
Characterizing the Features Encoded During Multiple Category Acquisition

Michael Fink*

Center for Neural Computation
The Hebrew University of Jerusalem
Jerusalem 91904, Israel
fink@huji.ac.il

Shimon Ullman

Faculty of Mathematics and Computer Science
Weizmann Institute of Science
Rehovot 76100, Israel
shimon.ullman@weizmann.ac.il

Abstract

Understanding how features are encoded during category acquisition is a fundamental challenge in human learning research. The current work proposes that in order to maintain accurate generalization, features that are informative for multiple categories must be actively preferred. The main contribution is in providing a novel methodology for empirically testing this hypothesis. It is shown that while acquiring a sequence of new categories, features that are informative for recognizing several categories are preferably encoded. Moreover, evidence is provided that after acquiring novel features, representations of categories learned in the past are actively reconstructed so that they comply with the newly encoded features. Finally, it is demonstrated that encoded features are efficiently utilized in facilitating future category acquisition. These results are observed using both perceptual and semantic stimuli. It is therefore suggested that preferring features that provide maximum information on multiple categories might be a global characteristic of human categorization systems.

1 Introduction

The dynamic environment in which humans live requires fast and accurate recognition of numerous categories. These requirements must often be met after only few examples of a novel category have been encountered. Yet humans operate quite well under these conditions, recognizing a substantial number of categories [4] with unparalleled speed and accuracy [13]. It remains a puzzle how the hindering effects of overfitting may be avoided when learning from a small sample. This paper follows an adaptive feature creation approach [10], postulating that efficient generalization might only be obtained by encoding features that are informative for recognizing many categories. The fact that categories tend to share common structures justifies why the proposed criterion enables knowledge to be transferred between known and novel categories. Thus for example, an expert Zoologist can quickly learn to classify unfamiliar species of mammals (Aardvarks or Armadillos), using features encoded through years of experience (e.g. skeletal pose). Two theoretical frameworks have suggested computational criteria as to when a feature set should be dynamically

* visit www.cs.huji.ac.il/~fink for an extended version of this paper

extended. The first theoretical framework emphasizes the importance of features that encode suspicious coincidences between input patterns (say x_i wings and x_j feathers) [2]. Formally, this theory encodes events where $\frac{p(x_i, x_j)}{p(x_i)p(x_j)} \gg 1$ (jointly encoding *feathers* and *wings*). It is by now a well established fact that features sensitive to input statistics might be created in the absence of any categorization feedback [9, 5]. In contrast to this *statistical learning* approach, the second theoretical framework stipulates that the encoded features are those which are maximally discriminative in the context of the given categorization task. Discriminative power is often quantified using mutual information [15]. Formally, let \mathcal{F} be the set of candidate features, and C be a class variable, the encoded feature is then:

$$F^* = \operatorname{argmax}_{F \in \mathcal{F}} \sum_{F, C} p(F, C) \log \frac{p(F, C)}{p(F)p(C)} \quad (1)$$

Several empirical findings are related to this theoretical framework [6, 11, 12, 1, 7]. Common to these different lines of research is the fact that the contribution of features is estimated with respect to a binary class variable (summing (1) over two states of C). As mentioned above, categorization systems must be capable of recognizing many thousands of categories. It therefore seems that examining the setting where feature creation is induced by an individual categorization task does not capture the essence of the generalization constraints in which categorization systems must operate. This research puts forth the hypothesis that in order to achieve efficient generalization in large categorization systems the underlying organizing principle must be the preferential encoding of features informative for multiple categories (Zoologists observing that a brushlike tail is informative in characterizing various animals like Brush-Tail Possum, Brush-Tail Wallaby, Brush-Tail Bettong etc.). Thus, novel features might be encoded in the absence of any suspicious coincidences in the input statistics and despite the fact that they do not maximize the information provided for any *individual* category. In summary, the proposed criterion is that encoded features should maximize information to all the categories collectively, summing C in (1) over all possible category assignments. It should be noted that similar distinctions appear in the machine learning literature. Coincidence detection manifests an unsupervised learning framework while maximizing distinctive value is a classical supervised learning task. The proposed criterion of measuring distinctive value for multiple categorization tasks collectively is analogous to the machine learning transition from binary to multi-label classification. Finally, the common goal of accurate generalization on many classes from only few examples is shared with the learning to learn approach to machine learning [14].

