Towards Context-aware Adjective Learning

Kadir Fırat Uyanık, Onur Yürüten, Sinan Kalkan, Erol Şahin

KOVAN Research Lab., Computer Engineering Dept., Middle East Technical University, Ankara, Turkey
{kadir, oyuruten, skalkan, erol}@ceng.metu.edu.tr
Electical and Electronics Engineering Dept., Middle East Technical University, Ankara, Turkey

***Abstract*— Context, which can be defined as the interrelated conditions in which something exists or occurs, is important for interpreting the environment, especially in determining which concept an object belongs to. In this paper, we are interested in how such context information can be used by a robot for constructing concepts that correspond to adjectives in language (i.e., adjective concepts). This is not only important for robot's "cognition" but also critical for natural human-robot communication which requires the robot to understand what an object means for us in the environment. For this reason, we are using adjective labels that eight participants assigned for objects in isolation and for objects in a context.**

In the first phase of the trainings, the robot iCub interacts with a set of objects and learns affordances by applying the behaviors in its behavior repertoire. After that, we evaluate appearance and affordance information of objects in isolation as well as in relation to other objects' attributes for adjective concepts using the adjective labels from humans. The results show that iCub can predict the adjectives of an object in a two-object environment with a confidence comparable to humans.

***Keywords*: context, affordances, adjectives, concept**

I. Introduction

Psychologists define concept as "the information associated with its referent and what the referrer knows about it" [1]. For example, the concept of an apple, which includes not only how an apple looks like but also how it tastes, how it feels etc., is more or less all the information that we know about apples.

It is becoming widely accepted that background information is very crucial for cognitive agents in forming concepts and concept acquisition processes, as e.g., identified by Yeh & Barsalou [2]:

> *"One of the most potent factors in cognition is the background situation that frames a stimulus (also called context)... When situations are incorporated into a cognitive task, processing becomes more tractable than when situations are ignored... By focusing on situations, the cognitive system simplifies many tasks. It becomes easier to recognize objects and events, to remember relevant information and skills, to understand language, to solve problems and perform reasoning, and to predict the actions of other agents."*

This background information affects the way we perceive the objects and the events in the environment. For example, an object that would normally be identified as being tall would be named short among taller objects. Called *relative attributes* in this paper, they are an integral part of the background information that may place an object into different categories, or concepts, based on what is available in the background.

The concept of affordances from J. J. Gibson [3] offers an ideal solution towards conceptualization since it naturally brings together perception, action and language. J. J. Gibson defined affordances as the action possibilities offered by objects to an agent: Firstly, he argued that organisms infer possible actions that can be applied on a certain object directly and without any mental calculation. In addition, he stated that, while organisms process such possible actions, they only take into account relevant perceptual data, which is called as perceptual economy. Finally, Gibson indicated that affordances are relative, and it is neither defined by the habitat nor by the organism alone but through their interactions with the environment.

This article studies how a robot, from its sensorimotor interactions with the environment, can conceptualize over affordances, appearance of objects and their immediate surrounding. We propose that the concepts from the appearance and the affordances of objects correspond to a subset of adjectives and these concepts are augmented by utilizing the context knowledge so that they can be more useful while interacting with humans. Experiments conducted with the human participants showed that a neighboring object as part of a spatial context has considerable influence on the value judgment for a particular object of interest.

As opposed to solely-appearance based learning methods, we propose a three-stage learning scheme which enables iCub to attribute adjectives to an object with close resemblance to what humans do.

First, iCub learns affordances of objects through interactions by applying the behaviors in its repertoire. Then, using the learned affordances and by utilizing the context dependent features -relative attributes of two objects-, iCub builds up more abstract and generic representations as adjectives for an object. These representations are more in-line with what humans ascribe to the objects in the environment.

In a previous study [4], we used the appearance and the affordances of objects for relating them to adjective concepts. In the current article, we extend this previous study by including context information as outlined above.

A. Related Work

Gibsonian affordances explains how inherent *values* and *meanings* of things in the environment can be directly perceived and how this information can be linked to the action possibilities offered to the agent by its environment. We follow the affordance formalization proposed by [5] (see [6, 7] for similar formalizations) who suggests that an affordance, a is a triple between an entity e, behavior b and an effect f, i.e.:

$$a = (e, b, f), \tag{1}$$

where f is the result of applying b on e. As an example, if a `grasp-with-right-hand` behavior is applied on a `blue-ball` leading to the `grasped` effect, the robot acquires one affordance relation like:

(`blue-ball`, `grasp-with-right-hand`, `grasped`).

