Perceptual Precedence or Increased Effort?: On the mechanism of the small-picture-size advantage in category learning

Toshiya Miyatsu
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Psychological and Brain Sciences

Dissertation Examination Committee:
Mark A. McDaniel (Chair)
Richard A. Abrams
Julie M. Bugg
Andrew C. Butler
Jeffrey M. Zacks

Perceptual Precedence or Increased Effort?:

On the mechanism of the small-picture-size advantage in category learning

by
Toshiya Miyatsu

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2019
St. Louis, Missouri

Table of Contents

List of Figures

Acknowledgments

Once a high school drop-out in Tokyo, I have never in my wildest dream imagined that I would receive a doctoral degree, let alone from a school like Washington University. My heart is filled with gratitude for every person who helped me to get here. There is not enough space to list all of you, but please remember that none of you are forgotten, and I will always be grateful to you.

The partial list start from my adviser, Mark McDaniel, who took a chance in me and brought me in to Washington University. He gave me the push that I needed when I needed and made me the scholar that I am today. When I look back at graduate school, the hardest but most meaningful work that comes to mind is the hundreds of hours I spent revising manuscripts based on his thoughtful feedback. Yes, I still have article confusions (English is hard!), but I enjoy writing now, and the things he told me will always echo in my head when I write in the future.

I also thank Richard Abrams, Julie Bugg, Jeff Zacks, and Andy Butler, for taking their time and serving in my dissertation committee. My committee meetings were filled with not only serious inquiry into psychological science but also cynical jokes (mostly Richard) and warm support. They were the best dissertation committee that I could ask for.

The members of the Memory and Complex Learning Lab present and past provided me with invaluable support and beautiful memories. When I picture the lab in my head, I recall many hours of subject testing, group lunches, and just pleasant hangouts with the faces of Francis Anderson, Carlee DeYoung, Reshma Gouravajhala, Madison Kasoff, Carolina Küpper-Tetzel, Ji Hae Lee, Jeri Little, Yiyi Liu, Amanda Mayer, Khuyen Nguyen, Walter Reilly, Sharda Umanath, and Emily Waldum. I was also blessed with many friends and mentors in the department, such as Hank Chen, Jason Finley, John Nestojko, and Victor Sungkhasettee.

During the challenging graduate school journey, what kept me sane was the community of people who Kyle Watson put together at Watson Martial Arts. Whenever I was on the mat training, the daily struggles went away and I was living my fight dream. I am as proud of my competitive accomplishment in Jiu Jitsu tournaments and Mixed Martial Arts fights here in St. Louis as I am about my academic accomplishments. Kyle and all the teammates at Watson Martial Arts made me not only a better martial artist and a better fighter but also a better person.

Many people from my undergraduate days gave me a continuing support and mentorship without which I could not even reach my graduate study. Bob and Elizabieth Bjork, Alan Castel, and the members of the Bjork lab, such as Doe Buchli, Monica Birnbaum, Colin Clark, Courtney Clark, Mikey Garcia, Saskia Giebl, Jeri Little, John Nestojko, Nick Soderstrom, Veronica Yan, and Carol Yue, all taught me the joy and intricacies of psychological sciences and groomed me into the scholar that I am today. Special thanks go to Dr. Brown who saw something in a kid lost in translation in a community college classroom and encouraged me to a career in psychology.

My parents have been my biggest supporters through good and bad. Their life-long dedication to education (they had long and celebrated careers as teachers) inspired me to pursue the subject that I studied in graduate school and I hope I made them proud.

Finally, the best part of my graduate school days was that I shared this journey with my wife Rose, who also completed her PhD in English Literature at Washington University. We grew together tremendously during these years, and because of it, I feel that we can overcome any challenges in life. I love you and I'm looking forward to many more adventures together!

Toshiya Miyatsu

Washington University in St. Louis

December 2019

Dedicated to my parents and Rose.

ABSTRACT OF THE DISSERTATION

Perceptual Precedence or Increased Effort?:

On the mechanism of the small-picture-size advantage in category learning

by

Toshiya Miyatsu

Doctor of Philosophy in Psychological & Brain Sciences

Washington University in St. Louis, 2019

Professor Mark McDaniel, Chairperson

I have previously identified a novel perceptual manipulation that enhances learning of some complex natural categories, and the current dissertation aims to uncover its mechanism. Specifically, learning of categories of tropical fish was enhanced when learned through small pictures (about 2º) compared to large pictures (about 19º). Through analyzing the previous results and extant theories in various domains, I identified two potential mechanisms through which this small-picture-size advantage manifested. The *perceptual precedence hypothesis* postulates that the processing of local dimensions is prioritized in large pictures and the processing of global dimensions is prioritized in small pictures. Therefore, small picture size should enhance category learning only when a global dimension is diagnostic (e.g., the exterior shape in the tropical fish categories). The *increased effort hypothesis* postulates that because small pictures are harder to process than large pictures, it creates a metacognitive sense of disfluency, and that perceptual disfluency engages learners in a more effortful and analytical processing of the stimuli. Thus, this theory predicts that small picture size should enhance category learning whether the diagnostic dimension is local or global. Two experiments directly pitted these unique predictions by the two theories against each other; participants studied category structure with either a global diagnostic dimension or local diagnostic dimensions. These experiments not only replicated the small-picture-size advantage, but also showed a large-picture-size advantage when local dimensions were diagnostic. The findings supported the perceptual precedence hypothesis and suggested that the picture-size effect is category-structure-specific rather than category-structure-general. The effects of the size manipulation on learners' metacognition is also discussed.

Chapter 1: Introduction

Recognizing a visually perceived object as a member of a certain category is a fundamental and ubiquitous component of human cognition that enables us to act upon the environment efficiently (Murphy, 2002). As such, category learning has been studied extensively in psychology and is involved in many meaningful contexts in the society, such as K-12 and higher education (e.g., teaching categories of rocks in geological science courses), physician training (e.g., radiologists learning to distinguish between malignant and benign tumors), and military training (e.g., learning to detect abnormalities on the ground surface that may signal a presence of a land mine). Accordingly, strong interest exists in discovering ways to improve category learning for various purposes, and researchers have recently made efforts towards discovering manipulations that enhance category learning (e.g., interleaving: Kornell & Bjork, 2008; test-enhanced learning: Jacoby, Wahlheim, & Coane, 2010; exemplar variability: Wahlheim, Finn, & Jacoby, 2012; fading of diagnostic features: Pashler, & Mozer, 2013; feature highlighting: Miyatsu, Gouravajhala, Nosofsky, & McDaniel, 2018; specific-level training: Miyatsu, Nosofsky, & McDaniel, in press; Nosofsky, Sanders, Gerdom, Douglas, & McDaniel, 2017). While these studies have examined factors, such as the selection of the training example set, the sequencing of examples during training, and the type of processing occurring during training, little research has been conducted to examine the potential benefit of manipulating simple perceptual characteristics of the way training examples are presented. In this dissertation, I will report five previous experiments that I have conducted to establish a novel perceptual manipulation that enhances category learning, sketch out two hypotheses regarding the mechanisms underlying this effect, and report two new experiments that tested unique predictions derived from these two hypotheses. In addition, I will investigate the effect of the size manipulation on learners' metacognition and provide two frameworks through which the results on metacognition can be interpreted.

In a typical category learning experiment, participants study example pictures of the categories to be learned one by one for a few seconds each in the training phase. Participants' learning is assessed later by asking them to classify new pictures into the learned categories. The critical manipulation that will be discussed throughout the current dissertation is to present these training examples in a small as opposed to a large size. This manipulation was motivated by recent technological advances that changed the way we consume and present information as well as by a recent literature on the effect of stimulus size on human metacognition. First, the variability of the size of the screens through which we consume information is greater than ever today. It is becoming increasingly more common to read texts, view photos, and watch videos from different devices with various screen sizes, such as a desktop monitor, a large flat screen TV, and a smartphone. That is, people today often consume the same information from screens of varying sizes. From a standpoint of enhancing category learning instruction and training, the size manipulation, if proven to be fruitful, carries great value because it could be easily applied to example pictures of any categories using photo-processing software, presentation software, or even just printing photos in various sizes. Second, as will be described below, a recent literature showed that learners' metacognition in word-list learning was affected by the stimulus size independent from their word-memorization performance, so it seemed possible that this stimulus-size effect on metacognition extends to a more complex learning situation like learning categories of natural objects.

Many may assume that the size in which a picture or any stimulus is presented would have no effects on human cognition. After all, it is the same picture that is presented, and the information contained in the picture does not change. However, a careful review of distant literatures spread across psychology, business and marketing, and human factor research showed that stimulus size affects perception, attention, emotion, object processing, and various subjective judgments (Miyatsu, in preparation). Although the stimulus-size manipulation has not been investigated in the context of category learning, some of these findings suggest that it may have effects on both the learning and the metacognitive aspects of category learning. First, stimulus size may affect the category learning itself. Larger stimuli are generally processed more quickly, and identification of objects are less accurate for smaller pictures (e.g., Breitmeyer & Breier, 1994; Schultz & Eriksen, 1978). A straightforward prediction from these findings is that people learn categories better from larger pictures. However, as I will elaborate in detail after presenting my previous experiments investigating the effect of the picture-size manipulation on category learning in the next chapter, some tasks are actually performed better when the stimuli are presented in smaller sizes. Thus, it is possible that small pictures promote better category learning than large pictures.

Second, metacognition, in particular learners' own understanding of how well they have learned the categories, may be affected by the picture-size manipulation. When learning lists of words and word pairs, people claim to have learned better from a larger compared to smaller font size although there is no memory performance difference between information presented in different font sizes (e.g., Rhodes & Castel, 2008). This is because people have a belief that larger stimuli are better for remembering (e.g., Mueller, Dunlosky, Tauber, & Rhodes, 2013) and because larger stimuli are processed more quickly and create a greater sense of fluency (e.g., Rhodes & Castel, 2008; Susser, Mulligan, & Besken, 2013). According to the cue-utilization framework (Koriat, 1997), both the belief and the perceptual fluency can be used as a cue to make judgements of learning resulting in the metacognitive illusion. Because people may have a similar belief that larger pictures are better for learning categories and because larger pictures are processed more quickly, it seems possible that a similar metacognitive illusion arises from the picture-size manipulation in category learning where people claim to have learned better from larger pictures regardless of whether they actually performed better with the larger pictures.

Chapter 2: Previous Experiments

Motivated by the potential picture-size effects on category learning and metacognition as described above, I conducted five preliminary experiments. In all five of these unpublished experiments reported in this section, the example pictures from the half of the categories were presented in a small size (e.g., about 2 degree of visual angle: a fish picture of 1 inch in width viewed from 30 inches) while the example pictures from the other half of the categories were presented in a large size (e.g., about 19 degree of visual angle: a fish picture of 10 inches in width viewed from 30 inches). These sizes were determined to achieve the smallest size in which all the features in the pictures could still be identified and the largest size that could be presented on laboratory computer screens without scrolling.

The five experiments differed in the category structure that participants studied. The category structure of a category learning experiment can be characterized by its dimensions and the diagnosticity of these dimensions while individual category can be characterized by its unique features. A *dimension* in category learning refers to an aspect of stimuli that varies from instance to instance. For example, in the tropical fish categories shown in Figure 1, the exterior shape, the colors, and the pattern expressed in the body are dimensions. *Diagnosticity* refers to how useful a given dimension is in defining the categories. For example, in each of the categories shown in Figure 1, the color and pattern vary widely among the instances within a category, but the exterior shape stays relatively consistent. That is, the exterior shape is a diagnostic dimension in this category structure. *Features* refer to the particular values expressed within dimensions. Each category expresses unique features within a diagnostic dimension. For example, in the fish categories shown in Figure 1, angelfish (left most column) have a dorsal fin, an anal fin (two fins on the top and the bottom of the body towards the tail), and a pointy mouth that form an arrowhead shape along the body, blenny (second to left most column) have a thin body without sharp fins, and triggerfish (right most column) have an oval shape body with distinctive dorsal and anal fins.

Figure 1. Example pictures from the tropical fish family classification experiment (Previous Experiment 1). All the pictures in the same column come from one family. From the left to right, angelfish, blenny, filefish, goby, and triggerfish.

The category structure in the experiments that will be described next was varied to assess which dimensions the picture-size manipulation is sensitive to and under what circumstances (i.e., the presence or absence of other dimensions and their diagnosticity) the manipulation affects learning. Because all five experiments had the same method except the materials that were learned, I first describe the general procedure that was uniform in all five experiments, and then describe the category structure used in each experiment and their results in turn.

2.1 General Procedure

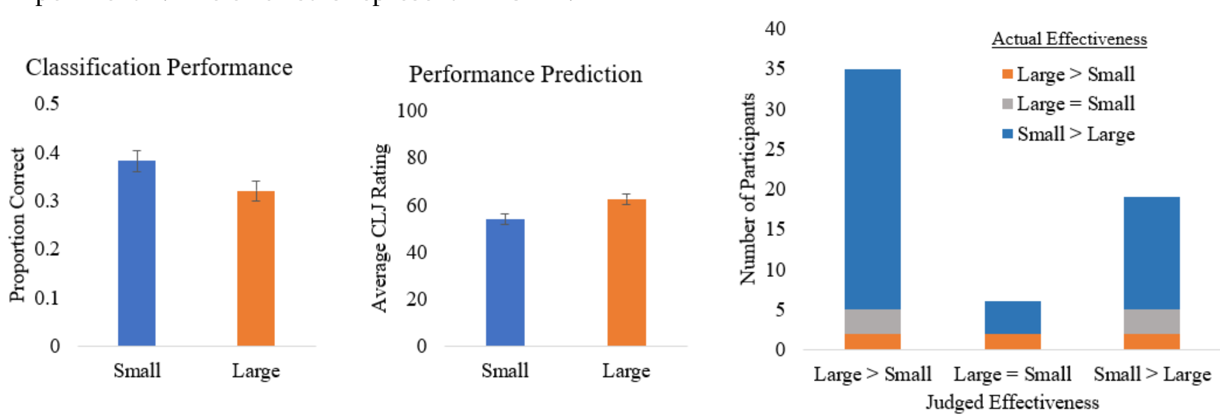
In all five experiments, participants studied and classified pictures of 12 categories of tropical fish. First, participants were presented with 72 pictures, six from each of the 12 categories, one by one for 5 seconds each. The order of the presentation was block-randomized, such that each block of 12 pictures included one picture from each of the 12 categories. Importantly, pictures from half of the 12 categories were presented in the small (about 2^{o} in the previous Experiments 1-3; about 1^{o} in the previous Experiments 4 and 5) while pictures from the other half of the categories were presented in the large size (about 19^{o}). After studying the 6th and final example from each category, the participants indicated how well they learned a given category on a scale of 0 – 100 (*category learning judgement*: CLJ). Then, the participants played tetris for 3 minutes as a distractor before completing the final test. In the final test, participants were presented with 48 new pictures of tropical fish one by one in a neutral size (about 11^{o}) and asked to classify them by clicking on one of 12 options labeled with the names of the 12 learned categories. They were given up to 10 seconds for each final test trial. The order of the final test was block-randomized, such that in a given block of 12 items, there was one item from each of the 12 categories. Upon completing the final test, participants answered a post-experimental question about which picture size they thought helped learn the fish categories better.

2.2 Previous Experiment 1

(family-level classification – the exterior shape diagnostic)

The initial experiment was run using the materials shown in Figure 1. In this experiment, participants learned the categories of tropical fish families (n = 60). As described above, the exterior shape is the predominant diagnostic dimension in this category structure. Figure 2 shows the results of this experiment. As shown in the left panel of the Figure 2, the classification performance was higher for the categories presented in the small size compared to the ones presented in the large size, $t(59) = 2.31$, $p < .05$, $d = 0.38$. Moreover, as indicated by both CLJs (i.e., learning judgement before the final test) and the post-experimental question, participants reported that they felt as if they learned better from the large pictures. To foreshadow, in all five experiments reported in this section, these patterns of results on CLJs and the post-experimental question held up. Because the current study primarily focuses on the classification performance aspect of the small-picture-size effect, I will omit the description of the results regarding the CLJs and post-experimental question for the remainder of this section. However, these results and their theoretical interpretations will be discussed in later chapters.

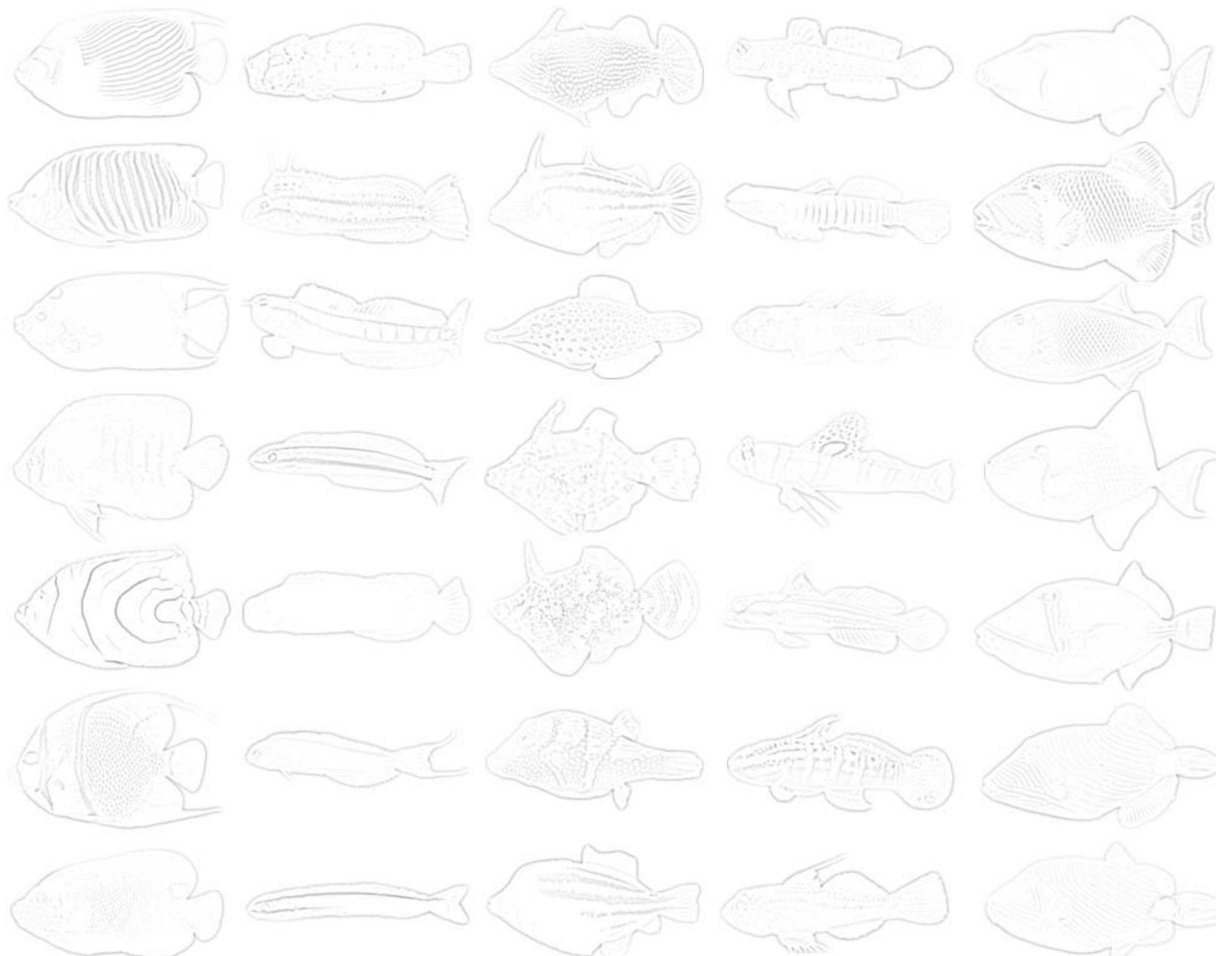
Figure 2. Participants' mean classification performance (left panel), mean CLJ ratings (middle panel), and post-experimental judgement on which picture size they learned better from, from Previous Experiment 1. The error bars represent ± 1 *SEM*.



2.3 Previous Experiment 2 (family-level vs abstract outline classification – the exterior shape diagnostic but no color)

In the next experiment, half the participants learned the family-level classification (i.e., identical to Previous Experiment 1; n = 57) while the other half learned the same categories without color (abstract outline; see Figure 3; n = 57).

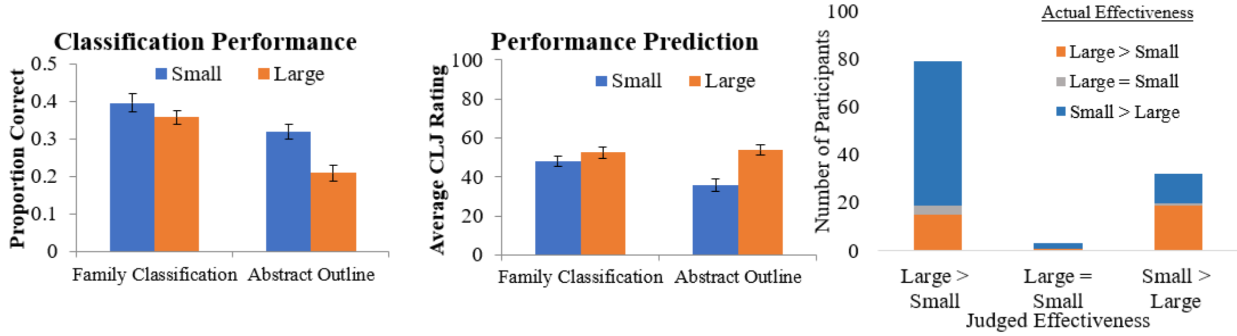
Figure 3. Example pictures from the tropical fish family classification experiment using abstract outline (i.e., no color: Previous Experiment 2). All the pictures in the same column come from one family. From the left to right, angelfish, blenny, filefish, goby, and triggerfish.