Biologically based systems could not be expected to accurately evaluate probability and information quantities, however the proposed computational criterion formally characterizes the fundamental factors governing feature creation and how they might derive novel predictions that differ from the previously proposed criteria. Thus, if indeed a categorization system is tuned to encode a common set of informative features the following three characteristics should be observed:

1. Features informative for recognizing many categories are preferentially encoded
2. Informative features can reconstruct former category representations
3. Novel features can facilitate acquisition of future categories

Four experiments were designed to test these predictions. Experiment I validated that features informative for many categories are preferentially encoded and can reconstruct former category representations. Experiment II demonstrated that features encoded in the past can facilitate acquisition of future categories. Both experiments utilized controlled sets of semantic stimuli (job candidates' descriptions). In order to test the three predictions in lower levels of the perceptual-conceptual continuum [8], Experiments III and IV replicated the first two experiments utilizing perceptual stimuli (configurations of colored cubes).

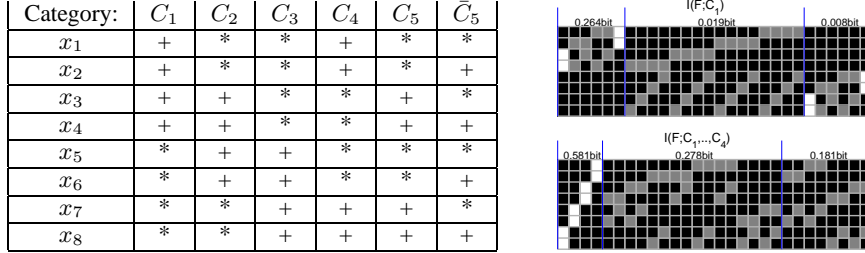


Figure 1: **Left:** Definitions of categories C_1, \dots, C_4 learned in Experiment I, and of categories C_5 and \bar{C}_5 learned in Experiment II. Input elements indicated by + must be in an *on* state for the category to be present, while * denotes the category’s indifference to certain input elements. **Right:** Each column represents a 2nd order conjunction feature by highlighting the two relevant input elements. Features are grouped by descending information content on category C_1 (top pane) and by information content on the four categories collectively (bottom pane). Although the *pair-features* (emphasized in white) do not provide the maximal information for category C_1 individually (providing 0.264 or 0.008 bits) they all appear as the maximally informative for the four categories collectively (providing 0.581 bits).

2 Exp. I: Semantic Features Informative for Multiple Categories

Experiment I was designed to examine whether features that are informative for recognizing multiple categories are preferentially encoded; and whether this process might actively reconstruct the representations of former categories acquired in the past.

Method: The experimental setting was based on a categorization task consisting of eight binary input elements $x_{i,i=1,\dots,8}$, jointly notated as \mathbf{x} . Four target categories $C_{n,n=1,\dots,4}$ were defined as a function of the input vector set $\{\mathbf{x}\}$. Each category was fully characterized by four specific input elements being in an *on* position (Fig. 1:Left). For example, $\mathbf{x} = [- - + + + - +]$ is a C_2 exemplar. It will now be shown why this specific category structure can validate the research predictions. Although many possible intermediate representations might help solve the categorization task, the subset including all 2nd order conjunctions of input elements (including $\binom{8}{2} = 28$ features) was selected as the focus of the proposed analysis. The role of features defined over higher order conjunctions is not ruled out, however, the goal of the proposed setting is to contrast the emergent representations of features with comparable complexity. Mutual information was evaluated between each of the 28 candidate features and the individual categories in addition to evaluating the information for the four categories collectively. As shown in Fig. 1:Left, categories C_1 and C_2 required two common input elements to be in an *on* state (x_3^+ and x_4^+). This common *pair-feature*, was termed v_2 . In fact, each of the four categories shares a common *pair-feature* with two other categories. These *pair-features* are formally defined as $v_1 \equiv x_1^+ \cap x_2^+$, $v_2 \equiv x_3^+ \cap x_4^+$, $v_3 \equiv x_5^+ \cap x_6^+$ and $v_4 \equiv x_7^+ \cap x_8^+$. As depicted in Fig. 1:Right, although the *pair-features* are not more informative than other pairs for category C_1 individually (providing 0.264 or 0.008 bits), they appear as the maximally informative features for the four categories collectively (each providing 0.581 bits). Thus, the first research prediction would be manifested if a salient representation of the *pair-features*: v_1, v_2, v_3 and v_4 emerges while acquiring the four categories. As described in the introduction section the proposed setting must also control the two alternative computational criteria that induce feature creation. This goal was achieved by finding an alternative representation that is comparable in all aspects to the *pair-feature* structure but is not as informative for the categorization tasks collectively. Such an alternative representation is generated by arbitrarily segmenting the eight input elements into four *incongruent-pairs*, that appear each in just one category i.e. $h_1 \equiv x_1^+ \cap x_3^+$, $h_2 \equiv x_4^+ \cap x_6^+$, $h_3 \equiv x_5^+ \cap x_7^+$ and $h_4 \equiv x_2^+ \cap x_8^+$.

