Is Word Generalization for Novel Concepts Modelled by Similarity or by Formal Concepts?

Sujith Thomas
Department of Computer Science and Engineering
Indian Institute of Technology Kanpur
Kanpur, India
Email: sujith@cse.iitk.ac.in

Harish Karnick
Department of Computer Science and Engineering
Indian Institute of Technology Kanpur
Kanpur, India
Email: hk@cse.iitk.ac.in

***Abstract*—In this paper, we conduct two word learning experiments to study the human word learning and generalization behaviour. The participants are shown abstract figures that form the positive examples of a word concept. All the positive examples have a common defining feature. The participants are then asked whether the word applies to various test examples. We vary several independent variables across our two word learning experiments. Our results show that the word generalization behaviour is based on the similarity of a test example with the positive examples of a word. The generalization behaviour is not based on the defining features. This is true even when enough examples exist from which the defining features can be inferred.**

***Keywords*—Human Word Learning; Hypothesis Space; Word Generalization; Formal Concepts; Object Similarity.**

I. Introduction

Humans have the ability to learn a new word after seeing a few examples of the word. Carey [1] named this phenomenon as *fast mapping*. Xu and Tenenbaum [2][3] demonstrate that fast mapping can be accurately modelled using a similarity-based generalization. In similarity-based generalization, the probability of generalization becomes higher for the test examples that are more similar to the positive examples of a word. Xu and Tenenbaum use a hierarchical clustering tree as their hypothesis space. Each node in the tree forms a hypothesis for a word concept.

Abbott, Austerweil and Griffiths [4] describe an approach for constructing a hypothesis space for word learning. The hypothesis space they use is also similarity based where the similarity is derived from the relationship between images and words in ImageNet [5] and WordNet [6] respectively. Abbott, Austerweil and Griffiths show that their similarity based generalization matches the empirical data.

The word categories used in the above papers [2][3][4] do not have a well-defined set of necessary and sufficient conditions. If the word categories have a set of necessary and sufficient conditions, will the generalization still be similarity based? Our hypothesis is that the generalization behaviour will be similarity based even when the word categories are well-defined. By well-defined, we mean that the necessary and sufficient conditions for the word category can be easily deduced from the positive examples.

We use artificial word concept categories defined using boolean features. Each word concept can be represented as a formal concept. We study the generalization behaviour of the human participants after showing them a few examples of a word concept.

In Section II, we describe our definition based hypothesis space. Section III explains how we construct our similarity based hypothesis space. The details of the two word learning experiments are given in Section IV and Section V. In Section VI, we discuss how our work relates to other works. Finally, we list our conclusions and point to future directions of our work in Section VII.

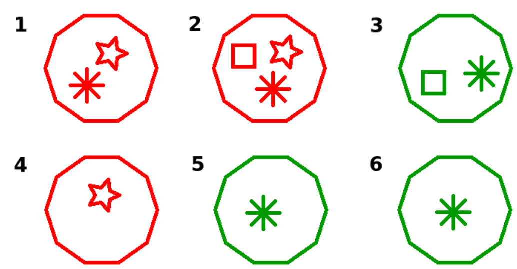

Fig. 1. The figure shows six objects that occur in a boolean world. Each figure can be represented using a set of boolean features.

TABLE I. The Table lists the objects shown in Figure 1. The 'X' mark denotes that an object has a boolean feature.

	star	red	spike-wheel	square	green
1	X	X	X		
2	X	X	X	X	
3			X	X	X
4	X	X			
5			X		X
6			X		X

II. Definition Based Hypothesis Space

Formal concepts form our hypothesis space of definition based concepts, i.e., the concepts that have a set of necessary and sufficient conditions. Here we follow the intuition that the set of boolean features that always occur together provide information about the categories present in the boolean world. For example, consider the six objects depicted in Figure 1. Table I lists these six objects along with the boolean features present in them. The 'X' mark denotes that an object 'has a' boolean feature.

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-340-7

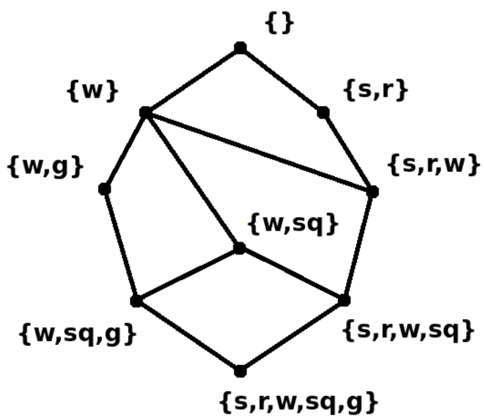

Fig. 2. The figure shows the complete lattice constructed from all the closed sets in Table I. Each node forms a hypothesis in the definition based hypothesis space.