After interacting with several objects of, e.g., different colors, the robot can generalize over the acquired affordance relations and realize, for example, that color of the ball is not a relevant feature for it to be `grasped`:

(`*-ball`, `grasp-with-right-hand`, `grasped`).

Moreover, after more interactions, the robot can generalize over its behaviors, realizing for example that `ball` objects can also be grasped with the other hand:

(`*-ball`, `grasp-with-*-hand`, `grasped`).

Recent studies successfully showed that affordance based robot control and learning architectures are useful in various scenarios, such as navigation [8], manipulation [9, 10, 11, 12, 13], conceptualization and language [14, 15], planning [11], imitation and emulation [6, 11, 15], tool use [16, 17, 18] and vision [15].

II. Methodology

A. Setup and Perception

We use the humanoid robot iCub [19] to demonstrate and assess the performance of the models we develop.

iCub perceives its environment through two Kinect cameras, one for perceiving the table and the other for the humans with the assistance of a motion capture system (Visualeyez II VZ4000) to detect gaze direction. In order to simplify perceptual processing, we assumed that iCub's interaction workspace is dominated (Figure 1) by an interaction table. We use PCL[20] to process raw sensory data. The table is assumed to be planar and is segmented out as background. After segmentation, the point cloud is clustered into objects and the following features extracted from the point cloud represent an object o (Eq. 1):

- *Surface features:* surface normals (azimuth and zenith angles), principal curvatures (min and max), and shape index. They are represented as a 20-bin histogram in addition to the minimum, maximum, mean, standard deviation and variance information.
- *Spatial features:* bounding box pose (x, y, z, theta), bounding box dimensions (x, y, z), and object presence.

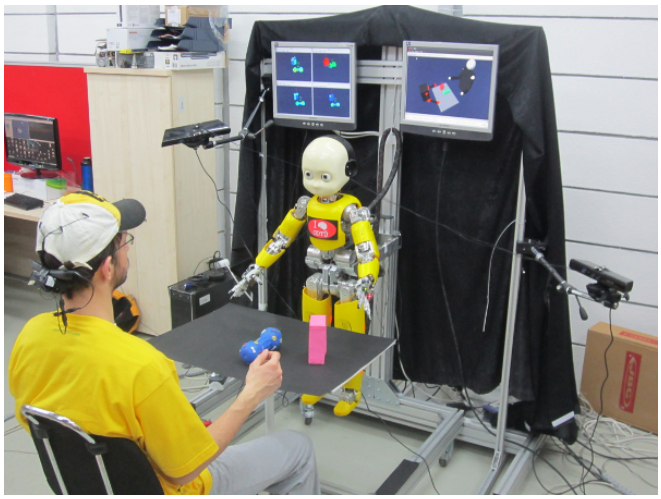
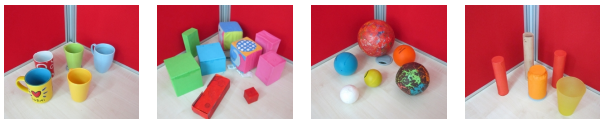


Fig. 1. Interaction environment is dominated by a table. There are at most two objects on the table. Kinect on the right of the iCub is dedicated to the table and tabletop-object processing and the kinect on the left is used to capture human body.

Fig. 3. The objects used during the context-aware adjective learning.



(a) cups (b) boxes (c) balls (d) cylinders

Fig. 4. The objects used during the affordance learning.

B. Data Collection

iCub interacted with a set of 35 objects of variable shapes and sizes to learn their affordances. Its behavior repertoire includes behaviors such as *push-left*, *push-right*, *push-forward*, *pull*, *top-grasp*, *side-grasp*, *say-pass-me*. It applied each behavior on each object to learn the affordance relations (eqn. 1), considering the effects generated on the object, such as *moved-left*, *moved-right*, *moved-forward*, *moved-backward*, *grasped*, *knocked*, *disappeared*, *no-change*[1].

iCub interacted with humans by asking them to describe an object that it points to. Participants presented with an adjective set so that they can assign adjectives that feels

[1]The *no-change* label means that the applied behavior could not generate any notable change on the object. For example, when iCub applies *say-pass-me* behavior and if there is no human around, this will not generate any change on the object.