The purposes of this experiment were to replicate the findings from the previous Experiment 1 as well as to examine the effect when the color, a salient dimension[1] that varied widely at the family level, was absent. Figure 4 shows the results of this experiment. A 2 (size: small or large) x 2 (category condition: family classification or abstract outline) mixed analysis of variance (ANOVA) on the classification performance data showed a significant main effect of the size, such that categories presented in the small size were learned better, $F(1, 112) = 14.78$, $p < .001$, $\eta p^2 = .12$, a significant main effect of the category condition, such that the participants in the family-level classification condition performed better than the ones in the abstract outline classification condition, $F(1, 112) = 24.61$, $p < .001$, $\eta p^2 = .18$, and a marginally significant interaction between these variables, $F(1, 112) = 3.44$, $p < .10$, $\eta p^2 = .03$. Post-hoc paired-sample t-tests showed that the small-picture-size advantage in the family-level classification condition did not reach significance, $t(56) = 1.47$, $p = .15$, $d = 0.23$, whereas it was significant in the abstract outline condition $t(56) = 3.87$, $p < .001$, $d = 0.72$. Thus, the small-picture-size advantage was extended to when the color, a salient dimension, was absent. However, the effect on the family-level classification did not fully replicate.

Figure 4. Participants' mean classification performance (left panel), mean CLJ ratings (middle panel), and post-experimental judgement on which picture size they learned better from, from Previous Experiment 2. The data for the post-experimental judgement were combined between the family-classification and abstract outline conditions because they showed the same pattern. The error bars represent ± 1 *SEM*.

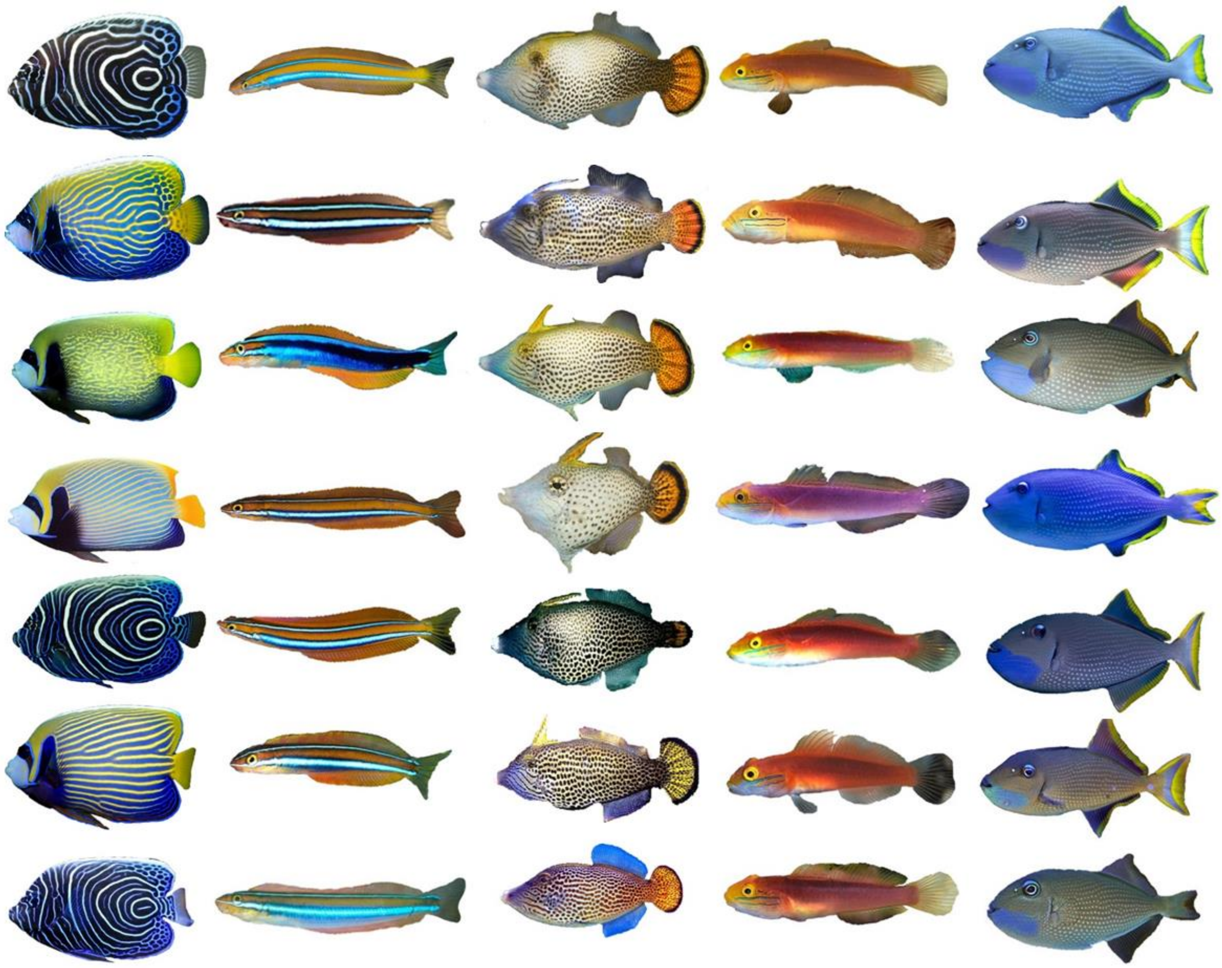
[1] For example, in a previous study involving classification of venomous and non-venomous snake categories, participants were unable to learn the categories without being instructed that the color was not diagnostic, demonstrating the salience of this dimension (Noh, Yan, Vendetti, Castel, & Bjork, 2014).



2.4 Previous Experiment 3 (family-level vs species-level classification – the exterior shape diagnostic vs the exterior & pattern/color diagnostic)

In the third experiment, half the participants learned the family-level classification (i.e., identical to the previous Experiments 1 and 2; n = 45) while the other half learned the species-level classification (n = 46). At the species level of classification, not only the exterior shape, but also the pattern/color expressed on the body is highly diagnostic (see Figure 5).

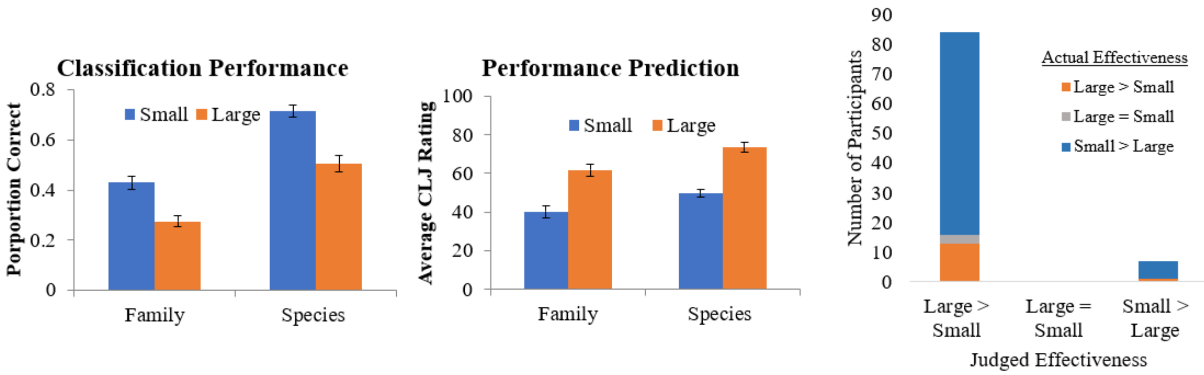
Figure 5. Example pictures from the tropical fish species classification experiment (Previous Experiment 3). All the pictures in the same column come from one species. From the left to right, angelfish, blenny, filefish, goby, and triggerfish.



The purposes of this experiment were to further assess the reliability of the small-picture-size effect on the family-level classification as well as to examine the effect when other highly diagnostic dimensions were available. A 2 (size: small or large) x 2 (category condition: family-level or species-level classification) mixed ANOVA on the classification performance data showed a significant main effect of the size, such that categories presented in the small size were learned better, $F(1, 89) = 73.57$, $p < .001$, $\eta p^2 = .45$, a significant main effect of the category condition, such that the participants in the species-level classification condition performed better than the ones in the family-level classification condition, $F(1, 89) = 67.00$, $p < .001$, $\eta p^2 = .43$, and a non-significant interaction between these variables, $F(1, 89) = 1.70$, $p > .05$, $\eta p^2 = .02$.

Thus, the small-picture-size effect at the family-level classification appeared reliable and the effect was also extended to when multiple diagnostic dimensions, the exterior shape and pattern/color, were present.

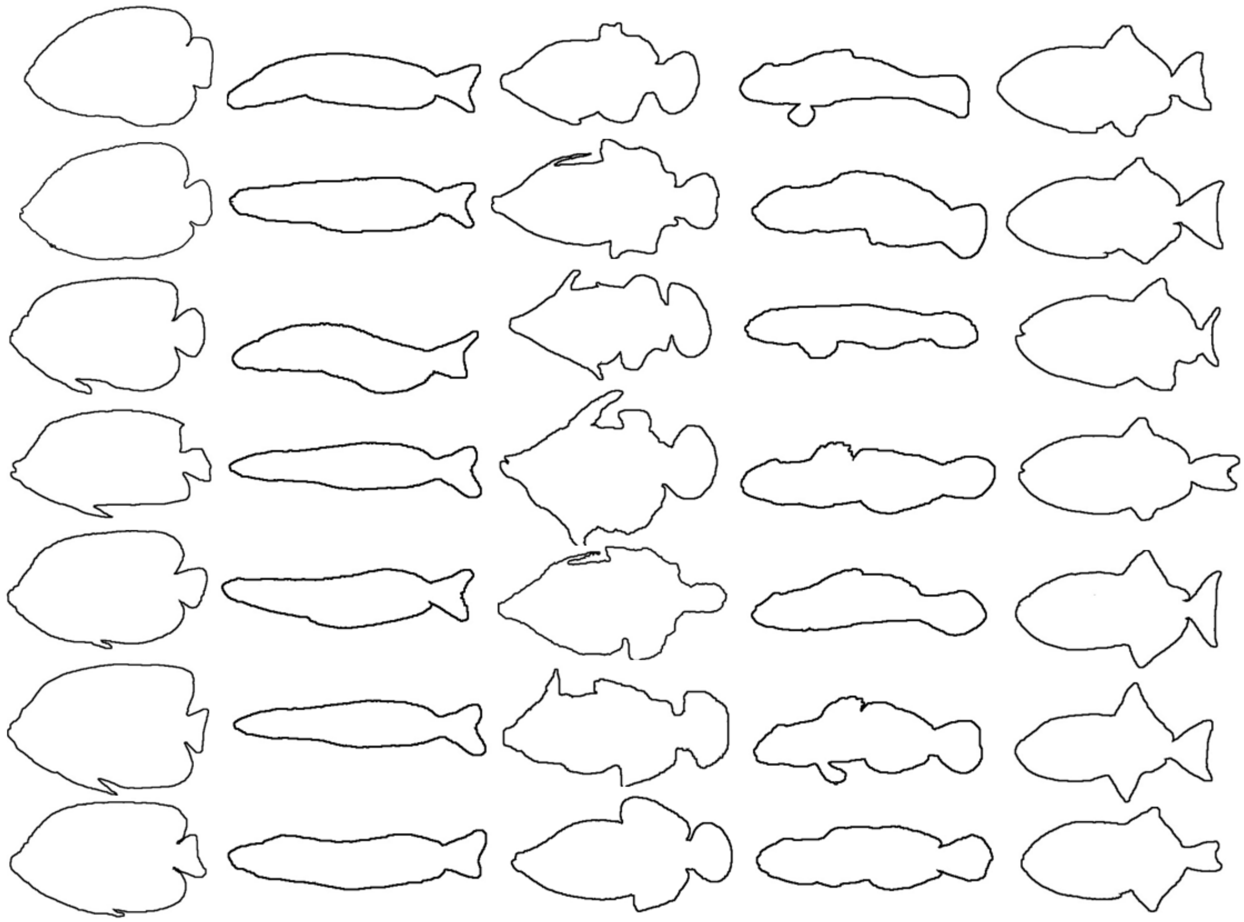
Figure 6. Participants' mean classification performance (left panel), mean CLJ ratings (middle panel), and post-experimental judgement on which picture size they learned better from, from Previous Experiment 3. The data for the post-experimental judgement were combined between the family and species conditions because they showed the same pattern. The error bars represent ± 1 *SEM*.



2.5 Previous Experiment 4 (species-level vs species outline classification – the exterior & pattern/color diagnostic vs the exterior shape was the only dimension present)

In the fourth experiment, half the participants learned the species-level classification (i.e., identical to Previous Experiment 3; n = 40) while the other half learned the same categories with only the outline (species outline; see Figure 7; n = 38).

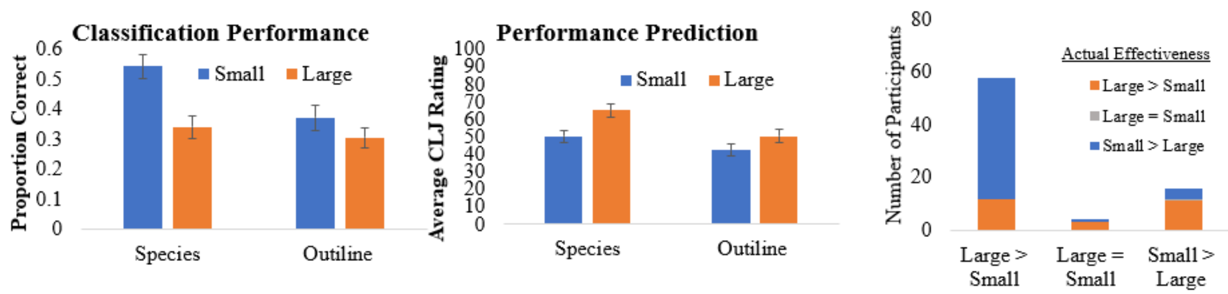
Figure 7. Example pictures from the tropical fish species outline classification experiment (Previous Experiment 4). All the pictures in the same column come from one species. From the left to right, angelfish, blenny, filefish, goby, and triggerfish.



The purposes were to replicate the findings from the previous Experiment 3 on the species-level classification as well as to examine the effect when all other dimensions, except the exterior shape, were absent. The working hypothesis at that time was that the small-picture size enhanced learning through shifting learners' attention from the periphery of the stimulus when small to the details inside of the perimeter when large. Thus, the idea was that if the attentional shift from the periphery (i.e., the exterior shape) to the details inside indeed was the underlying mechanism, the small-picture-size advantage should vanish in the outline condition because there were no other dimensions to pay attention to even when the pictures were presented in the large size. In addition, a smaller small-size manipulation was implemented in this and the following experiment to see the generality of this manipulation (about 2° in the previous Experiments 1-3; about 1° in the previous Experiments 4 and 5).

Figure 8 shows the results of this experiment. A 2 (size: small or large) x 2 (category condition: species classification or outline) mixed ANOVA on the classification performance data showed a significant main effect of the size, such that categories presented in the small size were learned better, $F(1, 76) = 23.23$, $p < .001$, $\eta p^2 = .23$, a significant main effect of the category condition, such that the participants in the species-level classification condition performed better than the ones in the species outline classification condition, $F(1, 76) = 4.92$, $p < .05$, $\eta p^2 = .06$, and a significant interaction between these variables, $F(1, 76) = 5.87$, $p < .05$, $\eta p^2 = .07$. Post-hoc paired-sample t-tests showed that the small-picture-size advantage in the species-level classification condition was significant, $t(39) = 5.69$, $p > .001$, $d = 0.83$, whereas it did not reach significance in the abstract outline condition, $t(37) = 1.54$, $p = .13$, $d = 0.29$. In short, the small-picture-size effect at the species level was replicated and was extended to an even smaller picture size, and the lack of (or at least diminished) effect in the outline condition suggested that some kind of attention shift from other dimensions to the exterior shape may underlie the effect.

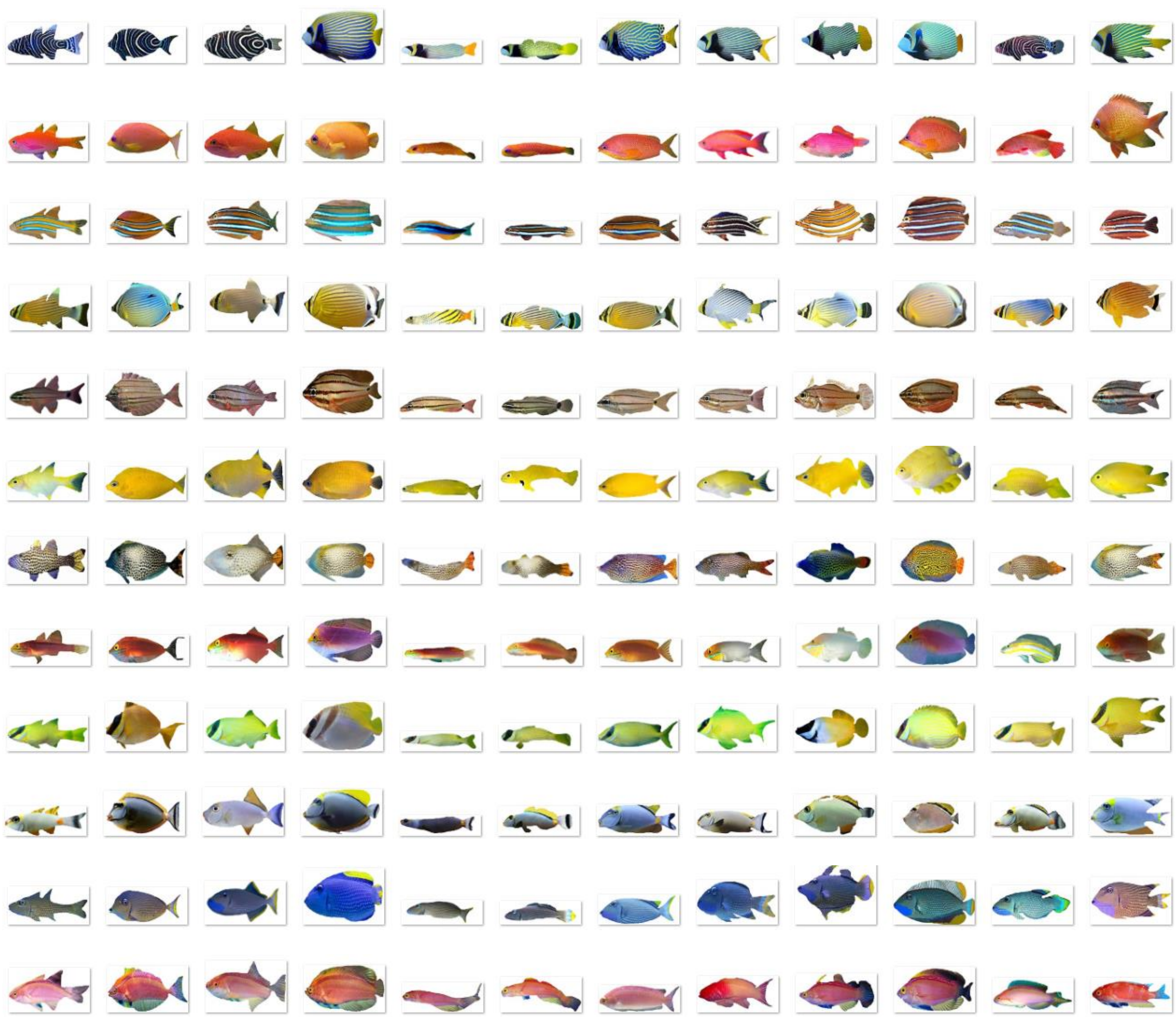
Figure 8. Participants' mean classification performance (left panel), mean CLJ ratings (middle panel), and post-experimental judgement on which picture size they learned better from, from Previous Experiment 4. The data for the post-experimental judgement were combined between the species and outline conditions because they showed the same pattern. The error bars represent ± 1 *SEM*.



2.6 Previous Experiment 5 (artificial category classification – the pattern/color diagnostic)

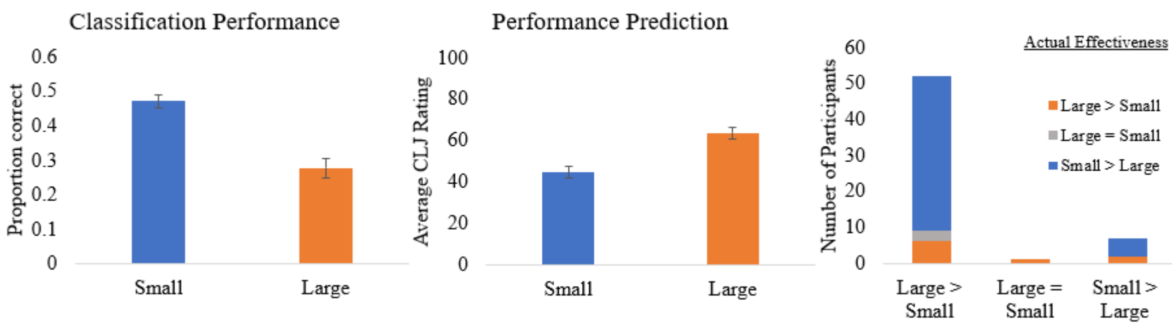
In the fifth experiment, participants learned an artificial fish category structure in which the pattern/color on the body, but not the exterior shape, was diagnostic. This category structure was created by tracing and cutting out the body pattern of twelve fish species in the shapes of twelve different fish species and presented the examples that had the consistent body pattern/color to be a category (see Figure 9; n= 60).

Figure 9. Example pictures from the tropical fish classification experiment using the artificial fish stimuli (Previous Experiment 5) in which the body pattern/color was the predominant diagnostic dimension. Each row represents a category.



The idea was to test the attentional shift hypothesis described above; if the small-picture size enhanced learning in the previous experiments through shifting learners' attention to the periphery of the stimulus when small and to the details inside of the perimeter when large, the large size should outperform the small size in learning this category structure because the details inside are now diagnostic. Figure 10 shows the results of this experiment. To my surprise, participants still performed better on the categories that were presented in the small compared to the large size, $t(59) = 7.44$, $p < .001$, $d = 1.04$. Thus, the picture-size manipulation turned out to be sensitive not only to the exterior shape but also to the pattern/color expressed in the body of the fish pictures.

Figure 10. Participants' mean classification performance (left panel), mean CLJ ratings (middle panel), and post-experimental judgement on which picture size they learned better from, from Previous Experiment 5. The error bars represent ± 1 *SEM*.



