A Connectionist Semantic Network Modeling the Influence of Category Member Distance on Induction Strength

Michael Vinos (michael.vinos@gmail.com)
Graduate Program in Basic and Applied Cognitive Science, University of Athens
Ano Ilissia University Campus, GR-15771 Zografos, Greece

Efthymios Tsilionis (eftsilio@gmail.com)
Graduate Program in Basic and Applied Cognitive Science, University of Athens
Ano Ilissia University Campus, GR-15771 Zografos, Greece

Athanassios Protopapas (aprotopapas@phs.uoa.gr)
Department of Philosophy and History of Science, University of Athens
Ano Ilissia University Campus, GR-15771 Zografos, Greece

Abstract

We present a model for inductive inference when both the premises and the conclusion are categorical. The phenomenon under investigation is that less similar categories in the premises lead to stronger conclusions. The model is based on the Rumelhart semantic connectionist network (Rogers & McClelland, 2004, 2008). Simulations addressed the main phenomenon and nine additional non-trivial phenomena of categorical induction (from Osherson, Smith, Wilkie, Lopez, & Shafir, 1990), providing support to the majority of the hypotheses.

Keywords: category-based induction; connectionist model; semantic network; categorization

Inductive inference is a distinctive attribute of human cognition and is distinguished as an important field in cognitive science research. Induction refers to our ability of deriving conclusions through generalization of pre-existing knowledge into new circumstances (Hayes, Heit, & Swendsen, 2010). For instance, we can predict that the sun is going to rise tomorrow morning and also we know that each new to us species of fish can swim and that it bears a set of other fish features. A well-studied form of inductive inference is categorical induction. Category-based induction concerns transfer of the properties of a category or group of categories to some other category (Hayes et al., 2010; Heit, 1997). Transfer is facilitated when the categories under comparison belong to the same superordinate category and therefore share a lot of common characteristics and features (Hayes et al., 2010; Heit, 1997).

Osherson et al. (1990) studied category-based induction presenting to participants problems like those in Box 1. Propositions above the line are called premises and are assumed to be valid. The task was to make judgments about the strength of the conclusion below the line. Properties in these examples (blood sodium concentration) are called "blank" because participants are unlikely to have prior knowledge about them. Use of blank properties permits study of the net effect of the categories themselves upon the strength of the inference, to avoid effects of other related properties (Feeney & Heit, 2011; Osherson et al., 1990).

Box 1. *Problems used by Osherson et al. (1990).*

Hippopotamuses have a higher sodium concentration in their blood than humans.
Hamsters have a higher sodium concentration in their blood than humans.

All mammals have a higher sodium concentration in their blood than humans. (1)

Hippopotamuses have a higher sodium concentration in their blood than humans.
Rhinoceroses have a higher sodium concentration in their blood than humans.

All mammals have a higher sodium concentration in their blood than humans. (2)

Osherson et al. (1990) presented thirteen phenomena of category-based induction, which they argued should be part of a theory of inference strength, based on findings from two studies. They proposed the *similarity-coverage model of argument strength*, suggesting that the strength of the conclusion depends on the degree of similarity among the premises and the conclusion and among the premises and members of the superordinate category that includes them.

Here we focus on phenomena 2 and 6, which concern the *diversity* of the premises. Specifically, the less similar the categories of the premises are, the more they confirm the conclusion. Phenomenon 2 concerns the *general* case, in which premise categories are included in the conclusion category, while Phenomenon 6 concerns the *specific* case, in which a single category includes premises and conclusion.

Phenomenon 2 is exemplified in Box 1. Argument (1) leads to a stronger conclusion than (2) because of the diversity of the premises, even though hamsters are less typical members of the category mammals than rhinoceroses. An example of the Phenomenon 6 is shown in Box 2. Argument (3) is stronger than (4) because lions are more similar to tigers although giraffes are not more similar to rabbits than tigers are (Osherson et al., 1990).

Box 2. *Examples of Phenomenon 6.*

Lions use norepinephrine as a neurotransmitter.
Giraffes use norepinephrine as a neurotransmitter.

Rabbits use norepinephrine as a neurotransmitter. (3)

Lions use norepinephrine as a neurotransmitter.
Tigers use norepinephrine as a neurotransmitter.