We use the abbreviations s, r, w, sq and g to denote the presence of boolean features 'has a star', 'is red', 'has a spike wheel', 'has a square' and 'is green', respectively. All the objects also have one feature in common, namely, the decagon shape of the figures. But this feature is omitted here for the sake of brevity. In Table I, we notice that the sets of features $\{s, r, w\}$, $\{w, g\}$ etc., occur together in the world. These sets of features should provide information about the concept categories present in the boolean world.

We call a set of features A, a *closed set* if there does not exist a feature f such that the set of features $A \cup \{f\}$ have the same frequency of occurrence. In other words, if A is a closed set then for any feature f the set of features $A \cup \{f\}$ occurs with a frequency strictly less than that of set A. By this definition the set of features $\{r, w\}$ is not a closed set because $\{r, w\}$ occurs with frequency two and so does the set $\{s, r, w\}$. On the other hand, $\{w, g\}$ is a closed set with frequency three. This is because we cannot add a boolean feature to $\{w, g\}$ such that its frequency of occurrence remains the same. Every closed set of features has an extension that contains the set of objects that have those features. The extension of closed set $\{w, g\}$ is the set $\{3, 5, 6\}$ because w and g is present in all the three objects. Every closed set of boolean features along with its extension form a *formal concept*.

The notion of a closed set of features is useful because they carry all the information about the features that co-occur. Considering only the closed sets in a boolean world is advantageous because the closed sets of features are usually much lesser in number compared to all possible subsets of boolean features.

The formal concepts in Table I are the following—$\{\}$:6, $\{w\}$:5, $\{s, r\}$:3, $\{w, g\}$:3, $\{w, sq\}$:2, $\{s, r, w\}$:2, $\{w, sq, g\}$:1, $\{s, r, w, sq\}$:1 and $\{s, r, w, sq, g\}$:0. The number after the colon represents the frequency of occurrence of each formal concept. Note that $\{\}$ (null set) and $\{s, r, w, sq, g\}$ (set of all features) are also closed sets in the boolean world shown in Table I. The formal concepts satisfy a partial order relation $\supset$ (superset of), and therefore, can be arranged in the form of a complete lattice [7]. The complete lattice in Figure 2 depicts all the formal concepts in Table I. Each node in Figure 2 forms a hypothesis in the definition based hypothesis space of word concepts.

Ganter and Wille [7] explain the theoretical aspects of formal concept analysis in great detail. For a brief introduction to formal concept analysis refer Krötzsch and B Ganter [8]. Formal concept analysis has been successful in a wide area of applications ranging from linguistics and software engineering to artificial intelligence [9].

The lattice of formal concepts discussed in this section forms our hypothesis space of definition based word concepts.

III. Similarity Based Hypothesis Space

In order to construct a similarity-based hypothesis space, we need a notion of similarity. Xu and Tenenbaum [2][3] construct a hierarchical clustering tree using the average similarity ratings between the various images. These ratings were provided by the human participants. Abbott, Austerweil and Griffiths also use a tree structured hypothesis space where the notion of similarity between the hypothesis is derived from WordNet and ImageNet [4].

Tversky [10] discusses the similarity measures that can be used to emulatex human similarity judgments. Tversky introduces a similarity measure called Ratio model which is as follows:

$$S(X, Y) = \frac{f(X \cap Y)}{f(X \cap Y) + \alpha f(X \setminus Y) + \beta f(Y \setminus X)} \qquad (1)$$

where X,Y are the sets of boolean features for the two objects to be compared and $\alpha, \beta \geq 0$. The α and β values determine the weights given to the features in X and Y while computing the similarity.

We use a simplified form of Tversky's Ratio model, where $\alpha = 1$, $\beta = 1$ and $f(Z) = |Z|$, where Z is a set. We let $\alpha = \beta = 1$ so that the features present in both the objects get equal weights while computing the similarity. The simplified form of Tversky's Ratio model becomes :

$$S(X, Y) = \frac{|X \cap Y|}{|X \cap Y| + |X \setminus Y| + |Y \setminus X|} \qquad (2)$$

This simplified form is just another way of writing the Jaccard similarity coefficient [10]:

$$J(X, Y) = \frac{|X \cap Y|}{|X \cup Y|} \qquad (3)$$

where X and Y are the sets of boolean features corresponding to the two objects to be compared.

