Proceedings ACS (2019) 37-54

Categorization in Real-World Tasks

Alisa Volkert ALISA.VOLKERT@UNI-TUEBINGEN.DE
Jona Schröder JONA.SCHROEDER@STUDENT.UNI-TUEBINGEN.DE
Department of Computer Science, Universität Tübingen, Tübingen, Germany

Alexandra Kirsch AK@ALEXKIRSCH.DE
Independent Scientist, Stuttgart, Germany

Abstract

We examine the role of category representations for decision-making in real-life tasks. To this end, we empirically examine how people categorize kitchen objects and make use of categories when storing objects in a kitchen. We then compare two computational models and their ability to represent the participants' mental models. We discuss the advantages and disadvantages of the models and point the way to further research.

1. Introduction

Mutual understanding between people and cognitive systems is needed, for example in language and for interaction. Typically, categories are modeled as sets of individual objects. This representation is practical for automatic reasoning, but it ignores many details of human categorization (Lakoff & Johnson, 1980). The affiliation of an object to a category can depend on the situational context. For example, a drinking glass can serve as a vase if no better container for flowers is available. There is also the phenomenon of typical vs. untypical representatives, for example a sparrow is a typical bird, while a penguin is a bird, but not a typical one (Rosch, 1975).

In this paper we show empirically that categorization can be a major piece of knowledge in decision-making tasks. In the context of a real kitchen situation we compare the results of a previous experiment, where participants stored objects in a real kitchen, with a new experiment where participants categorized objects.

We describe two feature-based models for representing categories in a more human-like way, both based on models from psychology. We explore how well they capture the mental models of our participants and discuss their potential applicability for decision-making.

2. Experiments

The following experiments were designed to gather data about categorization of typical kitchen objects. In the first experiment we asked participants to group objects according to any scheme they find appropriate, in the other one participants were asked to distribute the same objects into an empty kitchen in the way they would furnish their kitchen at home.

2.1 Experiment 1: Categorization

The categorization task was tested in a remote experiment with the software CatScan that was developed and used by the group of Alexander Klippel (Klippel et al., 2008, 2011, 2012, 2013; Mast et al., 2014) to examine categorization in spatial cognition. They gave us their source code and the permission to change it for our experiment. We made some adjustments for German language instructions and answers, adapted some parameters to fit our needs, and made some usability adjustments. We adjusted the original pre-trial that asks participants to categorize figures of cats, dogs and camels to make sure they performed the task seriously in the remote setup.

Materials We had 225 images of objects, each in the size of 150×150 pixels. The set contained images of identical objects, for example the five identical plates of a dinnerware set. Though, the set contained 157 different objects.

Participants 46 German-speaking participants were recruited from the University of Tübingen. Participants who passed the trial could win one of ten 20 € retail vouchers.

Procedure The participants downloaded our prepared CatScan software as a jar file, which they executed on their own Java Runtime Environment (Version 1.6d or higher) on a screen resolution of at least 800×600 pixels. After the experiment, CatScan generates a zip file containing the results. The participants uploaded this file with a unique identifier that they had received with the download.

After starting the CatScan software, the participants first answered some demographic questions. Then they received instructions to categorize given objects. Next they performed a trial task with images of cats, dogs and camels. This task served on the one hand to familiarize the participants with the procedure and the software, and on the other hand to ensure that they understood the task and performed the experiment seriously (which all of our participants did). The interface of CatScan for the categorization task is shown in Figure 1.

The experimental setup consisted of categorizing all 225 images into at most 224 groups (so there had to be at least one group with more than one object). In the last step, the participants were presented with each of their self-formed groups and were asked to 1) provide a label for the whole group, 2) identify one prototypical item in the group and 3) (optionally) describe their rationale of putting these objects into one group.

The participants had no time restrictions for any of the tasks.

2.2 Experiment 2: Decision-Making

In previous work (Schröder et al., 2019a,b) we asked 20 participants to sort the same 225 kitchen items as we used in Experiment 1 into a kitchen as one would do at home when organizing a new kitchen. On the whole there were 28 possible places (cupboards, drawers, surfaces) in the kitchen to store objects.

We then evaluated which objects were placed together, i.e. in the same group. We treated these groups in the same way as the categories of the previous experiment in the following evaluation.