		<i>on</i>	<i>off</i>
x_1	Languages	Spanish	French
x_2	Marital Status	Married	Single
x_3	College	Private	Public
x_4	Age	Thirties	Twenties
x_5	Gender	Male	Female
x_6	Profession	Lawyer	Accountant
x_7	Residency	New York	New Jersey
x_8	Department	Local	International

Figure 2: The eight binary semantic characteristics and the experimental setup.

Materials: The categories defined in Fig. 1 were implemented in a job assignment task, requiring participants to sort applicants to four business firms. Eight binary characteristics, encoded the input description of each candidate (Fig. 2). Thus, each of the four business firms required four specific characteristics to be in an *on* state. For example, Firm 3 required all candidates to be male, lawyers, living in New York and preferring the local department. Beneath the applicant sheet, an array of five target buttons was displayed. Each of the four peripheral buttons was associated with one of the four firms. The central button was reserved for applicants not qualified for any firm (see setting in Fig. 2). In order to control feature saliency the position of each field was randomly permuted in each trial. The intersections of the categories’ relevant characteristics defined the set of *pair-features*: $v_1 \equiv \text{Spanish and Married}$, $v_2 \equiv \text{Private-college and Thirties}$, $v_3 \equiv \text{Male and Lawyer}$, $v_4 \equiv \text{New-York and Local-department}$.

Procedure: Experiment I was composed of four training stages. At each stage, participants learned one additional category in a trial-and-error paradigm. In every trial a random subset of the eight input elements \mathbf{x} was activated. The resulting eight dimensional candidate description was displayed until the participants activated the category buttons. If the wrong category was indicated an error tone was triggered. In each stage, trials were generated until reaching a criterion of 100 *consecutive* successes. When the four training stages were concluded, a test was employed to validate whether the hypothesized *pair-features* have emerged. In this testing stage, each of the category buttons was highlighted in a sequential manner while requiring the participants to verbally report which characteristics were necessary for the associated firm. The proposed test assumes that if an internal representation of the *pair-features* had undergone a process of unitization [7], the verbal reports should be composed of two *pair-features*. This reporting pattern should not be exhibited if a representation based on any of the *incongruent-pairs* had emerged.

Results: Twelve participants completed the four training stages in 4-6 hours, requiring approximately 1000 trials. The verbal reports provided in the testing stage were then coded by observing whether the first two characteristics composed a *pair-feature* or an *incongruent-pair*. It was observed that the frequency of reporting *pair-features* was significantly higher than that of a comparable pattern of *incongruent-pairs* (binomial test, $p \leq 0.05$, $n=12$) in all of the four firm reports (Fig. 3). In addition to registering the report sequence, the participants of Experiment I were recorded while verbally reporting the requirements of each firm. The reports of each firm were manually annotated by marking the starting time and the ending time of the four characteristics. An annotated recording of category C_3 is presented in Fig. 3. By performing this annotation it becomes possible to measure the duration of the three gaps between the four reported characteristics and to assess whether the input elements composing each *pair-feature* are indeed temporally fused. The three gaps were scored according to the ascending order of their duration (the shortest gap was scored as 1, the intermediate gap was scored as 2 and the longest gap was scored as 3). It was observed that the second gap score ($M = 2.86$, $SD = 0.38$) was significantly larger ($t(6) =$