Figure 3 shows the hierarchical clustering tree constructed from all the objects in Figure 1 using the Jaccard similarity coefficient (3). Each node in the tree forms a hypothesis in the similarity based hypothesis space. The hypothesis that are present in Figure 2 need not be present in Figure 3, and vice versa. This is because objects that satisfy a definition in Figure 2 need not be very similar to form a cluster, and the objects that form a cluster in Figure 3 need not have a set of defining features.

Which hypothesis space do humans use for the learning and generalization of word concepts? We try to answer this question in the following two sections.

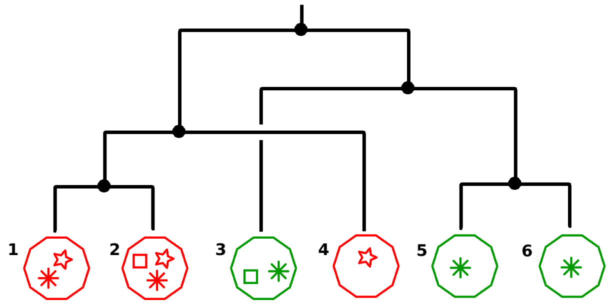

Fig. 3. The figure shows the hierarchical clustering tree constructed from all the objects in Figure 1 using the Jaccard similarity coefficient (3). Each node forms a hypothesis in the similarity based hypothesis space.

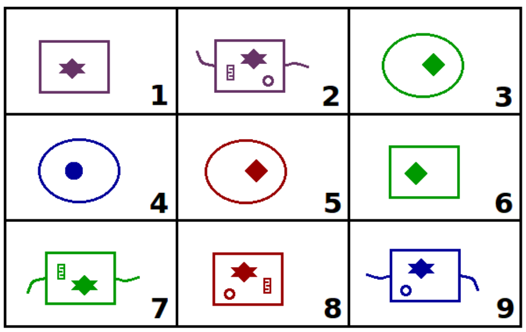

Fig. 4. A sample stimulus for the word learning experiment. Objects 7, 8 and 9 form the positive examples, and the objects 1, 2 and 3 form the test examples.

IV. First Word Learning Experiment

In this experiment, 23 participants were shown 3 positive examples of a word and asked whether the word generalizes to a test example. We used a random four letter word in our experiment having consonants and vowels in alternating positions. The positive examples in our experiment always correspond to a formal concept.

Figure 4 shows a sample stimulus. The participants were told that a novel word applies to objects 7, 8 and 9 in Figure 4. They were then asked whether the word applies to objects 1, 2 and 3. The rectangular bounding box and the star are the features that are common to objects 7, 8 and 9. If the generalization behaviour is based on a formal concept, then we would expect the participants to generalize a word to both objects 1 and 2. This is because both the objects 1 and 2 have a rectangular bounding box and a star. On the other hand, if the generalization behaviour is based on similarity, then we would expect the participants to generalize the word more often to object 2 compared to object 1. This is because object 2 is more similar to objects 7, 8 and 9.

Figure 5 shows another sample stimulus. The participants were told that a word applies to objects 7, 8 and 9. They were then asked whether the word generalizes to objects 4, 5 and 6. The only feature that is common to objects 7, 8 and 9 is the red colour. If the generalization behaviour is based on necessary and sufficient conditions then we would expect the participants to generalize the word to both objects 5 and 6 equally often.

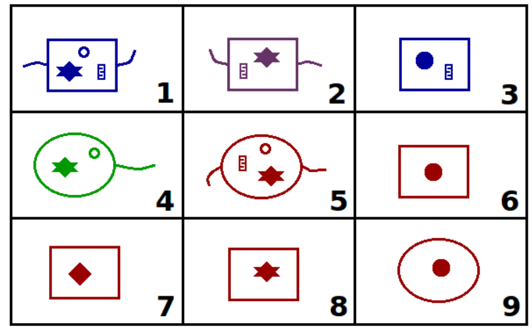

Fig. 5. Another sample stimulus for the word learning experiment. Objects 7, 8 and 9 form the positive examples, and the objects 4, 5 and 6 form the test examples.