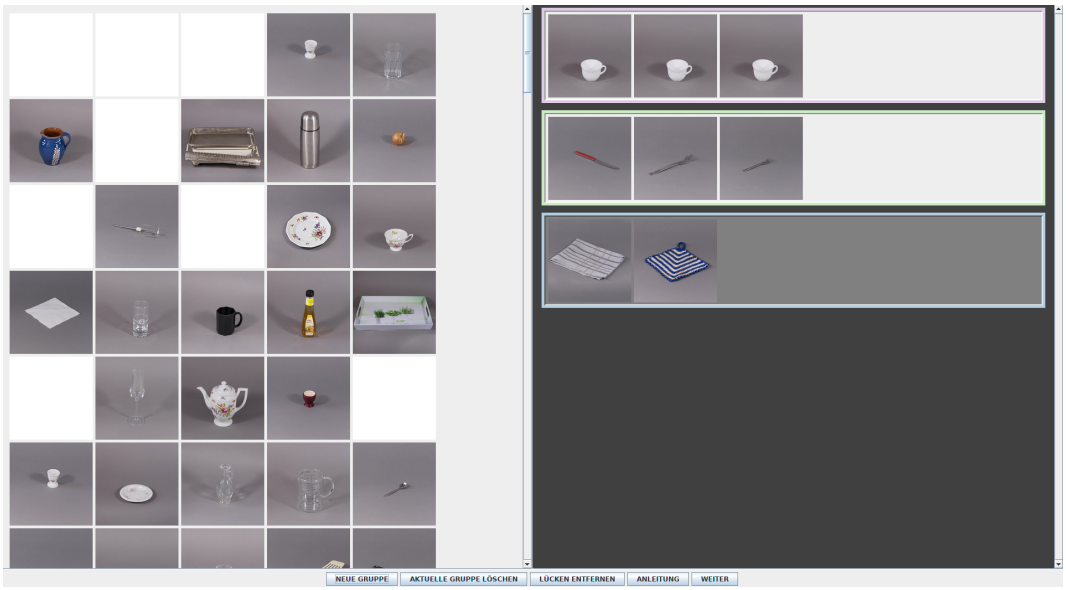


Figure 1: The CatScan user interface: The objects were given on the left, the categories are on the right. The buttons on the bottom allow participants to add a category, delete a category, remove empty spaces in the object images, re-read the instructions, and move to the next step in the experiment (only possible when all objects were assigned to categories).

3. Representation and Models

We assume that individual objects are categorized by perceptible features. To mirror the human trait of categorizing objects in the context of a situation, we use models of categorization that are based on similarities of features. In the following we use these models to assign one category to each object, but it is important to note that at least the prototype model could potentially also provide a "second best" category or check the acceptability of the association of an object to a class.

3.1 Objects

3.1.1 Representation

We represented an object as a vector with the following features

- on the ratio scale: height, width, depth, concavity, hardness, transparency, weight, capacity, handles,
- on the nominal scale: color, material, function.

For the nominal features color, material and function could receive several values, e.g. a metallic pot with a wooden handle.

The function feature was a label like "eat", "clean", etc. The assignment was performed by a consensus of our team. We added this feature because we had experienced that people often use

the function information. Our representation with labels is certainly too simplistic to represent the human association between an object and its possible functions, but it allows us to integrate this important aspect.

3.1.2 Similarity

To calculate the similarity, or conversely the distance, of two objects, we applied a distance measure to each feature. We defined distance functions for a feature of object 1 f_1 and the same feature of object 2 f_2 (Volkert & Kirsch, 2015; Schröder et al., 2019a):

- for the ratio scale, where the features are numbers: $d_{\text{ratio}}(f_1, f_2) = |f_1 - f_2|$
- for the nominal scale, where the features are sets of labels: $d_{\text{nominal}}(f_1, f_1) = 1 - \frac{|f_1 \cap f_2|}{|f_1 \cup f_2|}$

3.2 Exemplar Model

The exemplar model (Medin & Schaffer, 1978; Nosofsky, 1987; Jäkel, 2007; Nosofsky, 2011) assumes that categories emerge from the similarity to known object. So when I know one object that someone has called "plate" and I find another object that strongly resembles the first, I may call the second also "plate". The psychology literature proposes several specific models for the exemplar approach. But the basic idea is captured in the k-Nearest Neighbour classifier (Fix & Hodges Jr, 1952).

In this classification method, the learning consists only of specifying weights for a distance function over all features. In the classification step, the distance to all known objects is calculated and the majority class of the k closest objects is returned as the class.