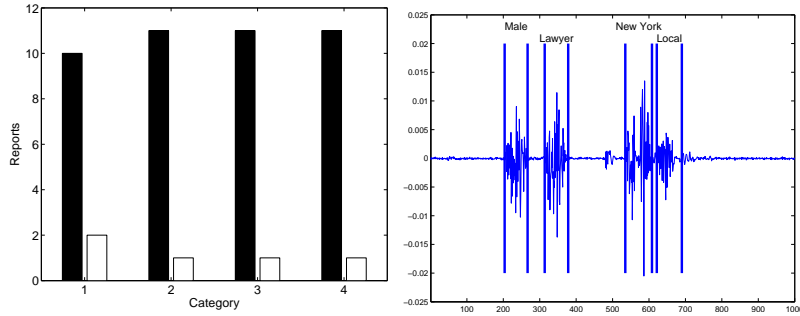


Figure 3: **Left:** Number of reports congruent/incongruent (black/white) with the *pair-feature* structure for categories C_1 through C_4 . **Right:** An annotated 10-second recording (time vs. amplitude) of a participant reporting category C_3 . The *pair-features* seem to be temporally fused: the shortest gap appearing between NY and Local, the second shortest gap between Male and Lawyer while the longest gap appears between Lawyer and NY.

4.86, $p \leq 0.05$, one tailed) than the first gap score ($M = 1.83$, $SD = 0.69$) and significantly larger ($t(6) = 10.49$, $p \leq 0.05$, one tailed) than the third gap score as well ($M = 1.43$, $SD = 0.53$). Similarly, the second gap scores were significantly larger than the first and third gaps in the reports of categories C_2 through C_4 . Thus, the pattern of results provides evidence that in the semantic system, features are preferably encoded if they provide information for multiple categories. However, Experiment I also addresses the second prediction, stating that encoding novel features might lead to a reconstruction of former category representations. This claim is based on observing the evolved representation of category C_1 . When participants pass the first training stage, they are equipped with a representation that enables perfect recognition of all category C_1 exemplars (due to the highly stringent training criterion of 100 consecutive successes). At this point the *pair-features* have no salient representation because they are by definition identical to the *incongruent-pairs* until learning at least one additional category. In fact when testing an additional group of participants that only learned category C_1 , no saliency of the *pair-features* was observed. However, when analyzing the reporting results of category C_1 , registered after concluding the learning procedure of all four categories, it is evident that the initial representation of category C_1 had been actively reconstructed. This reconstruction maintained that the representation of category C_1 complies with the *pair-features* acquired only later in the training process. Thus, the first two research predictions have been addressed while controlling the alternative factors that might have explained the emergent *pair-feature* structure.

3 Exp. II: Facilitated Semantic Category Acquisition

Experiment II examined whether the *pair-features* facilitate learning of future categories.

Method: Experiment II included an additional training stage to the four training stages of Experiment I. In this additional stage examples of a new category, C_5 , were presented. This new category was defined as $C_5 \equiv v_2^+ \cap v_4^+$ (see Fig. 1). Unlike the training procedure of Categories C_1 through C_4 , that continued until a criterion of perfect categorization performance had been reached, the fifth stage was restricted to a prefixed number of trials. It was predicted that when training a fifth category that is congruent with the *pair-feature* structure, a significant facilitation would be observed. Comparing the performance of participants that have already learned the four initial categories to the performance of naive participants is meaningless, due to history confounds. Therefore, Experiment II maintained a between subject paradigm by providing a control group that shares the same history as the

experimental group. This control group learned a different fifth category, $\bar{C}_5 \equiv h_2^+ \cap h_4^+$, that is equally complex as C_5 yet incongruent with the *pair-feature* structure.

Materials and Procedure: The stimuli and setting of Experiment II were similar to those described in Experiment I. Participants were first required to complete training stages 1 through 4 as in Experiment I. Next, both groups learned the requirements of a fifth firm using a limited set of 48 training trials. The stimuli presentation procedure of the fifth stage was identical to the first four training stages except that trials were limited to a prefixed duration of sixteen seconds. Providing 48 training trials was aimed at capturing an intermediate snapshot of the training process where the differential training rate of the experimental and control groups might be manifested. Rather than sampling 48 trials as in the first (unbounded) four training stages, the fifth stage displayed, in a random order, a predefined set of 24 negative examples and 24 positive examples of either the congruent category C_5 or the incongruent category \bar{C}_5 . Following the 48 training trials, participants were first tested on their capability to correctly categorize the 48 training examples, by repeating the presentation procedure in the absence of any feedback signal. Only then did participants verbally report the candidate characteristics composing the new firm’s requirements.