Rabbits use norepinephrine as a neurotransmitter. (4)

In addition to these two phenomena, we conducted a preliminary investigation of the remaining nine non-trivial phenomena in Osherson et al. (1990). All phenomena are based on the assumption that conceptual representations are hierarchically structured. A connectionist model that achieves such conceptual organization is the semantic network proposed by Rumelhart and Todd (1993; as cited in Rogers & McClelland, 2004, 2008), henceforth termed *Rumelhart network*. This network discovers similarity structures in its training environment, resulting in similar representations of items with common features, i.e., members of a category (Rogers & McClelland, 2004, 2008). Similarities are context-dependent, so two members with similar representations in one context may vary in another (Rogers & McClelland, 2004, 2008).

Our aim is to show that such a semantic network can exhibit inductive behavior with differentiated strength depending on the premises. More specifically, we examine whether the network will produce stronger conclusions when premise categories are more distant.

Modeling Framework and Method

The classic Rumelhart architecture, as described by Rogers and McClelland (2004, 2008), has been shown to form "coherent categories" of entities, simulating human conceptual organization and acquisition. It also exhibits "inductive projection" of new properties, acquired after initial training (Thibodeau, Flusberg, Glick, & Sternberg, 2013). Our implementation is a feedforward network trained with error back-propagation, displayed in Figure 1.

There are two groups of input nodes, namely *Item*, with 21 nodes, and *Relation*, representing "context constraints", with 4 nodes. There are two hidden layers: *Representation,* with 15 nodes, receiving connections from Item, and *Hidden*, with 28 nodes, receiving connections from Relation and Representation and sending connections to the output layer. There are 4 groups of output nodes, each one corresponding to one context, namely ISA, Is, Can, and Has, with 28, 8, 6 and 12 nodes, respectively. An additional output node was used for the blank property (termed *Queem* following Rogers and McClelland). The numbers of nodes in the hidden layers were found to be sufficient for category learning in a reasonable number of training epochs and with stable outcomes, that is, converging to the same category structure in most training trials.

The Queem entity plays the role of substitute/placeholder for every blank property in Osherson et al. (1990), for example, "use the neurotransmitter Dihedron", "require titanium for normal muscle development", "have a higher sodium concentration in their blood than humans", etc. As noted by Rogers and McClelland do (2004, 2008), node labels play no functional role but are merely descriptive tags in the simulation. What is important for the simulation are the structural relations.

An alternative architecture, lacking the Relation input layer and with only one hidden layer, was used during our initial simulations but even though it was found to form appropriate category structures, it failed to exhibit the phenomenon under investigation with the blank property. It seems that this particular structure, including the Representation, is required for the network to exhibit the desired richness of behavior beyond simple categorization.

