Perceptual learning of parametric face categories leads to the integration of high-level class-based information but not to high-level pop-out

Tim C. Kietzmann
Neurobiopsychology Laboratory, Institute of Cognitive Science, University of Osnabrück, Germany

Peter König
Neurobiopsychology Laboratory, Institute of Cognitive Science, University of Osnabrück, Germany

To date, the relative contribution of the different levels of the visual hierarchy during perceptual decisions remains unclear. Typical models of visual processing, with the reverse hierarchy theory (RHT) as a prominent example, strongly emphasize the role of higher levels and interpret lower levels as sequence of simple feature detectors. Here, we investigate this issue based on two analyses. Using a novel combination of perceptual learning based on two classes of parametric faces and a subsequent odd-one-out paradigm, we first test a vital prediction of RHT: high-level pop-out. With this experimental approach, we overcome the low-level confounds of previous studies while still introducing distinct high-level representations. Contrary to previous findings, our analyses show that there is no high-level pop-out, despite very early, near-perfect classification accuracy and extensive training of our subjects. Second, we explore the underlying form of category representation during subsequent stages of perceptual training. This is accomplished by including class-external and class-internal target–distractor combinations. Whereas the subjects’ responses during the first sessions are best explained instance-based and dependent on low-level metric differences, later patterns exhibit the inclusion of high-level, class-based information that is independent of target-stimulus similarity. Finally, we show that the utilized level of information is highly task-dependent.

Keywords: categorization, learning, object recognition, plasticity, visual cognition, pop-out, category learning, perceptual learning


Introduction

The search for the cortical mechanism underlying conscious object recognition and classification has been a long-standing focus of scientific interest. Although we are far from a clear picture, one aspect has become central to our understanding: the hierarchical nature of the visual system. Starting at low levels of the ventral pathway, which contains cells with relatively simple receptive field properties, representations become increasingly complex in subsequent levels of the hierarchy. Together with this shift in complexity, representations of the ventral pathway become more abstract and thus invariant to smaller changes in scale, viewpoint, and translation (Tanaka, 1996).

Despite the evolving consensus on the hierarchical view of visual processing, it is still an open question how the different levels of the hierarchy contribute to the overall perceptual process. Prominent models of visual processing strongly emphasize the role of high-level representations. That is, lower levels are interpreted as a sequence of feature detectors with the purpose to provide input to the higher levels (Bar, 2004; Serre, Oliva, & Poggio, 2007). The latter are then associated with the actual perception and are expected to contribute to remote processes including planning and action execution. This view is taken to an extreme by the reverse hierarchy theory (RHT; Ahissar & Hochstein, 2004; Hochstein & Ahissar, 2002). It argues that visual processing starts with an implicit pass of information through the lower levels of the hierarchy, leading to a first, broad categorization of the input based on the properties of high-level populations of neurons. Then, if required, the high levels access the information present at lower levels, where the receptive field sizes are smaller and more detailed information can be found. Importantly, both types of visual inference attribute the decisive part of processing to high-level representations and view lower levels as passive feature detectors. This view parallels earlier work ascribing visual awareness to high rather than lower levels of the visual hierarchy (Crick & Koch, 1995). Recently, however, it has been subject to intensive debate as studies using transcranial magnetic stimulation (TMS) closely linked visual awareness and recognition abilities to activity in V1, the area of lowest...

In order to advance our understanding of this issue, we performed two main analyses based on data collected from psychophysical experiments in which we combined perceptual learning with a subsequent odd-one-out task. First, we tested an important prediction of RHT: high-level pop-out. Pop-out is the ability to immediately spot a target stimulus in an array of distractors (spotting the “odd one out”), independently of the actual number of distractors. Contrary to earlier theories, RHT associates pop-out with high-level neuronal properties. This is based on the view that the initial feed-forward sweep pre-attentively activates abstract, translation-invariant representations on high levels, with generalized processing. Based on these properties, the odd-one-out can be immediately spotted. Thus, instead of interpreting pop-out as the source of parallel processing on lower levels (Treisman & Gelade, 1980), RHT explains the effect based on the large, translation-invariant receptive fields on higher levels. As a result, features that pop out are conceived to reflect properties present at high levels, rather than low-level ones. A direct consequence of this high-level interpretation of the phenomenon is the prediction that also high-level, conceptual information should pop out if high-level representations exist that can differentiate between target and distractor (Hershler & Hochstein, 2005; Hochstein & Ahissar, 2002).