2.7 Previous Experiments Summary

To sum up the previous findings, the small-picture size enhanced category learning when the exterior shape was the predominant diagnostic dimension (the previous Experiments 1, 2, and 3), when the color was absent (the previous Experiment 2), and when there were other dimensions that were highly diagnostic (i.e., pattern/color; the previous Experiments 3 and 4). However, the effect vanished, or at the very least it was decreased, when there were no other dimensions except the exterior shape, suggesting some kind of attentional shift elicited by the picture-size manipulation (the previous Experiment 4). Finally, the picture-size manipulation was shown to be sensitive not only to the exterior shape but also to the pattern/color expressed in the body

(Previous Experiment 5) when these dimensions were diagnostic. In addition, in all of these experiments, participants showed a large-size bias in metacognitive measures; they claimed to have learned better from the large pictures both after the learning and after the test. In the following section, I will propose two plausible hypotheses that can accommodate the findings on category learning and draw out unique predictions made by these hypotheses. Further, I will provide two frameworks that can shed light on how learners make metacognitive judgements and draw predictions from these viewpoints on the effect of the picture-size manipulation on category learning judgements.

Chapter 3: Potential Mechanisms

3.1 On the Picture-Size Effect on Category Learning

3.1.1 The Perceptual Precedence Hypothesis

What do the exterior shape and the body pattern, the dimensions that are sensitive to the picture-size manipulation, have in common? One possibility is that they are both dimensions that are expressed globally. That is, an entire stimulus needs to be scanned to encode these dimensions. For example, encoding of an entire fish picture is necessary to extract the exterior shape which informs the height-to-width ratio and the shapes of the features expressed in the periphery. Similarly, an entire stimulus needs to be encoded to extract the characteristic features that are expressed throughout the body, such as a category of fish having dots all over the body or having a few lines that run across the entire body horizontally.

One theory that was originated in a classic cognitive literature suggests that certain dimensions receives prioritized processing depending on the size of the stimulus. Navon (1977, 1981) claimed that people were attuned to extracting the global elements of an object before extracting the local elements, a tendency termed *global precedence*. In this line of research, subjects typically studied large letters made up of small letters (see Figure 11), and the reaction time to various tasks targeting the global and the local elements were assessed.

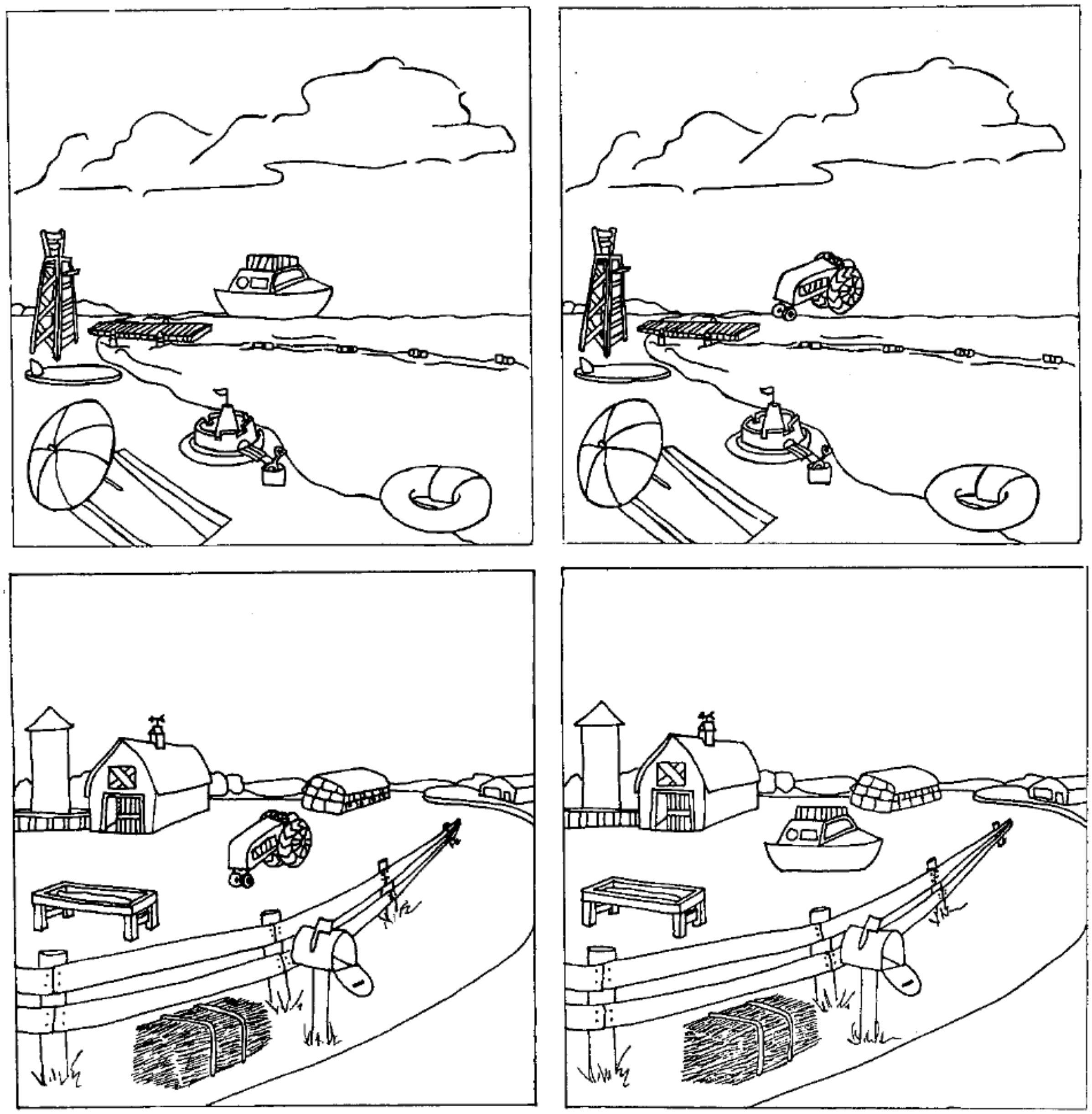
Figure 11. Typical stimuli used in experiments investigating the global precedence effects. Reprinted from "Do response time advantage and interference reflect the order of processing of global-and local-level information?" by Lamb, M. R., & Robertson, L. C. (1989), *Attention, Perception, & Psychophysics*, *46*(3), 254-258.

Two patterns of results were often taken as evidence for global precedence: a faster reaction time when targeting global information and a greater interference by global information to local target than the interference from local information to global target when the global and the local information conflicted (e.g., a large letter “A” is made up of small letters “S”).

However, a closer examination of this literature suggests a more nuanced interaction. Kinchla and Wolfe (1979) were the first to report that the global or the local advantage interacted with the size of the stimuli (see also, Lamb & Robertson, 1989; 1990; Mena, 1992; but see Navon & Norman, 1983). They had subjects indicate whether a specified letter was present or absent in a given stimulus that were varied in their size. Subjects responded faster to a large letter when the stimuli were small whereas they responded faster to small letters when the stimuli were large. In interpreting these results, they proposed a model of perceptual precedence based on the size or the spatial frequency of stimuli. The model postulates that there is a critical sampling bandwidth (range of size or spatial frequencies) from which the element that is initially processed is selected, and the processing of other elements (more global or local) occurs subsequently.

This interaction between the stimulus size and the element that receives a prioritized processing has been extended to a more complex, picture material as well. Antes and Mann (1984) presented subjects with a series of line drawings consisting of local elements (e.g., boat or tractor) and a global element (e.g., beach or farm). In half the drawings, the local and the global elements were thematically consistent while in the other half they were inconsistent (e.g., a boat in a farm; see Figure 12).

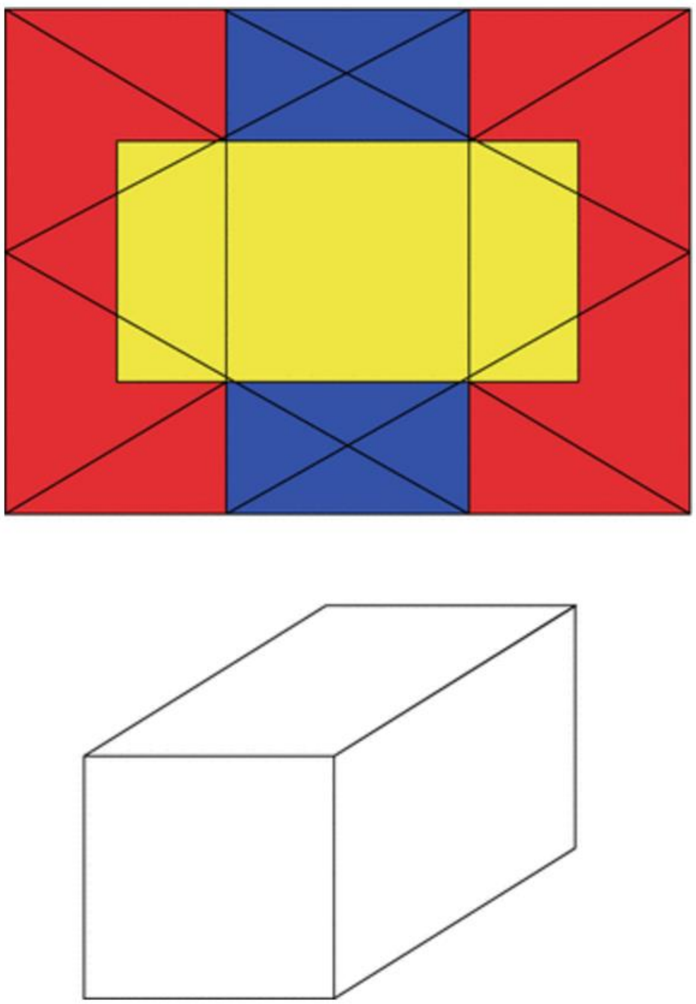
Figure 12. Examples of pictures used in Antes and Mann (1982). The global-local consistent pictures are on the left and the global-local inconsistent pictures are on the right. Reprinted from "Global-local precedence in picture processing," by Antes, J. R., & Mann, S. W. (1984), *Psychological Research*, *46*(3), 247-259.



Subjects responded to each drawing by answering a question, "Is this the (name of local or global element)?" The results showed a size-dependent local and global precedence, such that the subjects were faster to judge the local elements in large pictures, but the response time did not differ between when judging the local and the global elements in small pictures. In addition, the thematic inconsistency produced a greater interference when identifying the local element in small pictures (i.e., interference from the global element), but the opposite was true for large pictures; the thematic inconsistency did not produce any interference when identifying the local elements in large pictures while it greatly slowed down the reaction when identifying the global element.

A similar interaction has been also observed in other areas of cognitive psychology. For example, the performance in the Embedded Figure Test on which subjects identify a smaller simple figure embedded within a larger, more complex figure (see Figure 13) was enhanced in larger stimuli (Streibel & Ebenholtz, 1982).

Figure 13. A sample question from the Embedded Figures Test. The task is to identify the simple shape shown on bottom within the complex shape shown on top. Reprinted from "Embedded Figures Test (EFT)," by Happé, F. (2013), In *Encyclopedia of Autism Spectrum Disorders* (pp. 1077-1078). New York, NY: Springer.



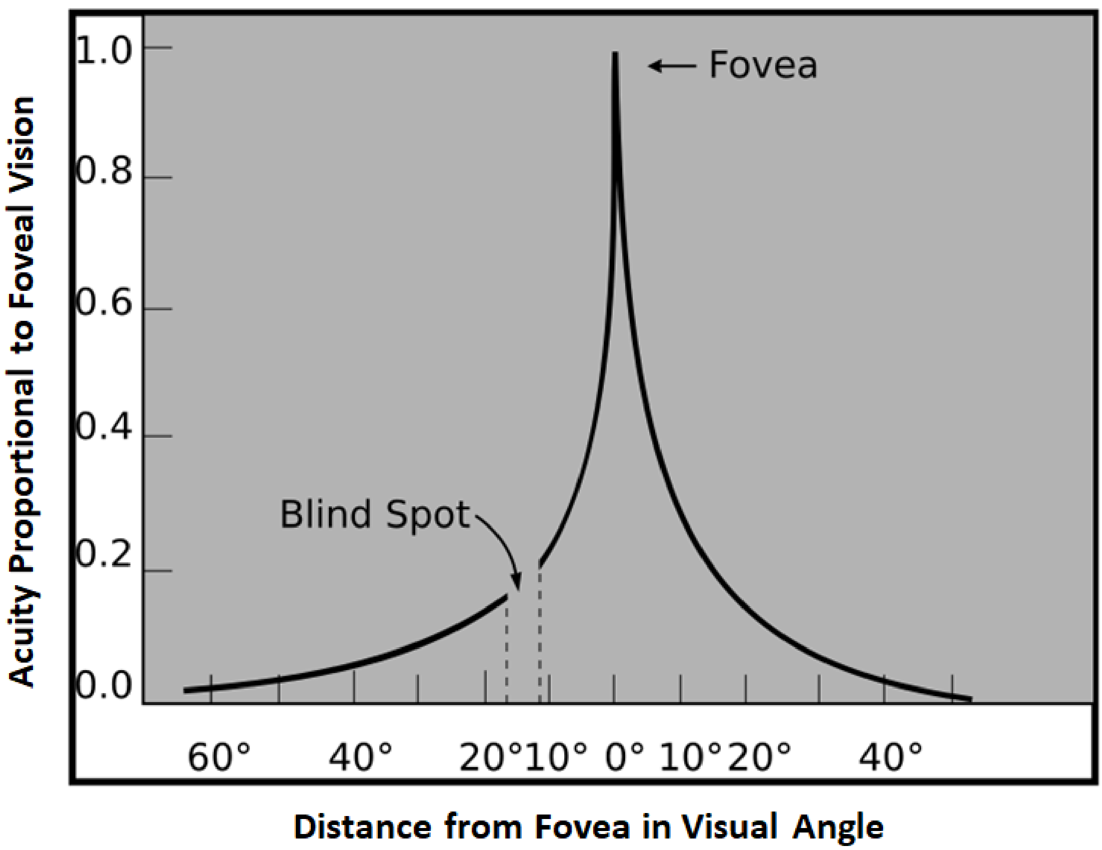
This finding can be interpreted as participants' attention drawn first to the local elements, shapes expressed in a part of a figure, in a larger size whereas the attention was first drawn to more encompassing shapes in a smaller size. Further, in the field of perceptual expertise using human face stimuli, smaller picture sizes have been shown to elicit holistic processing whereas larger sizes elicit processing of specific parts (Tanaka J. W., personal communication, February 2018).

This interaction between the stimulus size and the prioritized processing of the local or the global elements observed in the various fields and across many kinds of stimuli suggests the *perceptual precedence hypothesis* of the small-picture-size advantage on category learning. Specifically, when the fish pictures are presented in a large size, learners' attention is drawn first to the local dimensions (e.g., shape of a fin, color around the eye) because the size in which these features are expressed presumably fall within the critical bandwidth. On the contrary, when the fish pictures are presented in a small size, learners' attention is drawn first to the global dimensions (e.g., exterior shape, pattern on the whole body), and the processing of the local dimensions happens only after the processing of the global dimension is completed. Thus, within a given trial that lasted for only several seconds (5 seconds in the previous experiments), the global dimensions ended up being fully processed whereas the local dimensions were often only partially processed (but it did not matter because these dimensions were not diagnostic), resulting in the small-picture-size advantage in the previous experiments.

The original perceptual precedence theory (Kinchla & Wolfe, 1979) proposed a "middle-out" sequence (as opposed to the "top-down" sequence suggested by Navon; i.e., global precedence) through which different elements in a visual object are encoded and stated that an element that falls within "a critical bandwidth" is processed first. I expand this theory by postulating that this critical bandwidth is a product of some basic characteristics of human visual field and how the ease of extracting given features change because of them.

Human visual fields have two important characteristics that are relevant to the current, “middle-out” theory. First, the visual acuity decreases drastically as it gets further from the center of the visual field (i.e., fovea) and this is especially true for detecting finer features of an object (Hilz & Cavonius, 1974). Figure 14 shows the relative acuity of the human eye (left) on the horizontal meridian in degrees (visual angle) from foveal vision (Hunziker, 2006).

Figure 14. The relative acuity of the human eye (left) on the horizontal meridian in degrees visual angle from foveal vision. Reprinted from “Im Auge des Lesers: foveale und periphere Wahrnehmung-vom Buchstabieren zur Lesefreude. (The Eye of the Reader: Foveal and Peripheral Perception-from Letter Recognition to the Joy of Reading),” by Hunziker, H. W. (2006). Zürich, Switzerland: Stäubli.



At a distance of 30 inches (the distance between participants' eyes and the monitor in the previous experiments and a common distance between a desktop monitor and user's eyes), off-centering the target object by just 3 degree (i.e., 1.5 inches) would decrease the acuity in half. This limitation elicits gaze shifts in an attempt to capture objects that are significantly larger than the high acuity area on the center of the visual field, making it more difficult to encode such features compared to smaller ones. However, objects that are too small are also more difficult to encode because the visual acuity decreases as the target object becomes smaller (e.g., Mead, 1943), the characteristic that serves as the basis for the visual acuity test (i.e., the eye exam). Therefore, human visual field has a built-in middle-out system depending on the size of the target object.

Critically, the perceptual precedence hypothesis makes a unique prediction that has not been tested; if category learning took place using a category structure in which local dimensions were diagnostic, the large-picture size should be superior to the small-picture size (assuming that the size in which the diagnostic local dimension was expressed was within the critical bandwidth), or at the very least, the small-picture-size advantage should vanish. That is, this hypothesis postulates that the benefits of the picture-size manipulation is category-structure-specific. Specifically, category structures with a global diagnostic dimension(s) should benefit from the small-picture size because the order of processing elicited by the small size is appropriate for this category structure. Likewise, category structures with a local diagnostic dimension(s) should benefit from the large-picture size. These predictions will be tested in the experiments reported in the following chapter.

3.1.2 The Increased Effort Hypothesis

Another possible theoretical interpretation of the previous results is that the small-picture-size advantage emerged as a result of increased effort caused by perceptual disfluency. Perceptual disfluency manipulations, defined as perceptual manipulations that make the processing of the stimulus more difficult and slow down the encoding, have been shown to enhance learning of verbal materials. For example, presenting words upside down (i.e., uoịsɹǝʌuị: Sungkhasettee, Friedman, & Castel, 2011) and presenting texts in a more difficult to read font (e.g., Comic Sans MS in grey: Diemand-Yauman, Oppenheimer, & Vaughan, 2011) have been shown to enhance learning of word lists and text comprehension respectively (but see Eitel, Kühl, Scheiter, & Gerjets, 2014; Meyer et al. 2015; Yue, Castel, & Bjork, 2013). The idea is that the metacognitive sense of difficulty in reading a font that is slightly harder to read signals the learners that they do not have mastery over the material, and as a result, it invites a more effortful and analytical processing (Kuhl & Eitel, 2016; Yue, et al., 2013).

Although the perceptual disfluency has never been shown to enhance category learning, if a manipulation led to an increased effort, it seems possible that category learning would be enhanced. The small-picture size is a perceptual disfluency manipulation because in general, encoding of smaller objects is harder and slower than encoding of larger objects. For example, simple recognition of shapes, color patches, and letters are slower for smaller objects (Breitmeyer & Breier, 1994; Schultz & Eriksen, 1978; Sperandio, Savazzi, Gregory, & Marzi, 2009), and slower reaction time is a primary evidence of perceptual disfluency. In addition, the reduced sense of learning reported through CLJs and the post-experimental question in the previous experiments may be a reflection of this sense of disfluency experienced by the participants. Thus, this *increased effort hypothesis* postulates that the small-picture-size advantage emerges because the small pictures give a sense of disfluency, and that in turn engages learners in a more effortful processing of the stimuli.

Importantly, the increased effort hypothesis makes a contrasting prediction to the perceptual precedence hypothesis. It predicts that the small-picture size would enhance learning regardless of what the diagnostic dimension of a given category structure is. In other words, this hypothesis assumes that the benefit of the picture-size manipulation is category-structure-general. Specifically, the disfluency elicited by small pictures should lead to an increased effort regardless of whether the category structure has global or local diagnostic dimensions. In the next chapter, I will describe how these competing predictions by the two hypotheses were tested in the current experiments.

3.2 On the Metacognitive Accuracy in Category Learning

Because there has been only a handful of studies that investigated participants' online metacognitive understanding of their own category learning (i.e., CLJs; Doyle & Hourihan, 2016; Jacoby, et al., 2010; Wahlheim & DeSoto, 2016; Wahlheim, et al., 2011; Wahlheim, et al., 2012), theorization on the cognitive processes leading to metacognitive judgements in this situation is still limited. However, the well-established literatures on the nature of several types of metamemory judgements in the verbal learning tradition can give us ideas about the underlying processes when making similar metacognitive judgments in category learning.

3.2.1 The Direct-Access and Retrieval View

Historical views on the nature of metamemory assumed that learners had an internal monitor that could examine the degree of learning of a given material in a fairly unbiased manner (Hart, 1967; Burke, MacKay, Worthley, & Wade, 1991). For example, in experiments investigating the tip-of-the-tongue state (TOT: failure to recall information accompanied with successful retrieval of some surrounding information and a sense that the retrieval of the target information is imminent), learners can accurately predict which of the information that they could not recall could be correctly recognized if presented with the correct answers (e.g., Brown, 2011). That is, even when the target information was not successfully retrieved, learners could distinguish between the information that had a stronger representation in their memory from the weaker ones. This discriminative validity of such judgements suggests that learners can directly monitor the degree of learning. In addition to this subjective feeling of learning or strength of memory, many researchers have proposed a retrieval process to be a part of the metamemory judgements. For example, Benjamin (2008) described a two-process theory in which the subjective feeling (he called it matching) is followed by an explicit retrieval attempt of the target information. The retrieval fluency derived from such retrieval attempt (i.e., how readily the information comes to mind) has been shown to affect metacognitive judgments, and it ought to be useful in many situations because in general, information that is better learned is more readily retrievable (Benjamin & Bjork, 1994; c.f., Benjamin, Bjork, & Schwartz, 1998). Thus, the direct-access and retrieval view posits that learners can make accurate metacognitive judgments based on the subjective feeling of learning and the retrieval fluency derived from the explicit retrieval attempts. In the current scenario of category learning judgements (CLJs) in category learning, the feeling of familiarity and the sense of learning when presented with an example from a category and prompted to make a CLJ, combined with how readily another example or characteristic features of that category comes to mind, could provide sufficient information to make accurate CLJs. Therefore, this view suggests that although there seemed to be a persistent large-size bias in the previous experiments reported in the previous chapter, learners' metacognition can be accurate in some situations as will be investigated in the next chapter.