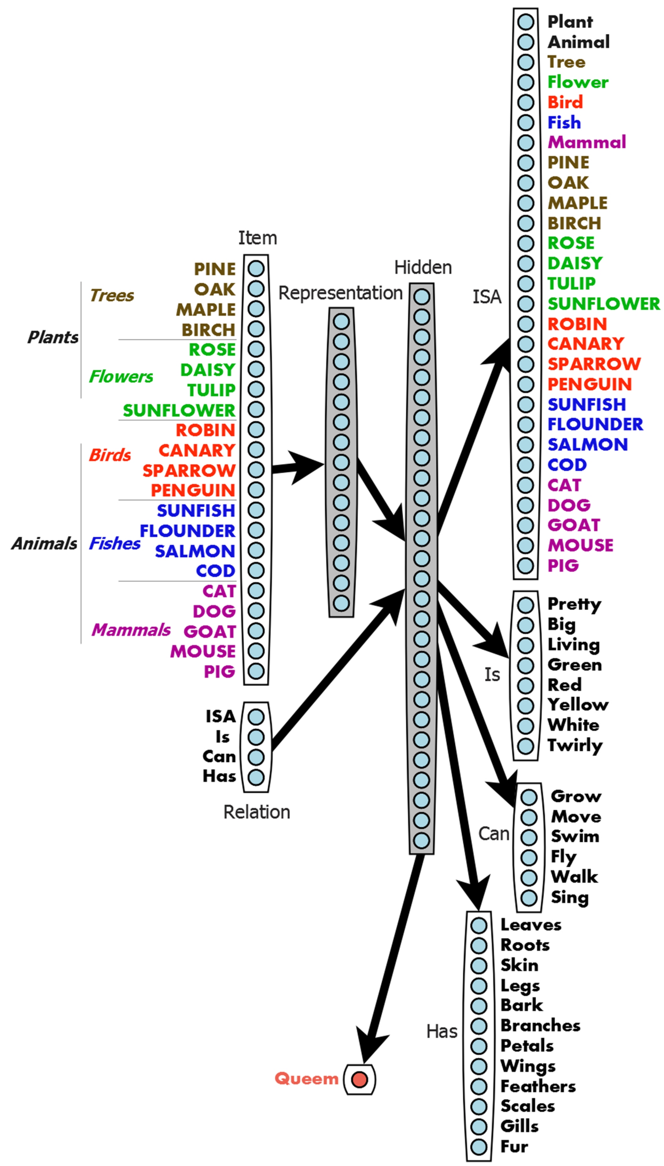

Figure 1: Network architecture used in the simulations.

We used the extended version of the Rogers and McClelland model (2004), with 21 input entities and one addition to the ISA output group. Specifically, we added the property (ISA)Mammal, activated for the training patterns of mammal species, similarly to the other classes. The input and output representations of the model are localist. During initial training, the Queem property has zero activation for all items in all four contexts.

There were 84 training patterns (21 items in 4 contexts). Here is a coding example for *Penguin* in the *Can* context:
Input: Item=(0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0)
|PENGUIN; Relation=(0,0,1,0)|Can
Output: ISA=(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0); Is=(0,0,0,0,0,0,0,0,0); Can=(1,1,1,0,1,0)|Grow; Move, Swim, Walk; Has=(0,0,0,0,0,0,0,0,0,0,0,0,0),
Queem=(0)

Simulations adopted the strategy of Rogers & McClelland (2004, 2008), with 2,500 epochs of initial training, leading to coherent categorization and modest generalization of the new property in the different contexts after additional training. Learning rate was set to 0.3 and momentum to zero. After the initial training, the learned categories were revealed in cluster analysis of the Representation layer.

Subsequently, pairs for the comparisons were chosen. Each comparison involved the generalization of the blank property Queem when the network learned that two members of a category, similar versus dissimilar to each other, possess the property. The comparison is meaningful when it concerns the learning of the property in the same context, Is, Can, or Has. The ISA context was excluded because it is only relevant to item classification within the hierarchy. The pairs were chosen in the following fashion: A computation of euclidean distances between all Hidden layer activation vectors was carried out for each context (having in mind that the clustering of representations on the Hidden layer differs in several respects from that on the Representation layer, which is context independent.) Based on that, the two most similar species were identified, for instance Goat–Dog from the category Mammals, to form the similarity condition. For the second pair, that is, the diversity condition, we keep one of the two chosen entities and computed the euclidean distances between it and all other entities in the superordinate category. Keeping to the above example, we sought the species most distant from Goat among all animals, which turned out to be Canary. So the pair for the diversity condition was Goat–Canary.

Subsequently, for each of the two pairs, the network was trained for additional epochs past the initial 2,500 to learn that the two species possess the Queem property in the particular context but not in the other three contexts (a total of eight training patterns). An example training pattern is:
Input: Item=(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0)
|GOAT; Relation=(0,0,0,1) |Has.
Output: ISA=(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0); Is=(0,0,0,0,0,0,0,0,0); Can=(0,0,0,0,0,0);
Has=(0,0,1,1,0,0,0,0,0,0,0,0,1)| Skin, Legs, Fur;
Queem=(1)| Queem

A criterion was established to terminate additional training when the new property has been acquired. An activation threshold of the Queem node for the relevant inputs might be appropriate but this would demand a large number of training epochs. Instead, after some experimenting a mean error criterion of 0.002 was adopted, which ensured activations for the members of both pairs greater than .9 and usually greater than .95 for at least one of them. With this termination criterion, 345–977 (*M*=579, *SD*=176) additional epochs of training were required.

