Impact of Features in Hierarchical Representation

Nazar Flome, Shuhang Zhou, Sunhee Jin, Amberly Tam

Professor Lu

Psych 186A

June 8, 2018

**Introduction**

Since the dawn of the cognitive revolution, great efforts have been made to understand how knowledge is represented within the mind. A sizable amount of research into the topic took a similarity approach to representation, arguing that similarity serves as an organizing principle for mental structures (Shepard, 1962). Shepard took a mental distance (spacing) approach to similarity, claiming that conceptualizations closer to each other in a mental space are more similar to each other than those that are more distant. This spacing approach allowed for a quantitatively convenient way to visualize representation data through techniques such as MDS.

Shepard’s spacing approach was not without limitations, however, as other researchers argued that similarity in cognition does not abide by the principles of symmetry and directionality in such a way as a mental distance model would (Tversky, 1977). Tversky proposed a feature based approach as a more accurate way to convey similarity. The feature approach saw similarity as a function of a list of properties.

While improving upon the limitations of the spacing approach, the feature approach to similarity had issues in that the information it utilizilized was not as straightforward to quantify and visualize as the information using the spacing approach. More recent research into the feature approach has sought better quantification capabilities by utilizing data such as typicality ratings and generation frequencies as a supplement to the standard feature generation tasks (De Deyene, 2008).

With rising interest in artificial intelligence, comparing human mental processing to those of algorithms has become a topic of discussion. Research has been conducted comparing human judgements of semantic similarity to the performance of deep learning models (Lee, 2005). As conducting such comparisons is still a relatively new frontier, much exploratory work remains to be done.

The ability to more intuitively convey human representation alongside access to advanced neural network models has created ample opportunities to see just how well some of these models align with human representation. For this project, we wanted to expand upon research in the domain of representation modeling by comparing two data sets of human hierarchical representation (Michael Lee & Leuven) to that of a deep learning model (Word2Vec).

**Question**

Considering the goals for this project, we constructed several points of inquiry of which to guide our research along. The first being, how accurate are hierarchical clustering algorithms in representing human representations of animals? In addition, we questioned how do features of animals affect hierarchical representations in humans and deep learning models?

**Method**

We collected animal data from Michael Lee’s dataset, Leuven’s dataset, and the Word2Vec Deep Learning Model.

Michael Lee’s data contains the similarity matrix of 21 animals, which represents human’s similarity ratings, on the scale of 0-1, of each pair of animals. Leuven’s data contains human judgement of features from different animals. Leuven’s 129x759 frequency matrix, with the rows representing different animals and columns representing different features, indicates how often people would think of some features when it comes to each animal. Word2Vec Deep Learning Model is a neural network model trained with the English GoogleNews Negative300 Database, which constructs different linguistic word contexts according to the input words. By feeding in two different animals, the model is able to give a similarity rating based on how likely these two animals occur in the same context. In our study, we collected our deep learning databy manually input two animals to the Word2Vec model on *Turku NLP Group* website. To summarize our dataset, Michael Lee’s dataset contains the human judgement of animals representation using similarity ratinging; Leuven’s dataset contains the human judgement of animal representation using features; the Word2Vec Model generates deep learning judgements of animals.

The visualization method we utilized is Classical Multidimensional Scaling (MDS). With the input of a similarity or distance matrix, MDS is able to graph the distances of those objects in a maplike “psychological space”.

We compared how different human judgement and deep learning model represent hierarchical structures. Note that Michael Lee’s and Leuven’s datasets contains different animals, so for each dataset we used Word2Vec Models to generate deep learning judgement of the same animals within each dataset, and then conducted two sets of comparisons: (1) Michael Lee’s (human judgement) vs. Michael Lee’s (Word2Vec Model). (2) Leuven’s (human judgement) vs. Leuven’s (Word2Vec Model). For the first set of comparison, since both Michael Lee’s(human judgement) and Michael Lee’s (Word2Vec Model) data are in the form of similarity matrix, we simply fed them into MATLAB function *cmdscale()* and got the graphs (see Figure 3a). For the second set of comparisons, since Leuven’s (human judgement) data is in the form of *AnimalsxFeatures* matrix, we had to transform it to a similarity/distance matrix in order to apply our MDS visualization method. MATLAB function *pdist()* computes distances between pairwise observations. By feeding in the *AnimalsxFeatures* matrix, *pdist()* is able to compute the pairwise distances between animals to a vector according to their feature ratings. Then we applied another MATLAB function *squareform()* to convert the distance vector into a distance matrix. It is noteworthy that for function *pdist()* there is the option to compute euclidean distance or cosine distance. We tried both options, the results of which we will discuss later in the discussion section.

