus similarity matrix $\mathbf{W}$). White color indicates high similarity, while black color indicates low similarity. b) A 2D embedding of the pairwise similarity matrix, produced by converting it into a distance matrix and applying the ISOMAP method for non-linear dimensionality reduction. This visualization clearly shows that the cowboy hat is the object in the group that is most distant from the remaining three. The distance between points in the 2D embedding approximates the distance in the matrix used as an input to the ISOMAP algorithm.

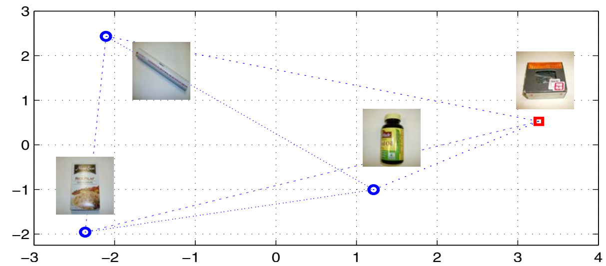
The robot's model selects the box with screws as the most different object, which is an incorrect choice, according to the human-labeled object category (i.e., *objects with contents*). Finally, Figure 5.c) shows a test in which the dumbbell is correctly selected as being different from the three plastic cups.

B. Success Rates Per Object Category

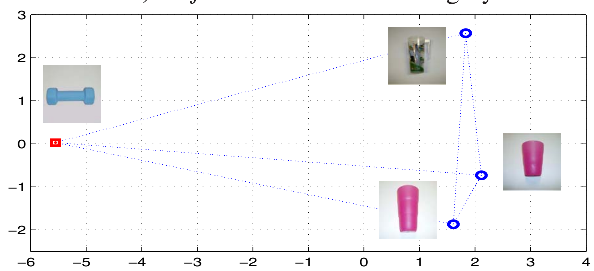
The performance rates for all six object categories are shown in Table I, averaged over all possible instantiations of the odd one out task for each category. Rates are shown for both the uniform weighting scheme as well as the weighting scheme in which contexts are weighted based on their usefulness for distinguishing between objects. In addition, for each category, the individual context that results in the highest success rate is determined and the resulting success rate is reported. The idea behind this test is that certain behavior-modality contexts may be better suited for detecting certain object categories than others. The expected success rate when randomly selecting the odd object is 25% (i.e., randomly choosing 1 out of 4).



a) *Empty bottles* category



b) *Objects with contents* category



c) *Cups* category

Fig. 5. Examples and solutions of the odd one out task with different object categories. See the text for more details.

TABLE I
SUCCESS RATES PER TASK CATEGORY

Category	Uniform Combination	Weighted Combination	Best Context
Pop Cans	100.00 %	100.00 %	100.00 %
Plastic cups	76.59 %	87.23 %	97.87 %
Metal objects	70.00 %	80.00 %	95.33 %
Empty bottles	62.96 %	66.47 %	63.97 %
Soft objects	50.44 %	67.78 %	97.33 %
Objects w/ contents	45.34 %	49.89 %	66.71 %

The results show that the robot's unsupervised model is substantially better than chance for all six object categories. The weighted combination scheme performs better than the uniform combination. The best results are achieved with the *pop cans* category, for which the robot was able to select the object that is not a pop can in 100% of the tests. These results indicate that the robot's behavioral and perceptual repertoire was able to capture the common properties that define pop cans (e.g., material type, specific sounds they generate, weight, etc.). The worst performance is for the *objects with contents* category. The only thing that these objects have in common is that they make noise when shaken (i.e., only 1 of 10 contexts, *shake-audio*, may be able to capture that). The robot's model, however, is completely unsupervised and does not know that the object similarity matrix extracted in the

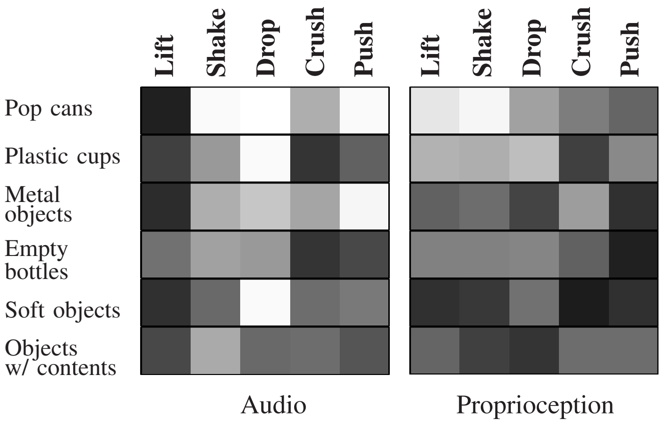


Fig. 6. Success rates for the odd one out task, shown for each category, and for each behavior-modality context. Light color indicates high success rates, while dark color indicates low success rates.

shake-audio context is the most relevant for this category type.