This experiment is designed to study human word learning ability.
You will be given three examples of a word. Based on this information you need to judge whether the word applies to another example.
Respond to the questions as quickly and accurately as you can.

Fig. 6. Instruction for the first word learning experiment.

On the other hand, if the generalization behaviour is based on similarity then we would expect that the word is generalized to object 6 more often.

There were six stimuli similar to Figure 4 and Figure 5. Each stimulus had three test questions associated with it. The test questions varied in their degree of similarity—very similar, less similar and not similar—with the positive examples of a word. The first two types of the test examples (i.e., very similar and less similar) also satisfied the formal concept.

Participants were first given five trial questions to familiarize them with the task. The participants were then asked to respond to 18 generalization questions. Figure 6 shows the instructions that were given to the participants.

A. *Similarity Measure Used*

We use the Jaccard similarity coefficient (3) to find the similarity between any two objects. We did the following experiment to see how well the Jaccard similarity coefficient matches the human similarity judgments. We asked a different set of 24 participants to rate the similarity between 50 pairs of abstract figures similar to those shown in Figure 4 and Figure 5. We found that the Spearman rank correlation coefficient [11] between the Jaccard similarity coefficient and average similarity ratings was significant $(r(48) = .77, p < .001)$.

B. *Results*

For each participant, if the generalization behaviour is based on formal concepts, then, we would expect that the word gets generalized to the less similar test example as often as it does to the more similar example. This is because both the types of test examples belong to the extension of the same formal concept.

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-340-7

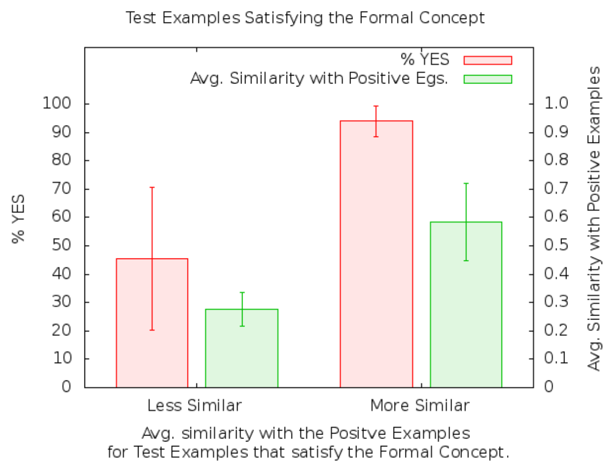

Fig. 7. The test example that satisfy the formal concepts were taken, and divided into two groups—less similar and more similar. More similar group received significantly higher percentage of word generalizations.

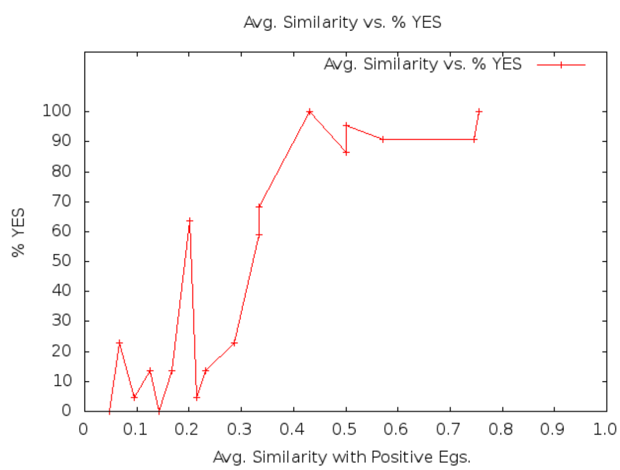

Fig. 8. The figure plots the percentage of participants that generalize a word to a test example against the average similarity of the test example with the positive examples. The correlation was found to be significant $(p < .005)$.

We divided the test examples that satisfy the formal concept into two groups—more similar and less similar. The test examples in the more similar group had a higher average similarity with the positive examples. Figure 7 shows the average percentage of trials in which a word was generalized to a test exemplar in each of the two groups. We see that this percentage is much higher for the more similar group. We also found the frequencies with which each participant generalized a word to a very similar and to a less similar test example group. The difference between the two frequencies for each participant was found to be statistically significant using the Wilcoxon signed ranks test [11] $(W(23) = 5.5, p < .001)$. This shows that the generalization behaviour was not based on formal concepts. The reason is that both the groups—more similar and less similar—satisfied the formal concept, and yet had a significantly different percentage of generalization.