Additionally, we are also interested in whether similarity ratings and feature ratings in human judgement would result in different hierarchical representations. We selected 18 animals that are measured in both Michael Lee’s and Leuven’s dataset, then compared 18 animals (human judgement: Michael Lee’s --- Similarity Rating) vs. 18 animals (human judgement: Leuven’s --- Features Rating) using MDS. Note that the human judgement data of Leuven involves the option of Euclidean and Cosine distance as well, which we will discuss in the following section.

**Results/Discussion**

***Cos vs Euclidean Distance***

First thing first, let’s start off with the effects of choosing “Euclidean” or “Cosine” distance when we convert Leuven’s “Animal x Features” matrix into a distance matrix. Figure 1a shows the representation of leuven’s data in cosine(left) and euclidean(right) distance. The clusters are very similar. Even after looking into the details of both graphs, it was difficult to identify any significant differences. Figure 1b shows the result of cosine(left) and euclidean(right) distance when visualizing the 18 animals from Leuven’s data. As one can see, again, the clusters are very similar and it is hard to determine which one works better. One may think cosine distance is better in a sense that, compared to euclidean distance, cosine distance clusters “shark” far away from land animals. On the other hand, euclidean distance may be the better one, because it clusters “elephant” away from smaller sized animals.

Overall, there are no noticeable differences in choosing either “Euclidean” or “Cosine” distance in our study. Therefore, we will just use “Euclidean” distances for the remainder of the study.

***Michael Lee (Human: Similarity Rating) vs Leuven (Human: Feature Rating)***

Humans represent animals using category knowledge. In Figure 2, the animals from Michael Lee’s and Leuven’s dataset are clustered by generic binary features. In Michael Lee’s data the animals are grouped into sea animals and land animals. Within the clusters, there are clusters of animals within a more specific category such as the cluster of land animals consists of a cluster of land animals that can fly: bee, butterfly, bat, dragon, and eagle. Although the animals from Leuven’s dataset are grouped into different categorizations, the clusters are categorized by binary features of winged and wingless animals. The different categorizations from these two datasets arise from different measures of human judgements. However, both representations are hierarchically categorized.

***Michael Lee (Human) vs Michael Lee (Word2Vec)***

Figure 3a shows the representation of animals from Michael Lee’s human similarity rating compared with the representation from the deep learning model’s similarity rating. Unlike humans who represent animals using category knowledge, deep learning models use contextual cues and consider features that people do not normally associate the animals with. In the deep learning model, there are two plausible categories which are winged animals and working animals. Other than the two plausible categories, there are no identifiable categories within the deep learning model. Figure 3b highlights the animals across both representations have similar distances; however, they are in different clusters. For instance, the chimpanzee and koala have similar distances from each other across both representations, but in the human judgement model chimpanzee is clustered with elephant, zebra, and camel whereas in the deep learning model chimpanzee is grouped with koala and shark. Although the deep learning model similarly simulates the distances of animals in the human judgement representation, the deep learning model does not rely on categories to represent hierarchies of animals.

Although humans and the deep learning model clustered these animals quite differently, the cluster at the bottom left of the human judgements and the cluster at the top right of the deep learning model are quite similar. These clusters both contain animals that can fly. Both clusters contained the following animals: bee, butterfly, bat, and eagle. The only difference between these clusters is that humans rated dragons to be similar to these animals, which places the dragon in this cluster, while the deep learning model excluded dragon but included chicken. While dragons are mythological creatures, they are said to have huge wings that enable them to fly. This may be why humans think they are similar to these other flying animals. However, the deep learning model only takes into account of how these animal words are used contextually. Since dragons are normally not discussed in the same context as these other flying animals (bee, butterfly, bat, eagle), the model excluded dragon, but included chicken in this cluster, since chickens also have wings that enable them to fly.