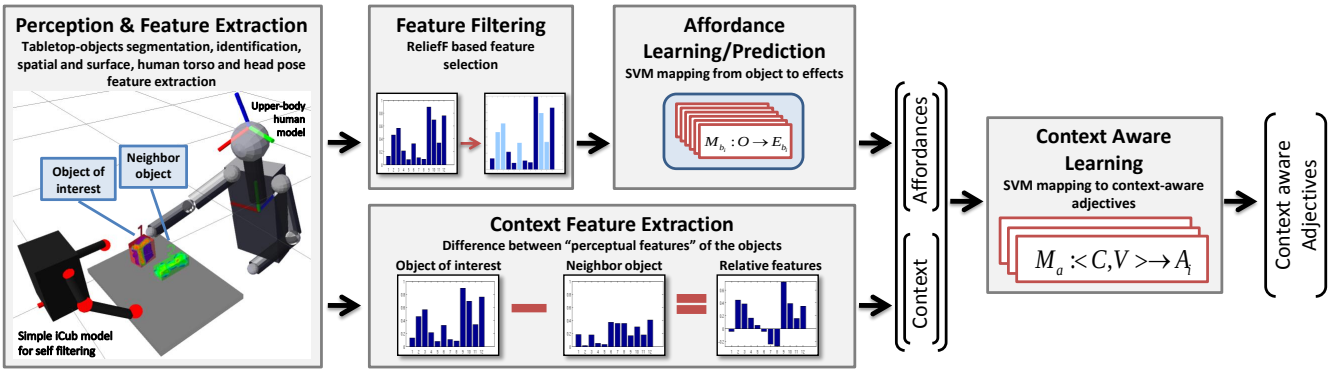


Fig. 2. Overview of the system. iCub perceives the environment and learns the affordances. It learns adjectives based on the affordance predictions and contextual features.

most intuitive for them. Participants were provided with enough time till they decide on the adjectives among three adjective-pairs, namely *tall/short*, *round/edgy*, and *thin/thick*. We collected 224 instances from 8 participants, three of them attended to the experiments by physically interacting with the robot, and five of them presented with the images obtained during these experiments via Google+ hangouts[2]. We shared documents over *Google docs* with each participant so that they weren't distracted from each others' choices.

III. Developing Concepts

A. Learning Affordances

Using the effect labels $E \in \mathcal{E}$, we train a Support Vector Machine (SVM) classifier for each behavior b_i to learn a mapping $\mathcal{M}_{b_i} : \mathcal{O} \rightarrow \mathcal{E}$ from the initial representation of the objects (i.e., $\mathcal{O}$) to the effect labels ($\mathcal{E}$). The trained SVMs can be then used to predict the effect (label) $E_{o_l}^{b_k}$ of a behavior b_k on a novel object o_l using the trained mapping $\mathcal{M}_{b_k}$. Before training SVMs, we use ReliefF feature selection algorithm [21] to filter out irrelevant features (weight < 0), and relevant features are passed to the training phase.

B. Context Aware Adjectives

We train SVMs for learning the context-aware adjectives of objects from their affordances and context related features (see Fig. 2). We use trained SVMs for affordances (i.e., $\mathcal{M}_b$ in Sect. III-A) to form a **48-dimensional** space, $\mathcal{V} = (\hat{E}_1^{b_1}, ..., \hat{E}_8^{b_1}, ..., \hat{E}_1^{b_6}, ..., \hat{E}_8^{b_6})$, where $\hat{E}_i^{b_j}$ is the confidence of behavior b_j producing effect E_i on the object o. We train an SVM for learning the mapping $\mathcal{M}_a : \mathcal{V} \rightarrow \mathcal{A}$. After learning, iCub can predict the adjective labels for an object in a novel context.

IV. Results

To compare our model, we have also trained a baseline classifier that maps the combination of pure perceptual and context information to adjective labels (as opposed to ours that map affordance and context information). Table I shows the results for prediction accuracies of our model, the baseline classifier and the confidence of participants.