Figure 8 shows how the percentage of trials in which a word was generalized to a test example varies with the average

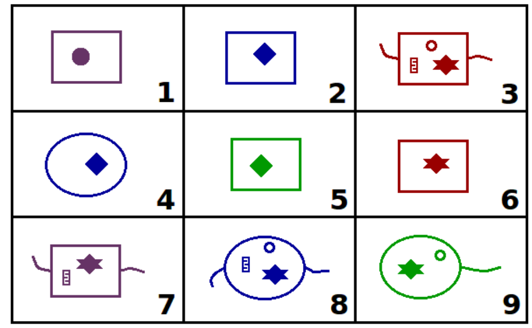

Fig. 9. Figure shows the stimulus corresponding to the anomalous spike in Figure 8 at X=0.2.

Jaccard similarity with the positive examples. The figure shows that the percentage increases with the increase in the average similarity of a test example. The Spearman rank correlation coefficient between the two variables in the figure was found to be statistically significant $(r(16) = .85, p < .005)$.

The above two results show that the generalization behaviour for the experiment is better modelled using the similarity based generalization compared to a formal concept based generalization.

C. Discussion: The Anomalous Spike in Figure 8

Figure 8 shows the data for the first word learning experiment. The figure plots the percentage of participants that generalize a word to a test example against the average similarity of the test example with the positive examples. In Figure 8, we see that there is a sudden spike in the graph at the value X=0.2 along the X-axis.

Figure 9 shows the stimulus corresponding to the anomalous spike in Figure 8. In the word learning experiment, participants were told that a word applies to objects 7, 8 and 9 in Figure 9. The formal concept corresponding to the positive examples was the presence of a star shape. After showing the positive examples the participants were asked whether the word applies to object 6 in Figure 9. Sixty-four percent of participants preferred to generalize the word to object 6 even though its average Jaccard similarity coefficient was only 0.2 (See Figure 8).

One reason for this could be the fact that the star shape present in object 6 and in the positive examples is visually more salient. This is because the star feature has more edges and corner points compared to the other boolean features. The Jaccard similarity coefficient in (3) does not take into account the visual saliency of any of the boolean features. The Tversky's ratio model in (1) also does not allow individual boolean attributes to take on different weights depending on its visual saliency. For this reason, we used a different similarity measure for our second word learning experiment in Section V.

V. Second Word Learning Experiment

Our first word learning experiment in Section IV used only a few boolean features. Due to this, we could not have test examples that were very similar to the positive examples,

Copyright (c) IARIA, 2014.
ISBN: 978-1-61208-340-7

TABLE II. THE FOUR TYPES OF TEST EXAMPLES FOR EACH STIMULUS

	Similar to positive examples?	Satisfies the formal concept?
Type 1	Yes	Yes
Type 2	Yes	No
Type 3	No	Yes
Type 4	No	No

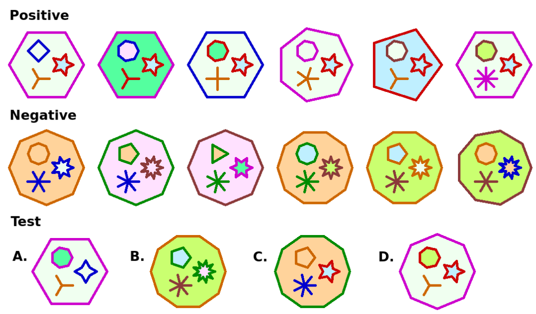

Fig. 10. The figure shows a sample stimulus. The top row, middle row, and the bottom row list the positive, negative, and test examples respectively. All the positive examples have a five pointed blue star with a red border.

and yet did not satisfy the formal concept. In our second word learning experiment we increase the number of boolean features, so that we can have four distinct types of test examples as shown in Table II. We also increase the number of positive examples and introduce negative examples of a word. We wanted to investigate whether increasing the number of positive examples would help the participants infer the defining features, and generalize based on it.

The second word learning experiment was conducted as follows. Each participant was shown six positive and six negative examples of a word. The participant was then asked whether the word generalizes to a test example. We used a random four letter word having consonants and vowels in alternating positions. Just like the previous experiment, the positive examples always corresponded to a formal concept. We had 24 participants in our experiment.