We used a Nearest Neighbour classifier with $k = 1$ and the distance function

$$d_{\text{exem}}(o_1, o_2) = \sum_{f \in F} w_f \cdot d_f(o_1^f, o_2^f)$$

according to Minda & Smith (2011) where o_1, o_2 are feature vectors representing two objects, F is the set of features, w_f is the weight of feature f, o_1^f, o_2^f the values of feature f for each object, and d_f is the distance function d_{ratio} or d_{nominal} depending on the scaling level of the feature.

In the learning phase the weights w_f are optimized to best represent the training data set.

3.3 Prototype Model

The prototype model (Posner & Keele, 1968; Rosch, 1975; Volkert & Kirsch, 2015; Volkert et al., 2018; Schröder et al., 2019a) assumes that instead of memorizing every object one has ever seen, only a prototype is stored for each category and category membership is determined by the maximum similarity of an object to a prototype.

Prototype Formation A prototype of objects in category c is a vector with the same features as the objects. The feature values of a prototype p are determined for features

- on the ratio scale: $f_p = \frac{\sum_{o_i \in c} o_i^f}{|c|}$

- on the nominal scale: $f_p = \text{mode}(o_i^f), o_i \in c$

where f_p is the feature value of the prototype, o_i^f the values of feature f for object o_i (Hampton, 1993; Minda & Smith, 2011; Volkert & Kirsch, 2015; Schröder et al., 2019a). The mode function returns the most frequent value occurring in the input.

Categorization Analogously to the exemplar model, the category of an object is determined by calculating the distance to all prototypes and returning the category of the closest prototype. But the reduction of a group of objects to a single point, i.e. the prototype, in the vector space loses the information of how much variation inside a category is tolerable. Therefore we normalize the distance function by the standard deviation s_c^f of feature f over the instances that formed the prototype p_c (Volkert & Kirsch, 2015):

$$d_{\text{proto}}(o, p_c) = \sum_{f \in F} \frac{w_f \cdot d_f(o^f, p_c^f)}{s_c^f}$$

Instantiation To determine weights for the features in the prototype model, we identified the classificatory significance of each feature by measuring the correct classification of an object to its own group, based on one specific feature. The better a feature could predict the corresponding group, i.e. the closer the object's feature value is to its prototype feature value, the stronger its classificatory significance. Based on that significance we calculated weights for each feature.

4. Evaluation

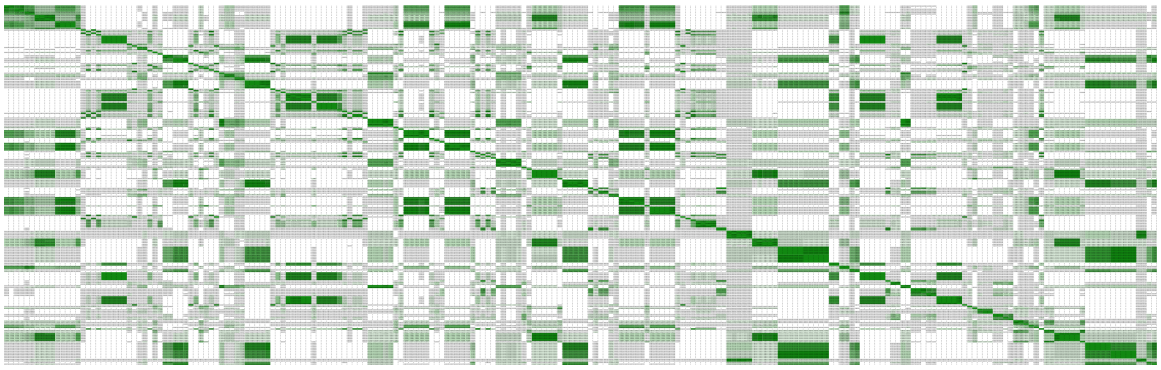
In the following we first compare the classification results of the two experiments to establish the connection between categorization and decision-making. Then we present classification results for the exemplar and prototype model when instantiated with data from our experiments.

4.1 Categorization and Decision-Making

For both experiments we calculated for each participant a similarity matrix, where the rows and columns represent one of the 225 objects each. The value in each field is one if the objects were grouped together (either in a self-defined group in Experiment 1 or in the same place in Experiment 2) (Tang & Heymann, 2002; Klippel & Montello, 2007; Klippel et al., 2013). We then added the similarity matrices over all participants into an overall similarity matrix per experiment (Wallgrün et al., 2002; Klippel et al., 2013). The result is shown in Figure 2.