After the completion of the additional training cycles we followed Rogers and McClelland (2004) and modified the network weights, retaining only those from the Hidden layer to the Queem node, and replacing the rest from the initial training (2,500 epochs). Thus the remainder of the originally trained network was unaffected by training the Queem node. This network was submitted to a test with 21 input lines containing all species/items in the current context, to obtain activation values of the Queem node. Separate values for some species, or mean values for a category (at any level in the hierarch) or for all living things, as appropriate, were compared to test the experimental hypothesis. These activation values were used to quantify the strength of the argument and the corresponding "confirmation score" (comparing to Osherson et al., 1990). In this manner, the strength of the conclusion (and hence of the whole argument) concerning a certain species corresponds to the activation of the Queem node for this species in this context. When the conclusion concerns a category (e.g., fish), its strength corresponds to the mean activation for all species belonging to the category.

Results

Results are divided in two parts: The first part corresponds to the main phenomenon, that is, Premise Diversity. The second part concerns the other nine phenomena described by Osherson et al. (1990). Due to space limitations, we present results only for the Has context/relation (except when examining the context effect) but the results were similar for the other contexts. In addition, the patterns of results do not alter even with different premises, given that the latter obey the selection rules.

Premise Diversity Simulations

We have put together graphs in Figure 2 depicting the results for all the types of simulations for the two forms of the phenomenon, general and specific. Accompanying each bar chart is the respective argument formed under the general paradigm proposed by Osherson et al. (1990). Indicative colors are used for the bars and their corresponding arguments (either diverse or similar). In general for all simulations the results supported the initial hypothesis that the arguments with distant premises are stronger than the arguments with close premises. Note that, in the graph depicting the results for the specific version of the phenomenon, the three species, which correspond to the