The extant literature on category learning using CLJs provides some support for this view. All three studies that examined the resolution of CLJs by correlating CLJs and classification performance at the participant-level showed a significant correlation between these measures when examined through Pearson correlation (Wahlheim & DeSoto, 2016: range: .24-.29) as well as Goodman and Kruskal's gamma (Jacoby, et al., 2010: range: .31-.33; Wahlheim, et al., 2011: range: .47-.49). In addition, participants' CLJs were sensitive to some manipulations that enhanced category learning, such as testing (Jacoby, et al., 2010) and interleaving (Wahlheim, et al., 2011). Thus, it seems possible that participants' CLJs will show some degree of discriminative validity when examined through correlation and be sensitive to the benefit of the picture-size manipulation when examined in different category structures as I will examine in the next chapter.

3.2.2 The Cue-Utilization View

Various modern frameworks of how metacognitive judgements are made put emphasis on the involvement of factors outside of the learning material itself as cues to infer the degree of learning. For example, Koriat's (1997) cue-utilization framework (see also Dunlosky & Metcalfe, 2009; Kelley & Jacoby, 1996) postulates that when making metacognitive judgments, learners incorporate extrinsic cues that pertain to the condition of learning and beliefs associated with them (e.g., the belief that items that were studied twice are more memorable than items that were studied once), in addition to intrinsic cues that are unique to the learning material (e.g., the difficulty of learning a particular item) and mnemonic cues that were experienced by the learners themselves (e.g., retrieval fluency). In line with this view and as briefly described earlier, a literature on the effect of the font-size manipulation on word-list learning and metacognition showed that learners gave higher ratings on judgments of learning (JOLs: post-learning prediction on later test performance) for the words that were presented in a larger font than a smaller font. This is because words in a larger font are processed more rapidly (i.e., increased perceptual fluency) and/or people have a belief that larger fonts are better for learning, and these types of information are used as cues in making inference on the degree of learning.

The results from the previous experiments clearly supported this view in that all studies showed that participants gave a higher CLJ ratings for categories that were presented in the large than the small size despite the actual performance was better for the categories that were presented in the small size. However, whether this effect applies to other situations, as will be investigated in the next chapter, is unclear. Specifically, does this effect extend to categories other than fish and when the category structure is designed so that the large-picture size produces better learning? On the one hand, it is possible that the utilization of the extrinsic and mnemonic cues, such as the perceptual fluency and the belief associated with the large-picture size, continue to influence CLJs even when the actual performance is better for the large size. If that is the case, the CLJ ratings will be higher for the large size above and beyond the degree of the final test performance. On the other hand, it is also possible that when these cues and the actual performance point to the same direction, their effects will be sub-additive, so that CLJ ratings will not significantly differ from the classification performance.

Chapter 4: Current Experiments

To reiterate, the core idea of the current experiments is to test the unique predictions made by the perceptual precedence hypothesis and the increased effort hypothesis regarding the small-picture-size advantage in category learning. All participants in the following two experiments will study half of the example pictures in the small and the other half in the large size. Critically, they will be randomly assigned to study a category structure with different types of diagnostic dimensions; in each experiment, one group of participants will study a category structure in which a local dimension(s) is diagnostic while another group of participants will study a category structure in which a global dimension is diagnostic[2]. I attempted to demonstrate the generality of the hypothesized findings by testing categories across two classes: animate (fish and tern) and inanimate (orchid and rocks). This is because people process animate and inanimate objects differently (e.g., animate pictures are remembered better: Bonin, Gelin, & Bugaiska, 2014; see Nairne, VanArsdall, & Cogdill, 2017 for a review), and thus category learning involving these classes of objects could be qualitatively different. As outlined above, the perceptual precedence hypothesis predicts that the small-picture-size advantage to emerge when the categories are defined by a global dimension, but it also predicts a large-picture-size advantage to emerge when the categories are defined by a local dimension(s). On the contrary, the increased effort hypothesis predicts that the small-picture size should enhance learning regardless of the globality of the diagnostic dimension(s).

---

[2] I initially attempted to examine the same hypotheses by using artificial fish category structures created by combining the shapes and patterns of different fish in which the diagnostic dimension was either global (Experiment 1: exterior shape; Experiment 2: pattern on the body) or local dimension (Experiment 1: the shape of the caudal fin; Experiment 2: the pattern on the caudal fin; see Appendices A and B for these materials and brief descriptions of the results). However, several pilot studies failed to show the picture-size effect, and thus I decided on using the naturalistic stimuli as presented here.

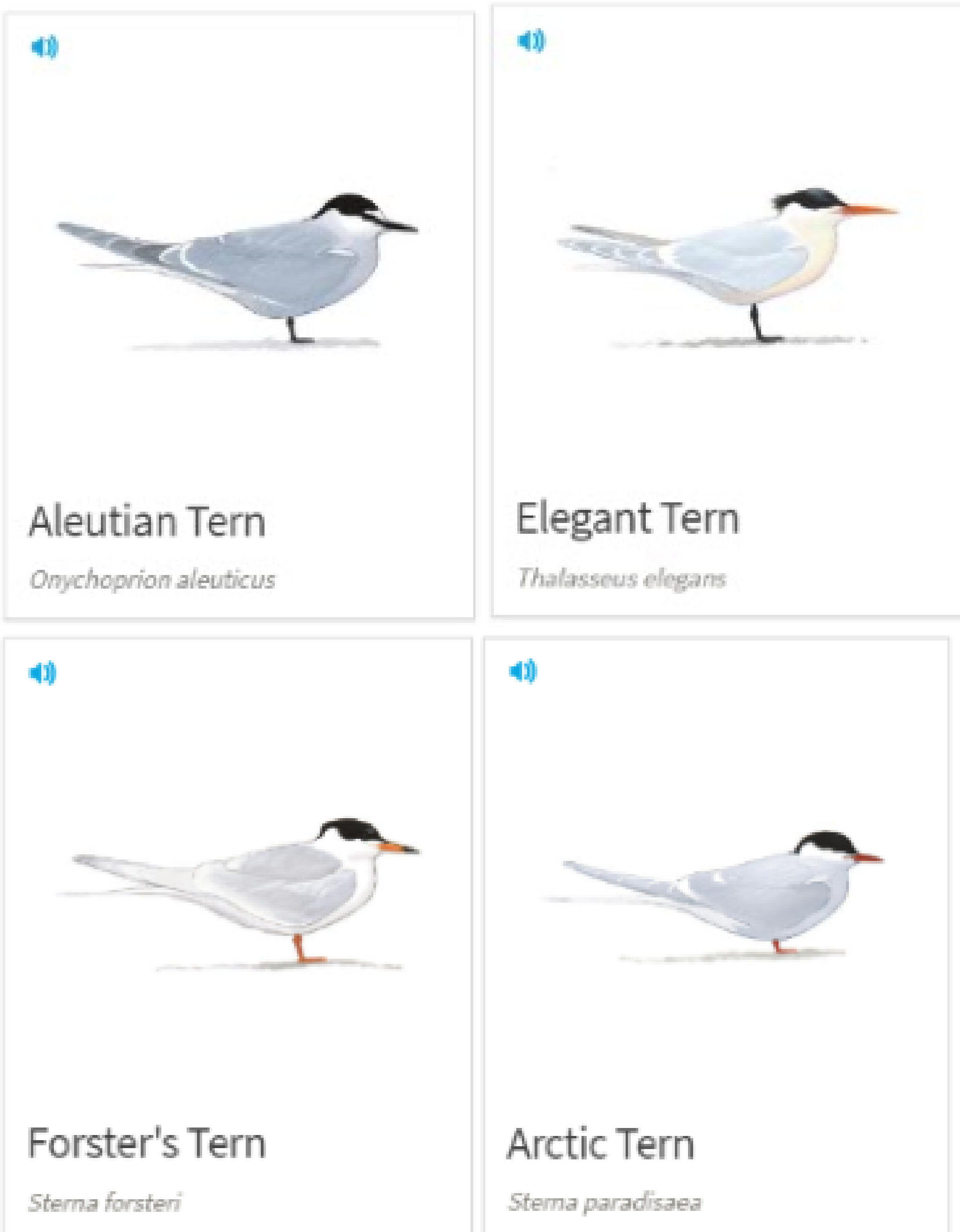
Four natural category structures (animate-global diagnostic, animate-local diagnostic, inanimate-global diagnostic, and inanimate-local diagnostic) were carefully assembled to serve as the learning material for the current experiments. In addition to the fish categories which will be studied in the animate-global diagnostic condition, I attempted to identify three more natural category structures with the most desired distribution of diagnosticity across their dimensions (i.e., a clear global or local diagnostic dimension or dimensions) by consulting classification experts in the field who had intimate knowledge of how visual classification was made in a given domain.

First, in order to identify an animate category structure with a local diagnostic dimension, I contacted dozens of biological scientists specialized in animal behavior and classification. Of whom, Dr. Zuleyma Tang-Martinez[3], a professor emeritus in the Biology Department of University of Missouri, St. Louis, and an animal behavior expert, gave me the most helpful information. She pointed out that the exterior shape is diagnostic in most animal classification, but some species of terns look very similar in shape and the details expressed in certain parts of their body (i.e., forehead, bill, legs, the shape of the back of the head) determine the categorization. Figure 15 shows a few examples of the tern categories included in the current experiment. Although they appear very similar at a glance, local features define each category; Aleutian terns are primarily identified by the white patch on the forehead, elegant terns by the elongated feathers on the back of the head, Forster's terns by the black tip on a beak, and arctic terns by not having these features.

Figure 15. Examples of the tern categories that were used in Experiment 1 in which local features (e.g., beak, forehead, back of the head) are diagnostic.

---

[3] https://www.umsl.edu/~biology/About%20the%20Department/Faculty/tang.html



For the inanimate category structure with a global diagnostic dimension, I contacted scientists specialized in plant biology and identification. Of whom, Dr. Peter Bernhardt[4], a professor in the Department of Biology at St. Louis University, and a pollination biology expert,

[4] https://www.slu.edu/arts-and-sciences/biology/faculty/bernhardt-peter.php

pointed to some hybrid orchid species among which visual classifications are primarily made according to their exterior shape[5]. Figure 19 shows the orchid stimuli. Note that as these are hybrid species, the color and pattern vary widely among the instances within a category, making the exterior shape the primary diagnostic dimension.

Finally, for the inanimate category structure with a local diagnostic feature, I chose rock categories with which I have previously conducted several studies (e.g., Miyatsu, et al., 2019; Miyatsu, et al., in press). Rock categories are complex natural categories that have many dimensions as identified by computational modeling based on similarity rating, and learners seem to use all these dimensions to classify new instances (e.g., Nosofsky, et al., 2017; Nosofsky, Sanders, Meagher, & Douglas, 2018). However, the classification scheme provided by geo-science experts (e.g., Miyatsu, et al., 2019) indicated that the most prominent diagnostic dimensions are color, grain-size, and texture. One may consider these dimensions as global because these dimensions are often expressed throughout a rock. However, information represented in these dimensions is highly consistent across a stimulus. That is, any small part of a rock would look very similar to other small parts from the same rock, and only a small part of a rock needs to be encoded in detail to extract that the rock has small grains or glossy texture. Thus, I consider these dimensions to be local in the current framework.

[5] He also pointed out that the shape of the sexual organs can be diagnostic in these species, but these organs are often concealed and not readily visible in a single, front-view picture.

4.1 Experiment 1

4.1.1 Method

Design. A 2 x 2 mixed-factorial design, with the size of the example pictures (small or large) being the within-subjects variable and the categories to be learned (fish or tern) being the between-subjects variable, was employed.

Participants. Sixty undergraduates from Washington University in St. Louis (30 each in the local and the global diagnostic conditions; 57% female, M_{age} = 20.0) participated in each experiment. The sample size was determined by a priori power analysis to give me an extremely high power (.99) to detect a medium size main effect of the size manipulation (f = 0.31, r = .44)[6] as well as a high power (.95) to detect a medium size interaction (f = .25) and an adequate power (.82) to detect a medium-small interaction (f = .20).

Materials. The material was 100 tropical fish pictures (10 examples each for 10 categories) and 100 tern pictures[7]. Figures 16 and 17 show the pictures of fish and terns used in Experiment 1. The fish pictures were taken from the previous experiments dealing with family-level classification. The tern pictures were assembled through a web search according to the information provided by the animal classification expert as described above. All pictures were scaled similarly and pasted on a white background.

Figure 16. Examples of the fish stimuli that were used in Experiment 1. Each row represents a category (i.e., species).

[6] The effect size and the correlation between the performance on small and large items were calculated by meta-analyzing data from the previous experiments.

[7] A high-resolution version of all the pictures used in the current experiments and the previous experiments as well as data from all of these experiments can be accessed at Open Science Framework at osf.io/r6k4t

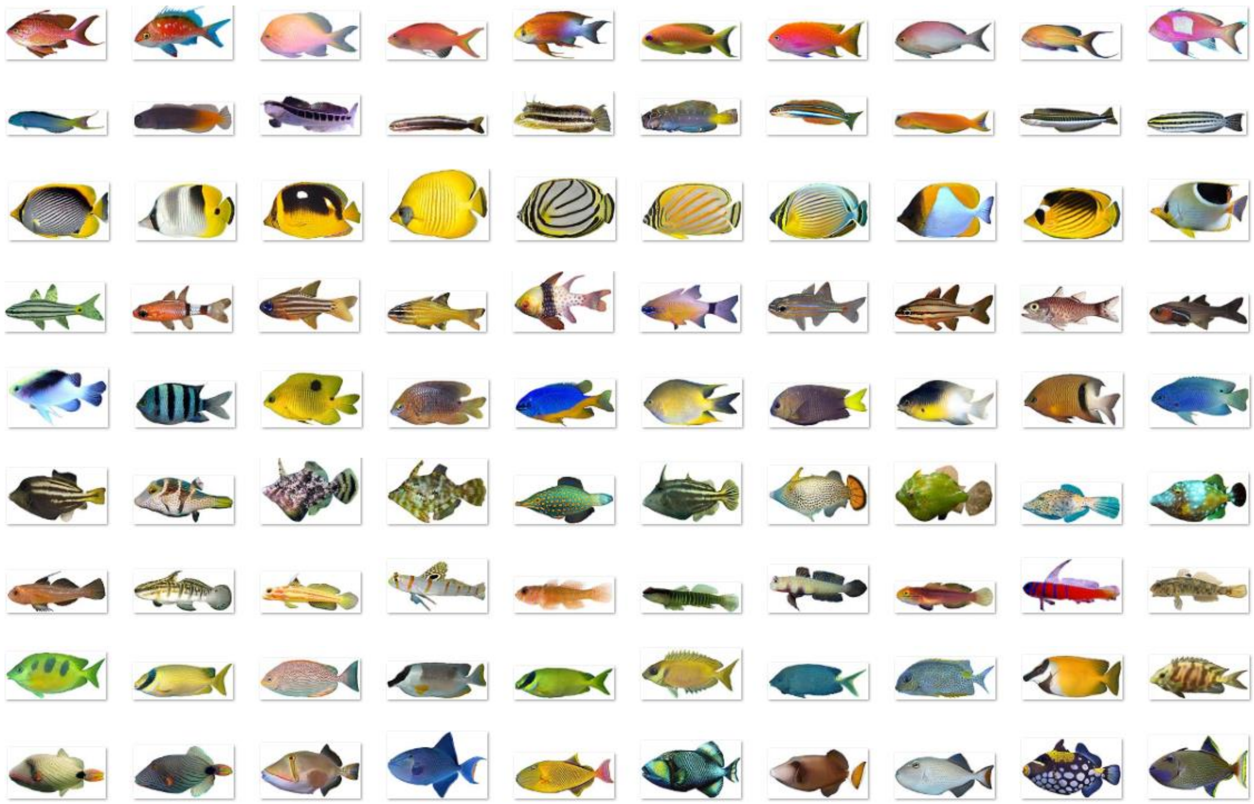


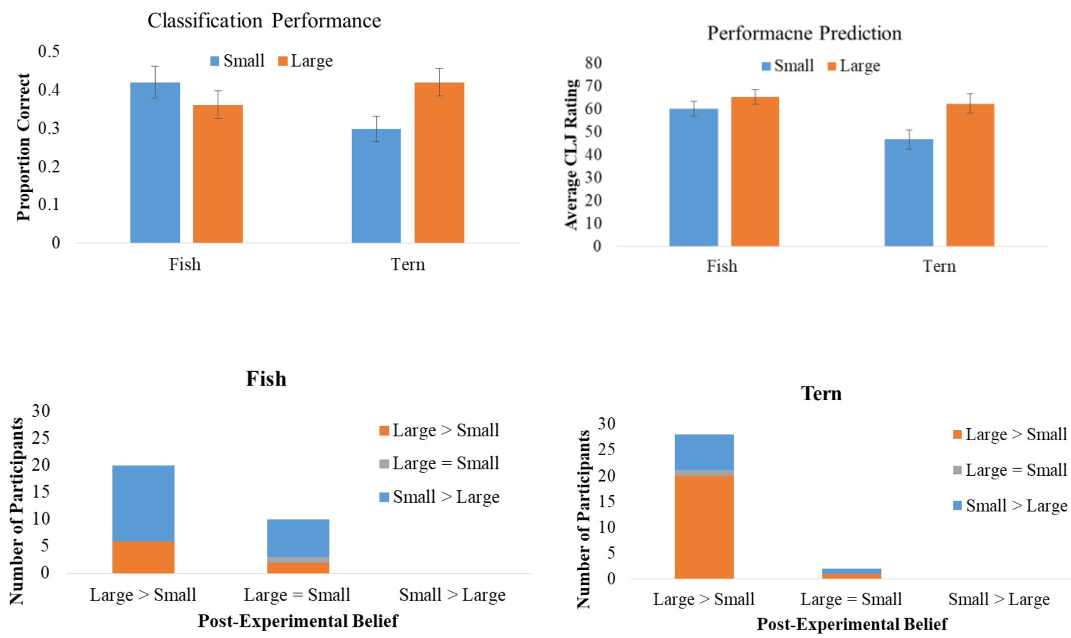
Figure 17. Examples of the tern stimuli that were used in Experiment 1. Each row represents a category (i.e., species).

Procedure. First, all participants were seated in front of a computer and asked not to change posture drastically for the entire duration of the experiment, such as putting the legs on the table or get closer to the screen to see the details of small pictures, because these actions can compromise the intended visual angle at which the presented pictures are encoded. And then, the participants studied ten categories of fish or tern by observing six examples from each category one by one for 5 seconds each. For each participant, six training examples out of ten available examples from each category were randomly chosen, and the reminding four examples served as the test items. The example pictures from half of the categories were presented in a small size (about 2 degree of visual angle: a fish or a tern picture of 1 inch in width viewed from 30 inches) whereas the other half were presented in a large size (about 19 degree of visual angle: a fish or a tern picture of 10 inches in width viewed from 30 inches). The assignment of each category to the small or the large size were counterbalanced so that each category was presented in the small and the large size equally often. The pictures were presented in a block-randomized interleaved sequence, such that each block consisted of an example from each of the ten categories in a random order uniquely created for each participant. Upon studying the sixth and last example from each category, the participants were asked to make category learning judgments (CLJs) by answering the following question on a scale of 0 to 100: *On a scale of 0 to 100, how confident are you that you will be able to correctly categorize a new member of this particular category* *during a later test?* After studying and making the CLJs for all ten categories, the participants played tetris for 3 minutes as a distractor task. The participants then completed the final test in which they were presented with 40 new pictures, four pictures each from the ten categories, in a neutral size (about 11º) and were asked to classify them by clicking one of the ten options labeled with the names of the ten learned categories. The order of the final test was block-randomized, such that each of the four test blocks consisted of one example from each of the ten categories. The participants were given 10 seconds for each test trial. Finally, the participants were asked which picture size they thought helped them learn the categories better before being debriefed and left the laboratory.

4.1.2 Results

Classification performance. The top-left panel of Figure 18 shows participants' mean performances on the final test according to their conditions.

Figure 18. Participants' mean classification performance (top-left panel), mean CLJ ratings (top-right panel), and post-experimental judgement on which picture size they learned better from (bottom panels), from Experiment 1. The error bars represent ± 1 *SEM*.



A 2 X 2 mixed factorial analysis of variance (ANOVA), with the picture size (small or large) as the within-subjects variable and the condition (fish or tern) as the between-subjects variable, was conducted on these data. Neither the main effect of picture size, $F(1, 58) = 2.39$, $p > .05$, $\eta p^2 = .04$, nor the main effect of condition, $F(1, 58) = 0.44$, $p > .05$, $\eta p^2 = .01$, was significant. However, there was a significant interaction between these two variables, $F(1, 58) = 19.37$, $p < .001$, $\eta p^2 = .25$. Post-hoc paired-samples t-tests showed that the participants in the fish condition performed significantly better in small ($M = .42$, $SD = .23$) compared to large pictures ($M = .36$, $SD = .19$), $t(29) = 2.65$, $p < .05$, $d = 0.50$. Twenty-one out of 30 participants showed this small-picture-size advantage. On the contrary, the participants in the tern condition performed better in large ($M = .42$, $SD = .20$) compared to small ($M = .30$, $SD = .18$) pictures, $t(29) = 3.53$, $p < .01$, $d = 0.64$. Also, 21 out of 30 participants showed this large-picture-size advantage.