TABLE I

Adjective Prediction Accuracy of Two Methods and Confidence of Human Participants

Adjectives	iCub Affordance+Context	iCub Perceptual+Context	Humans
Tall-Short	88.02%	83.85%	73%
Edgy-Round	98.43%	98.43%	93%
Thin-Thick	83.33%	72.08%	65%

In the round-edgy case, we see that the context dependence does not have any noticeable influence on the adjective decision for each methods and also for participants. This suggests that this pair of adjectives have a strict distinction for humans, and iCub were able to reach to the same confidence rate with them. In the case of short/tall adjectives, we have observed the relatively different preferences among the participants: for cases when there are two tall objects are present, some subjects labeled the taller one as "tall", and the other one as "short". Some subjects, when presented with the same case, labeled both objects as "tall". Similar cases also occurred with the thin/thick adjectives.

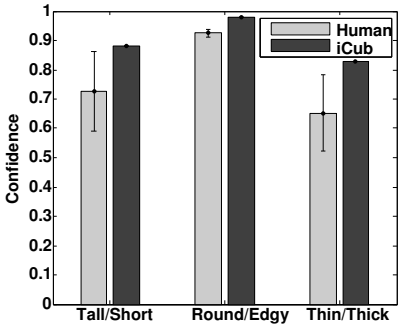


Fig. 5. Human and robot prediction accuracies are shown with the mean and variance information of the confidence of the human across the whole dataset.

As it is shown in the table II, participants were confident

[2]Google+ hangout is a unique feature of *Plus* service of Google Inc. that facilities video chatting, sharing documents and video at the same time among a group of people.

only when *round/edgy* assignment is done. For other two adjectives, *thin/thick* and *tall/short*, their confidences varied with the variance of 13% and 14% respectively. We observed 4 major effects why people tend to give different responses while deciding on their adjective categories:

A. *Effect of spatial context*

Participants reflected their need for making a comparison when they are asked to describe an object at the very first moments of the experiments.

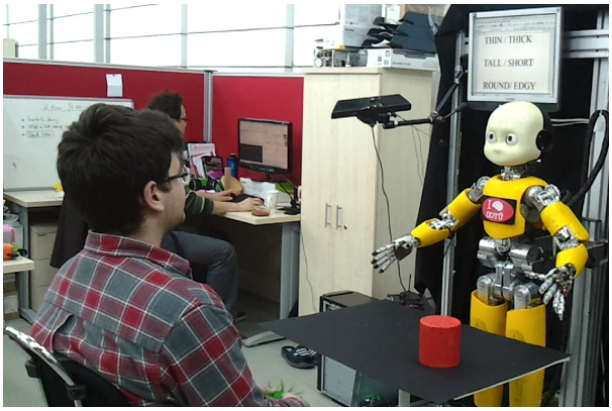
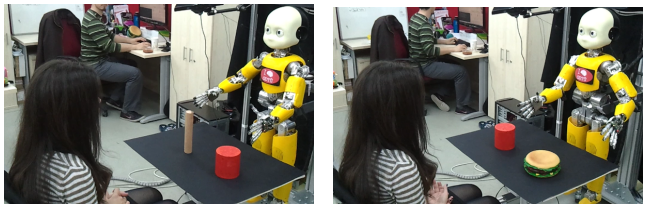


Fig. 6. Four participants asked *With respect to what?* and similar questions when they are presented with a single object at the very first moments of the experiments.

When participants are shown with the instances and asked for assigning the adjective labels of the objects, we noted 8 cases when they clearly expressed that they were trying to remember what they had said about that particular object when it had already been shown earlier but in a different context.



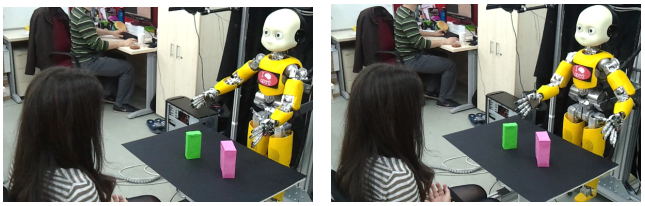
(a) Labeled as *short* (b) Labeled as *tall*

Fig. 7. Participants influenced by the neighbor object during the adjective assignments.