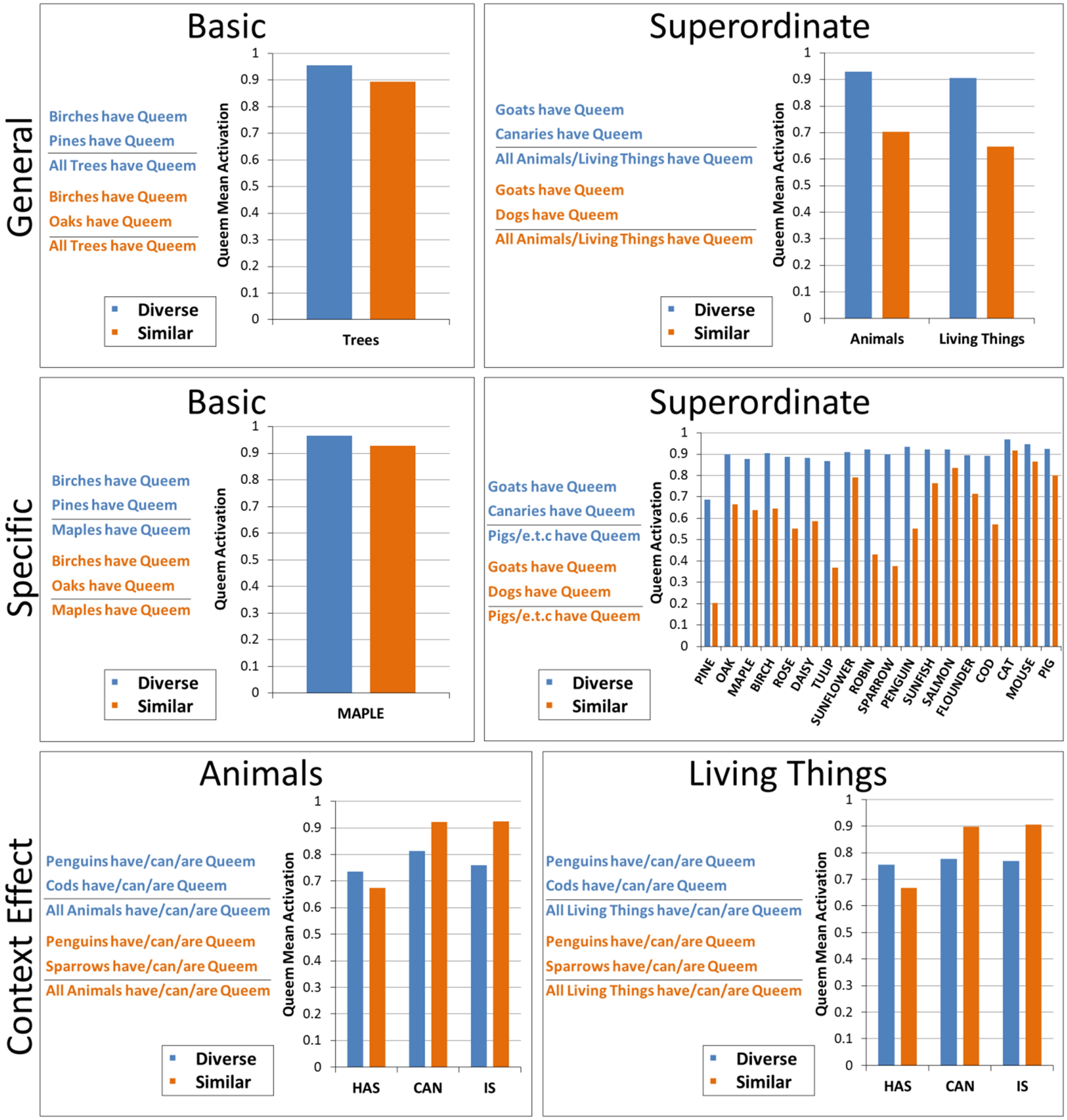

Figure 2: Activation of the Queem node for the different simulations of the phenomenon of premise diversity in each condition (diverse and similar). Inside each panel we give an example of the arguments in each condition. Above the line we present the premises that were used for the additional training and under the line the conclusion that was used for testing. The top two rows of panels refer to the two versions of diversity (general and specific) and the bottom row to the simulation with the different contexts. The columns of panels refer to which category the simulation was performed.

premises used, are not included.

Two kinds of premises were used in the diverse condition: one where the entities/items both belong to the same (basic) class as the conclusion (Trees) and another where the items belong to the same superordinate category but not to the same class (Mammals and Birds), as shown in Figure 2. This has to do with the limited range of distances between the activation vectors of the Hidden layer for a single class. The differences in activation values for the Queem node were larger for the *superordinate* simulations.

We finally conducted a series of simulations to examine the differentiating impact of context on the generalization of blank properties. Simulations revealed the paradoxical effect of *reversal* of the phenomenon. That is, the arguments of the diversity condition that exhibited greater strength than the ones of the similarity condition in one context may display a