Category learning judgments (CLJs). The accuracy of CLJs was assessed in three ways. First, the participants' sensitivity to the size manipulation and the category structure (i.e., fish or tern condition) was assessed by a size-by-condition mixed-ANOVA to see if the pattern lined up with the same analyses performed on the classification performance. Second, the match between the patterns of results from classification performance and CLJs in each condition was assessed by separate size-by-outcome (i.e., classification or CLJ) within-subjects ANOVAs[8] for the fish and the tern condition. Third, monitoring resolution was assessed by computing mean within-participant gamma correlation between CLJs and classification performance for each category (Wahlheim, Dunlosky, & Jacoby, 2011). In essence, Goodman and Kruskal's gamma is preferred over other approaches (e.g., Pearson product-moment correlation) in this and other situations dealing with similar metacognitive judgements, such as judgments of learning (JOLs;

[8] The CLJ ratings, which were given in the scale of 0 to 100, were divided by 100 in this analysis to match the scale with the classification performance.

e.g., Nelson, & Dunlosky, 1991) and feeling of knowing (FOKs; e.g., Metcalfe, 1986) because the gamma correlation is a rank correlation, and as such it is immune to the effect from some participants using the rating scale differently from others (e.g., liberal or conservative use of the upper or the lower range; see also Nelson, 1984).

The top-right panel of Figure 18 shows participants' mean performance prediction (CLJs) according to their conditions. A 2 X 2 mixed factorial ANOVA, with the picture size (small or large) as the within-subjects variable and the condition (fish or tern) as the between-subjects variable, was conducted on these data. There was a significant main effect of picture size, such that categories presented in the large size were judged to be learned better, $F(1, 58) = 23.31$, $p < .001$, $\eta p^2 = .29$, but the main effect of condition was not significant, $F(1, 58) = 3.42$, $p > .05$, $\eta p^2 = .06$. In addition, there was a significant interaction between these two variables, $F(1, 58) = 5.66$, $p < .05$, $\eta p^2 = .09$. A glance at the means shows that the participants gave higher CLJ ratings for large than small pictures in both fish (large: $M = 65.90$, $SD = 17.97$ vs small: $M = 60.80$, $SD = 17.37$) and tern conditions (large: $M = 61.99$, $SD = 22.35$ vs small: $M = 46.99$, $SD = 22.62$). Post-hoc paired-samples t-tests showed that while this difference was only marginally significant in the fish condition, $t(29) = 1.74$, $p < .10$, $d = 0.32$, it was fully significant in the tern condition, $t(29) = 5.08$, $p < .001$, $d = 0.93$. Seventeen out of 30 participants in the fish condition and 25 out of 30 participants in the tern condition gave higher CLJ ratings for large pictures.

A 2 X 2 within-subjects factorial ANOVA, with the picture size (small or large) and the outcome (classification or CLJs) as the independent variables, was conducted separately for the fish and the tern conditions. For the fish condition, the main effect of picture size was not significant, $F(1, 29) < 1$, $p > .05$, $\eta p^2 = .00$, but the main effect of outcome was significant, such that the participants gave higher CLJ ratings ($M = .63$, $SD = .22$) than the actual classification performance (M = .39, SD = .29), $F(1, 29) = 59.50$, $p < .001$, $\eta p^2 = .67$. In addition, there was a significant interaction between these two variables, $F(1, 29) = 8.01$, $p < .01$, $\eta p^2 = .22$. For the tern condition, there was a significant main effect of picture size, such that the average of the classification performance and the CLJ ratings were higher for the large (M = .52, SD = .26) compared to the small pictures (M = .38, SD = .25), $F(1, 29) = 23.15$, $p < .001$, $\eta p^2 = .44$, as well as a significant main effect of outcome, such that the participants gave higher CLJ ratings (M = .55, SD = .29) than the actual classification performance (M = .36, SD = .23), $F(1, 29) = 26.89$, $p < .001$, $\eta p^2 = .48$. Importantly, the interaction between these two variables was not significant, $F(1, 29) < 1$, $p > .05$, $\eta p^2 = .03$, indicating that the degree to which the participants assigned higher CLJ ratings for large pictures did not exceed the degree to which they performed better in large pictures.

The average gamma correlation between participants' CLJs for each category and the corresponding classification performance was .35 (SD = .38). A one-sample t-test indicated that it was significantly above chance, $t(59) = 7.12$, $p < .001$, $d = 0.93$. An independent-sample t-test indicated that the degree of gamma correlation did not differ between the fish (M = .35, SD = .37) and the tern conditions (M = .35, SD = .40), $t(59) = 0.02$, $p > .05$, $d = 0.00$.

The post-experimental question. The bottom panels of Figure 18 show the number of participants who claimed to have learned better from large pictures, small pictures, or both the same, in the fish and the tern conditions respectively. In the fish condition, 20 participants claimed to have learned better from the large pictures, 10 participants claimed that there was no difference between the small and the large pictures, and 0 participant claimed that they have learned better from the small pictures. In the tern condition, 28 participants claimed to have learned better from the large pictures, 2 participants claimed that there was no difference between the small and the large pictures, and 0 participant claimed that they have learned better from the small pictures.

4.1.3 Discussion

To begin, in line with both the perceptual precedence and the increased effort hypotheses, the small-picture-size advantage in the classification of the fish categories was replicated. Importantly, however, in line with the perceptual precedence hypothesis but not with the increased effort hypothesis, there was a large-picture-size advantage in the tern condition. Before I discuss the implications of this critical interaction, I will present another experiment dealing with inanimate categories to assess the generality of this finding. As described before, prior research has shown that people process animate objects differently from inanimate objects (e.g., Bonin, et al., 2014; Nairne, et al., 2017), and thus it is possible that this pattern of results is restricted to animate categories.

The results from the metacognitive measures replicated some key findings in the literature and extended them further. In the fish condition, a large-picture-size bias akin to the font-size bias in word-list learning (e.g., Rhodes & Castel, 2008) was observed; the participants gave higher CLJ ratings for the categories that were presented in the large size despite the small size being better for actual learning. Further, similar to Kornell and Bjork's (2008) study investigating the effect of interleaving and blocking, the participants in the fish condition also reported that they learned better from the large pictures after the final test despite many of them actually performing better in the small pictures. After several replications of this effect in the previous experiments, it is still stunning that none of the 21 participants who actually performed better in the small pictures in the fish condition (the blue part of the bars in the bottom-left panel of Figure 18) believed that small pictures were better for learning the categories. In the tern condition, the participants also claimed to have learned better from the large pictures both after learning (i.e., CLJ) and after the final test (i.e., the post-experimental question). However, unlike the fish condition, these metacognitive assessments were in line with the actual performance because the tern categories were learned better in the large size. Interestingly, the non-significant size-by-outcome interaction in the within-subjects ANOVA run separately for the tern condition indicated that the pattern of results did not differ between the classification performance and CLJs in this condition. That is, the degree to which the participants assigned higher CLJ ratings to large pictures did not exceed the degree of the large-picture-size advantage in the actual learning.

Despite the illusory prediction in the fish condition that the large pictures were better for learning, CLJ clearly had predictive validity as demonstrated by its above-chance average within-participant Goodman and Kruskal's gamma correlation. The overall gamma of .35 was comparable to the previous studies using naturalistic categories and CLJs (Jacoby, et al., 2010: range: .31-.33; Wahlheim, et al., 2011: range: .47-.49) as well as the same measure computed from the data from the previous experiments (range: .33-.38) and indicated that the participants' CLJs were sensitive to the difference in difficulty of learning each category.

In sum, the classification performance showed the category-structure-specific small- and large- picture-size advantage, providing a preliminary support for the perceptual precedence hypothesis. In addition, in line with the cue-utilization view, persisting metacognitive illusion was observed through CLJs and the post-experimental question in one condition; the participants in the fish condition claimed to have learned better from the large pictures both before and after the final classification test. However, in line with the direct access and retrieval view, some degree of metacognitive accuracy was also demonstrated; the participants in the tern condition correctly claimed to have learned better from large pictures both in CLJs and in the post-experimental question, and CLJs predicted the actual performance at a rate that was well-above the chance.

4.2 Experiment 2

4.2.1 Method

Design. A 2 x 2 mixed-factorial design, with the size of the example pictures (small or large) being the within-subjects variable and the categories to be learned (orchid categories with global diagnostic dimension or rock categories with local diagnostic dimension) being the between-subjects variable, was employed.

Participants. Sixty undergraduates from Washington University in St. Louis (30 each in the local and the global conditions; 62% female, M_{age} = 20.1) participated in each experiment.

Materials. The material was 120 (10 examples each for 12 categories) orchid pictures and 120 rock pictures. The orchid pictures were assembled through a web search according to the guidance provided by a plant biology expert as described above. The rock pictures were taken from previous studies dealing with rock classification (Miyatsu, et al., 2019; Miyatsu, et al., in press). All pictures were scaled similarly and pasted on a white background. Figure 19 and 20 show the pictures of orchid and rocks used in Experiment 1.

Figure 19. Examples of the orchid stimuli that were used in Experiment 2 in which the exterior shape is the primary diagnostic dimension. Each row represents a category (i.e., species).

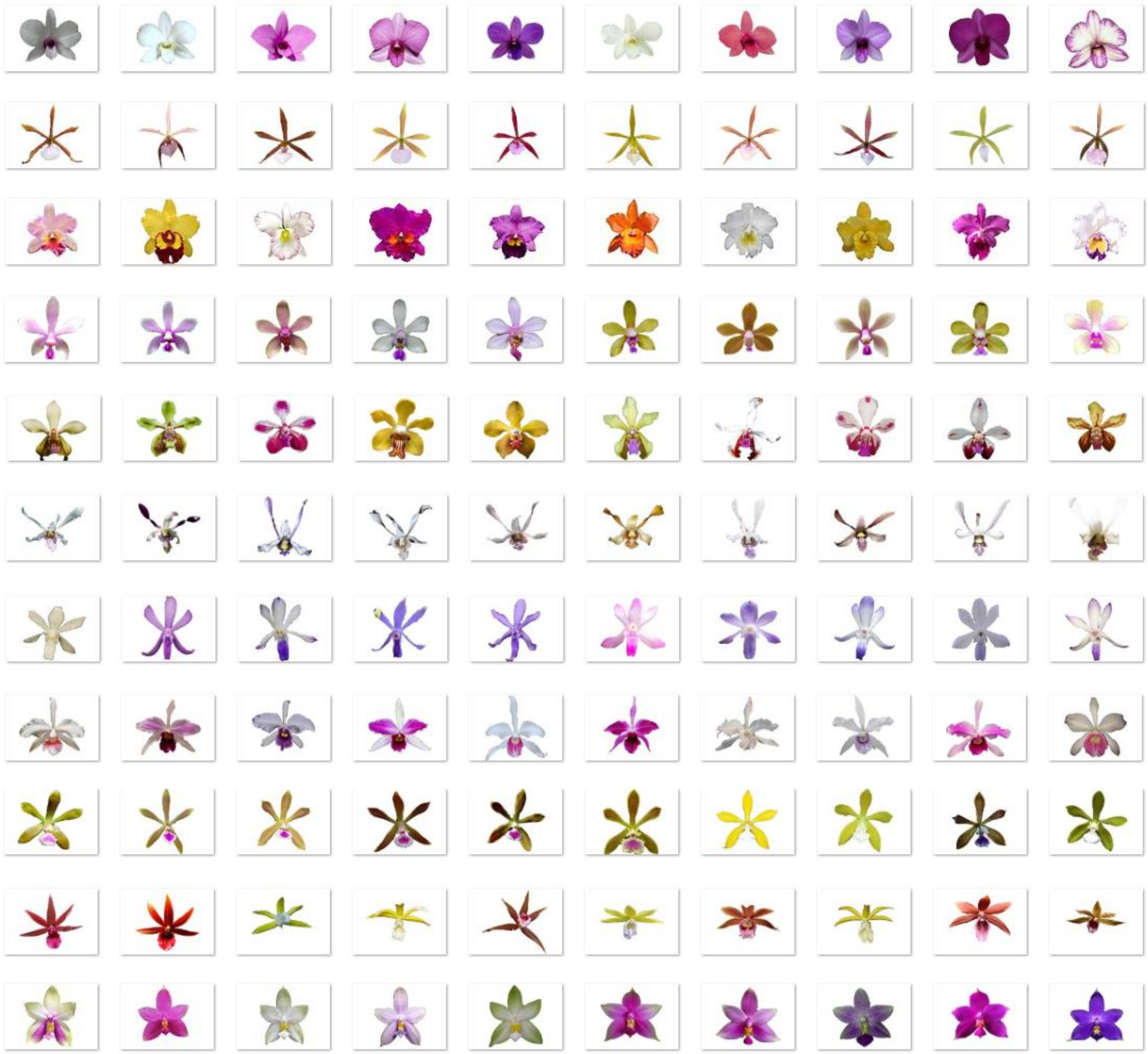
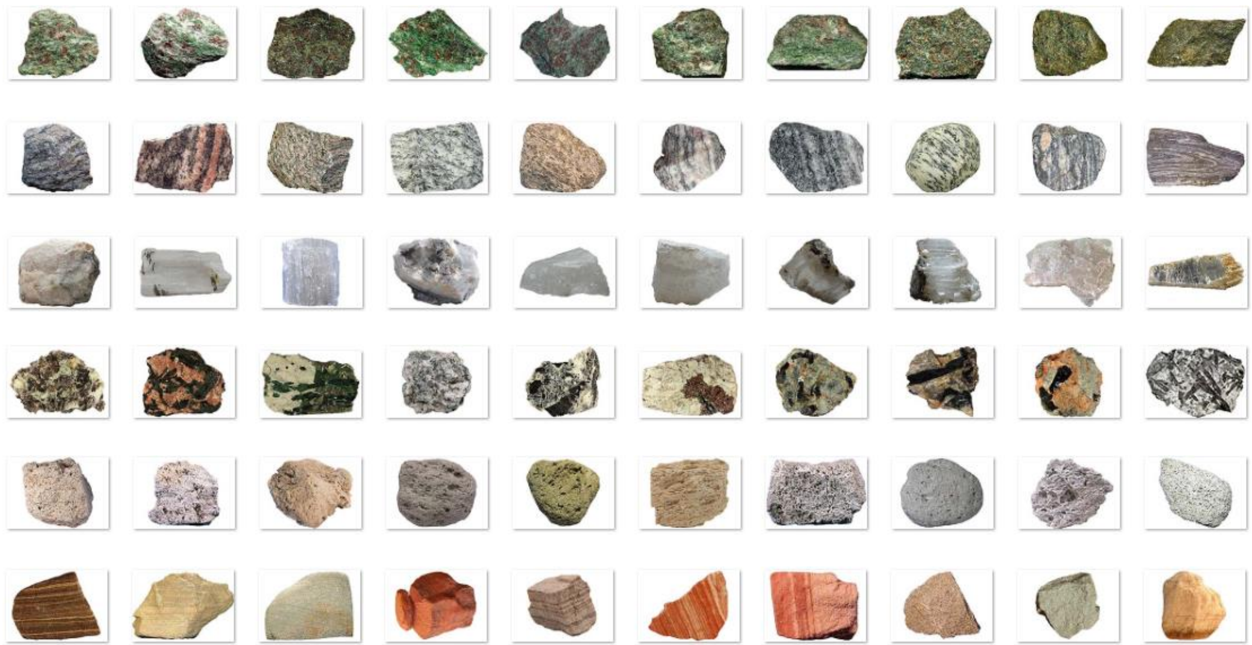


Figure 20. Examples of the rock stimuli that were used in Experiment 2. Each row represents a category (i.e., species).

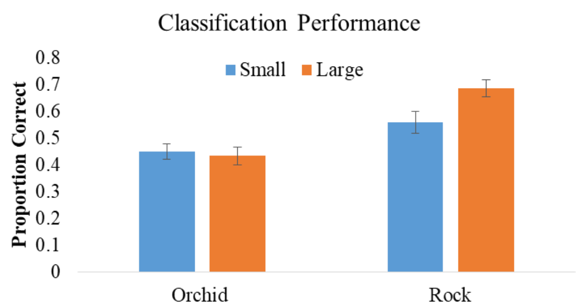


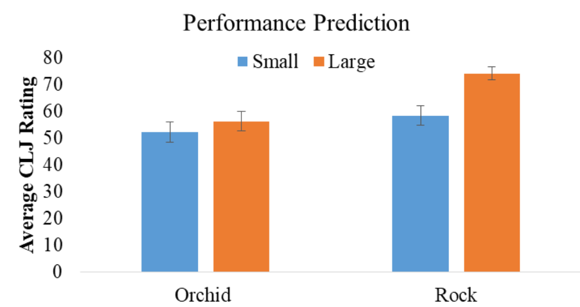
Procedure. Experiment 2 procedure was identical to Experiment 1 except the participants learned orchid or rock categories instead of fish or terns, and the total number of categories in each condition was 10 instead of 12.

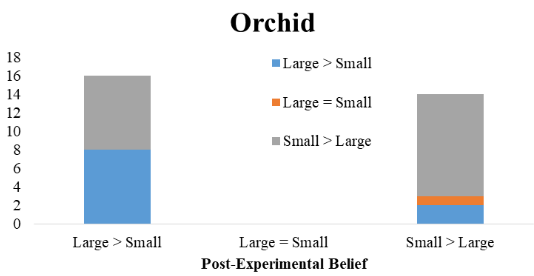
4.2.2 Results

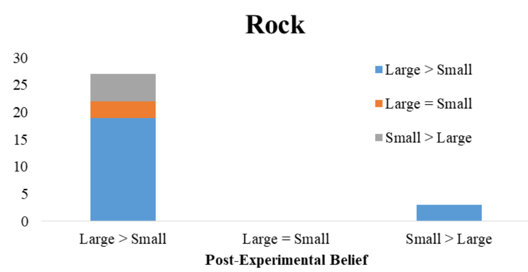
Classification performance. The top-left panel of Figure 21 shows participants' mean performances on the final test according to their conditions.

Figure 21. Participants' mean classification performance (top-left panel), mean CLJ ratings (top-right panel), and post-experimental judgement on which picture size they learned better from (bottom panels), from Experiment 2. The error bars represent ± 1 *SEM*.









A 2 X 2 mixed factorial ANOVA, with the picture size (small or large) as the within-subjects variable and the condition (orchid or rock) as the between-subjects variable, was conducted on these data. There was a significant main effect of picture size, such that categories presented in the large size were learned better, $F(1, 58) = 6.63$, $p < .05$, $\eta p^2 = .10$, as well as a significant main effect of condition, such that the participants in the rock classification condition performed better than the ones in the orchid classification condition, $F(1, 58) = 16.88$, $p < .001$, $\eta p^2 = .23$. However, these main effects were qualified by a significant interaction between the two variables, $F(1, 58) = 10.72$, $p < .01$, $\eta p^2 = .16$. Post-hoc paired-samples t-tests showed that the participants in the orchid condition performed similarly in small ($M = .45$, $SD = .17$) and large ($M = .43$, $SD = .18$) pictures, $t(29) = .54$, $p > .05$, $d = 0.13$. In contrast, the participants in the rock condition performed better in large ($M = .69$, $SD = .18$) compared to small ($M = .56$, $SD = .22$) pictures, $t(29) = 3.84$, $p < .01$, $d = 0.72$. Nineteen out of 30 participants in the orchid condition performed better in small pictures while 22 out of 30 participants in the rock condition performed better in large pictures.

Category learning judgments (CLJs). The top-right panel of Figure 21 shows participants' mean performance prediction (CLJs) according to their conditions. A 2 X 2 mixed factorial ANOVA, with the picture size (small or large) as the within-subjects variable and the condition (orchid or rock) as the between-subjects variable, was conducted on these data. There was a significant main effect of picture size, such that categories presented in the large size were judged to be learned better, $F(1, 58) = 34.84$, $p < .001$, $\eta p^2 = .38$, as well as a significant main effect of condition, such that the participants in the rock condition gave higher CLJ ratings than the ones in the orchid condition, $F(1, 58) = 7.10$, $p < .01$, $\eta p^2 = .11$. However, these main effects were qualified by a significant interaction between the two variables, $F(1, 58) = 11.98$, $p < .01$, $\eta p^2 = .17$. Similarly to Experiment 1, a glance at the means shows that the participants gave higher CLJ ratings for large than small pictures in both orchid (large: $M = 56.36$, $SD = 19.63$ vs small: $M = 52.25$, $SD = 20.15$) and rock conditions (large: $M = 74.23$, $SD = 13.53$ vs small: $M = 58.48$, $SD = 20.57$). Post-hoc paired-samples t-tests showed that while this difference was only marginally significant in the orchid condition, $t(29) = 1.94$, $p < .10$, $d = 0.35$, it was fully significant in the rock condition, $t(29) = 6.02$, $p < .001$, $d = 1.10$. Sixteen out of 30 participants in the orchid condition and 27 out of 30 participants in the rock condition gave higher CLJ rating for categories that were presented in large pictures.

A 2 X 2 within-subjects factorial ANOVA, with the picture size (small or large) and the outcome (classification or CLJs) as the independent variables, was conducted separately for the fish and the tern conditions. For the orchid condition, the main effect of picture size was not significant, $F(1, 29) < 1$, $p > .05$, $\eta p^2 = .01$, but the main effect of outcome was significant, such that the participants gave higher CLJ ratings (*M* = .54, *SD* = .27) than the actual classification performance (*M* = .44, *SD* = .22), $F(1, 29) = 8.45$, $p < .01$, $\eta p^2 = .23$. In addition, there was a significant interaction between these two variables, $F(1, 29) = 4.59$, $p < .05$, $\eta p^2 = .14$. Post-hoc paired-samples t-tests showed that while the classification performance did not differ between small and large pictures (small: *M* = .45, *SD* = .17 vs large: *M* = .43, *SD* = .18), $t(29) = 0.54$, $p > .05$, $d = 0.13$, the CLJ ratings showed a marginally significant large-size bias (small: *M* = .52, *SD* = .20 vs large: *M* = .56, *SD* = .20), $t(29) = 1.94$, $p = .06$, $d = 0.34$. For the rock condition, there was a significant main effect of picture size, such that the average of the classification performance and the CLJ ratings were higher for the large (*M* = .71, *SD* = .19) compared to the small pictures (*M* = .57, *SD* = .28), $F(1, 29) = 31.80$, $p < .001$, $\eta p^2 = .52$, but the main effect of outcome did not reach significance, $F(1, 29) = 2.56$, $p = .12$, $\eta p^2 = .08$. Importantly, the interaction between these two variables was not significant, $F(1, 29) < 1$, $p > .05$, $\eta p^2 = .03$, indicating that the degree to which the participants assigned higher CLJ ratings for large pictures did not exceed the degree to which they performed better in large pictures.