High-level pop-out has been intensively investigated using faces as targets in arrays of distractors taken from other basic level object categories (Figure 1). However, despite its conceptual simplicity, the phenomenon is heavily debated (Brown, Huey, & Findlay, 1997; Hershler & Hochstein, 2005, 2006; Purcell, Stewart, & Skov, 1996; VanRullen, 2006). While the existence of the overall effect is generally undisputed (a face seems to pop out of an array of distractors), it is argued that low-level differences can provide a sufficient explanation: although target and distractor belong to different high-level categories, they could additionally differ systematically in features that are explicitly represented on lower levels. These low-level confounds presently prohibit a pure high-level interpretation of the results, as required for RHT.

An unambiguous way of demonstrating high-level pop-out has to meet two conditions. First, the low-level properties of target and distractor need to be controlled, i.e., two categories of stimuli need to be used that are not distinguishable based on simple low-level properties. Second, target and distractor need to belong to different high-level classes, with respective, separate high-level neuronal representations. Recently, Sigala and Logothetis (2002) have illustrated both of these properties for a stimulus set consisting of two subordinate face categories with parametrically varying features. In parameter space, two categories were defined such that only complex conjunctions of feature properties, but no single feature, could be used for categorical decisions. After training macaque monkeys, single cell recordings in inferotemporal cortex (IT) revealed neurons that were specifically responsive to the diagnostic features and therefore discriminated the two arbitrarily defined categories. Using comparable category training, other studies have provided additional support for the existence of high-level representational changes in sensory and prefrontal cortex. Baker, Behrmann, and Olson (2002) report an enhanced selectivity to learned stimuli of neurons in IT, whereas Jiang et al. (2007) attribute the underlying changes to the lateral occipital cortex as well as prefrontal cortex. As RHT predicts high-level pop-out independently of whether the representations are located in IT or PFC (Hochstein & Ahissar, 2002), all of the above studies show that the training paradigm fulfills both requirements: the two classes of parametric stimuli are similar in low-level features (only complex compositions of these features allow for a categorization) but nevertheless lead to different high-level representations.

In addition to the first analysis addressing high-level pop-out, our second analysis investigates the underlying form of category representation. It provides an estimate on how much low-level metric and high-level class-based information, that is, “stimulus similarity” and “class membership,” contribute to the perceptual decision and
is therefore directly related to the question for the relative contribution of the different levels of the visual hierarchy. To be able to distinguish between the two options, we used the same experimental paradigm as in the case of high-level pop-out but extended it with a condition in which target and distractor were taken from the same category instead of different ones. The reasoning was that if the representations were purely instance-based, then the response times should be independent of the class membership of the distractor but dependent on the low-level similarity of the two. However, if a more abstract class membership was used, we should find significant differences between the two conditions. Importantly, since the experimental setup included successive sessions of perceptual learning of the two parametric categories, it is possible to monitor the different modes of category representation during the different stages of training.

Finally, we included a perceptual similarity judgment task and tested the subjects' responses to intermediate stimuli. The recorded similarity data were tested for converging evidence with regard to high-level class effects and utilized as a measure of perceptual similarity in other analyses. The responses to intermediate stimuli are of interest, as they test the results of the pop-out paradigm in a rather different task and can provide further insights into the utilized representation and level of processing that underlie the categorical dissociation.

Based on above considerations, the current study combines several sessions of training of two categories of parametric face stimuli with subsequent tests for high-level pop-out. This allows for a close monitoring of the subjects' categorization- and high-level pop-out performance during the different phases of learning. The training procedure was kept identical to the one described by Sigala and Logothetis (2002), showing one stimulus at a time and providing audio feedback indicating the validity of the classification (see Tasks and procedure section for more details). In the test phase, a target face was presented together with different numbers of distractor faces, the distractors being all identical. The distractor could either be taken from the same or the other category (class-internal and class-external conditions), as required for the analysis of the underlying form of representation.

**Methods**

**Participants**

Four volunteers took part in the experiment and completed 16 sessions each (3 females, 1 male, ages ranging from 23 to 49). All subjects were informed of their right to withdraw from the experiment at any time without the need to state a reason and gave written informed consent to participate. Furthermore, all subjects were informed of the experimental procedure and were naive to the purpose of the study. Upon completion of the overall experiment, the subjects were debriefed.

**Stimuli**

The experiments were based on parameterized line drawings of faces, as also used in previous studies (Sigala, Gabbiani, & Logothetis, 2002; Sigala & Logothetis, 2002).
The faces varied in four dimensions: eye separation, eye height, mouth height, and nose length. As can be seen in Figure 2b, only the combination of two of the dimensions was diagnostic for the category membership. No single low-level feature could be used to distinguish between the two categories. Five stimuli of each category were used during the standard training and test phases. In addition to these ten, eight intermediate stimuli were created, which were used in a later “intermediate stimuli” test (Figure 7a). These stimuli were selected to contain all combinations of close/far from the trained instances and close/far from the abstract decision boundary. By this, we expected to be able to gain further insights into the nature of the representation used for the perceptual decisions.