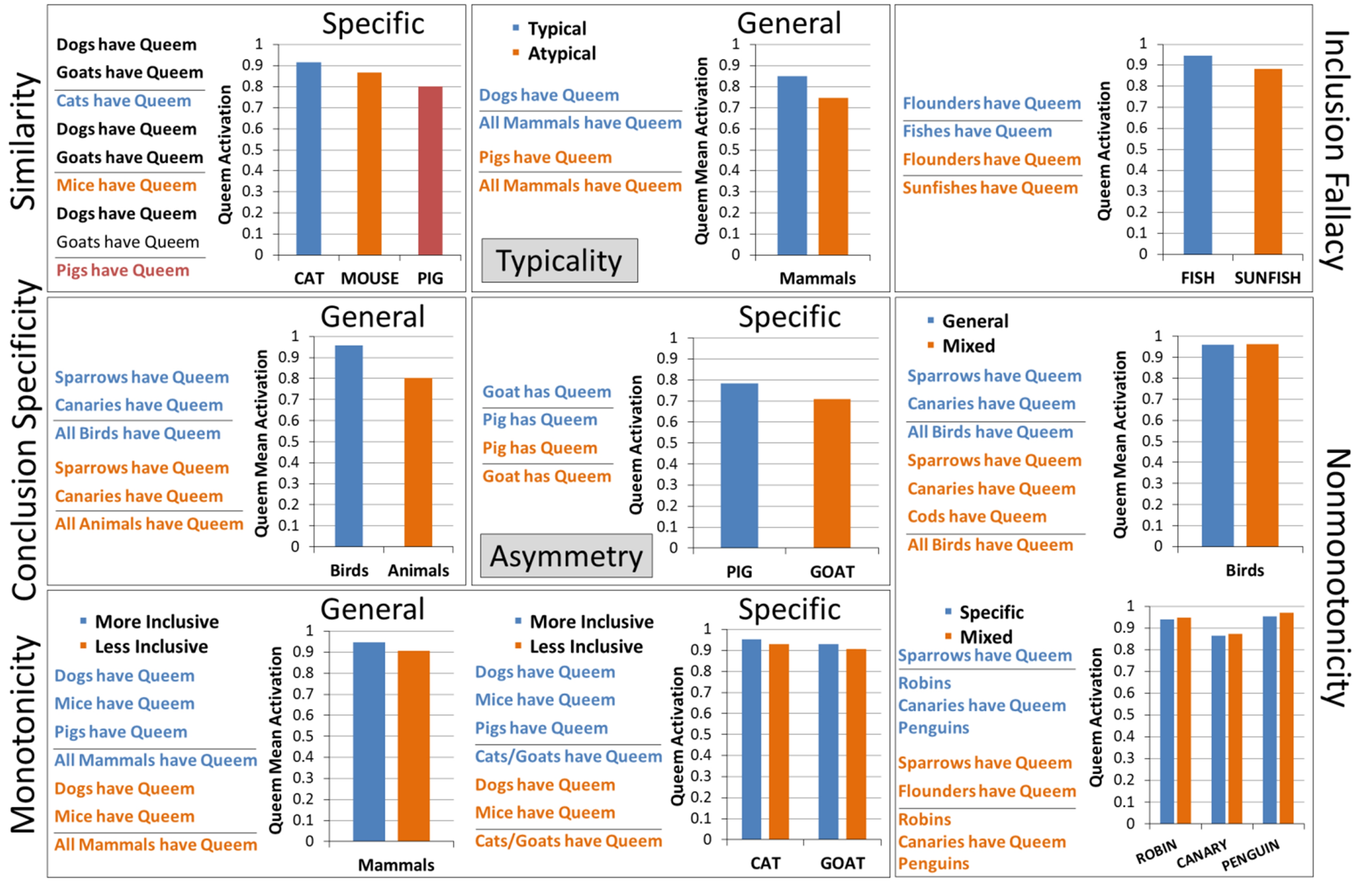

Figure 3: Activation of the Queem node for the remaining nine phenomena. As in Figure 2, inside each panel we give an example of the arguments in each condition. On the left side of each panel the name of the specific phenomenon is presented. The two columns of panels refer to the two types of arguments (general and specific).

reversed strength pattern when examined in another context: the corresponding arguments of the similarity condition are now stronger than the ones of the diversity condition.

Simulations of the remaining Phenomena

For the rest of the phenomena, the results are displayed in Figure 3, accompanied by the respective arguments and using indicative colors for their types. Although there are in total eleven phenomena, two of them are trivial (the premise-conclusion identity and the one in which the conclusion category is included in the premise category) and hence require no investigation or validation. The majority of hypotheses posed from these phenomena have been consistently supported in our simulations although the differences in activation values were not large. The only hypotheses not supported were the two referring to non-monotonicity. Brief descriptions derived from Osherson et al. (1990) and clarifications are given below.

Premise Typicality: "the more typical the premises are of the conclusion, the more they confirm it". We selected species for the premises from one class computing the mean value of the activation vectors for all class members over the Hidden layer and then choosing the one closest to it as the typical member and the most distant one as less typical.

Conclusion Specificity: "the more specific is the conclusion, the more it is confirmed by the premises"

Premise Monotonicity (general and specific versions): "more inclusive sets of premises yield more strength than less inclusive sets"

Premise-Conclusion Similarity: "the more similar the premises are to the conclusion, the more they confirm it"

Premise-Conclusion Asymmetry: "single-premise arguments are not symmetric, in the sense that premise/conclusion may not have the same strength as conclusion/premise"

Non-Monotonicity (general and specific versions): "arguments can be made weaker by adding a premise that converts them into mixed arguments". Note, that these two phenomena involve mixed arguments in the sense that they are neither general nor specific.