The average gamma correlation between participants' CLJs for each category and the corresponding classification performance was .42 (SD = .33). A one-sample t-test indicated that it was significantly above chance, $t(59) = 9.95$, $p < .001$, $d = 1.28$. An independent-sample t-test indicated that the degree of gamma correlation did not differ between the orchid (*M* = .38, *SD* = .33) and the rock conditions (*M* = .46, *SD* = .33), $t(59) = 0.96$, $p > .05$, $d = 0.25$.

The post-experimental question. The bottom panels of Figures 21 show the number of participants who claimed to have learned better from large pictures, small pictures, or both the same, in the orchid and the rock conditions respectively. In the orchid condition, 17 participants claimed to have learned better from the large pictures, 7 participants claimed that there was no difference between the small and the large pictures, and 6 participants claimed that they have learned better from the small pictures. In the rock condition, 22 participants claimed to have learned better from the large pictures, 8 participants claimed that there was no difference between the small and the large pictures, and 0 participant claimed that they have learned better from the small pictures.

4.2.3 Discussion

Similarly to Experiment 1 using the animate categories, the data from the classification performance showed the critical interaction between the picture size and the category structure (i.e., global or local diagnostic) demonstrating that the effect of the picture-size manipulation depends on the category structure to be learned. However, unlike the cross-over interaction from Experiment 1, there was no statistically significant difference between the small- and the large-picture size in the global diagnostic condition (i.e., orchid). Thus, the current experiment failed to extend the small-picture-size advantage to a category other than fish.

Despite the lack of the small-picture-size advantage for the orchid condition, the metacognitive illusion was extended to a different class of category structure (inanimate). Participants in the orchid condition gave marginally higher CLJ ratings on average to the categories presented in the large size comparted to the ones presented in the small size although there was no difference in the actual performance between the large and the small size. Replicating the results from Experiment 1, the participants' CLJ ratings were mostly in line with their actual performance in the local-diagnostic (i.e., rock) condition. The non-significant size-by-outcome interaction in the separate within-subjects ANOVA for the rock condition indicated that, again, the degree to which the participants assigned higher CLJ ratings to large pictures did not exceed the degree of the large-picture-size advantage in the actual learning.

Despite the illusory performance prediction regarding the picture size in the orchid condition, CLJs once again showed its predictive validity. The overall gamma of .42 was above chance and comparable to that of Experiment 1 (.35), the previous experiments (range: .33-.38), and the previous studies of similar nature (Jacoby, et al., 2010: range: .31-.33; Wahlheim, et al., 2011: range: .47-.49).

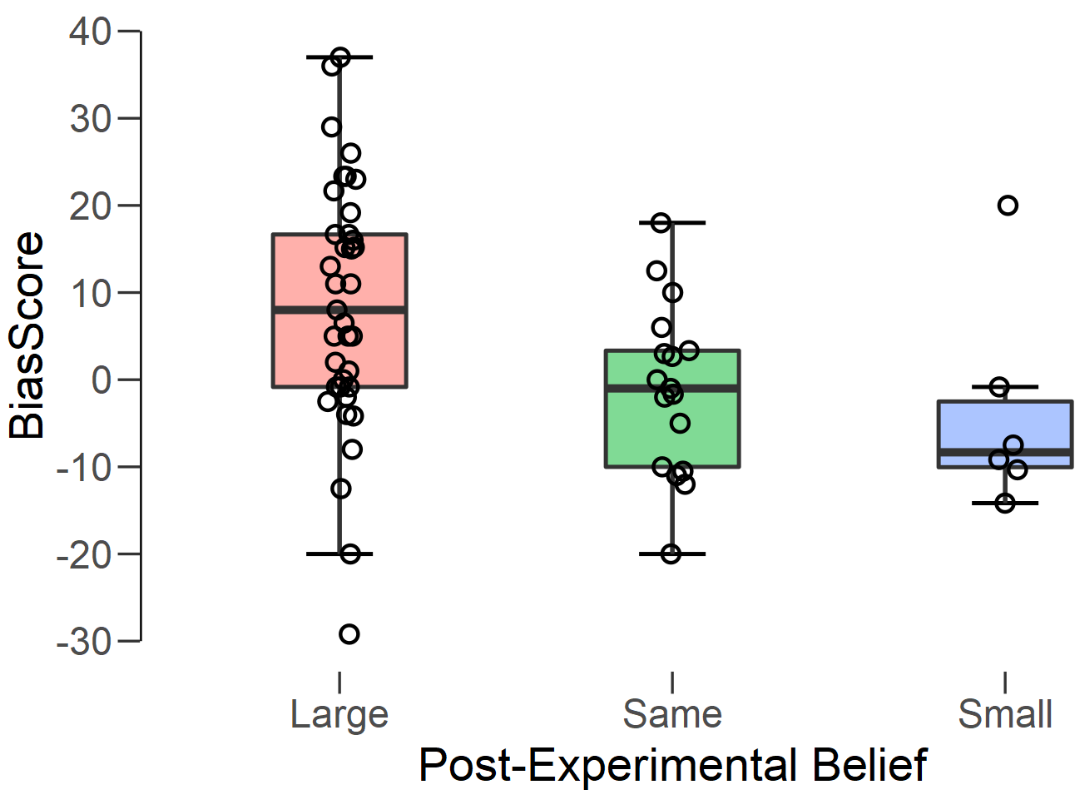
4.2.4 Supplemental Analyses on the Sources of the Metacognitive Illusion

Before discussing the implications of the above findings in the next chapter, I will present supplemental analyses combining Experiments 1 and 2 as well as the previous Experiments 1 through 5 that could inform the sources of the large-picture-size bias in category learning judgment (CLJs: post-learning performance prediction). This metacognitive illusion parallels with the font-size illusion (e.g., Rhodes & Castel, 2008) wherein learners gave higher ratings in judgements of learning (JOLs: post-learning performance prediction in word-list learning) for words presented in a large compared to a small font despite there was no difference in the actual memory test performance. As developed in the introduction briefly, there are two camps in conceptualizing the font-size bias arguing about the degree of contributions from two sources: belief and fluency. On the one hand, researchers have argued that this font-size bias arises because learners have a general belief that large fonts are better for learning and that belief is reflected in the higher JOL ratings for words presented in a lager font (e.g., Mueller, et al., 2014). On the other hand, researchers have argued that the font-size bias manifests because words in a larger font are processed more quickly, and this fluency is attributed as a sign of learning when making JOLs (e.g., Rhodes & Castel, 2008). Importantly, some researchers in the belief camp claim that the belief is the primary source of the font-size bias and there is little contribution from fluency (e.g., Mueller, et al., 2014) whereas the fluency camp acknowledges the contribution from belief but still argues that fluency contributes to this effect (e.g., Price, McElroy, & Martin, 2016). Given the similarity between the font-size bias and the large-picture-size bias, as well as the plausibility of belief and fluency as the sources of the large-picture-size bias, I will present below analyses aimed at assessing the contribution from these two sources in the large-picture-size bias observed in the current experiments (in the fish and orchid conditions) and the previous experiments.

Contribution from belief – large-picture-size bias plotted by post-experimental belief. Many participants in the current and the previous experiments held the belief that the large pictures were better for learning categories at least after the experiment as indicated by the post-experimental question. One might wonder how well this post-experimental belief aligned with the pre-existing belief the participants held when they made CLJs. Theoretically, the post-experimental belief is a combination of the pre-existing belief and the belief that arose from the experience during the experiment. However, an analysis of the literature on this issue suggests that the post-experimental belief predominantly reflects the pre-existing belief. First, experience-based modification of belief is difficult such that learners are often unable to mend metacognitive illusions just by going through learning and test (e.g., Koriat, & Bjork, 2006). Second, the ineffectiveness of this purely experience-based de-biasing (as opposed to theory-based de-biasing in which participants are told that one method of studying is more effective than others) has also been demonstrated in a category learning study that is very similar to the current study (Yan, Bjork, & Bjork, 2016). Thus, it seems reasonable to use the post-experimental question in the current and the previous experiments as a proxy of the pre-existing belief that was held at the time of CLJs and see if the large-picture-size bias differs as a function of that belief.

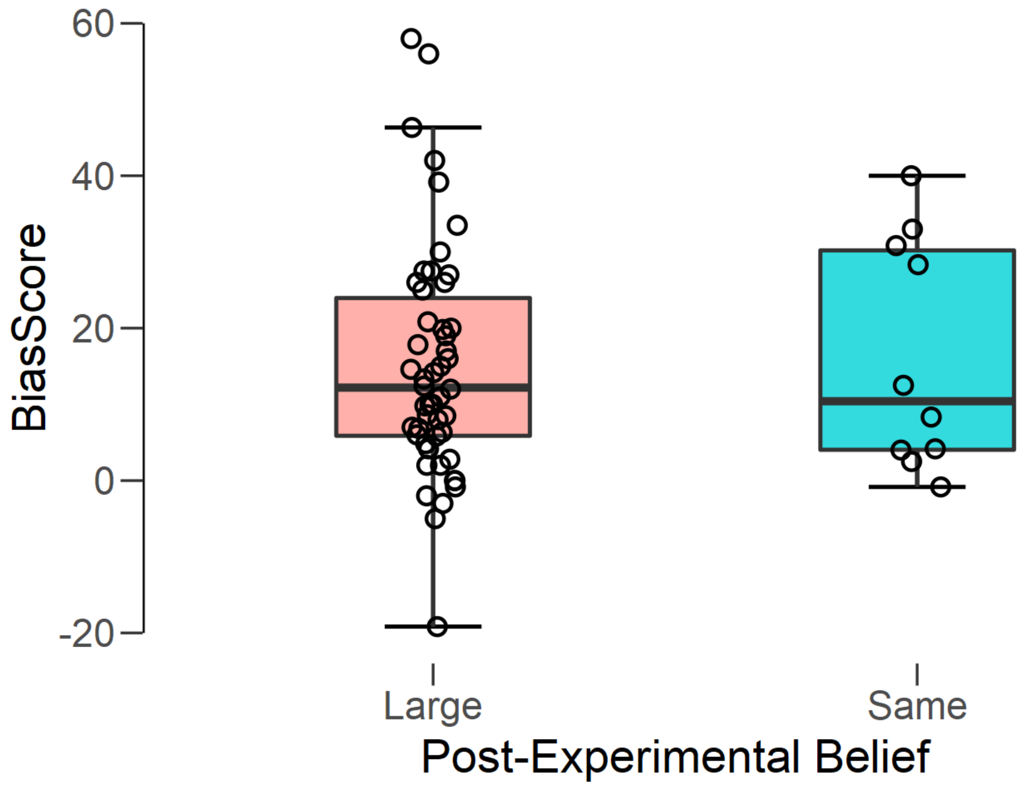
In the post-experimental question, participants indicated whether they believed that the large pictures were better for learning, small pictures were better for learning, or both the same. If these beliefs contributed to the way the participants made CLJs, there should be greater large-picture-size bias (i.e., $M_{large} - M_{small}$) among the "believers" who indicated that the large pictures were better for learning than the "non-believers" who indicated that the small pictures were better or both the same. Figure 22 shows the average large-picture-size bias as a function of the belief in the fish and the orchid conditions from Experiments 1 and 2.

Figure 22. Participants' mean bias score on CLJs from the fish and orchid conditions in Experiments 1 and 2 plotted by a function of post-experimental belief.



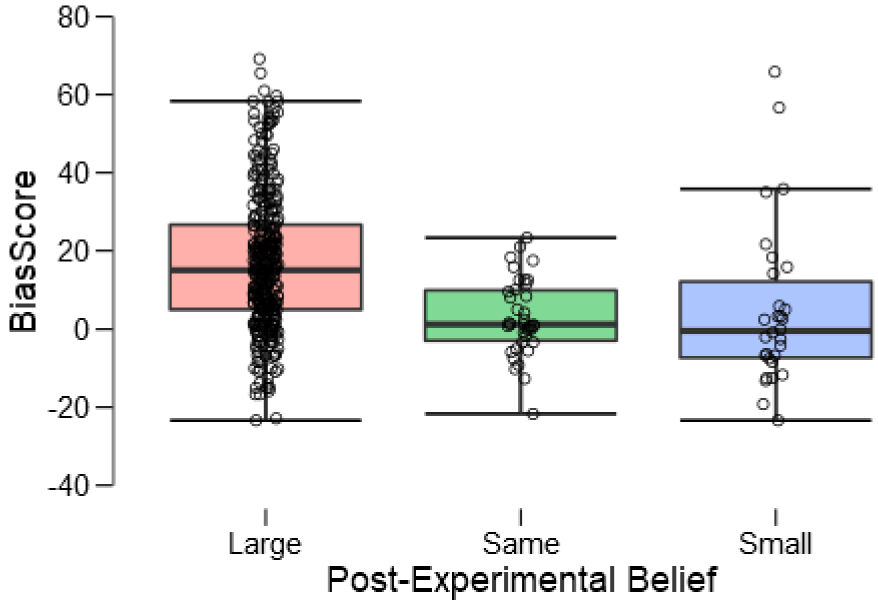
An independent-sample t-test showed that the mean bias score of the believers (M = 8.54, SD = 14.49) was indeed higher than that of the non-believers (M = -1.72, SD = 10.29), $t(58) = 2.96$, $p < .01$, d = 0.82. Interestingly, as you can see on Figure 23, this pattern did not hold up for the tern and the rock conditions from Experiments 1 and 2 where large pictures also produced better learning.

Figure 23. Participants' mean bias score on CLJs from the tern and rock conditions in Experiments 1 and 2 plotted by a function of post-experimental belief. Note that there were no participants in these conditions who believed that small pictures were better for learning, and thus that box could not be plotted.



An independent-sample t-test showed that the mean bias score of the believers (*M* = 15.19, *SD* = 15.32) was not significantly different from that of the non-believers in these conditions (*M* = 16.28, *SD* = 15.12), $t(58) = 0.21$, $p > .05$, $d = 0.07$. However, as you can see on Figure 24, this pattern of results held up when the same analysis was applied to the previous Experiments 1 through 5 in which all conditions showed statistically significant or numerical small-picture-size advantage in the classification performance.

Figure 24. Participants' mean bias score on CLJs from the previous Experiments 1 through 5 plotted by a function of post-experimental belief.



An independent-sample t-test showed that the mean bias score of the believers (M = 12.36, SD = 15.25) was again higher than that of the non-believers (M = 3.73, SD = 14.41), $t(401) = 5.69$, $p < .001$, $d = 0.79$. Thus, both in the current experiments and in the previous experiments, the belief contributed to the large-picture-size bias, higher CLJ ratings given to categories presented in the large compared to the small size, when small pictures produced better learning.

Did the large pictures produce increased fluency compared to the small pictures?[9]

As the reaction time is the primary measure of fluency, one may expect a faster reaction time to CLJ trials for categories presented in the large compared to the small size if the large pictures indeed produced a greater fluency. Interestingly, this prediction did not pan out neither in the current experiments nor in the previous experiments. In the current experiments, there was no

[9] In the analyses involving the reaction time measure reported in this and the following section, the reaction time for each trial was standardized at the participant level, and the trials with a z-score of less than -3 and greater than 3 were excluded from the analyses.

statistically significant difference between the small pictures (*M* = 8648.15, *SD* = 2221.08) and the large pictures (*M* = 8764.32, *SD* = 2882.53), $t(119) = 0.42$, $p > .05$, $d = 0.05$. In the previous experiments, participants' reaction time for CLJ trials were actually faster for the small pictures (*M* = 8667.62, *SD* = 2485.02) than for the large pictures (*M* = 9079.52, *SD* = 4094.23), $t(401) = 2.75$, $p < .01$, $d = 0.12$.

Was there a relationship between fluency and CLJ ratings at all? Increased fluency (measure by faster reaction time) has been associated with greater ratings on various kinds of metacognitive judgements (see Alter & Oppenheimer, 2009 for review), and thus it is possible that it also affected participants' CLJ ratings in the current and the previous experiments. To test this prediction, Pearson product-moment correlation between CLJ ratings and the reaction time for these CLJ trials was computed. There was a small but significant negative correlation between the two variables both in the current experiments, $r(1307) = -.072$, $p < .01$ (Experiment 1: -.032; Experiment 2: -.174), and in the previous experiments, $r(4822) = -.036$, $p < .05$ (range: -.106-.005). These analyses showed that increased fluency was associated with greater CLJ ratings. However, this is likely a different type of fluency than the perceptual fluency (i.e., the speed at which a picture is processed) as I will discuss more in the following chapter.

Chapter 5: General Discussion

5.1 The Picture-Size Effect on Category Learning

In the current dissertation, I investigated the mechanism through which the small-picture-size advantage on category learning manifested. The current experiments not only replicated the effect (Experiment 1 fish condition) but also showed a large-picture-size advantage in some conditions (Experiment 1 tern condition & Experiment 2 rock condition). Along with the previous experiments, these are the first empirical demonstrations of how a simple perceptual manipulation of picture size affects learning of complex natural categories. Critically, there was a significant interaction between the picture size and the category structure (local or global diagnostic dimensions), such that small pictures produced better learning only when a global dimension was diagnostic of the category structure whereas large pictures were superior when local dimensions were diagnostic. In addition, this critical interaction was observed across two important classes of natural categories (Experiment 1: animate; Experiment 2: inanimate).

In Chapter 3, I sketched out two potential hypotheses regarding the picture-size effect on category learning: the increased effort hypothesis and the perceptual precedence hypothesis. The findings enumerated above are inconsistent with the increased effort hypothesis which assumed a category-structure-general mechanism. Specifically, it predicted that the small-picture size to be superior regardless of the globality of diagnostic dimensions of a given category structure. Under this hypothesis, a large-picture-size advantage should occur in no situation. On the contrary, the large-picture-size advantage found in the tern and the rock conditions, as well as the critical size-by-category interaction observed in both Experiments 1 and 2, supported the perceptual precedence hypothesis and demonstrated that a category-structure-specific mechanism underlies the picture-size effect on category learning. However, the results posed some challenges to the perceptual precedence hypothesis as will be discussed later; The small-picture-size advantage was not observed in the orchid condition in Experiment 2, and as a result, the interaction pattern was not a full crossover as anticipated by the hypothesis.

The category-structure-specific nature of the picture-size effect is in line with the material-appropriate-processing framework (MAP: Einstein, McDaniel, Owen, & Cote, 1990; McDaniel, & Einstein, 1989) which was developed in the text and the word-list learning research. In general terms, MAP postulates that the efficacy of a given manipulation depends on the material to which it is applied. A framework akin to MAP in the context of category learning, *category-appropriate-processing* framework (CAP) could be useful in conceptualizing manipulations that can enhance category learning. The core idea of CAP is that the effectiveness of a given manipulation to enhance category learning depends on the category structure to be learned, and it emphasizes the importance of analyzing the match between the cognitive process elicited by the manipulation and the processing that is beneficial to the given category structure. This framework accommodates the present findings as well as several findings from recent research. In the case of the current study, the encoding of instances in category learning can be seen as a sequential encoding of various dimensions, and the first dimensions to be encoded vary depending on the size manipulation: global dimensions first in the small pictures and local dimensions first in the large pictures. Accordingly, the small-size manipulation enhanced learning when the category structure had global diagnostic dimensions (the previous Experiments 1-5 and the fish condition in the current experiments; but see the orchid condition), and the large-size manipulation enhanced learning when the category structure had local diagnostic dimensions (the tern and the rock conditions in the current experiments).

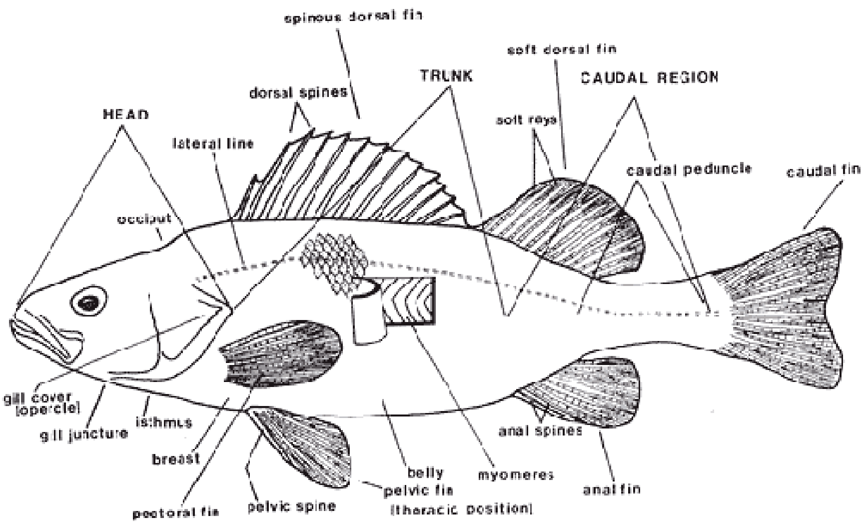
Diagnostic dimensions are not the only way to characterize category structures and to determine the kind of processing that is beneficial. For example, in order to learn category structures that are high in both within- and between- category similarity (i.e., instances belonging to the same category look similar and the categories within the structure look similar), the differences between categories need to be learned well. This is because why the instances belong to the same category is clear (because they look similar) in this case but discriminating between categories is challenging (because instances from different categories look similar). On the contrary, in order to learn category structures that are low in both within- and between- category similarity (i.e., instances belonging to the same category look different and the categories within the structure look different), the commonalities within each category need to be learned. This is because why the instances belong to the same category is hard to grasp in this case (because they look different) but discriminating between categories is easy (because instances from different categories look different). Accordingly, interleaving (i.e., presenting examples from several categories in succession to emphasize the processing of the differences between categories) has been shown to enhance learning of many complex category structures with high within- and between- category similarities (e.g., Kornell & Bjork, 2008), and blocking (i.e., presenting examples form the same category consecutively to emphasize the processing of the similarity within a category) has been shown to enhance learning of category structures with low within- and between- category similarities (e.g., Carvalho & Goldstone, 2014, 2015).