**Apparatus**

The experiment was conducted on an Apple Mac Pro (4 × 2.66 GHz, 4-GB RAM) running Linux. The distance to the screen was 60 cm. Stimuli were presented on a 19-inch flat screen monitor (Sync Master 971p, Samsung Electronics, Seoul, South Korea) with a native screen resolution of 1280 × 1024 pixels and a refresh rate of 75 Hz. Each face was presented at a width of 2° of visual angle, with the entire display being 35.56° × 28.4°. The subject responses were recorded using the arrow keys and numpad of the keyboard.

**Tasks and procedure**

Each of the 16 experimental sessions contained a training phase and a test phase. In addition, sessions 1 and 16 contained an initial perceptual similarity task and sessions 2 and 15 contained an intermediate stimulus categorization task. Each session was conducted on a separate day, with 3–4 sessions per week. After session 8, there was a 2-week pause. Each of the different phases (training, test, similarity, and intermediate stimuli) was preceded by on-screen instructions and 5 preliminary trials in which the subjects could get used to the upcoming procedure.

**Training phase**

Each session contained a training phase in which the subjects learned the class membership of the individual faces. For this, the faces were shown separately, and the subjects were asked to indicate, via left or right button press, to which category the presented face belonged. The association of classes and buttons was randomized across subjects. The stimulus presentation ended immediately after the button press. If the answer was correct, a high-pitch feedback tone was provided. If the answer was incorrect, a low-pitch feedback tone was played, followed by a pause of 2 s with an empty screen. The pause was used to motivate the subjects to give correct responses and to ensure that the setup was identical to the previous study by Sigala and Logothetis (2002). Specifically, at no point in time the assignment of stimuli to categories was defined or explained. Each training session contained three blocks, with intermediate pauses. Each block contained 240 stimulus presentations, such that each experimental training phase contained 720 trials. The order of stimulus presentation was randomized, and each stimulus was shown equally often in each block.

**Test phase**

In order to test for pop-out effects, an odd-one-out paradigm with different numbers of distractors (3, 7, and 11) was used following the training phase of each experimental session. Each trial started with a fixation cross in the middle of the screen. After fixating it, the subjects started the trial by pressing the arrow-up key. In each trial, a target stimulus was shown together with different numbers of distractors. The distractors were either taken from the same category (class-internal condition) or the other category (class-external condition). The distractors were homogeneous, such that each display contained only two different stimuli: one being the target and the other being the distractors. The task of the subject was to indicate via a corresponding button press whether the odd face was present at the left or right half of the display. They were asked to respond as fast and accurately as possible, staying at about 90% accuracy (in line with VanRullen, 2006).

In the visual array, target and distractors were randomly positioned on a 4 × 3 (horizontal × vertical) grid, with 1° spacing between the stimuli and 0.5° random jitter (see Figure 3 for an example). The width of the individual stimuli was 2° and therefore identical to the training presentations. The grid position of the target was selected randomly and it was ensured that an equal amount of stimuli was presented on both sides of the display. In order to allow for reliable left/right dissociation even in cases with small numbers of distractors, the fixation cross stayed visible in the center of the screen during the complete trial. Moreover, the combination of target and distractor was randomized across trials. For each session, it was ensured that all stimuli were used as target in the same amount of trials. Moreover, each target was shown equally often in the class-external and class-internal conditions. Each test session contained 60 trials.

**Similarity judgment**

In addition to the training and test phases, sessions 1 and 16 contained an initial block, in which subjects 2 to 4 were asked to rate the perceptual similarity of two
concurrently presented stimuli. The rating was given on a scale from 1 to 5, with 1 being dissimilar and 5 being highly similar. Before each trial, the subjects were asked to fixate a central fixation cross, but they could freely explore the stimuli upon trial onset. Each combination of two stimuli was shown twice, such that each of the two compared stimuli appeared once at each position (left and right of the fixation cross). Each similarity judgment block contained 90 trials in total. No feedback and no category information were provided.

Intermediate stimuli

After sessions 2 and 15, subjects 2 to 4 were asked to classify the trained plus previously unseen intermediate stimuli. In parameter space, the latter stimuli were positioned such that all combinations of “close/far from boundary” and “close/far from instance” were covered (gray circles in Figure 7). Similar to the training phase, the stimuli were presented individually, and the subjects were asked to indicate the category of the shown instance via a left or right button press, corresponding to the previously learned mapping. However, no feedback was provided. All stimuli were shown twice in randomized order. This resulted in 36 trials for each intermediate stimuli block.