Remember that the tasks were described differently: categorizing objects vs. storing them in a real kitchen. The physical environment in the kitchen had constraints such as overall space at one place or the size of single objects that were absent in the categorization task. In the decision-making experiment in the kitchen, however, participants never used all possible shelves. In addition, some participants in this experiment reported that they had grouped objects according to their aesthetic appeal (e.g. like putting the ugly plates apart from the nice-looking ones).

Despite these and other differences, the overall similarity matrices look remarkably similar. Certain groups of objects such as cutlery were always put into the same group. In order to check for



(a) Experiment 1: Categorization

(b) Experiment 2: Decision-Making

Figure 2: Overall similarity matrices (OSM) from the two experiments. Those are obtained by summing up the similarity matrices of all participants for each experiment.

the similarity between those two matrices, we calculated the difference between them. Since both matrices contain percentages of participants grouping two objects together, the resulting difference matrix returns percentages as well. Having a look on the absolute values of the difference matrix, revealed that only few spots show a larger difference than 60 %. In order to evaluate the difference between the two experiments, we calculated the mean of differences for each object. Here, we want to discuss those objects having a mean difference larger than 13 %.

The huge soup plate No. 110 got the biggest mean difference (14 %) when calculating the differences between the two matrices (s. Tab. 1). When the threshold in the decision-making experiment was bigger than 65 % it became a standalone. Having a closer look on the overall similarity matrix (OSM) of the decision-making experiment revealed that 65 % of the participants each grouped it together with another soup plate (No. 114), 60 % grouped it together with soup plate No. 113 (s. Fig. 3) and 50 % grouped it together with the large plates (No. 34–36, s. Fig. 4), the soup plates (No. 111 and 112) and a deep cake plate (No. 225)(s. Fig. 5). These subsets of our participants do not have to be the same individuals though. It is conspicuous, that plate No. 114 (Fig. 3) rather looks like a pasta plate as plate No. 110 (Tab. 1) does. The grouping of these two plates in the decision-making experiment has as a consequence that soup plate No. 110 falls into the same category as the other soup and large plates when the threshold is lower than 66 %.

Having a closer look on the differences between the two experiments concerning plate No. 110 revealed that anybody in the decision-making experiment has neither grouped the plates (No. 212–216, Fig. 6a) nor the saucers (No. 217–221, Fig. 6b) of the dinnerware set together with plate No. 110. When decreasing the threshold in the decision-making experiment, plates of the dinnerware set are grouped together with the other small plates (No. 48–52, Fig. 6). This goes in line with the finding that only 10 % of the participants in the decision-making experiment have grouped plate No. 110 together with the small plates, whereas at least 89 % of the participants in the categorization experiment grouped them together. There was also a huge difference concerning the other saucers. More than 70 % difference in the grouping behavior is a consequence of the different grouping behavior in the two experiments. While in the decision-making experiment only 5 % of the participants grouped plate No. 110 together with the saucers, at least 76 % grouped them together in the categorization experiment.

Another object type with a relatively high mean difference (13 %) was the espresso cup saucer (No. 152–156, Tab. 1). There were especially not grouped together with the large plates (No. 34–36, Fig. 4): nobody in the decision-making experiment grouped these two types of items together. Another large difference shows up considering the grouping with—or rather—without the plate No. 110 discussed above. Only 5 % of the participants in the decision-making experiment grouped the espresso saucers together with plate No. 110. In the categorization experiment more than 76 % of the participants grouped them together. Same holds for the espresso saucers and the pasta plate No. 114 (Fig. 3). Here, only 10 % of the participants grouped these two types of items together. More than 76 % of the participants in the categorization experiment grouped the espresso saucers together with the small plates (No. 48–52, Fig. 7), whereas in the decision-making experiment only 10 % grouped these two types of objects together. Another 65 % difference shows up when looking on the grouping with the plates of the dinnerware set. While more than 80 % in the categorization experiment grouped them together with the espresso saucers, only 15 % in the decision-making

(a) Item No. 114

(b) Item No. 113

Figure 3: Item No. 114 that was grouped together with the large soup plate (s. Tab. 1) by 65 % of the participants and plate No. 113 that was grouped with it by 60 % of the participants in the decision-making experiment. Pictures: Vanessa Bernath

(a) Item No. 34

(b) Item No. 35

(c) Item No. 36

Figure 4: Three large plates that were grouped together with the large soup plate (s. Tab. 1) by 50 % of the participants each in the decision-making experiment. Pictures: Vanessa Bernath

experiment grouped these two types of objects together.
When decreasing the threshold in the decision-making experiment down to 65 %, the espresso cup saucers were grouped together with the espresso cups.
Also one vase (No. 202, Tab. 1) was grouped differently depending on the experiment (13 %). While in the categorization experiment it has been grouped together with all other glasses, in the decision-making experiment it was a standalone. We had to decrease the threshold down to 45 % of the participants in order to identify a category, where this vase might belong to. In this case it was grouped together with the other vases and jugs. This might be because of slightly different sensory input participants got when doing the categorization experiment. Here, they only could see the pictures of the objects, whereas in the decision-making experiment they even could and had to touch the objects. So in the categorization experiment the size of the objects might have been a bit unclear to them.