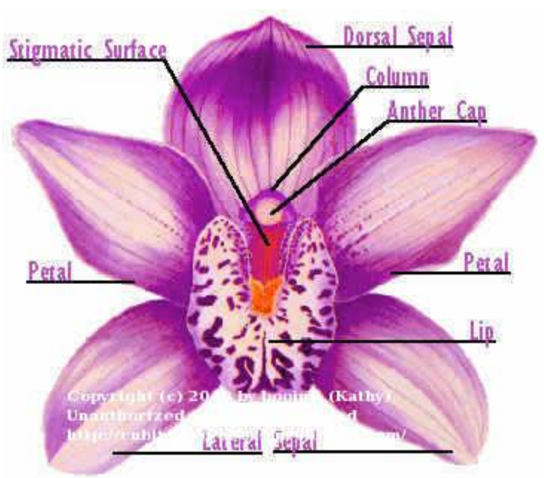
CAP can accommodate many other findings from the recent trend in the field of category learning which identified factors that can enhance category learning as well. For instance, the efficacy of specific-level training in teaching broad-level categories (e.g., in teaching broad-level rock categories of igneous, sedimentary, and metamorphic, teaching the specific categories first, such as andesite and obsidian under igneous, breccia and chert under sedimentary, and gneiss and migmatite under metamorphic) has been shown to be dependent on the between-specific-level category similarity (Miyatsu, et al., in press; Nosofsky, et al., 2017). Taking these findings together (see also Pashler & Mozer, 2013, on the category-structure-specific nature of fading, training that uses an exaggerated version of stimulus discrimination; Miyatsu et al., 2019, Wahlheim, et al., 2012, for boundary category structure of the efficacy of exemplar variability), CAP postulates that the benefit of any manipulation on category learning depends on the category structure to be learned.

It is important to note that the term “category structure” in the CAP framework is used in the broadest sense. In the current study, category structures were characterized by their dimensions and the diagnosticity of each dimension. I also provided above examples of how the similarity at various levels (e.g., within- and between- category similarities) can be used to characterize category structures. These and any other ways to characterize a category structure by its distinctive properties should be considered when analyzing category structures in the CAP framework. For example, in the classic category learning literature, category structures are often characterized as either rule-based or information integration according to the presence (or absence) of clear, verbalizable rules that can define the categorization (e.g., Ashby & Ell, 2001). In general, CAP puts emphasis on analyzing the category structure and the cognitive processing that is elicited by a manipulation of interest and has both theoretical and practical utility. For example, CAP generates an interesting prediction that a similar category-structure-specific process may be at play in other manipulations that are assumed to enhance category learning of all category structures (e.g., test-enhanced learning: Jacoby, et al., 2010; see also the following section on practical implications).

The current study is not without its limitations. The most prominent of which is the failure to extend the small-picture-size advantage to categories other than fish. Despite the orchid categories having the general characteristics that should be benefitted from the small-picture size according to the perceptual precedence hypothesis (a global diagnostic dimension), there was no small-picture-size advantage in this condition. There are a few possible related reasons for why there was no small-picture-size advantage in the orchid condition. First the orchid categories are less complex than the fish categories (compare Figures 16 & 19). Specifically, there is a considerably fewer number of dimensions in the orchid categories compared to the fish categories. For example, while there are various physical characteristics of fish that vary from instance to instance (i.e., dimensions), such as mouth, eye, six different kinds of fins, and pattern that can vary depending on the part of the body, orchid flowers vary in just a few dimensions and the color and pattern are mostly consistent throughout (see Figure 25 for schematic illustrations of tropical fish and orchid flower demonstrating this difference in the anatomical complexities).

Figure 25. Schematic illustrations showing the anatomical complexity of tropical fish (top) and orchid flower (bottom). Retrieved from www.fishlore.com and www.garden.org.





Relatedly, while orchid flowers are symmetrical, fish have various features distributed asymmetrically throughout. This difference makes the encoding of the orchid flowers less complex because scanning of both sides of an orchid picture is not necessary. In contrast, the presence of unique features in both the left-most (e.g., mouth, eye) and right-most part of the fish (e.g., caudal fin) invites far more complex encoding involving sideways gaze shifts. Importantly, the reduced complexity matters because the perceptual precedence hypothesis assumes an encoding process in which a number of features compete for attention. When there are only a few features to be encoded, despite the large-picture size initially guiding learners' attention to local dimensions like the shape of column in an orchid picture, learners could still encode the rest of the features including the diagnostic feature of the exterior shape in the time given for a study trial (5 seconds).

In addition, animate and inanimate objects have been shown to be processed differently, and the processing difference might have prevented a small-picture-size advantage to manifest in the orchid condition. Researchers have argued that human cognition is attuned to prioritize the processing of animate over inanimate objects because animate objects have greater biological and survival significance (see Nairne, et al., 2017 for review). For example, perceived animacy both by the appearance or by motions of the objects have been shown to capture attention (e.g., Lipp, Derakshan, Waters, & Logies, 2004; Pratt, Radulescu, Guo, & Abrams, 2010). People also detect changes more quickly and accurately for animate compared to inanimate objects (New, Cosmides, & Tooby, 2007). However, the large-picture-size advantage was demonstrated in an inanimate category structure (i.e., the rock condition), so animacy is not a strict boundary condition for the picture-size effect per se. Nonetheless, it is possible that these processing differences contributed to the failure to obtain a small-picture-size advantage in the orchid condition.

Related to the issues raised above, one way to summarize the results from the previous and the current experiments is that small pictures significantly enhanced learning for only fish categories. As contrasted against the orchid categories above, fish categories were complex, asymmetrical, and animate, making them unique among all the categories in which the picture-size manipulation has been tested. In addition, people could have richer past experience with fish, such as pets or food sources, or have greater prior knowledge on and association with fish as this is a very common category in American (and other) culture. For example, one recent word frequency norm in American English derived from 51 million words (SUBTLEX-US: Brysbaert, New, & Keuleers, 2012) indicated that the word, fish (1138^{th} most frequent, $SUBTL_{WF}$[10]: 83.49), has appeared more frequently than the words orchid (15172th, $SUBTL_{WF}$: 2.16) or flower (2646^{th}, $SUBTL_{WF}$: 22.76). Careful considerations of these parameters can lead to an identification of categories other than fish that would benefit from small pictures which is paramount in demonstrating the generality of the picture-size effect as well as in advancing our understanding of the picture-size-dependent attentional shift mechanism.

5.2 Alternative Accounts of the Picture-Size Effect

A part of the perceptual precedence hypothesis that has been implied throughout my description of the proposed mechanism and deserves additional elaboration is the *distraction view*. Simply put, this view considers the small-picture-size advantage as a large-picture-size disadvantage. Specifically, performance was better in the small pictures in the fish experiments because the large-picture size enhanced the extractability of local features that had little diagnostic value. The

[10] $SUBTL_{WF}$ is the word frequency per million words.

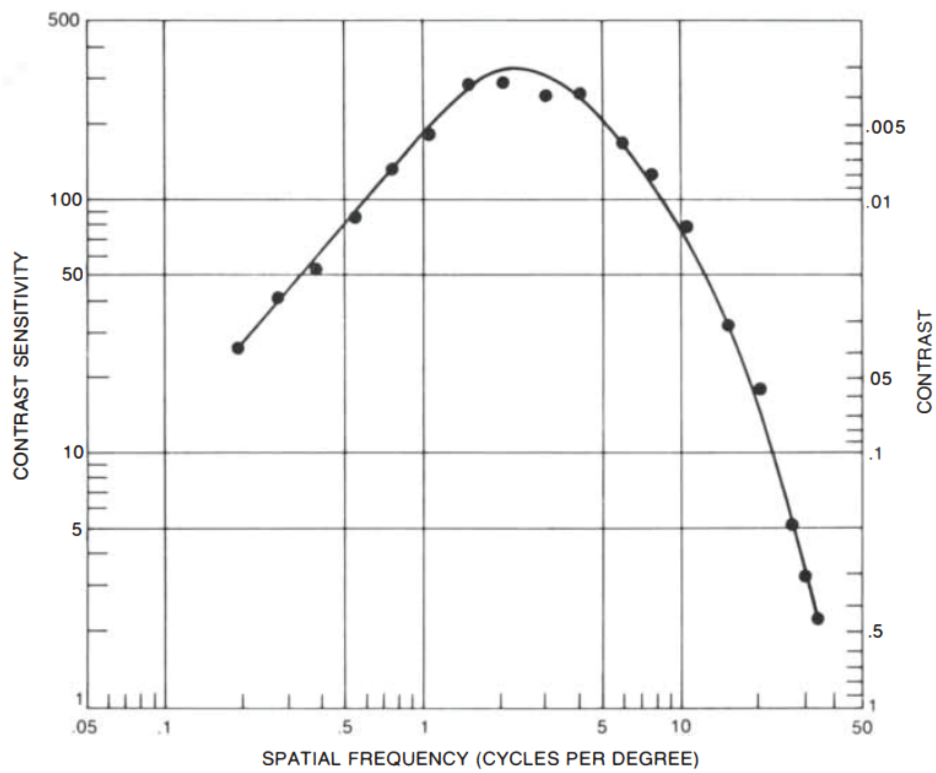
reverse is also true in that the distraction view considers the large-picture-size advantage in the rock and tern conditions as a small-picture-size disadvantage. In these cases, performance was better in in the large pictures because the small-picture size enhanced the extractability of global features that had little diagnostic value. It is important to note that the attentional shift mechanism assumed in the perceptual precedence hypothesis hinges upon the *relative* extractability of features. For example, if there were a local feature and a global feature and the extractability of the local feature went up because of a size manipulation, extractability of the global feature relative to the local feature would go down regardless of the effect of the size manipulation to the global feature. Therefore, it is possible to explain the small-picture-size advantage only by this distraction view. Whether the distraction-based mechanism accounts for all of the picture-size effect or the emphasis of global diagnostic dimensions by the small-picture size and the emphasis of local-diagnostic dimensions by the large-picture size accounts for some of the effect is an empirical question. Regardless, these are descriptions of the two sides of the same coin of the attentional shift that is assumed by the perceptual precedence hypothesis.

Another perspective of the picture-size effect that deserves a careful consideration is the one based on spatial frequency. This point was also discussed by Kinchla and Wolf (1974) in their original perceptual precedence theory, "…"most recognizable" forms[11] may be thought of in terms of size in the visual field or in terms of an optimal band of spatial frequencies" (p.230). In the current study, the size manipulation led to a systematic change in the detectable spatial frequency. Specifically, only the lower frequencies were visible in small pictures whereas the higher frequencies were also visible in large pictures. Crucially, the analysis of the literature concerning spatial frequency shows a striking resemblance to that of the size literature. First,

[11] The word form means features expressed at different levels, such as the large letter and small letters in the global precedence material shown in Figure 11.

there is evidence for a top-down sequential processing like that of global precedence (Navon, 1977, 1981) but based on spatial frequency. For example, Schyns and Oliva (1994) reported that in quickly recognizing complex scenes (e.g., a picture of highway), people rely first on coarse scale represented by low spatial frequencies and then move on to fine spatial scales. Importantly however, there is also a middle-out characteristic in human visual system concerning spatial frequency. As Figure 26 shows, in experiment using grating patches, human visual system is most sensitive to moderate spatial frequencies and the contrast sensitivity drops off for higher and lower frequencies (Campbell & Maffei, 1974). Thus, it is possible that the attentional shift between different features assumed by the current perceptual precedence hypothesis is based on spatial frequency.

Figure 26. Contrast sensitivity of a human subject plotted as function of spatial frequency. The scales are logarithmic. Very high contrast is given a value of 1, and contrast sensitivity is the reciprocal of contrast. Reprinted from “Contrast and spatial frequency,” by Campbell, F. W., & Maffei, L. (1974), *Scientific American, 231*(5), 106-115.



It is also important to point out that the change in the extractability of different features based on the change in spatial frequency is very similar to that based on the change in size. Perhaps the most intuitive way of thinking about this is to draw a parallel between progressively shrinking a picture and progressively blurring a picture (i.e., progressively removing lower and lower spatial frequency components of an image). When only the highest frequencies are removed (slightly blurred), fine details (i.e., local features) become unrecognizable first. In contrast, global features (most global of which being the exterior shape) remain recognizable much later in the blurring process when only lowest-frequency components are present. This is very similar to how in a large size, all the fine details are available, but as the image becomes smaller, the extraction of these details becomes more difficult and the extraction of global features becomes easier. Thus, the spatial frequency perspective of the picture-size effect aligns closely with the size perspective as articulated in the current perceptual precedence hypothesis.

5.3 The Picture-Size Effect on Metacognition

The current study was also the first to demonstrate the effect of the simple perceptual manipulation of picture size on metacognition in the context of category learning. Category learning judgement is clearly a valid measure of participants' discriminative ability according to their own learning. This was evident in the adequate level of the within-participant correlation (i.e., gamma) between the CLJ ratings and the classification performance for particular categories in the current experiments, the previous experiments, and the previous literature (e.g., Wahlheim, et al., 2011). Given this predictive validity of CLJs, the degree of the large-picture-size bias was surprising. In all conditions in the current and the previous experiments in which small pictures produced significantly or numerically better classification performance (10 conditions total), participants gave higher CLJ ratings on average to the categories that were presented in the large compared to the small size.

However, as the direct access and retrieval view anticipated, participants' CLJs appeared to be accurate in some situations. In addition to the moderately high gamma correlations between the CLJs and the classification performance reported above, the participants' CLJs correctly predicted the large-picture-size advantage in the tern and the rock conditions. Interestingly, the non-significant size-by-outcome (classification or CLJs) interaction in these conditions indicated that the degree to which the participants assigned higher CLJ ratings to the large pictures did not exceed the degree of the large-picture-size advantage in the classification. Therefore, it appears that the large-picture-size bias was not present when the direction of the bias aligned with the condition that promoted better performance. Certainly, this lack of large-picture-bias when large pictures produced better learning could represent a combination of an accurate assessment of own learning and a successful prevention of the extrinsic and the mnemonic cues (e.g., the processing fluency, the belief that large pictures are better for learning) to influence the inference process. However, it is also possible that it represented a lack of an accurate assessment of own learning and the persisting influence from the same extrinsic and mnemonic cues that caused the large-picture-size bias when small pictures produced better learning. Although teasing apart the influence from these cues is beyond the scope of the current study, the current results warrant future investigations aimed at dissociating the contributions from different types of cues in making the metacognitive judgments.

The current study also informed the sources of this metacognitive illusion and contributed to the high-interest topic in the field (e.g., a forthcoming special issue on this topic at Zeitschrift fur Psychologie) both empirically and methodologically. As mentioned before, the on-going discussion in the field of the stimulus-size effect on metacognitive judgement is whether this effect is underlie solely by the belief or both by the belief and the fluency. The supplemental analyses plotting the mean bias score on CLJs (i.e., $M_{large} - M_{small}$) by the belief indexed by the post-experimental question clearly showed that the belief contributed to this effect. In contrast, there was no evidence of the picture-size manipulation producing an increased fluency for large pictures as the reaction time for the CLJ trials for the large and the small pictures did not differ. However, a methodological limitation might have concealed the potential increased fluency for the large pictures. Specifically, in the current and the previous experiments as well as in all the past studies using CLJs, a CLJ is made after a set duration of study time. For example, in my studies participants studied the last example from a category for five seconds before being prompted to make a CLJ. This practice of allocating a set time for studying the example before making a CLJ for that category is intended to equate the study time between categories and is a standard practice in the field because category learning itself is the focus of these studies in most cases. However, in order to measure the fluency caused by the size or any other manipulations at a given CLJ trial, the CLJ trial should not be preceded by a set study time because the reaction time difference of 40-200 *milliseconds* would likely fall within the set study time and would not be reflected in the latency for the CLJ trail. Such methodological change is also applicable for other paradigm (e.g., word-list learning) and is necessary for future investigations focusing on teasing apart the effect of belief and fluency on metacognitive judgements.

Interestingly, there was evidence that some kind of fluency contributed to higher CLJ ratings as there was a weak but significant negative correlation between the latency for a CLJ trial and the rating given in that trial (i.e., the faster to make a CLJ, the higher the rating: r = -.07 and -.04 in the current experiments and the previous experiments respectively; according to

Cohen ,1988, r = -.10 represents a small correlation). It is important to note that this is likely a different kind of fluency from the perceptual fluency (i.e., the ease or speed at which the stimulus is processed) that larger pictures presumably produce. Rather, this is a retrieval or inference fluency (i.e., the ease or speed at which information is retrieved or inference is made: Benjamin & Bjork, 2014; Oppenheimer, 2008). When prompted to indicate the likelihood of correctly identifying a member of a given category (i.e., CLJ) in the current study, participants likely have taken the ease or speed at which they recalled the previous examples from that category or the speed at which they inferred what the characteristic features were for that category as the evidence of learning and assigned the rating accordingly. This is the first evidence that such fluency contributed to CLJ ratings and adds to the emerging literature.

5.4 Practical Implications

The current study has various practical implications. First, as briefly mentioned in the introduction, methods that can optimize category learning instruction are of interests to many fields, such as K-12 education, physician training, and military training. The size-dependent attentional shift demonstrated in the current study gives a clue to what the optimal size of example-picture presentation would be depending on whether the diagnostic dimensions of the target category structure are global or local. In addition, the category-appropriate-processing (CAP) framework emphasizes the importance of analyzing the category structure in terms of its diagnostic dimensions, similarity between categories, and the type of processing that are beneficial for learning the given category structure. Such analyses allow a selection of appropriate techniques to be applied, and it provides an excellent start point from which an optimal category instruction can be built.

Lastly, the metacognitive illusion produced by the large-picture size has various implications as well. Pointing out situations in which learners' metacognitive awareness is in a stark contrast to the actual learning outcome is extremely important from an educational perspective because when left to their own devices, learners will always choose the method of studying that they *think* would produce better learning. As studying through the monitor of various sizes is also becoming increasingly popular (i.e., reviewing lecture slides on a desktop computer, watching a recorded lecture on a smartphone), disseminating the potential pitfall of studying on a larger monitor may be useful. Of course, although I think it is very likely, whether a large picture or monitor size produces a similar metacognitive illusion (i.e., inflated confidence in learning) is an empirical question and awaits future research.

5.5 Summary and Concluding Comments

In the current dissertation, I reported the first empirical demonstrations of how the picture size affected category learning and metacognition. The picture-size effect was category-structure-specific in that category structures with a global diagnostic dimension benefited from the small-picture size whereas category structures with local diagnostic dimensions benefited from the large-picture size. I proposed that this size-by-category-structure interaction was a product of the size-dependent attentional shift between the features represented in different sizes (or spatial frequencies) based on the basic characteristics of human visual system. Further, a novel framework (i.e., category-appropriate-processing framework: CAP) encompassing the current and the past findings was presented. This framework should serve as a good starting point in constructing optimal category instruction of any kind and also generates interesting research questions. The picture-size manipulation also affected metacognition; learners' judgments on their own learning were inflated with the large picture size in some situations. Given the wide-spread practical implications of both the size-dependent attentional shift and the metacognitive illusion as well as the increasing variability of the size of monitors through which information is consumed, future research on the effect of stimulus size on a wide range of task performance and subjective judgements are warranted.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219-235.

Antes, J. R., & Mann, S. W. (1984). Global-local precedence in picture processing. *Psychological Research, 46*(3), 247-259.

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, *5*(5), 204-210.

Benjamin, A. S., & Bjork, R. A. (2014). Retrieval fluency as a metacognitive index. In *Implicit Memory and Metacognition* (pp. 321-350). London, UK: Psychology Press.

Blum, H. (1967). A transformation for extracting new descriptors of shape. In W. Wathen-Dunn (Ed.), Models for the Perception of Speech and Visual Form (pp. 362-380). Cambridge, MA: The MIT Press.

Bonin, P., Gelin, M., & Bugaiska, A. (2014). Animates are better remembered than inanimates: Further evidence from word and picture stimuli. *Memory & Cognition*, *42*(3), 370-382.

Breitmeyer, B. G., & Breier, J. I. (1994). Effects of background color on reaction time to stimuli varying in size and contrast: Inferences about human M channels. *Vision Research, 34*(8), 1039-1045.

Brown, A. S. (2011). *Tip of the tongue state*. Hove, U.K.: Psychology Press.

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, *44*(4), 991-997.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language*, *30*(5), 542-579.

Campbell, F. W., & Maffei, L. (1974). Contrast and spatial frequency. *Scientific American*, *231*(5), 106-115.

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, *42*(3), 481-495.

Carvalho, P. F., & Goldstone, R. L. (2015). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review*, *22*(1), 281-288.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the (): Effects of disfluency on educational outcomes. *Cognition*, *118*(1), 111-115.

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks, CA: Sage.

Einstein, G. O., McDaniel, M. A., Owen, P. D., & Cote, N. C. (1990). Encoding and recall of texts: The importance of material appropriate processing. *Journal of Memory and Language*, *29*(5), 566.

Eitel, A., Kühl, T., Scheiter, K., & Gerjets, P. (2014). Disfluency meets cognitive load in multimedia learning: does harder-to-read mean better-to-understand? *Applied Cognitive Psychology*, *28*(4), 488‑501.

Hart, J. T. (1967). Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior*, *6*(5), 685-691.

Hilz, R., & Cavonius, C. R. (1974). Functional organization of the peripheral retina: Sensitivity to periodic stimuli. *Vision Research*, *14*(12), 1333-1337.

Hunziker, H. W. (2006). *Im Auge des Lesers: foveale und periphere Wahrnehmung-vom Buchstabieren zur Lesefreude*. (The Eye of the Reader: Foveal and Peripheral Perception-from Letter Recognition to the Joy of Reading). Zürich, Switzerland: Stäubli.

Jacoby, L.L., Wahlheim, C.N., & Coane, J.H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1441-1451.

Kelley, C. M., & Jacoby, L. L. (1996). Memory attributions: Remembering, knowing, and feeling of knowing. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 287–307). Hillsdale, NJ: Erlbaum

Kinchla, R. A., & Wolfe, J. M. (1979). The order of visual processing: “Top-down,” “bottom-up,” or “middle-out”. *Perception & Psychophysics*, *25*(3), 225-231.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349.

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1133.

Kornell, N., & Bjork, R.A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science, 19*(6), 585-592.

Kühl, T., & Eitel, A. (2016). Effects of disfluency on cognitive and metacognitive processes and outcomes. *Metacognition and Learning*, *11*(1), 1-13.