Data analysis

Data cleaning

The analyses were only conducted on valid trials, i.e., trials in which the correct response was given. In order to remove outliers that could occur in cases in which the subjects would pause or talk, only trials with reaction times within 2 standard deviations around the subject mean were considered. After these two steps, 89.3% of training trials, 79.4% of test trials, and 81.5% of trials of the intermediate stimulus test remained for further analyses.

Pop-out analysis

To check whether the performance of the subjects could be interpreted as pop-out, the slope of reaction time per distractor was computed separately for data from the class-external and class-internal conditions. In the literature, a reaction time slope below 10 ms per item is regarded as pop-out (VanRullen, 2006).

Analysis of the underlying form of representation and implied contributing levels

In addition to the test for pop-out, we used the odd-one-out setup to infer whether the abstract class boundary was used for classification or whether the decision was only based on the similarity to the trained category instances. The first option, a decision using abstract, high-level class information, predicts a reaction time difference between class-internal and class-external distractors and an independence of low-level metric differences in the class-external condition. The alternative, a representation that is purely based on instances with their own decision boundaries, predicts no differences between the conditions. Specifically, it should not matter whether a target belongs to the same or different category as long as it is a different instance.

Although it was ensured that no single low-level feature could be used as basis for category discrimination, it is still the case that the similarity of stimuli within a category is on average higher than the similarity across categories. Although this does not affect the test for high-level pop-out, which should only occur in the class-external condition, it must be taken into account before differences between the class-internal and class-external conditions can be attributed to high-level differences because reduced reaction times in the class-external condition could simply be due to the on average decreased similarity between target and distractor. To control for this, a measure of low-level similarity is required. Since the stimuli are parametric, a straightforward option is to use a Euclidean distance metric in the underlying four-dimensional space. In addition to this, we included the subject responses from the similarity judgment task as a measure of perceptual similarity during a later analysis. To check for a reaction time difference in the two
conditions, while accounting for low-level similarity effects, we used an analysis of covariance (ANCOVA) with “class membership” (external/internal) as main factor, “stimulus distance” as covariate variable, and a two-way interaction between main factor and covariate. For the statistical analysis, the 16 experimental sessions were divided into four analysis blocks, with four sessions each (four ANCOVAs with $p < 0.05$, Bonferroni corrected). To verify the validity of this grouping, we conducted an ANOVA for each group with the main factors “class membership” and “session.” None of the four blocks exhibited a significant interaction of session and class membership ($p > 0.45$).

Results

Fast category learning, long-lasting improvements

The analysis of the training performance shows that the two categories are learned very quickly (see Figure 4). Already after the first session, subjects correctly classify the stimuli in more than 80% of the cases. After session 3, the subjects have successfully inferred and learned the categories and reach more than 95% accuracy. During the following sessions, performance increases even further, reaching 99% accuracy in the mean over sessions 6 to 16. In addition to the near-perfect classification accuracy, the reaction times continue to decrease up to the final session (starting at 772 ms in the first session, and reaching 530 ms in the last). This indicates continuing improvements until the end of the experiment.

No high-level pop-out for parametric faces

The high training accuracy indicates that a high-level discrimination is possible rather early. This is in line with the results of the human subjects in the study by Sigala et al. (2002), in which the subjects learned to categorize the stimuli in about 500–1000 trials (Sigala, personal communication). This corresponds roughly to the training phase of a single session in our experiment. Do the implied high-level representations also lead to high-level pop-out effects, as predicted by RHT? For this, the required result would be an independence from the number of distractors for the class-external condition and a dependence in the case of class-internal targets. As can be seen in Figures 5a–5d, the reaction times for both conditions, class internal as well as class external, increase across conditions and are therefore not independent of the number of distractors. This picture is further clarified by looking at the reaction time slopes (Figure 5e). In the literature, values below 10 ms per item are interpreted as pop-out. With a mean slope across sessions of 61 ms/distractor, the slopes are far beyond this threshold until the end of the experiment. To statistically verify the dependence on the number of distractors, we performed an ANOVA with “condition” and “analysis block” as main factors. The results show

![Figure 4](http://jov.arvojournals.org/) Training results. The average training accuracy and reaction times in the different training sessions are shown. The subjects reach near-perfect accuracy very early and their reaction times improve until the end. After session 8, there was a 2-week pause for the subjects.
Figure 5. Pop-out results. (a–d) The reaction times in the different conditions, with targets being either class internal or class external. (e) Mean slope for class-external targets across sessions. The resulting slopes are far from the threshold of pop-out at 10 ms/item, indicating that no pop-out occurred during the experiment.
that both main effects are significant ($p < 0.001$), indicating the dependence on the number of distractors and the overall performance improvement with training in the pop-out task. Nevertheless, no significant interaction could be found ($p > 0.2$), which implies that the dependence on the number of distractors does not change with training (i.e., there are no significant changes in reaction time slopes with training), despite steadily improving classification performance (see also Figure S1). Thus, we do not find any evidence of high-level pop-out.