On the whole one can say, that participants differentiated much more in the decision-making experiment than in the categorization experiment. This makes sense, since putting saucers to the large plates is something people would rather not do in their kitchen.

(a) Item No. 111 *(b)* Item No. 112 *(c)* Item No. 225

Figure 5: Three other objects that were grouped together with the large soup plate (s. Tab. 1) by 50 % of the participants each in the decision-making experiment. Pictures: Vanessa Bernath

(a) Item No. 212–216 *(b)* Item No. 217–221

Figure 6: Anybody in the decision-making experiment has neither grouped the plates (No. 212–216) nor the saucers (No. 217–221) of the dinnerware set together with the large soup plate No. 110 (s. Tab. 1) in the decision-making experiment. Pictures: Vanessa Bernath

(a) Item No. 48–49 *(b)* Item No. 50 *(c)* Item No. 51 *(d)* Item No. 52

Figure 7: When decreasing the threshold in the decision-making experiment, plates of the dinnerware set (Fig. 6) were grouped together with the other small plates in the decision-making experiment. Pictures: Vanessa Bernath

Table 1: Comparison of the categorization and the decision-making experiment. Categorization of some objects differed more than that of other objects in the two experiments. Here we show three objects where categorization differed the most. For each object we show the category it was in when we set the threshold to 75 % of participants in each experiment. Some objects do not belong to any category using this threshold.

Ambivalent Object	Category containing this object in the categorization experiment	Category containing this object in the decision-making experiment
		No single category

Table 1: Comparison of the categorization and the decision-making experiment. Categorization of some objects differed more than that of other objects in the two experiments. Here we show three objects where categorization differed the most. For each object we show the category it was in when we set the threshold to 75 % of participants in each experiment. Some objects do not belong to any category using this threshold.

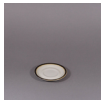
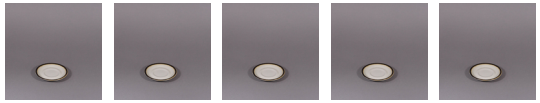
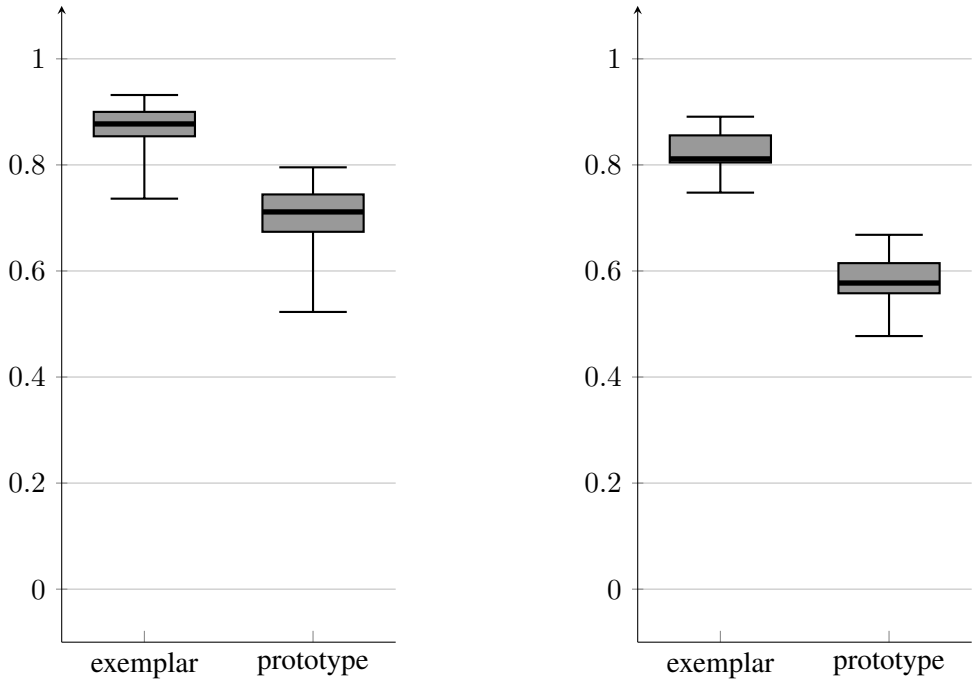
Ambivalent Object	Category containing this object in the categorization experiment	Category containing this object in the decision-making experiment
		