Figure 10 shows a sample stimulus that was used. The top row lists the positive examples of a word. The middle row lists the negative examples. The bottom row contain the test examples for which we want to study the generalization behaviour. The test examples are shown to the participant one at a time.

In Figure 10, all the positive examples contain a five pointed star with a blue background and a red border. This is the only feature that is common across all the positive examples. If the generalization behaviour is based on formal concepts then we would expect that the word will be generalized to both test examples C and D with equal frequency. This is because both C and D have a five pointed star with a blue background and a red border. On the other hand, if the generalization behaviour is based on similarity then we would expect that the word is generalized to A and D more frequently compared to B and C. This is because A and D are more similar to the positive examples of the word.

Figure 11 shows the positive, negative and test examples for another stimulus. The positive examples in Figure 11 contain

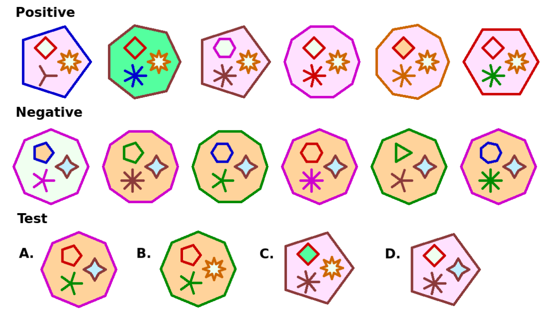

Fig. 11. The figure shows another sample stimulus. The top row, middle row, and the bottom row list the positive, negative, and test examples respectively. All the positive examples have an eight pointed star with a white background and orange border.

> The experiment is designed to study the generalization behaviour in human word learning task. You will be shown three sets of figures on the screen. First set will contain six positive examples of a novel word. The second set will contain six negative examples of the same novel word. The third set will contain a single figure for which you need to decide whether the word applies or not. You will have 12 seconds to answer each question. Please be as fast and as accurate as you can be.

Fig. 12. Instruction for the second word learning experiment.

an eight pointed star with a white background and an orange border. This is the only feature that is common to all the positive examples. If the word generalization is based on formal concepts then we would expect the word to be generalized to examples B and C with the same frequency. On the other hand, if the generalization behaviour is based on similarity then we would expect the word to be generalized to examples C and D more often.

There were six stimuli as shown in Figure 10 and Figure 11. Each stimulus had four test examples, and hence there were 24 test examples in total. Table II shows the four types of test examples that were used in each stimulus. There were 24 test examples in total. The participants were given five trial questions to familiarize them with the task. Figure 12 shows the instruction that was given to the participants.

A. Similarity Measure Used

We need a notion of similarity for this experiment also. As discussed in Section IV-C, the Jaccard similarity coefficient does not take into account the visual saliency of any of the boolean features. To solve this problem, we use a linear regression model. A linear regression model can learn the appropriate weights that needs to be assigned to each of the boolean features to account for its visual saliency.

In order to train the linear regression model, we conducted the following experiment. We asked a different set of 20 participants to rate the similarity between two figures that were randomly generated by our Python program. The figures were the same as those used as objects for the second word learning experiment. Each participant was asked to provide similarity

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-340-7

measures for 40 pairs of randomly generated figures. This gave us 800 (20×40) data-points to train our linear regression model.

The Spearman rank correlation coefficient between the user similarity ratings and those predicted by the trained linear regression model was found to be .70. The trained linear regression model was used to obtain the similarity measures for our second word learning experiment.

B. Result

If the similarity based generalization dominates the formal concept based generalization then we would expect a word to generalize to type 2 test examples more often than type 3 test examples (See Table II). The participants generalized a word to a type 2 test example 75% (108 of 144) of the trials but this percentage was only 37% (53 of 144) for the type 3 test examples. The difference between the type 2 and type 3 generalization frequencies for each participant was found to be statistically significant using the Wilcoxon signed ranks test ($W(21) = 24, p < .001$). Here $df = 21$ because for 3 participants the difference between the frequencies was zero. The statistical significance of the difference in frequencies shows that the generalization behaviour is not based on formal concepts.

Figure 13 shows the 24 test examples divided into two groups—those that satisfy the formal concept and those that do not. We find the average percentage of participants who generalize a word to the test examples in each of these two groups. If the generalization behaviour is based on formal concept, then we would expect the average percentage to be closer to 100% for one group and closer to zero for the other. Figure 13 shows the average percentage of participants who generalized a word to the test examples in each group. We see that for both the groups the average percentage is closer to 50%. Spearman rank correlation coefficient between the generalization made by the participants and the generalization based on definition was found not to be statistically significant ($r(574) = .15, ns$). Here $df = 574$ because we have 24 participants and 24 trial questions ($N = 24 \times 24 = 576$).