### Late inclusion of abstract category information: When class information becomes a feature

Because the class-internal condition was included in the odd-one-out setup, the recorded data allow for further investigations of the type of representation underlying the perceptual decision. The results of the performed ANCOVA for the different analysis blocks can be seen in Table 1. As main factor, class membership (external/internal) was included and the Euclidean distance between the stimuli in parameter space was taken as covariate. The log-transformed reaction times were included as dependent variable (note that the resulting patterns of significance remain unchanged when the reaction time data are directly used). Normality and homoscedasticity were verified using the Kolmogorov–Smirnov and Hartley’s $F_{\text{max}}$ tests, respectively. In the first three analysis blocks, the only significant factor is stimulus distance, there are no significant class-membership and interaction effects. Thus, despite near-perfect classification accuracy in the first three analysis blocks, the high-level class membership does not play a significant role in the odd-one-out task. This allows for the interpretation that the underlying form of representations is instance-based and that the metric-based target–distractor similarity is the important aspect governing performance.

In analysis block 4 (corresponding to the final four experimental sessions), things change. In addition to the significant effect of stimulus distance, the main factor class membership and the interaction between the distance and class membership become significant. First and foremost, this indicates that, after correcting for low-level similarity, a significant difference in reaction times in the pop-out paradigm exists, depending on the relative class memberships of target and distractor. On average, spotting the target is significantly faster if it is taken from a different category than if it belongs to the same category as the distractor (mean $\text{RT}_{\text{internal}} = 1947 \text{ ms} \pm 44 \text{ ms SEM}$; $\text{RT}_{\text{external}} = 1694 \text{ ms} \pm 30 \text{ ms SEM}$). Still, due to the significant interaction effect, which implies that the dependence on target–distractor similarity differs in the two class-membership conditions, it is important to also analyze the resulting regression models (Figure 6). In the current case, the slopes of the regression functions depict the dependence of reaction times on low-level similarity. The underlying data are pooled across all conditions and therefore provide information that cannot be inferred from the previous pop-out analyses and figures. In blocks 1 to 3, the two functions have a negative slope, indicating that increased stimulus similarity (decreased stimulus distance) correlates with increased reaction times. This shows the reaction time dependence on metric-based target–distractor similarity. In addition, the two lines are mostly identical in slope and intercept, which would be expected from an instance-based decision process that is ignorant to abstract, high-level class information. In block 4, the class-external fit is mostly below the class-internal fit. This is the visualization for the significant main effect of class membership. Moreover, the slope of the two fits is different, visualizing the interaction effect. While the slope of the class-internal fit remains negative, indicating an unchanged dependence on target–distractor similarity, the class-external fit becomes rather flat. This shows that, if the target is taken from a different class than the distractor, the dependence of the reaction times on stimulus similarity is significantly reduced. Note that this is not a sign of pop-out, which is defined as an independence of reaction times and the number of distractors. Instead, it signifies an independence from the metric-based target–distractor similarity and is a sign of a high-level distinction. Finally, the difference between the class-internal and class-external conditions is most prominent in cases of high similarity (small Euclidean distances between target and distractor). If target and distractor are very dissimilar, there is no difference between the two conditions. In order to exclude the possibility that the found class-membership effects are due to the similar motor responses during the training and pop-out parts, we performed an ANOVA with the factors “class membership” and “congruency.” Congruent trials were the ones in which the target was presented at the side that was also associated with the categorical

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main effect: Group</strong></td>
<td>$p &gt; 0.10$</td>
<td>$p &gt; 0.10$</td>
<td>$p &gt; 0.10$</td>
</tr>
<tr>
<td><strong>Main effect: Distance</strong></td>
<td>$p &lt; 0.01^{**}$</td>
<td>$p &lt; 0.01^{**}$</td>
<td>$p &lt; 0.01^{**}$</td>
</tr>
<tr>
<td><strong>Interaction effect</strong></td>
<td>$p &gt; 0.10$</td>
<td>$p &gt; 0.10$</td>
<td>$p &gt; 0.10$</td>
</tr>
</tbody>
</table>

Table 1. Significance results of the ANCOVA analysis. In blocks 1–3, only the covariate “distance” is significant, in line with an instance-based interpretation. In block 4, the group membership (class-internal vs. class-external) becomes a significant factor. Finally, there is a significant interaction effect, indicating a different slope of the linear regression fit in the different group-membership conditions.
response during training. As expected, there was a significant effect of class membership ($p < 0.001$) but no significant main effect of congruency ($p > 0.5$) and no significant interaction effect ($p > 0.5$). This indicates that the found class-membership effects did not originate from the selected type of motor responses.