Table 1: Comparison of the categorization and the decision-making experiment. Categorization of some objects differed more than that of other objects in the two experiments. Here we show three objects where categorization differed the most. For each object we show the category it was in when we set the threshold to 75 % of participants in each experiment. Some objects do not belong to any category using this threshold.

Ambivalent Object	Category containing this object in the categorization experiment	Category containing this object in the decision-making experiment
		No single category



(a) Categorization *(b)* Decision-Making

Figure 8: Prediction scores from the 10-fold cross validation for each experiment.

4.2 Model Comparison

Not surprisingly, in both experiments we see a large overlap between the grouping of the participants, but still there was individual divergence. This observation confirms the findings of Lakoff (1987) that categorization is individual, but not arbitrary. We all have our own mental model of the world, but we share large parts by culture. To test whether the prototype and exemplar models can represent individual mental models, we instantiated each model for each participant. In a 10-fold cross validation, we used 90% of the objects a participant had classified to instantiate the model, and 10% as a test set. Each model was evaluated by a score that counted the number of test items where the closest prototype or the best category of the exemplar model was the same group that the participant had grouped the item into.

Figure 8 shows that with the data of both experiments the exemplar model has a higher score, indicating a better representation of the person's mental model. The advantage of the exemplar model is more pronounced in the decision-making task. This indicates that the prototype model is stronger when the categories are "clean", i.e. the categorization had no physical or aesthetic constraints. The exemplar model seems to be more robust if information is used for the grouping that is not directly related to the object properties.

One drawback of the exemplar model is the runtime for classification. The exemplar model needs about 8–9 times longer to determine a category.

Another argument for the prototype model could be a wider applicability to different tasks. As the figure shows, the prototype model can reflect the participants' mental model to an acceptable level, while the exemplar model seems to be better adapted to the specific task. In the context of a full decision procedure (like the one by Kirsch (2019)), other aspects such as available space could be added by other knowledge modules.

In all, both models have their benefits and drawbacks. It remains to be explored how they behave in the context of a decision-making model. The results also confirm that categorization is extremely related to the real-world task.

5. Related Work

Prototype theory is one of the most important categorization theories in psychology (Posner & Keele, 1968; Rosch, 1973; Minda & Smith, 2011). The prototype view assumes that a category of things in the world (animals, objects, shapes, etc.) is represented mentally by a prototype, which captures the common features of the members of a category. New stimuli are classified by comparing them with the prototypes, regardless of whether that prototype actually exists in the real world or is simply an abstract idea of a non-existing average stored in our mind. Posner and Keele (1968) discovered significantly faster categorization response for stimuli near an average that had never been seen before and a very fast response for the average stimulus itself, thus providing evidence, that humans use prototypes to classify entities. While well fitting stimuli in regard to the prototype facilitate responses, poor stimuli hinder responses (Rosch, 1975).

Another psychological model is the exemplar theory, which is often contrasted with the prototype theory. It denies that there is an abstract summary of all entities belonging to a category. Instead, it is assumes that a new stimulus is compared with all exemplars already stored in mind (Medin & Schaffer, 1978; Nosofsky, 2011).

A computational framework for concept representation in cognitive systems and architectures is DUAL PECCS (Prototypes and Exemplars-based Conceptual Categorization System). It is a cognitive system for conceptual representation and categorization and relies on a combination of the prototype and the exemplar theory (S1) (Lieto, 2014; Lieto et al., 2017). It uses a hybrid representation of concepts called heterogeneous proxytypes. A *proxytype* is an element of a representation network in long-term memory corresponding to a specific category that can be activated in working memory (Prinz, 2002). *Heterogeneous* proxytypes can be prototypes, exemplars or classical representations. A concept is represented by all of these. DUAL PECCS also uses an ontology (S2) to categorize input sentences with both systems: S1 and S2. If the two systems do not categorize equally, both proposed categories are provided (Lieto et al., 2017).