While these clusters contained similar flying animals, the distances between the animals differed between human judgements and the deep learning model. The distances between these flying animals are smaller in human judgements while the distances between these animals are larger in the deep learning model. This shows that humans perceive these flying animals to be more similar to each other than they might actually be. Humans may be rating these flying animals solely on their ability to fly and neglecting other aspects of these animals that may not be similar to each other. The deep learning model seems to preserve the differences between the animals by clustering them less tightly.

Instead of placing the dragons with the flying animals, the deep learning model placed it with another cluster of animals. This cluster, at first sight, seemed to resemble another cluster from the human judgements, the sea animal cluster. The human judgements grouped the animals shark, frog, and goldfish together, forming an apparent sea animal cluster. However, the deep learning model clustered the frog and goldfish with scorpion, dragon, and snake, resulting in a lack of a definite cluster of sea animals. This may be explained by the fact that there exists an animal fable named “The Scorpion and the Frog” and a type of goldfish called dragon eye goldfish. This may contribute to the clustering of these animals as these animals are not similar in much features, but rather associated with each other in these ways.

***Leuven (Human) VS Leuven (Word2Vec)***

Similar to the representation from the Michael Lee’s data, the representation from Leuven’s dataset shows that humans have a clearer hierarchical representation of animal categories than deep learning models. For example, birds have their own cluster in the human judgements while birds are mixed together with other mammals in the deep learning model representation. In the human judgement representation, animals within categories are clustered closer together compared to the animals in the deep learning model. This once again reinforces the idea that humans perceive these animals within clusters to be more similar to each other than they may be.

In the previous representations, humans have shown to have a more accurate hierarchical representation of animals. However, the deep learning model has a more accurate hierarchical representation of sea animals as shown in Figure 4b. The deep learning model distinguishes sea mammals from fish. Although humans seem to cluster within the fish category, we cannot explain these clusters. The bottom cluster in the human representation appears to represent sea mammals, but there are some fish such as the goldfish and anchovy that are not within that category. One fish that was included in the fish cluster from the human judgements, but was not included in the fish cluster in the deep learning model, was the fish, sole. The similarity ratings from the deep learning model between sole and all the other animals were very low. This can be explained by the fact that sole has many meanings to it. While humans are aware that the most appropriate definition to use for this word, amongst all the other animal names, is the fish definition, the deep learning model is unaware of that.

Both human judgements and the deep learning model yielded a cluster of insects. Humans seem to rate these insects by certain keywords as all of the insects with the word “fly” in the name were grouped together. Some birds were included in the insect cluster for the deep learning model representation. One possible explanation is that birds also fly just like insects do.

**Additional Explorations**

We are also interested in the effects of adding higher hierarchical animal-type words to our animal dataset. We did so by adding 4 higher hierarchical words (“vertebrate”, “animal”, “mammal”, and “insect”) to Michael Lee’s human judgement data. After graphing it, we compared it to the original Michael Lee’s graph, as shown in Figure 5.

First of all, as we can see, the overall layout of the graph is shifted after adding new words. Second, the placements of the higher hierarchical words are not very ideal. For example, the word “mammal” is placed far away from mammals like camel and horse. We infer that this issue might be due to small size of animal types in Michael Lee’s data. Further future study is need to test this hypothesis.

**Conclusion**

Overall, we can conclude that: (1) For human judgement, different measurement methods (in this case, similarity rating versus feature rating) can result in different hierarchical representation. (2) Humans cluster animals using more simple features and rules while the deep learning model uses more complicated and detailed features. (3) Within a specific cluster, humans are not good at determining how different the animals are in that cluster; the deep learning model places more different animals further apart within the cluster. (4) The deep learning model does not cluster animals hierarchically as clear as humans do because, unlike humans who only consider simple features, the deep learning model considers all features which makes its clusters hard for humans to understand.

**Limitation**

The limitations from this study comes from the Word2Vec Deep Learning model. The deep learning model is based on the English GoogleNews Negative300 Database, so the similarity rating of two animals is calculated based on how often they appear in the same linguistic context. Humans are able to interpret animal names which are rare and ambiguous (e.g., sole, ray); however, the deep learning model is unable to dissociated words that can be used in multiple context. For instance, participants who rated similarities between animals were given a list of animals, so they were aware that they had to dissociate the animal sole from a foot sole. Unlike the participants, the deep learning model had to consider soles from all contexts within its database.