Thus, despite not finding any sign of high-level pop-out, the results of the ANCOVA argue that, after extensive training, the abstract class information becomes a vital part in the perceptual decision—a change in processing, which can also be interpreted as an effect of a developing expertise of the subjects. Importantly, the class information is only used if applicable, whereas a dependence on low-level similarity remains, if the class feature is not expressive.

**Similarity judgments underline the found high-level effects**

The results of the ANCOVA indicated that high-level class information is an integrated part of the perceptual
process. From this, it can be expected that it also influences the subjective judgment of perceptual similarity. To test this hypothesis, we asked the subjects before sessions 1 and 16 to judge the similarity of two concurrently presented stimuli. With training, the class-external similarity judgment decreased significantly ($s_{\text{ext early}} = 2.71$, $s_{\text{ext late}} = 2.38$; $p < 0.01$, Wilcoxon paired test). For the judgment of the class-internal data, a reverse tendency could be found ($s_{\text{int early}} = 3.38$, $s_{\text{int late}} = 3.58$; $p = 0.118$, Wilcoxon paired test). A decrease in similarity for class-external judgments and an increase for class-internal judgments is an implicit prediction of an integrated high-level class feature in the perceptual process. To ensure that the similarity test had no effect on the later pop-out performance, we compared the accuracy and reaction times of the three subjects who performed the similarity test with the data from the subject for which the test was not included. Neither the accuracy nor the reaction times differed significantly indicating that the similarity judgment did not affect the later pop-out performance.

As a measure of low-level stimulus similarity, we so far used Euclidean distances in parameter space, which assign identical weights to the different parameter dimensions. This selection is valid as a measure of physical differences between stimuli. With regard to perceptual differences, however, it cannot be excluded that the subjects do not attribute equal weights to the different parameter dimensions. To exclude this possibility as an explanation for the found effects, we reanalyzed the data collected during the test phases of analysis blocks 1 and 4 and used the average perceived similarities for the stimulus pairs, recorded during the similarity judgment task, as a measure of stimulus distance. The results of the ANCOVA are similar to our previous analyses and conclusions. During analysis block 1, only the perceived stimulus similarity had a significant effect on the reaction times, whereas analysis block 4 exhibits significant effects of class membership and a significant interaction. Again, the underlying model fits showed that the dependence on stimulus similarity was significantly reduced in the class-external condition as compared to the class-internal condition.

Intermediate stimuli exhibit a mixed mode combining exemplar and high-level effects

Using the pop-out paradigm, we compared the levels of processing underlying two distinct cases: class-internal and class-external targets and distractors. To address whether similar results can also be observed during the classification of singular stimulus instances, we presented intermediate stimuli to the subjects during an early and a late session. These stimuli were explicitly selected to test for an instance- vs. boundary-based form of perceptual decision and were positioned accordingly. Figure 7 shows the resulting reaction time patterns (the reaction times and accuracies are presented in textual form in Table 2). In
patterns are therefore in line with the reaction time results. The accuracy seems to be dependent on the proximity to the training instances. In session 15, the only misclassified stimulus is far from a training instance and close to the decision boundary. The accuracy patterns are therefore in line with the reaction time results.

In session 2 (Figure 7b), the longest reaction times correspond to stimuli far from the category instances (1, 3, 1, 3), whereas the ones close to the trained stimuli (2, 4, 2, 4) are classified faster. This is in line with the significant factor distance and nonsignificant class-membership results of the ANCOVA analysis in the pop-out paradigm. After session 15 (Figure 7c), an additional effect can be observed. Now, not only the proximity to the trained instances but also the distance to the decision boundary becomes a relevant factor. If stimuli are close to a training instance (2, 4, 2, 4), or far from the decision boundary (1, 4, 1, 4), they are classified faster. If the tested stimulus is close to the boundary and far from the training instances, it tends to be classified more slowly. This change of pattern can best be observed in stimuli 1 and 3. After training, stimulus 1, being far from the decision boundary, is classified faster, whereas stimulus 3, close to the boundary, is classified slower than before. A similar pattern can be observed in the accuracy results (Table 2). In session 2, most of the errors are being made for stimuli that are far from the training instances, whereas the only errors in session 15 are made for stimulus 3 of category 1, which is far from the training instances and close to the decision boundary.

The late “intermediate stimulus” response patterns are therefore best described as a “logical or” interaction in which low-level metric and high-level class-based information play a joint role in the decision process. These results show a development from an instance-based to an integrated high- and low-level form of representation for the classification of singular, intermediate stimuli.