Rouder & Ratcliff (2006) have proposed a rule-based categorization, which might be used when a new stimulus may be confused with stored exemplars. For example a platypus may not be easy to classify as a mammal or a bird relying on the exemplars stored in mind since it is a mammal laying eggs. Using a rule—or definition—that all animals nursing their young with milk are mammals, the new stimulus can be classified correctly.

From an engineering perspective, several approaches for categorization have been constructed using the psychological theories just mentioned. Hepner et al. (1990) proposed a connectionist

method using artificial neural networks, while Madsen & Thomson (2009) used a symbolic approach with ontologies. Gupta et al. (2011) also used a symbolic representation to choose object places where people will expect to find the objects. The robot *Dora* (Hanheide et al., 2011) had to perform the inverse task: retrieving objects in an apartment that were placed there by people. Their probabilistic representation covers the possibility of situation-dependent classification to some extent, however without representing reasons. Jacobsson et al. (2008) proposed a model of shared understanding that integrates different symbolic and subsymbolic representations to represent situation-specific knowledge for a robot.

6. Conclusion

Our experiments show that categorization can be tightly coupled to the decisions people make in real-world tasks. Therefore, the representation of categories beyond sets needs more research. The exemplar and prototype models are both promising starting points, but they leave potential for improvement. Both aggregate the feature differences by a weighted sum. In decision-making it is well-known that people hardly ever use weighted sums to integrate different pieces of information and non-compensatory combination methods may also be a good option or categorization (Kirsch, 2019).

To really appreciate the value of a categorization method, it will have to be put into the context of decision-making. In our example, the next step is to reproduce human behavior when putting objects into a kitchen and check for the acceptability of determined places. Finding intuitive and acceptable solutions is a necessary skill for household robots.

Acknowledgements

We thank Alexander Klippel and his team for providing their CatScan software for our experiment.

References

Fix, E., & Hodges Jr, J. L. (1952). *Discriminatory analysis-nonparametric discrimination: Small sample performance*. Technical report, California University Berkeley.

Gupta, K., Schneider, A., Klenk, M., Gillespie, K., & Karneeb, J. (2011). Representing and reasoning with functional knowledge for spatial language understanding. *Proceedings of the Workshop on Computational Models of Spatial Language Interpretation-2, CogSci-2011*.

Hampton, J. (1993). Prototype models of concept representation. In I. Van Mechelen, J. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts - theoretical views and inductive data analysis*, 67 – 95. London: Academic Press, 1st edition.

Hanheide, M., Gretton, C., Dearden, R., Hawes, N., Wyatt, J. L., Pronobis, A., Aydemir, A., Göbelbecker, M., & Zender, H. (2011). Exploiting probabilistic knowledge under uncertain sensing for efficient robot behaviour. *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI'11)*.

Hepner, G. F., Logan, T., Ritter, N., & Bryant, N. (1990). Artificial Neural Network Classification Using a Minimal Training Set - Comparison to Conventional Supervised Classification. *Photogramm Eng Remote Sensing*, *56*, 469–473.

Jacobsson, H., Hawes, N., Kruijff, G.-J., & Wyatt, J. (2008). Crossmodal content binding in information-processing architectures. *Proceedings of the 3rd international conference on Human robot interaction - HRI '08* (pp. 81–88). New York, New York, USA: ACM Press. From `http://portal.acm.org/citation.cfm?doid=1349822.1349834`.

Jäkel, F. (2007). *Some theoretical aspects of human categorization behavior: similarity and generalization*. Doctoral dissertation, Eberhard-Karls University Tübingen. From `http://tobias-lib.ub.uni-tuebingen.de/volltexte/2008/3196/`.

Kirsch, A. (2019). A unifying computational model of decision making. *Cognitive Processing*, *20*, 243–259. From `https://link.springer.com/article/10.1007/s10339-019-00904-3`.

Klippel, A., & Montello, D. R. (2007). Linguistic and Nonlinguistic Turn Direction Concepts. *9th International Conference, COSIT 2007, Melbourne, Australia* (pp. 354–372).

Klippel, A., Wallgrün, J. O., Yang, J., Li, R., & Dylla, F. (2012). Formally grounding spatio-temporal thinking. *Cognitive processing*, *13 Suppl 1*, S209–S214. From `http://www.ncbi.nlm.nih.gov/pubmed/22806649`.