**Future Studies**

There are ample opportunities for further research to build off of this study. For one, different deep learning models may be utilized to explore how well different models may compare to human representations. For example, instead of using a Bag-of-Words model, further research might utilize an n-gram model. Just as different deep learning models may be explored, exploring different human representation datasets hat utilize different data gathering techniques is likely to yield further insights. Furthermore, future studies may explore representation in different mediums. Instead of just comparing the representations of words, studies may explore representation with image or audio formats.

**References**

De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M. J., Voorspoels, W., & Storms,

G. (2008). Exemplar by feature applicability matrices and other Dutch normative data for

semantic concepts. Behavior Research Methods, 40, 1030 –1048.

doi:10.3758/BRM.40.4.1030

Lee, M.D., Pincombe, B.M., & Welsh, M.B. (2005). An empirical evaluation of models of text

document similarity. In B.G. Bara, L.W. Barsalou & M. Bucciarelli, (Eds.), Proceedings

of the 27th Annual Conference of the Cognitive Science Society, pp. 1254-1259.

Mahwah, NJ: Erlbaum.

Shepard, Roger N. (1962). "The analysis of proximities: Multidimensional scaling with an

unknown distance function. I.". *Psychometrika*. **27** (2): 125–140.

doi:10.1007/BF02289630

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327-352.

[http://dx.doi.org/10.1037/0033-295X.84.4.327](http://psycnet.apa.org/doi/10.1037/0033-295X.84.4.327)