Table 2. Reaction times and accuracies for the different intermediate stimuli depicted in Figure 6. In session 2, the accuracy seems to be dependent on the proximity to the training instances. In session 15, the only misclassified stimulus is far from a training instance and close to the decision boundary. The accuracy patterns are therefore in line with the reaction time results.

<table>
<thead>
<tr>
<th>Category</th>
<th>Session 2</th>
<th>Session 15</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Category 1</td>
<td>1 681.9 (100%)</td>
<td>696.4 (100%)</td>
</tr>
<tr>
<td></td>
<td>2 629.2 (100%)</td>
<td>406.6 (100%)</td>
</tr>
<tr>
<td></td>
<td>3 2382.7 (67%)</td>
<td>917.7 (83%)</td>
</tr>
<tr>
<td></td>
<td>4 726.0 (100%)</td>
<td>712.8 (100%)</td>
</tr>
<tr>
<td>Category 2</td>
<td>1 2333.8 (83%)</td>
<td>559.5 (100%)</td>
</tr>
<tr>
<td></td>
<td>2 965.9 (83%)</td>
<td>457.3 (100%)</td>
</tr>
<tr>
<td></td>
<td>3 1711.0 (100%)</td>
<td>769.0 (100%)</td>
</tr>
<tr>
<td></td>
<td>4 470.7 (100%)</td>
<td>488.9 (100%)</td>
</tr>
</tbody>
</table>

In response to earlier studies that did not find pop-out for faces (Brown et al., 1997; Kuehn & Jolicour, 1994; Nothdurft, 1993), Hershler and Hochstein note that the distractors used were too similar to the target. This way, they argue, one single high-level representation was activated, which could not differentiate target and distractors after the initial feed-forward sweep of information (Hershler & Hochstein, 2005). This “one population” argument does not apply here, since it was previously shown that the same training setup as the current one leads to distinct high-level representations responsive to the two categories. Additionally, the argument introduces a dependence on low-level dissimilarity between target and distractor, which could again act as explanation for the found effect.

An additional argument in favor of high-level pop-out could be that the perceptual training was not long enough in order to lead to well-separated neuronal representations. For two reasons, this argument does not apply. First, the number of training trials for the fish stimuli in the study conducted by Sigala and Logothetis (2002) was below 10000 and therefore smaller than the one in the current experiment (Sigala, personal communication). Still the neuronal effects could reliably be shown. Second, a recent study by Hershler and Hochstein (2009) tested car and bird experts for high-level pop-out. Despite years of training, pop-out was only visible for faces but not for target stimuli corresponding to their area of expertise (birds or cars). This is a particularly strong case, as comparable experts were previously shown to exhibit significant activation of FFA upon perceptual decisions involving these types of stimuli (Gauthier, Skudlarski, Gore, & Anderson, 2000).

In our setup, we avoided the low-level confounds of earlier studies, by utilizing two subordinate stimulus categories that cannot be classified on the basis of simple low-level features but only via feature conjunctions. Additionally, perceptual training based on these stimuli was previously shown to give rise to distinct high-level representations. Therefore, they fulfill both requirements of high-level pop-out according to RHT. For future studies, it might nevertheless be interesting to test two basic-level categories of parametric stimuli, with higher (but still similar) inter- and intra-category differences.

With the second major analysis, we addressed the underlying form of category representation and implied
contributions of the different levels used during the perceptual decisions. Here, we found a shift from an early low-level instance-based to a late, abstract, and class-based form of representation. For the latter, we demonstrated a dependence on stimulus similarity if target and distractor are taken from the same category, whereas response times were shown to be significantly less dependent of low-level differences if the high-level class membership of target and distractor is different. Again, this is no case of high-level pop-out, as this would require independence of the number of distractors, which is never the case. These results show that, dependent on the available information, different modes of processing are required and used for the perceptual decision.

With regard to an early instance-based representation, as reported here for the first three analysis blocks, converging evidence comes from a recent fMRI-adaptation study that investigated response properties in the lateral occipital cortex (Gillebert, Op De Beeck, Panis, & Wagemans, 2009). Before the fMRI measurements, the subjects were trained to categorize two classes of parametric stimuli in a paradigm comparable to ours. The subjects were trained for two sessions, 50 min each, which is roughly equivalent to our first analysis block. As measure, they used the amount of BOLD adaptation resulting from the presentation of two subsequent stimuli that could either be taken from the same or different categories. The analyses show that there is no significant difference between their class-internal and class-external conditions. In line with more recent work by Sigala, this indicates that the early form of representation is best explained instance-based (Gauthier & Palmeri, 2002; Sigala, 2004). This is in agreement with our analyses. However, one prediction of our class-membership effect in analysis block 4 is that prolonged training should result in a significant difference in adaptation based on class-internal vs. class-external stimulus presentations.