Results The twelve participants of both the experimental and the control groups reported that the training procedure of the fifth category was difficult. This effect is probably due to the limited time provided for the 48 training trials. The participants’ performance on the 48 testing trials (presented without feedback), was scored by averaging the proportion of the correctly classified examples of the fifth category with the proportion of the correctly classified remaining fillers so that chance level is 0.50. It was observed that while the control group performed slightly above chance level ($M = 0.59$, $SD = 0.11$) the accuracy of the experimental group ($M = 0.81$, $SD = 0.13$) was substantially higher ($t(10) = 2.38$, $p \leq 0.05$, one tailed). The reporting results were scored by subtracting the number of any incorrectly reported characteristics from the number of all correctly reported characteristics. Here too, a significant difference ($t(10) = 4.65$, $p \leq 0.05$, one tailed) was observed between the experimental group scores ($M = 2.33$, $SD = 1.51$) and the control group scores ($M = -0.33$, $SD = 1.21$). It could therefore be concluded that the emerged *pair-feature* representation was a functional tool in facilitating the acquisition of a novel semantic category.

4 Exp. III: Perceptual Features Informative for Multiple Categories

The goal of Experiment III was to examine whether the results of Experiment I were specific to the semantic domain or whether evidence for encoding features that are informative for recognizing many categories might also be observed in a perceptual classification task.

Method: Due to the fact that Experiment III was aimed at replicating the structure of Experiment I, the applied method was similar to that described above.

Materials: To test the first research prediction the input vector set, $\{\mathbf{x}\}$, was implemented into a visual categorization task. The eight-dimensional binary inputs were translated into images composed of eight *color cubes*. For each cube, one color was selected to function as the *on* state and another color as the *off* state (Fig. 4). In the *color cube* implementation, each category required four specific neighboring cubes to be in an *on* position. Categories based on neighboring cube configurations were chosen, because they are easier to acquire. Exemplars of each category were generated by using color combinations of the remaining four non-relevant cubes. The intersections of the categories’ relevant cubes defined the set of *pair-features*: $v_1 \equiv \text{Red-Blue}$, $v_2 \equiv \text{Brown-Purple}$, $v_3 \equiv \text{Black-Orange}$ and $v_4 \equiv \text{Yellow-Green}$. Since the stimuli probabilities were identical to those described in Experiment I, the mutual information measurements from Fig. 1 remain valid, leaving the *pair-features* as the maximally informative representation.

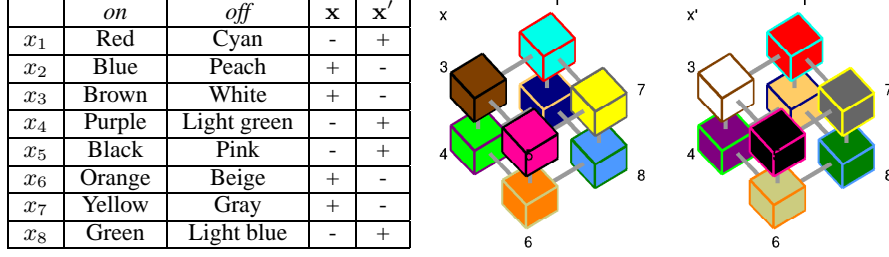


Figure 4: An example of two stimuli \mathbf{x} and \mathbf{x}' composed of eight binary *color cubes*.

Procedure: The training procedure of Experiment III was similar to that described in Experiment I. When the four training stages were concluded, participant were required to verbally report the *color-cubes* relevant for each of the four categories.

Results: Twenty participants completed the four training stages in 2-3 hours. The relatively short training duration resulted from the fact that processing the perceptual stimuli required approximately half the time than in the analogues stages of Experiment I. All twenty participants succeeded in reporting the four colors relevant to each category. The frequency of reporting according to the *pair-feature* pattern was observed to be significantly higher than that of a comparable *incongruent pair* pattern in all four categories (binomial test, $p \leq 0.05$, $n=20$). Thus, Experiment III validates the first research predication. As in Experiment I, the fact that the report of category C_1 reflects the *pair-feature* structure, validates the second prediction regarding the reconstruction of former category representations.

5 Exp. IV: Facilitated Perceptual Category Acquisition

Experiment IV examined whether the perceptual system may utilize the informative feature set to facilitate learning of additional future categories.

Method: Experiment IV replicated Experiment II, utilizing the perceptual stimuli described in Experiment III. Thus, the fifth training stage required learning a novel perceptual category $C_5 \equiv v_2^+ \cap v_4^+$ from just a few training examples. The facilitation of acquiring category C_5 was assessed in comparison to a control fifth category, defined as $\bar{C}_5 \equiv h_2^+ \cap h_4^+$.