Michael Lee’s Dataset

http://faculty.sites.uci.edu/mdlee/similarity-data/

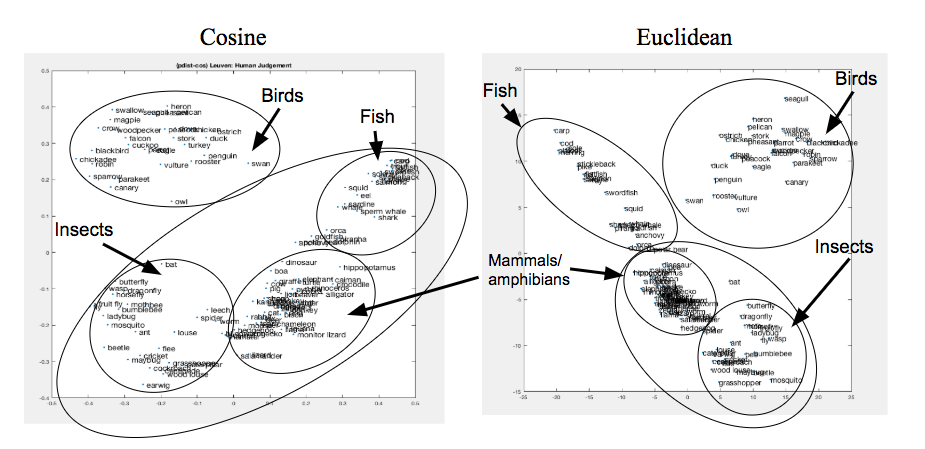
Leuven’s Dataset

https://ppw.kuleuven.be/apps/concat/datasets/brm\_concepts/

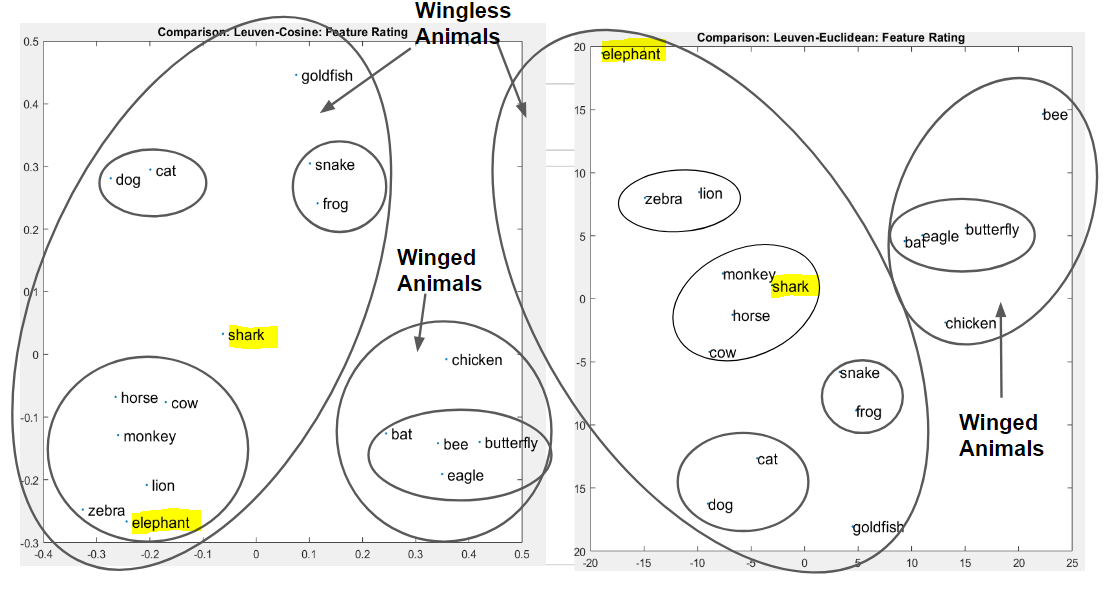
Turko BioNLP Word2Vec Model:

<http://bionlp-www.utu.fi/wv_demo/>

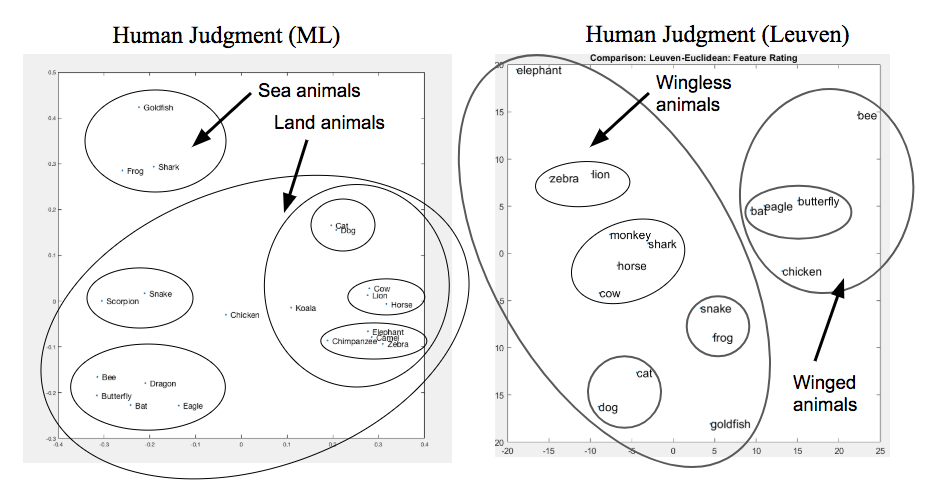
**Appendix**

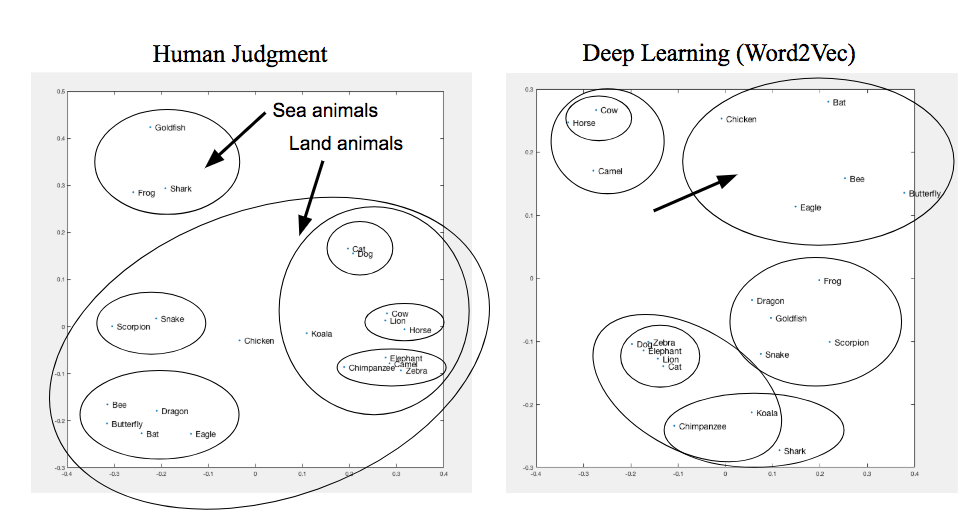
****

*Figure 1a:* Representation of animals based on Leuven’s data calculated with cosine(left) versus Euclidean(right).

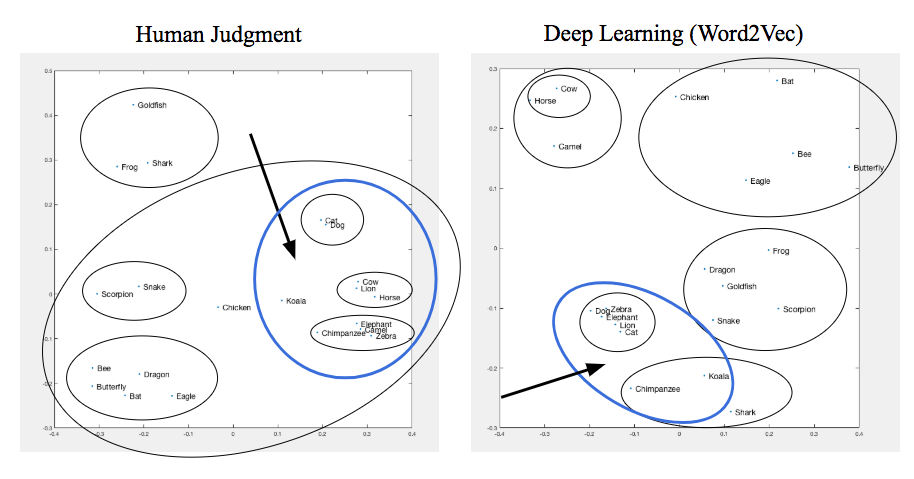
****

*Figure 1b:* Leuven’s data (18 animals) calculated with cosine(left) versus Euclidean(right).

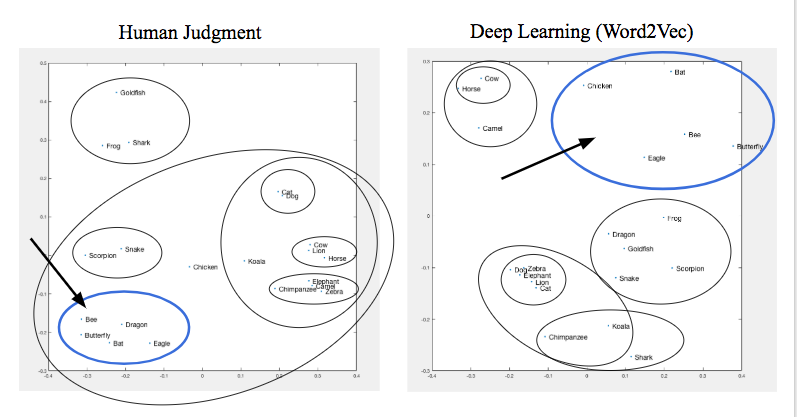
*Figure 2:* Representation of animals based on Michael Lee’s similarity rating (left) and Leuven’s feature ratings (right).