Inclusion Fallacy: "a specific argument can be made stronger by increasing the generality of its conclusion". To investigate the general conclusion for fish, we did not compute the mean value for fish but instead created a new item (Fish), for which nodes 12–15 of the Item layer had a value of .25, and in the other layers activation values were the same as for the other fish.

Discussion

The simulations supported both the general and specific versions of the diversity phenomenon. The strength of the arguments in the diversity condition was higher than in the similarity condition, and this was observed with many different initial parameters. The cause of this effect lies in the distance of the representations. The greater the distance between two items the more the blank property is generalized across the remaining items. Thus, the distance determines the degree of generalization. Another important conclusion is that the distances between items in the representational space change according to the context; as a result, the diversity effect can be inverted. The power of the context is that it transforms dissimilar objects into similar ones.

Although our main goal was the investigation of diversity we also performed a preliminary exploration of the other phenomena described by Osherson et al. (1990). Our results supported all phenomena except nonmonotonicity. Even though the differences in Queem activation were small, they were consistent. Undoubtedly, there is space for more extensive investigation of these phenomena, including diversity, to discover other aspects of network function.

A number of concerns emerged during our investigation. One issue was the small number of items and subsequently the small size of the categories. The consequence of this, and the fact that members of a category share many common properties, was that the euclidean distances between members of the same class (e.g., Mammals) were not big enough for the diversity phenomenon to appear to a large extent. For this reason we chose to use in our simulations mainly items from the same superordinate category where the number of items is greater and as a result the representational space increases. A possible improvement might be to construct basic categories with more objects by introducing more and appropriately selected properties in the network. This will also lead to an increase in the number of nodes.

Another issue concerns the method of replacing the weights of connections to the Queem node after the additional training. To minimize interactions with the existing knowledge of the network, an alternative method would be to freeze all connection weights during the additional training except the weights to the Queem node. Interesting questions that arise are whether learning could be facilitated with the proposed method (fewer epochs of training) and if the connection weights would be similar across the two methods. The implementation of this method would be a subject of future research.

As Rogers and McClelland (2004) indicate, the semantic network captures all the different kinds of human developmental phenomena about inductive reasoning supported by empirical findings (specificity, coalescence, differentiation etc.). Thus, an interesting new route of investigation could relate to these particular aspects of the categorical phenomena examined in this study.

Finally, in our simulations the different contexts regard the activation of specific properties for the different items. The network discovers the similarity relations between the items and creates the proper representations. However, items in the real world also exhibit causal relations. Causal relations are used very often in inductive inference and they are even preferred over similarity relations (Feeney & Heit, 2011; Hayes et al., 2010). Hence, if the premises and the conclusion share a causal link the argument is judged stronger and the diversity effect almost disappears (Feeney & Heit, 2011). Connectionist models are capable of discovering other forms of relations in the training data besides similarity structures. Recently, a network using the Rumelhart architecture displayed analogical inference (Thibodeau et al., 2013). This network generalized relations to untrained data. Therefore, a potential improvement of our model would be to examine whether learning causal relations would lead to stronger induction for a blank property in comparison to similarity relations.

Acknowledgments

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Programs: THALIS–UOA–375737, ARISTEIA–COGNIMUSE–GSRT–1770 and ARISTEIA–BabyAffect–GSRT–3610

References

Feeney, A., & Heit, E. (2011). Properties of the diversity effect in category-based inductive reasoning. *Thinking & Reasoning*, *17*(2), 156–181.

Hayes, B. K., Heit, E., & Swendsen, H. (2010). Inductive reasoning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(2), 278–292.

Heit, E. (1997). Features of similarity and category-based induction. In *An Interdisciplinary Workshop On Similarity And Categorisation (SimCat)*. Edinburgh, UK. Retrieved from http://faculty.ucmerced.edu.sci-hub.org/sites/default/files/eheit/files/simcat.pdf

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*(2), 185.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. London, England: MIT press.

Rogers, T. T., & McClelland, J. L. (2008). Précis of semantic cognition: A parallel distributed processing approach. *Behavioral and Brain Sciences*, *31*(06), 689–714.

Thibodeau, P. H., Flusberg, S. J., Glick, J. J., & Sternberg, D. A. (2013). An emergent approach to analogical inference. *Connection Science*, *25*(1), 27–53.