B. *Effect of object geometry*

Results show that we are more stable against the spatial context while attributing *round* or *edgy* adjectives to the objects. *Tall* adjective assignment, on the other hand, were influenced by the ratio between the height and the cross-section of the object. However, it was observed that this effect could be overridden by the effect of the context.

However, the effect of dimension-wise ratios can become dominant features if the relative difference between objects are not easily recognizable. And it takes longer time for participants to make decision since the details are less visible.



(a) Labeled as *thick-tall-edgy* (b) Labeled as *thin-tall-edgy*

Fig. 9. Dimension-wise ratios of the object can be influential if there is not noticeable difference between the objects and this multi-object case turns into single-object case since there is not comparable information to be used.

C. *Effect of temporal context*

When people are shown with the instances and asked for assigning the adjective labels of the objects, they tried to remember what they had said about that particular object when it was already shown earlier but in a different context.

Some of the participants asked if they are going to be presented with a taller object when they are shown a tall object for the first time. This shows that we tend to make use of temporal information while assessing the attributes of the objects in the environment.

D. *Results on Novel Instances*

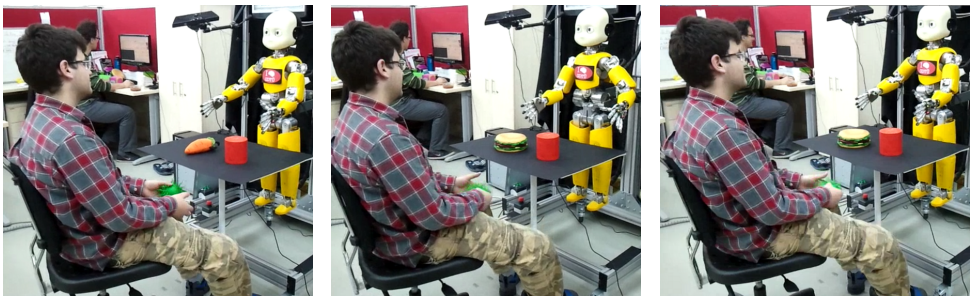
Table II shows the results on novel context instances that were not included within the training set. Among them, there was also a novel object (bulb-box) available.The results show that our model has developed an adequate representation that matches with the general preferences of the human participants. In mostly trivial cases, the model responded with high confidences (consider the context formed by bulb-box and green box). In non-trivial cases, the decision of the model coincided with the majority vote of human participants. For example, in comparison with bulb-box and the soft box, nearly half of the human participants attributed the soft box short. Our model performed likewise with the accuracy of 63 %.

V. Conclusion

In a previous study [4], we proposed linking affordances as well as appearance of objects with adjectives. In the current article, we proposed promising attempts towards including context information into learning adjectives. iCub learned the affordances of the objects and from these, it learned different types of SVM models for predicting the adjectives for the objects. Taking these learned adjective labels and how the affordance and appearance features change in the environment, iCub learned adjectives in context with a performance similar to humans.

Acknowledgments

The authors K. F. Uyanik and O. Yuruten, acknowledge the support of TUBITAK 2210 scholarship program.

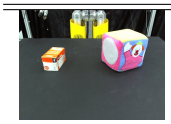
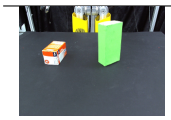
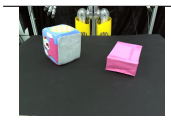
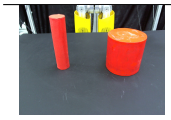


(a) Labeled as *thick-short-round* (b) Labeled as *thick-short-round* (c) Labeled as *thick-tall-round*

Fig. 8. Participants influenced by the dimension-wise ratios of an object, but this was overridden by the effect of spatial context.

TABLE II

ADJECTIVE PREDICTIONS FOR NOVEL CONTEXT INSTANCES. FOR EACH OF THE ADJECTIVES, THE FIRST PREDICTION IS FOR THE OBJECT TO THE LEFT. THE CONFIDENCE OVER THE PREDICTED LABEL DRAWS PARALLELS WITH THE DECISION OF HUMAN PARTICIPANTS.