Converging evidence for the integration of categorical information into the perceptual process is provided by the similarity judgment task. With training, the perceived similarity of stimuli from the same class increased whereas the similarity judgment across categories decreased. These results are in line with earlier studies describing the effect of categorization on perceived similarities (Goldstone, 1994; Goldstone, Lippa, & Shiffrin, 2001). Moreover, by using the recorded perceptual similarities as a measure of stimulus distance, we could show that the found high-level class effects could not be explained by a differential weighting of dimensions in parameter space. The question of whether the found high-level representations are based on an additionally learned decision hyperplane or rather on the additional storage of appropriate prototypes is beyond the scope of the current study (see Sigala et al., 2002 for an argument against a representation based on single prototypes). In any case, both approaches form an abstract representation of the learned categories.

In line with the analyses of the odd-one-out paradigm, the results of the intermediate stimulus classification task showed that early reaction time patterns resemble an instance-based form of category representation with a dependence on the training exemplar distances. However, the later response patterns were shown to resemble a “logical or” mixture of both metric- and class-based information. This further exemplifies the task-dependent contribution from levels of different complexity during the late stage of training. In pop-out, if high-level information is expressive it is used exclusively, whereas low-level target–distractor similarity is predominant in the class-internal condition. In the intermediate stimulus categorization task, in which no distractors were present, we find a mixed mode integrating both types of information.

Having found the two modes of perceptual processing in the pop-out task, being either based on target–distractor similarity or more abstract class information, the resulting question is which physiological levels in the visual hierarchy might be underlying these effects. Based on previous results in the literature, we expect the abstract class information to originate from high-level visual areas (Kiani, Esteky, Mirpour, & Tanaka, 2007; Sigala & Logothetis, 2002) or from the lateral prefrontal cortex (LPFC; Freedman, Riesenhuber, Poggio, & Miller, 2003 but see Minamimoto, Saunders, & Richmond, 2010). Note that this distinction does not matter for the predictions of RHT, as both areas were described as being potential sources of the categorization during the initial vision at a glance (Hochstein & Ahissar, 2002). The dependency on stimulus distance, however, is directly related to metric-based target–distractor similarity. This type of effect is expected to originate from lower levels of processing, which are mostly stimulus-driven and contain the relevant and more detailed stimulus information. However, a clear decision cannot be based on psychophysical measurements, especially in the case of long perceptual training, and future work is required to fully address this issue.

In our experiments, no high-level pop-out could be observed, despite the near-perfect classification performance of our subjects. Instead, we have found clear evidence for an integrated use of low-level metric- and high-level class-based information, dependent on the task and the information content of the individual levels of representation. As a result of these findings, we suggest that the visual hierarchy only provides a means for increasingly complex and nonlinear response properties, or modules, whereas the evaluation of the resulting representations is highly distributed. Although a hierarchical view on visual processing might be applicable with regard to the representational complexity, it does not automatically imply an increasing level of perceptual importance with subsequent levels of processing. The proposed view is in line with the results of recent experiments, which underline the importance of lower level representations during perceptual decisions (Kamitani & Tong, 2005; Silvanto et al., 2005; Tong, 2003).
Acknowledgments

The authors gratefully acknowledge the support of the Niedersächsisch-Israelische Gemeinschaftsvorhaben “Does the study of simple visual stimuli assess the primitives of natural vision” (Prof. König, Prof. Ahissar), and the Research Training Group “Adaptivity in Hybrid Cognitive Systems” of the Institute of Cognitive Science at the University of Osnabrück. Furthermore, we would like to thank Frank Jaekel, Niklas Wilming, Alper Acik, Orit Hershler, Konrad Körding, and Merav Ahissar for helpful discussions and comments on an earlier version of the manuscript. Finally, we would like to thank Natasha Sigala for valuable advice and help with the stimulus set.

Commercial relationships: none.
Corresponding author: Tim C. Kietzmann.
Email: tkietzma@uos.de.
Address: Albrechstraße 28, Osnabrück 49069, Germany.

Footnotes

1This presupposes that similar representations are present in monkey and human upon identical training procedures. Evidence for the validity of this assumption was provided by Sigala et al. (2002) who showed that monkeys and humans follow the same categorization strategies in the currently used task.

2It should be noted that similarity tests involving the judgment of categorical stimuli cannot exclude the possibility that subjects explicitly judge stimuli as being more similar if they realize that the stimuli belong to the same category (Goldstone et al., 2001).

References