Klippel, A., Wallgrün, J. O., Yang, J., Mason, J. S., Kim, E.-K., & Mark, D. M. (2013). Fundamental Cognitive Concepts of Space (and Time): Using Cross-Linguistic, Crowdsourced Data to Cognitively Calibrate Modes of Overlap. *COSIT* (pp. 377–396). Cham: Springer International Publishing. From `http://link.springer.com/10.1007/978-3-319-01790-7`.

Klippel, A., Weaver, C., & Robinson, A. C. (2011). Analyzing Cognitive Conceptualizations Using Interactive Visual Environments. *Cartography and Geographic Information Science*, *38*, 52–68.

Klippel, A., Worboys, M., & Duckham, M. (2008). Identifying factors of geographic event conceptualisation. *International Journal of Geographical Information Science*, *22*, 183–204. From `http://www.tandfonline.com/doi/abs/10.1080/13658810701405607`.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. The University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lieto, A. (2014). A computational framework for concept representation in cognitive systems and architectures: Concepts as heterogeneous proxytypes. *Procedia Computer Science*, *41*, 6–14. From `http://dx.doi.org/10.1016/j.procs.2014.11.078`.

Lieto, A., Radicioni, D. P., & Rho, V. (2017). Dual PECCS: A Cognitive System for Conceptual Representation and Categorization. *Journal of Experimental & Theoretical Artificial Intelligence*, *29*, 433–452. From `https://www.tandfonline.com/doi/full/10.1080/0952813X.2016.1198934`.

Madsen, B. N., & Thomson, H. E. (2009). Ontologies vs. Classification Systems. *NEALT Proceedings Series* (pp. 27–32). Northern European Association for Language Technology (NEALT).

Mast, V., Wolter, D., Klippel, A., Wallgrün, J., & Tenbrink, T. (2014). Boundaries and Prototypes in Categorizing Direction. *Spatial Cognition IX*. From http://link.springer.com/chapter/10.1007/978-3-319-11215-2{_}7.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.

Minda, J. P., & Smith, J. D. (2011). Prototype Models of Categorization: Basic Formulation, Predictions, and Limitations. In E. M. Pothos (Ed.), *Formal approaches in categorization*, 40–64. Cambridge: Cambridge Univ. Press, 1st edition.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Nosofsky, R. M. (2011). The Generalized Context Model: an Exemplar Model of Classification. In E. M. Pothos (Ed.), *Formal approaches in categorization*, 18–39. Cambridge, MA, US: Cambridge Univ. Press, 1 edition.

Posner, M. I., & Keele, S. W. (1968). On the Genesis of Abstract Ideas. *Journal of Experimental Psychology*, *77*, 353–363.

Prinz, J. J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, Massachusetts: MIT Press, 1st edition.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*, 192–233.

Rosch, E. H. (1973). On the Internal Structure of Perceptual and Semantic Categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*, 308pp. Oxford, England: Academic Press.

Rouder, J. N., & Ratcliff, R. (2006). Comparing Exemplar- and Rule-Based Theories of Categorization. *Current Directions in Psychological Science*, *15*, 9–13.

Schröder, J., Volkert, A., Hornuff, S., & Kirsch, A. (2019a). Human-Like Prototypes Representing Categories of a Real-World Setup. *Kognitive Systeme* (pp. 20–21). Duisburg: DuEPublico. From https://duepublico.uni-duisburg-essen.de/servlets/DocumentServlet?id=48470.

Schröder, J., Volkert, A., Hornuff, S., & Kirsch, A. (2019b). Human-Like Prototypes Representing Categories of a Real-World Setup. *Procedia Computer Science*, *00*, 1–11. Under review.

Tang, C., & Heymann, H. (2002). Multidimensional Sorting, Similarity Scaling and Free-Choice Profiling of Grape Jellies. *Journal of Sensory Studies*, *17*, 493–509. From http://doi.wiley.com/10.1111/j.1745-459X.2002.tb00361.x.

Volkert, A., & Kirsch, A. (2015). Prototype-based Knowledge Representation for an Improved Human-robot Interaction. *Kognitive Systeme*, *3*, 9. From http://duepublico.uni-duisburg-essen.de/servlets/DocumentServlet?id=40717.

Volkert, A., Müller, S., & Kirsch, A. (2018). Human-like Prototypes for Psychologically Inspired Knowledge Representation. *Procedia Computer Science*, *123*, 501–506. From http://linkinghub.elsevier.com/retrieve/pii/S1877050918300772.

Wallgrün, J. O., Klippel, A., & Mark, D. (2002). A New Approach To Cluster Validation in Experimental Investigations of (Geo) Spatial Concepts.