Items	Edgy-Round	Short-Tall	Thin-Thick
	Edgy 99%, Edgy 99%	Short 80%,Short 63%	Thin 96%, Thick 90%
	Edgy 99%, Edgy 99%	Short 63%,Tall 65%	Thin 82%, Thick 58%
	Edgy 99%, Edgy 85%	Tall 95%,Short 95%	Thick 53%, Thin 65%
	Round 97%, Round 95%	Tall 89%,Short 93%	Thin 92%, Thick 89%

REFERENCES

[1] A.M. Borghi. Object concepts and embodiment: Why sensorimotor and cognitive processes cannot be separated. *La nuova critica*, 49(50):90–107, 2007.

[2] W. Yeh and L.W. Barsalou. The situated nature of concepts. *The American journal of psychology*, pages 349–384, 2006.

[3] J.J. Gibson. *The ecological approach to visual perception*. Lawrence Erlbaum, 1986.

[4] Yuruten O., Uyanik K. F., Caliskan Y., Bozcuoglu A.K., Sahin E., and Kalkan S. Learning adjectives and nouns from affordances on the icub humanoid robot. In *Simulation of Adaptive Behavior*. Submitted, 2012.

[5] E. Şahin, M. Çakmak, M.R. Doğar, E. Uğur, and G. Üçoluk. To afford or not to afford: A new formalization of affordances toward affordance-based robot control. *Adaptive Behavior*, 15(4):447–472, 2007.

[6] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor. Learning object affordances: From sensory–motor coordination to imitation. *Robotics, IEEE Tran. on*, 24(1):15–26, 2008.

[7] D. Kraft, N. Pugeault, E. Baseski, M. Popovic, D. Kragic, S. Kalkan, F. Wörgötter, and N. Krüger. Birth of the object: Detection of objectness and extraction of object shape through object action complexes. *International Journal of Humanoid Robotics*, 5(2):247–265, 2008.

[8] E. Ugur and E. Şahin. Traversability: A case study for learning and perceiving affordances in robots. *Adaptive Behavior*, 18(3-4):258–284, 2010.

[9] P. Fitzpatrick, G. Metta, L. Natale, S. Rao, and G. Sandini. Learning about objects through action - initial steps towards artificial cognition. In *IEEE ICRA*, 2003.

[10] R. Detry, D. Kraft, A.G. Buch, N. Kruger, and J. Piater. Refining grasp affordance models by experience. In *IEEE ICRA*, pages 2287 –2293, 2010.

[11] E. Ugur, E. Oztop, and E. Şahin. Goal emulation and planning in perceptual space using learned affordances. *Robotics and Autonomous Systems*, 59(7-8), 2011.

[12] E. Ugur, E. Şahin, and E. Oztop. Affordance learning from range data for multi-step planning. *Int. Conf. on Epigenetic Robotics*, 2009.

[13] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor. A computational model of object affordances. In *Advances In Cognitive Systems*. IET, 2009.

[14] I. Atil, N. Dag, Sinan Kalkan, and Erol Şahin. Affordances and emergence of concepts. In *Epigenetic Robotics*, 2010.

[15] N. Dag, I. Atil, S. Kalkan, and E. Sahin. Learning affordances for categorizing objects and their properties. In *IEEE ICPR*. IEEE, 2010.

[16] J. Sinapov and A. Stoytchev. Learning and generalization of behavior-grounded tool affordances. In *Development and Learning, 2007. ICDL 2007. IEEE 6th International Conference on*, pages 19 –24, july 2007.

[17] J. Sinapov and A. Stoytchev. Detecting the functional similarities between tools using a hierarchical representation of outcomes. In *Development and Learning, 2008. ICDL 2008. 7th IEEE International Conference on*, pages 91 –96, aug. 2008.

[18] A. Stoytchev. Learning the affordances of tools using a behavior-grounded approach. *Towards Affordance-Based Robot Control*, pages 140–158, 2008.

[19] G. Metta, G. Sandini, D. Vernon, L. Natale, and F. Nori. The icub humanoid robot: an open platform for research in embodied cognition.

In *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, pages 50–56. ACM, 2008.

[20] R.B. Rusu and S. Cousins. 3d is here: Point cloud library (pcl). In *IEEE ICRA*, pages 1–4. IEEE, 2011.

[21] K. Kira and L. A. Rendell. A practical approach to feature selection. In *Proc. 9th Int. Workshop on Machine Learning*, pages 249–256, 1992.