Lamb, M. R., & Robertson, L. C. (1989). Do response time advantage and interference reflect the order of processing of global-and local-level information? *Attention*, *Perception, & Psychophysics*, *46*(3), 254-258.

Lamb, M. R., & Robertson, L. C. (1990). The effect of visual angle on global and local reaction times depends on the set of visual angles presented. *Perception & Psychophysics*, *47*(5), 489-496.

Lipp, O. V., Derakshan, N., Waters, A. M., & Logies, S. (2004). Snakes and cats in the flower bed: fast detection is not specific to pictures of fear-relevant animals. *Emotion*, *4*(3), 233.

Lloyd-Jones, T. J., & Luckhurst, L. (2002). Outline shape is a mediator of object recognition that is particularly important for living things. *Memory & Cognition*, *30*(4), 489-498.

Meyer, A., Frederick, S., Burnham, T. C., Guevara Pinto, J. D., Boyer, T. W., Ball, L. J., & Schuldt, J. P. (2015). Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology: General*, *144*(2), 16–30.

McDaniel, M. A., & Einstein, G. O. (1989). Material-appropriate processing: A contextualist approach to reading and studying strategies. *Educational Psychology Review*, *1*(2), 113-145.

Mead, L. C. (1943). The influence of size of test stimuli, interpupillary distance, and age on stereoscopic depth perception. *Journal of Experimental Psychology*, *33*(2), 148-158.

Metcalfe, J. (1986). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(2), 288.

Miyatsu, T., Gouravajhala, R., Nosofsky, R. M., & McDaniel, M. A. (2019). Feature Highlighting Enhances Learning of a Complex Natural-Science Category. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *45*(1), 1-16.

Miyatsu, T., Nosofsky, R. M., & McDaniel, M. A. (in press). Effects of Specific-level Versus Broad-level Training for Broad-Level Category Learning in a Complex Natural Science Domain. *Journal of Experimental Psychology: Applied.*

Miyatsu, T. (in preparation). *The Effects of Stimulus size on Human Cognition.*

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language*, *70*, 1-12.

Murphy, G.L. (2002). *The Big Book of Concepts.* The MIT Press; Cambridge, Massachusetts.

Nairne, J. S., VanArsdall, J. E., & Cogdill, M. (2017). Remembering the living: Episodic memory is tuned to animacy. *Current Directions in Psychological Science, 26*(1), 22-27.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*(3), 353-383.

Navon, D. (1981). The forest revisited: More on global precedence. *Psychological Research*, *43*(1), 1-32.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The “delayed-JOL effect”. *Psychological Science*, *2*(4), 267-271.

New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences*, *104*(42), 16598-16603.

Noh, S. M., Yan, V. X., Vendetti, M. S., Castel, A. D., & Bjork, R. A. (2014). Multilevel induction of categories: Venomous snakes hijack the learning of lower category levels. *Psychological Science*, *25*(8), 1592-1599.

Nosofsky, R. M., Sanders, C. A., Gerdom, A., Douglas, B. J., & McDaniel, M. A. (2017). On Learning Natural-Science Categories That Violate the Family-Resemblance Principle. *Psychological Science 28*(1), 104-114.

Nosofsky, R. M., Sanders, C. A., Meagher, B. J., & Douglas, B. J. (2018). Toward the development of a feature-space representation for a complex natural category domain. *Behavior Research Methods*, *50*(2), 530-556.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237-241.

Pashler, H., & Mozer, M.C. (2013). When does fading enhance perceptual category learning? *Journal of Experimental Psychology: Learning, Memory and Cognition, 39*(4), 1162-1173.

Pratt, J., Radulescu, P. V., Guo, R. M., & Abrams, R. A. (2010). It's alive! Animate motion captures visual attention. *Psychological Science*, *21*(11), 1724-1730.

Price, J., McElroy, K., & Martin, N. J. (2016). The role of font size and font style in younger and older adults' predicted and actual recall performance. *Aging, Neuropsychology, and Cognition*, *23*(3), 366-388.

Schultz, D. W., & Eriksen, C. W. (1978). Stimulus size and acuity in information processing. *Bulletin of the Psychonomic Society*, *12*(6), 397-399.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological Science*, *5*(4), 195-200.

Sperandio, I., Savazzi, S., Gregory, R. L., & Marzi, C. A. (2009). Visual reaction time and size constancy. *Perception*, *38*(11), 1601-1609.

Streibel, M. J., & Ebenholtz, S. M. (1982). Construct validity of perceptual style: Role of stimulus size in the embedded-figures test and the rod-and-frame test. *Attention, Perception, & Psychophysics*, *31*(2), 128-138.

Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for inverted words: Illusions of competency and desirable difficulties. *Psychonomic Bulletin & Review*, *18*(5), 973.

Susser, J. A., Mulligan, N. W., & Besken, M. (2013). The effects of list composition and perceptual fluency on judgments of learning (JOLs). *Memory & Cognition*, *41*(7), 1000-1011.

Tarr, M. J. (2003). Visual object recognition: Can a single mechanism suffice? In M. A. Peterson & G. Rhodes (Eds.), *Advances in visual cognition. Perception of faces, objects, and scenes: Analytic and holistic processes* (pp. 177-207). New York, NY, US: Oxford University Press.

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition*, *39*(5), 750-763.

Wahlheim, C. N., Finn, B., & Jacoby, L. L. (2012). Metacognitive judgments of repetition and variability effects in natural concept learning: evidence for variability neglect. *Memory & Cognition*, *40*(5), 703-716.

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General*, *145*(7), 918.

Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is—and is not—a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition*, *41*(2), 229-241.

Appendix A

A Brief Description of the Pilot Study with Artificial Fish Material

In the original iteration of the dissertation, I attempted to test the unique predictions from the perceptual precedence and the increased effort hypotheses by having participants learn and classify artificial fish categories. This material was created by combining the shapes of fish that were used in the previous experiments and geometric patterns using a raster graphic editor. Critically, all instances were defined by 4 dimensions – the shape of the body (global), the shape of the tail (local), the pattern of the body (global), and the pattern on the tail (local) – and in a given experiment and condition, only one of the dimensions was diagnostic and all others had zero diagnosticity. I took this approach initially because in the natural material many features are correlated so that even in a material in which one dimension is predominantly diagnostic, other dimension can carry some, albeit low, diagnosticity. Therefore, constructing a category structure in which strictly one dimension was diagnostic offered an opportunity for the most stringent test of the hypotheses.

Figures A1 (shape diagnostic) and A2 (pattern diagnostic) show the material used in the first set of the pilot study. In Pilot Experiment 1 using the shape diagnostic material shown in Figure A1, half the participants studied and classified categories organized by each row which were defined by the shape of the body (global diagnostic condition) while the other half of the participants studied and classified categories organized by each column which were defined by the shape of the tail (local diagnostic condition). In Pilot Experiment 2 using the pattern diagnostic material shown in Figure A2, half the participants studied and classified categories organized by each row which were defined by the pattern on the body (global diagnostic condition) while the other half of the participants studied and classified categories organized by each column which were defined by the pattern on the tail (local diagnostic condition).

Several iterations of pilot experiments were run using these materials both online with participants from Amazon Mechanical Turk and in the lab with Washington University undergraduate participants. Figure A3 show illustrative examples of the results. In short, the small-picture-size advantage was not replicated, and the overall performance was too low; in some cases the performance was barely above the chance (8% in classifying 12 categories). Even when the performance was adequately high (the global diagnostic condition in the lab sample of Experiment 2: ~30%), the picture-size effect did not emerge. Participants seemed to have especially hard time using the information from the local dimension (the shape of and the pattern on the tail).

Why did the picture-size effect fail to emerge in this category structure? Two notable changes from the previous material are the introduction of the geometric pattern and the reduction of the material complexity that came with it. First, participants' attention seemed to be drawn more to the geometric pattern than to the natural shape. This was evident in the near-floor performance when the shape was diagnostic and the relatively higher performance when the pattern was diagnostic. In retrospect, this inclination towards the pattern processing is understandable because the extraction of the pattern features was easier and taxonomical terms were readily available (e.g., thick vertical lines throughout the body, small circular dots on the tail) whereas the shape features were more complex and harder to distinguish (e.g., thinner and longer body with a fin on top, a sharp boomerang-shaped fin with equally long top and bottom parts). Second, the geometric patterns inside the shape outline of the fish had only two features (i.e., the pattern of the body and the pattern of the tail) whereas the natural material offered several features that were more complex, such as the pectoral fin, ear flap, and cheek. If the picture-size effect hinged upon attentional shift between these complex features or increased effort in processing these features, it is possible that the introduction of the easily extractable geometric pattern and the reduction in the material complexity have compromised these processes.

Figure A1. Examples of the fish stimuli that were used in pilot experiments. Each row represents a category that is defined by the shape of the body whereas each column represents a category that is defined by the shape of the caudal fin (i.e. tail).

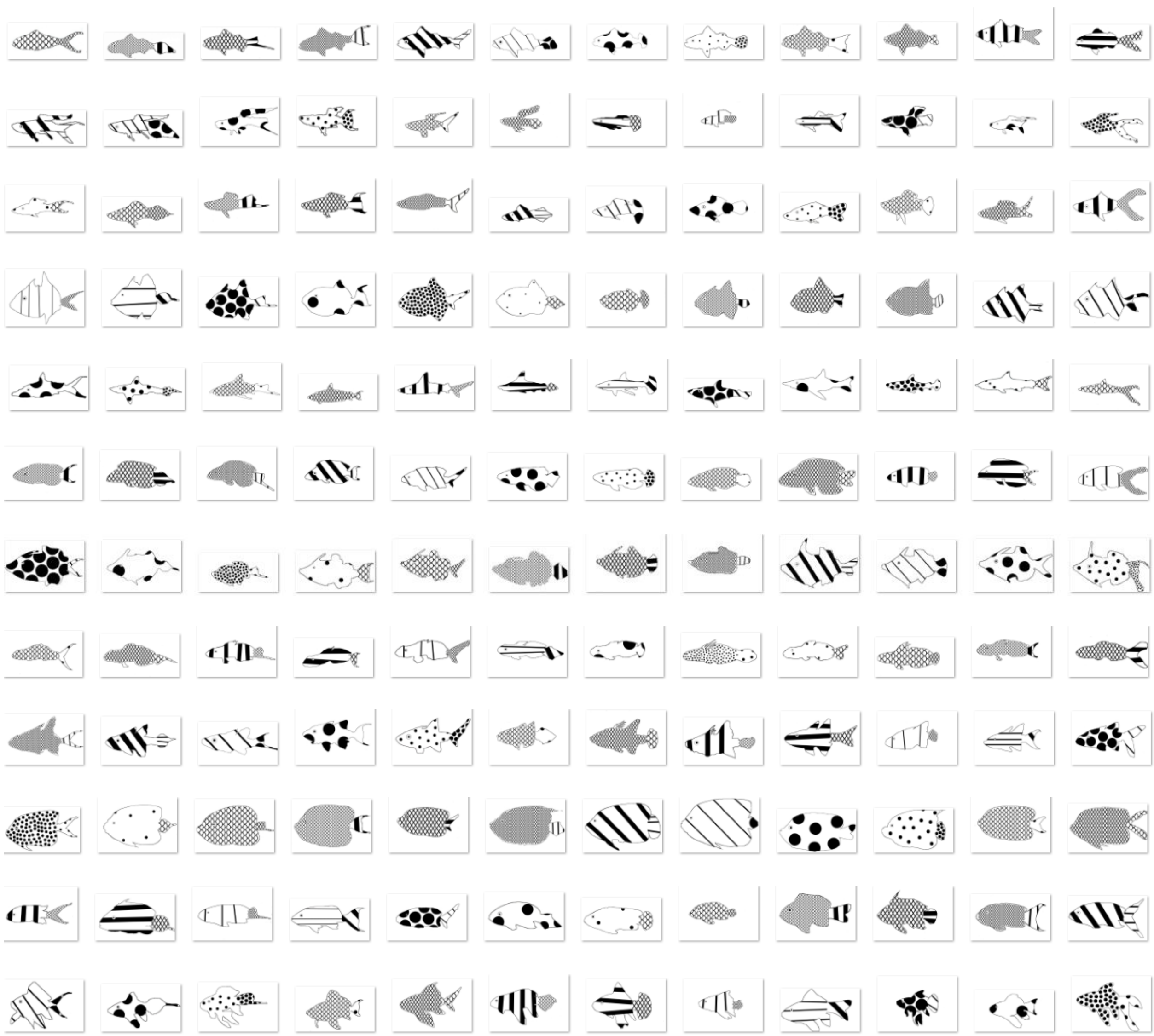


Figure A2. Examples of the fish stimuli that were used in pilot experiments. Each row represents a category that is defined by the pattern of the body whereas each column represents a category that is defined by the pattern of the caudal fin (i.e. tail).

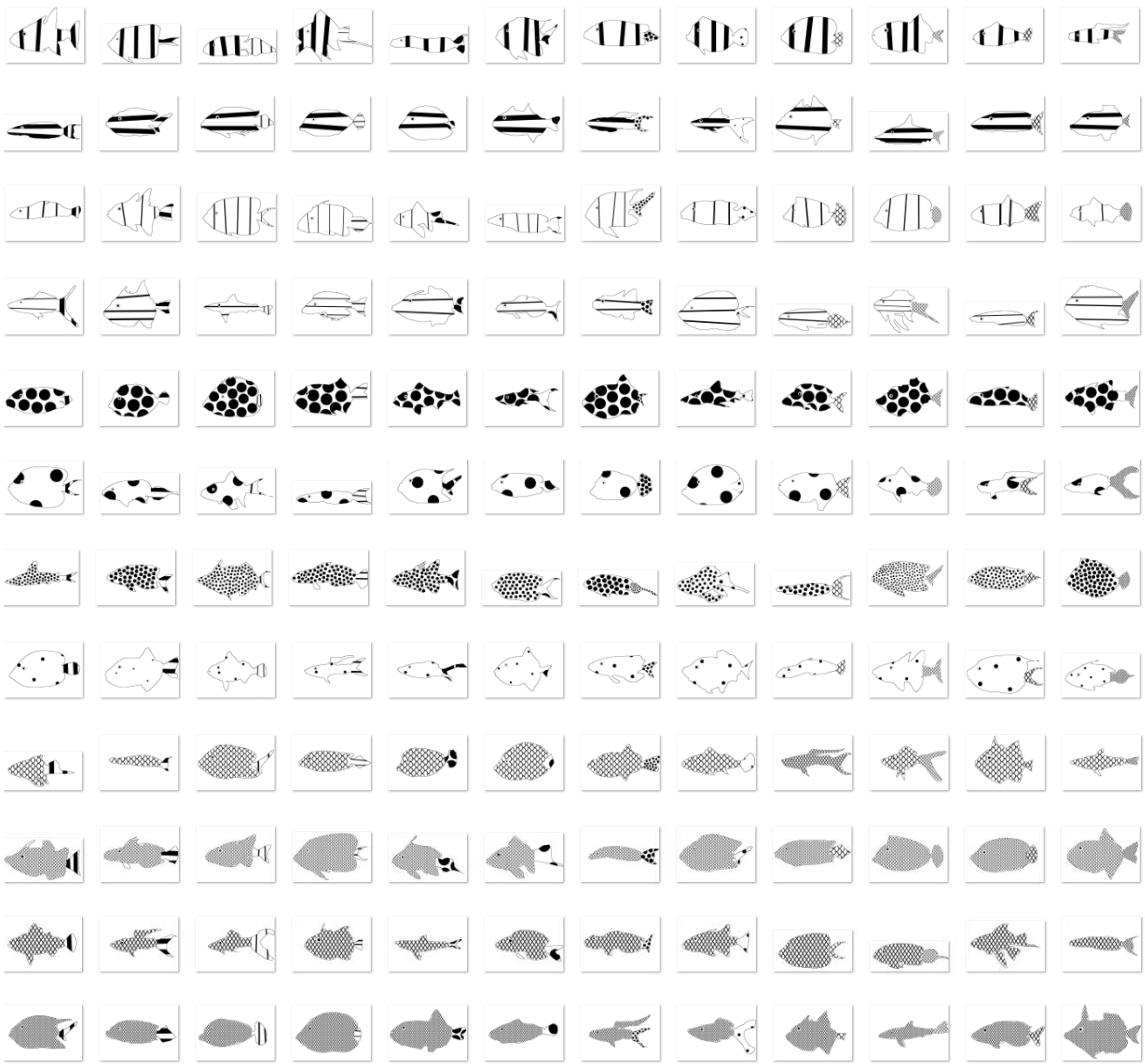


Figure A3. The results from the first set of pilot experiments. The error bars represent ± 1 *SEM*.

Appendix B

A Brief Description of the Second Set of Pilot Study with Artificial Fish Material

Given that the introduction of geometric patterns and the significant reduction in material complexity may have compromised the processes underlying the picture-size effect, in the second set of pilot study, I created an artificial fish category structure in which the shape of the body (global) or the tail (local) was strictly diagnostic by combining natural shapes and natural patterns using a raster graphics editor.

Figure B1 shows the material used in the second set of the pilot study. In Pilot Experiment 2, half the participants studied and classified categories organized by each row which were defined by the shape of the body (global diagnostic condition) while the other half of the participants studied and classified categories organized by each column which were defined by the shape of the tail (local diagnostic condition). Importantly, because all the features are naturalistic, the material complexity is similar to the fish material with which the picture-size effect has been previously demonstrated.

Figure B2 shows results from two pilot experiments using the artificial fish with natural features. In the first iteration shown in the left panel, participants studied and classified 12 categories of artificial fish shown in Figure B1. Given the near floor performance in this experiment, in the following experiment participants studied and classified six categories that seemed to be as distinguish from each other. However, the picture size effect was not replicated even when the performance was adequately high (in the global diagnostic condition).

The reason for why the picture-size effect failed to emerge using this material is unclear. In the process of creating this artificial material, some crucial aspects of natural category likely have been disrupted. Nonetheless, given the difficulty in obtaining the picture-size effect using artificial material, I decided to test the unique predictions form the two hypotheses by identifying natural categories with clear global or local diagnostic dimension as reported in the current dissertation.

Figure B1. Examples of the fish stimuli that were used in pilot experiments. Each row represents a category that is defined by the shape of the body whereas each column represents a category that is defined by the shape of the caudal fin (i.e. tail).

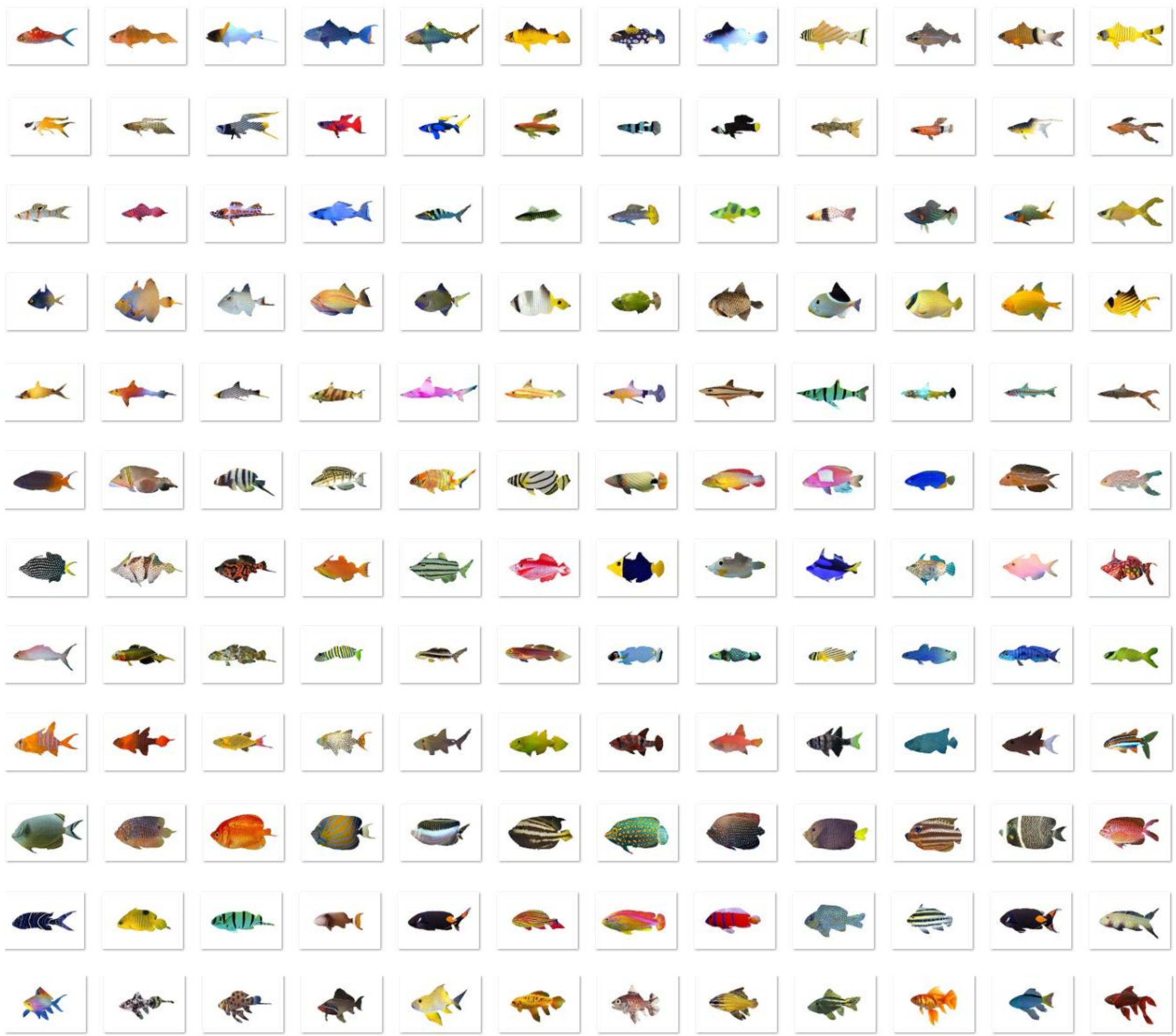


Figure B2. The results from the second set of pilot experiments. The error bars represent ± 1 *SEM*.