****

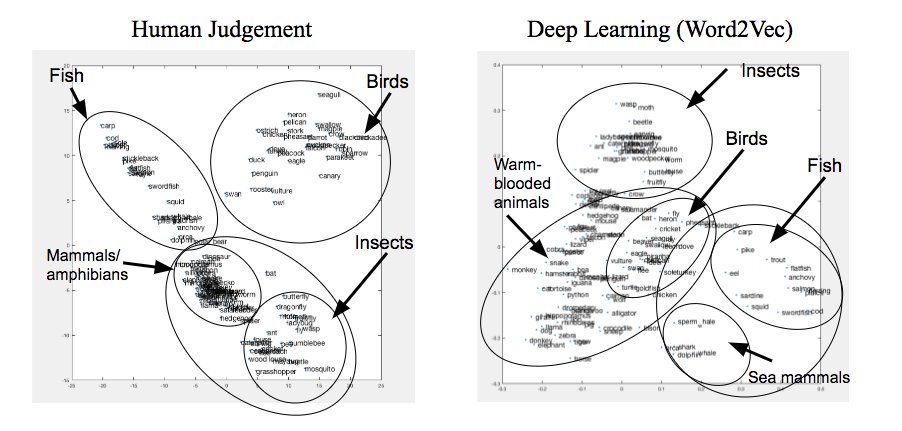
*Figure 3a:* Representation of animals based on Michael Lee’s similarity rating (left) and deep learning model similarity rating (right).



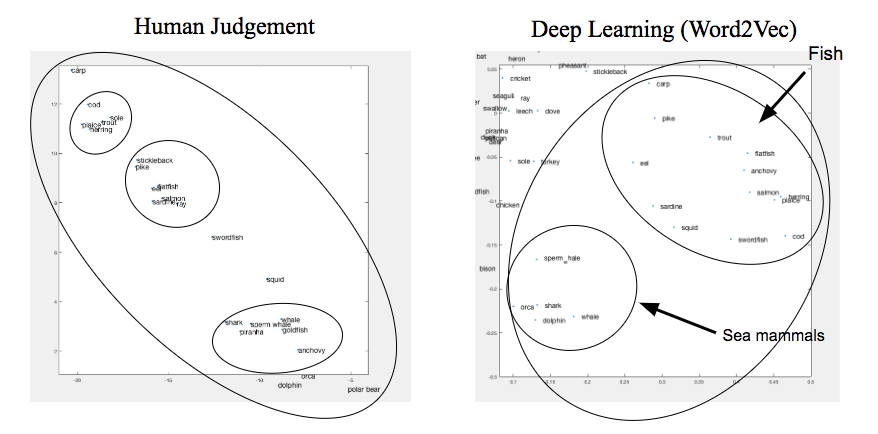
*Figure 3b:* Focus on animals with similar distances but clustered differently across Michael Lee’s similarity rating (left) and deep learning model similarity rating (right).

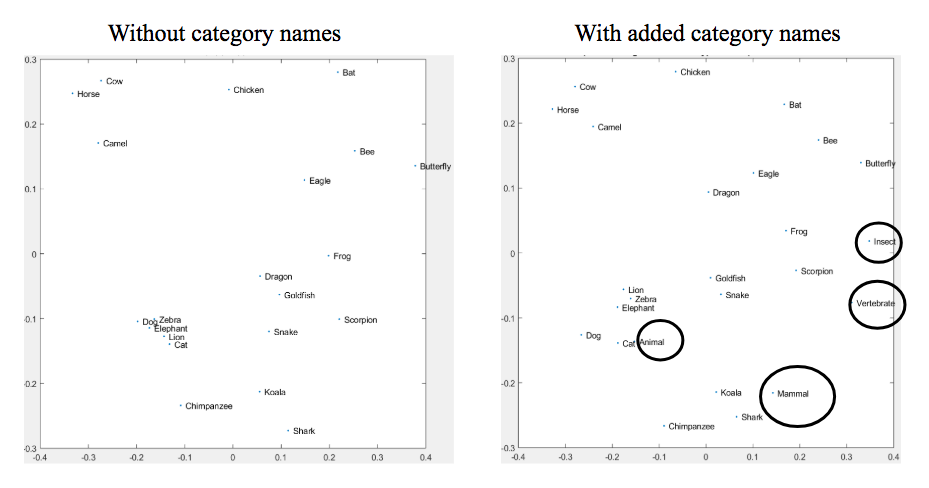


*Figure 3c:* Focus on flying animals within Michael Lee’s similarity rating (left) and deep learning model similarity rating (right).



*Figure 4a:* Representation of animals based on Leuven’s similarity rating(left) and deep learning model similarity rating (right).

*Figure 4b:* Close-ups of fish cluster from Leuven’s similarity rating (left) and deep learning model similarity rating (right).



*Figure 5:* Michael Lee’s human judgement graph with(left) and without(right) adding higher hierarchical animal-type words.