The last column of Table I shows that for most object categories, there exists a specific behavior-modality context that results in a success rate that is higher than the one achieved when using the consensus similarity matrix. For example, when using only the object similarity matrix extracted from the *shake-audio* context, the success rate for the *objects with contents* category jumps to 66.71%. The *empty bottles* category was an exception - in that case, using the weighted consensus similarity matrix results in a higher success rate than with any individual context-specific similarity matrix.

Figure 6 visualizes the success rates for each category when using each context-specific similarity matrix $\mathbf{W}^c$. Light color indicates high success rates. The results show that the properties of different categories are best captured by different behaviors and modalities. For example, the *plastic cups* category is best captured by the *drop-audio* behavior-modality context. This context is also very useful when the robot is evaluated on the *soft objects* category, likely because the robot detects an absence of a loud noise when these objects are dropped on the table. As expected, the *objects with contents* category is best captured by the *shake-audio* behavior-modality combination, since the contents make distinct sounds when the objects are shaken.

VI. Conclusions and Future Work

This paper demonstrated an interactive framework and a theoretical model that allow a robot to solve the *odd one out* task by estimating the similarity relations between objects in different behavior-modality contexts. The experimental evaluation showed that the robot's choice for the odd object was consistent with human-defined object categories, with success rates varying from 45% to 100%, depending on the category. Certain behavior-modality combinations produced object similarity relations that were able to better capture the target category. These results show that sensorimotor interaction can capture many of the physical properties of objects that define an object category.

One limitation of this paper is that the objective function for deciding which of the objects does not belong in the group was pre-defined. Future work can address this by incorporating some amount of human supervision into the overall framework. If the robot knows whether its choice for the odd object is right or wrong, it could potentially estimate which behavior-modality combinations are most suitable for capturing the properties of a target object category. This information can also be used to estimate specific weights for each context in order to learn a new object similarity relation that better captures a given human-defined category. Pursuing this line of research would allow robots to solve a variety of additional tasks, including sorting and ordering objects.

References

[1] M. Buckley, M. Booth, E. Rolls, and D. Gaffan, "Selective perceptual impairments after perirhinal cortex ablation," *Journal of Neuroscience*, vol. 21, no. 24, p. 9824, 2001.

[2] D. Roberson, J. Davidoff, and N. Braisby, "Similarity and categorisation: Neuropsychological evidence for a dissociation in explicit categorisation tasks," *Cognition*, vol. 71, no. 1, pp. 1–42, 1999.

[3] R. Stephens and D. Navarro, "One of These Greebles is Not Like the Others: Semi-Supervised Models for Similarity Structures," in *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, 2008, pp. 1996–2001.

[4] A. R. Luria, *Cognitive development, its cultural and social foundations*. Harvard University Press, 1976.

[5] M. Snowling, C. Hulme, A. Smith, and J. Thomas, "The effects of phonetic similarity and list length on children's sound categorization performance," *Journal of Experimental Child Psychology*, vol. 58, no. 1, p. 160, 1994.

[6] S. Nolfi and D. Marocco, "Active perception: A sensorimotor account of object categorization," in *From Animals to Animats 7: Proc. of the Sixth International Conf. on Simulation of Adaptive Behavior*, 2002.

[7] L. Natale, G. Metta, and G. Sandini, "Learning haptic representations of objects," in *Proceedings of the International Conference on Intelligent Manipulation and Grasping*, 2004.

[8] T. Nakamura, T. Nagai, and N. Iwahashi, "Multimodal object categorization by a robot," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2007. IROS 2007*, 2007, pp. 2415–2420.

[9] S. Takamuku, K. Hosoda, and M. Asada, "Object Category Acquisition by Dynamic Touch," *Advanced Robotics*, vol. 22, no. 10, pp. 1143–1154, 2008.

[10] J. Sinapov, M. Weimer, and A. Stoytchev, "Interactive learning of the acoustic properties of household objects," in *Proc. of the IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2009, pp. 2518–2524.

[11] S. Griffith, J. Sinapov, M. Miller, and A. Stoytchev, "Toward interactive learning of object categories by a robot: A case study with container and non-container objects," in *Proc. of the 8th IEEE Intl. Conf. on Development and Learning*, 2009.

[12] J. Sun, J. Moore, A. Bobick, and J. Rehg, "Learning Visual Object Categories for Robot Affordance Prediction," *The International Journal of Robotics Research*, 2009.

[13] J. Sinapov and A. Stoytchev, "Detecting the functional similarities between tools using a hierarchical representation of outcomes," in *Proc. of the 7th IEEE Intl. Conf. on Development and Learning*, 2008.

[14] J. Modayil and B. Kuipers, "The initial development of object knowledge by a learning robot," *Robotics and Autonomous Systems*, vol. 56, no. 11, pp. 879–890, 2008.

[15] T. Bergquist, C. Schenck, U. Ohiri, J. Sinapov, S. Griffith, and A. Stoytchev, "Interactive object recognition using proprioceptive feedback," in *Proceedings of the 2009 IROS Workshop: Semantic Perception for Robot Manipulation, St. Louis, MO*, 2009.

[16] T. Kohonen, *Self-Organizing Maps*. Springer, 2001.

[17] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.

[18] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.