Materials and Procedure: The stimuli and setting of Experiment IV were similar to those described in Experiment III. Participants were first required to complete training stages 1 through 4. Then, both the experimental and the control groups learned a fifth category using a limited set of 48 training images. The stimuli presentation process of this stage was similar to that described in the first four training stages except for the fact that the trial duration was fixed to six seconds. Following the 48 training trials, participants were required to verbally report the *color cubes* composing the new category.

Results: As in Experiment II, the ten participant of both the experimental and the control groups reported that the training procedure of the fifth category was difficult. It was observed that the learning rate was significantly higher in the congruent condition (Fisher Exact Probability Test, $p \leq 0.05$, $n=10$). Members of the experimental group reported on average 2.2 correct *color cubes*, i.e. participants learned most of the new category's characteristics. Members of the control group, reported on average only 0.8 out of the four *color cubes* present in category \bar{C}_5 . Thus, the emerged *pair-feature* representation was indeed a functional tool in facilitating the acquisition of a novel perceptual category.

6 Summary and Discussion

The proposed experimental setting required learning four categories, each based on a conjunction of four input elements. Pairs of input elements, termed *pair-features*, were suggested as the preferred internal representation due to their information content for the target categories collectively. Although no direct feedback was provided for the *pair-feature* structure, it was experimentally found that participants' reporting patterns corresponded to an internal structure based on these pairs. The existence of the *pair-features* cannot be attributed to their perceptual salience or frequency of appearance, since these factors were carefully controlled. It was also observed that the *pair-feature* structure actively re-constructed the previously acquired representation of the first category. Finally, it was demonstrated that the emergent *pair-features* can significantly facilitate learning of future categories. The three experimental predictions have been validated using both semantic and perceptual stimuli, thus suggesting, that preferring features that are distinctive for multiple categories might be a general factor guiding feature creation across the perceptual-conceptual continuum. Several questions regarding the generalization of the proposed theory might be raised. The learning task stripped the category acquisition and feature creation processes to the bare minimum, two conjunctions of binary inputs deterministically defined each category. The effects of more complex settings, like continuous input dimensions, complex features and other (more natural) learning schemes is the goal of future research. This research suggests that categorization systems must actively encode features that are informative for many categories. A similar observation has been recently employed by researchers in the computer vision community for deriving algorithms aimed at learning many categories from few examples [3]. Thus, the algorithmic approaches and the research of human categorization capacities seem to be mutually inspired.

References

- [1] A. Archambault, C. O'Donnell, and P. G. Schyns. Blind to object changes: When learning the same object at different levels of categorization modifies its perception. *Psychological Science*, 10(3):249–255, 1999.
- [2] H. B. Barlow. Unsupervised learning. *Neural Computation*, 1:295–311, 1989.
- [3] E. Bart and S. Ullman. Learning novel classes from a single example using cross-generalization. *Proceedings of the Computer Vision and Pattern Recognition conference (CVPR) 2005*, 2005.
- [4] I. Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94:115–147, 1987.
- [5] J. Fiser and R. N. Aslin. Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12(6):499–504, 2002.
- [6] R. L. Goldstone. Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, 123(2):178–200, 1994.
- [7] R. L. Goldstone. Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1):86–112, 2000.
- [8] R. L. Goldstone and L. W. Barsalou. Reuniting perception and cognition. *Cognition*, 65, 1998.
- [9] O. Rosenthal, S. Fusi, and S. Hochstein. Forming classes by stimulus frequency: Behavior and theory. *Proceedings of the National Academy of Science*, 98:4265–4270, 2001.
- [10] P. G. Schyns, R. L. Goldstone, and J. Thibaut. Development of features in object concepts. *Behavioral and Brain Sciences*, 21:1–54, 1998.
- [11] P. G. Schyns and G. L. Murphy. The ontogeny of part representation in object concepts. *The Psychology of Learning and Motivation*, 31:305–349, 1994.
- [12] P. G. Schyns and L. Rodet. Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 23(3):681–696, 1997.
- [13] S. Thorpe, D. Fize, and C. Marlot. Speed of processing in the human visual system. *Nature*, 381(6582):520–522, 1996.
- [14] Sebastian Thrun. *Learning To Learn: Introduction*. Kluwer Academic Publishers, 1996.
- [15] S. Ullman, M. Vidal-Naquet, and E. Sali. Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5(7):682–687, 2002.