Figure 14 shows the 24 test examples divided into two groups based on their similarity to the positive examples of a word. If the participant generalization behaviour is based on similarity, then, we would expect this percentage to be closer to 100% for one group and closer to zero for the other. The data in Figure 14 confirms this. Spearman rank correlation coefficient between the generalization made by the participants and the generalization based on similarity was found to be statistically significant $r(574) = .51, p < .001$). Here $df = 574$ because we have 24 participants and 24 trial questions ($N = 24 \times 24 = 576$).

Figure 15 shows how the percentage of participants who generalized a word to a test example varies with the average similarity between the test example and the positive examples. In the figure, we see that the percentage increases with the average similarity. The Spearman rank correlation coefficient between the two variables was found to be significant ($r(22) = .87, p < .005$). Here $df = 22$ because there are 24 test examples.

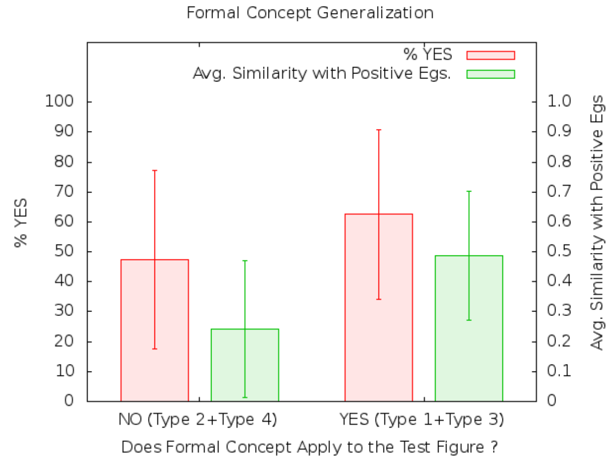

Fig. 13. The 24 test examples are divided into two groups—those that satisfy the formal concept and those that do not. There is no significant difference between the two groups when it comes to the average percentage of participants, who generalized the word to the test example.

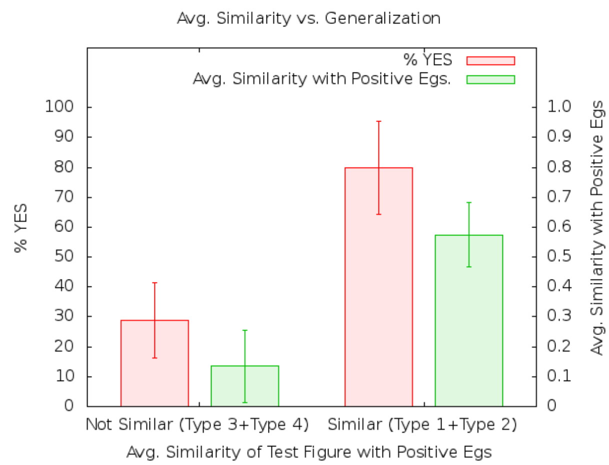

Fig. 14. The 24 test examples are divided into two groups based on their average similarity with the positive examples. The figure shows that the more similar group has a significantly greater percentage of word generalization.

The above results show that the generalization behaviour for the second word learning experiment is better modelled using the similarity based generalization compared to a formal concept based generalization.

VI. Discussion

Our results show that the word generalization behaviour is similarity based even when the word category has a set of defining features. This result is consistent with earlier results [2][3][4] that show that the word generalization in fast-mapping is similarity based. Our word learning experiments are different from previous works because we ensure that a word in our experiment always corresponds to a formal concept.

Tenenbaum et al. [12] discuss the importance of abstract knowledge in helping humans do fast learning. The representation of this abstract knowledge is domain specific, and varies

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-340-7

Fig. 15. The figure shows how the percentage of generalization for at test example varies with its average similarity with the positive examples. The correlation was found to be significant ($p < .005$).

widely from a tree structure to a directed acyclic graph [12]. In our work, we have tried to explore the nature of this knowledge representation for human word learning. Human word learning is usually modelled using Bayesian inferencing on a structured hypothesis space [3][4]. Selecting the right knowledge representation is important because it implies qualitatively different set of hypotheses, for a probabilistic model to choose from [13].

Laurence and Margolis [14] review the various, major theories on concept formation. Formal concepts are more like the classical theory of concepts, while the similarity-based generalization conforms more to the prototype theory [15]. Laurence and Margolis [14] also discuss other theories of concepts that try to combine the classical and prototype theory. M. Freund [16] proposes a formal model that combines the notion of *typicality* from the prototype theory with the formal concept analysis.

We have used abstract figures as examples in our word learning experiments. This ensures that the generalization behaviour is not influenced by the background knowledge that the participants might have about the examples used in our experiments. The features present in the examples can only be visual features; therefore, it becomes easier to ensure that the positive examples correspond to a formal concept.

We have changed several independent variables in our two word learning experiments. These include the stimuli, number of boolean features, number of positive examples, and the presence of negative examples. Despite changing several independent variables, we found that the generalization behaviour (dependent variable) was based on similarity, and not on defining features. We speculate that the reason for this is that abstracting out a definition from a set of positive examples puts a greater cognitive load on the participants, compared to judging the similarity between a test example and the positive examples.

VII. Conclusion and Future Work

We have conducted two word learning experiments that show that the generalization behaviour during fast mapping is based on similarity, and not on a set of defining features. This is true even when enough positive examples exist from which the defining features can be inferred.

The two experiments used different sets of stimuli and had different sets of participants. We increased the number of boolean features, the number of positive examples and introduced negative examples, for our second word learning experiment. Despite changing several independent variables, our results consistently show that the generalization behaviour is not based on formal concepts.

Currently, the knowledge representations used in the literature are domain specific, and not stimuli specific. In our future work, we want to investigate whether the same word concept representation can be used to model word learning across different stimuli conditions.

References

[1] S. Carey, “The child as word learner,” in *Linguistic theory and psychological reality*, M. Halle, J. Bresnan, and G. A. Miller, Eds. Cambridge, MA: MIT Press, 1978, pp. 264–293.

[2] J. Tenenbaum and F. Xu, “Word learning as bayesian inference,” in *Proceedings of the 22nd annual conference of the cognitive science society*, 2000, pp. 517–522.

[3] F. Xu and J. Tenenbaum, “Word learning as bayesian inference.” *Psychological review*, vol. 114, no. 2, pp. 245–272, 2007.

[4] J. Abbott, J. Austerweil, and T. Griffiths, “Constructing a hypothesis space from the web for large-scale bayesian word learning,” in *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, 2012, pp. 54–59.

[5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” in *CVPR09*, 2009.

[6] G. A. Miller, “Wordnet: a lexical database for english,” *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[7] B. Ganter and R. Wille, *Formal concept analysis*. Springer Berlin, 1999.

[8] M. Krötzsch and B. Ganter, *A Brief Introduction to Formal Concept Analysis*. Chapman & Hall/CRC, 2009, vol. Conceptual Structures in Practice, ch. 1, pp. 3–16.

[9] U. Priss, “Formal concept analysis in information science,” *Annual Review of Information Science and Technology*, vol. 40, no. 1, pp. 521–543, 2006.

[10] A. Tversky, “Features of similarity,” *Psychological review*, vol. 84, no. 4, pp. 327–352, 1977.

[11] J. Greene and M. d’Oliveira, *Learning to use statistical tests in psychology*. McGraw-Hill International, 2005.

[12] J. B. Tenenbaum, C. Kemp, T. L. Griffiths, and N. D. Goodman, “How to grow a mind: Statistics, structure, and abstraction,” *science*, vol. 331, no. 6022, pp. 1279–1285, 2011.

[13] T. L. Griffiths, N. Chater, C. Kemp, A. Perfors, and J. B. Tenenbaum, “Probabilistic models of cognition: exploring representations and inductive biases,” *Trends in Cognitive Sciences*, vol. 14, no. 8, pp. 357–364, 2010.

[14] S. Laurence and E. Margolis, “Concepts and cognitive science,” *Concepts: core readings*, pp. 3–81, 1999.

[15] E. Rosch and C. Mervis, “Family resemblances: Studies in the internal structure of categories* 1,” *Cognitive psychology*, vol. 7, no. 4, pp. 573–605, 1975.

[16] M. Freund, “On the notion of concept i,” *Artificial Intelligence*, vol. 172, no. 4-5, pp. 570–590, 2008.

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-340-7