1

On the Future of Object Recognition: The Contribution of Color

David J. Therriault
University of Florida
USA

1. Introduction

Cognitive theories of object recognition have traditionally emphasized structural components (Biederman, 1987; Grossberg & Mingolla, 1985). The idea that object recognition is largely driven by shape was advantageous to theory building because of its economy (i.e., only a single dimension needed to be attended and there are a finite number of mutually exclusive components). However, recent work provides evidence that surface level information (e.g., object color) is readily used in object recognition (Rossion & Pourtois, 2004; Tanaka & Presnell, 1999; Therriault et al., 2009; & Naor-Raz et al., 2003). The purpose of this chapter is two-fold: to present results from experiments that more closely examine color's influence on object recognition and to reconcile these results with traditional theories of object recognition.

Section 2 contains a historical overview of the claims made between strucutral (i.e., edge) and view-point dependent (i.e., surface + edge) characterizations of object recognition. Although the debate may be subsiding over the status of viewpoint invariance, many open questions remain concerning how color contributes to the processing and recognition of objects.

Section 3 reviews conflicting research on the role of color in object recognition. Some studies fail to find any effects of color upon recognition, others find evidence for only high color diagnostic objects, and still others find that color readily influences recognition. This section concludes by offering some explanations for differences in obtained results.

Section 4 presents a recent set of experiments from my lab exploring the role of color in recognition, conceptualization, and language use. Most striking, the results from four different experiments are identical with respect to color. The presentation of correctly colored items always enhanced recognition and conceptualization of the objects.

In Section 5, the early conceptual analogy used in object recognition (i.e., speech segmentation) is reviewed and updated. I propose that object recognition is more anlagous to word recognition in reading. This is a more apt analogy because it can accomodate both structural and view-point evidence.

Finally, Section 6 argues that evidence calls for a more nuanced, flexible and integrated theory of object recognition, one that includes both bottom-up and top-down processing. The chapter concludes that the study of color vision is a fruitful area from which to gain a deeper understanding of object recognition generally; and that this pursuit would benefit greatly from the contribution of disciplines beyond cognition (e.g., neuroscience, biology, and linguistics).

2. Structural and view-based accounts of object recognition

Research examining human object recognition has historically been polarized between two views (Hayward, 2003; Hummel, 2000; Tarr & Bülthoff, 1995). The first view, and still the predominant one, argues that a structural approach best characterizes how we recognize objects in our environment. A quick review of three introductory cognitive textbooks confirms the solid footing of structural approaches in the field (i.e., all of these textbooks' coverage of object recognition ends with example figures of structures). The most prominent structural theory remains Beiderman's (1987) RBC (i.e., recognition by components) theory. According to this theory, a finite set of mutually exclusive structural components called geons are the mainstay of object recognition and representation (Biederman, 2007; Biederman, 1987; Biederman and Bar, 2000, Biederman and Gerhardstien, 1995; Biederman and Ju, 1988). Geons are volumetric structures created from the contrasts of two dimensional edges based upon symmetry, curvature, parallelism, and co-termination. Figure 1 contains a sample of geons.

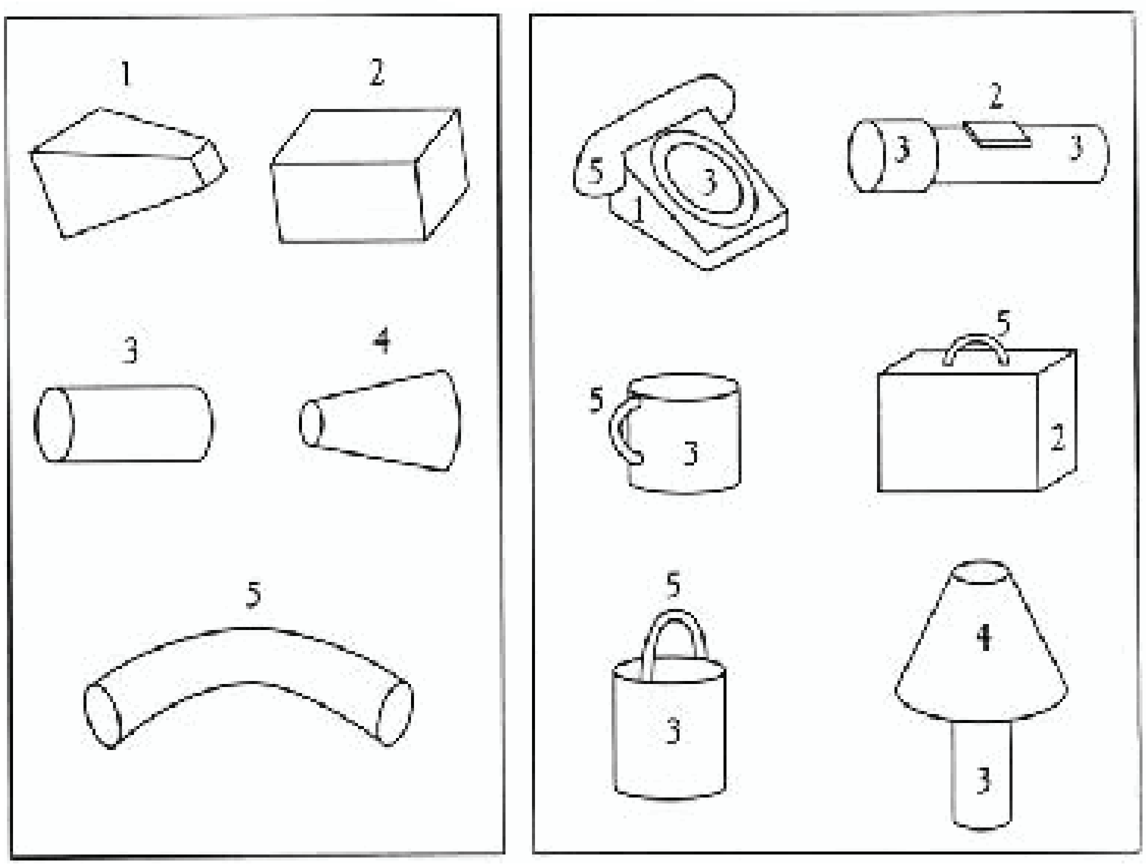


Fig. 1. A sampling of geons (left panel) and common objects with their constitute geons labelled (right panel). (From Biederman, 1990).

These structures are thought to underpin our ability to represent objects, in that, to recognize an object we must first decompose it into its constituent parts and "build" our representation. Geons are the smallest unit upon which elements of an object can be differentiated. One of the stronger claims of RBC theory is that these structures are processed without respect to surface features (they are said to be *invariant* to viewpoint, size, texture, or color). Evidence suggests that these structures are also fairly resistant to occlusion and interference from visual noise. Researchers who adopted the strong version of this theory typically documented the contribution of edge-based information in recognizing objects.

In a view-dependent or an edge + surface account of object perception, elements other than geons contribute in meaningful ways to object recognition. Some structural approaches, for example, Marr (1982) and Marr and Nishihara (1978) argue that surface level information is a necessary step in the process of recognition but only in the service of shape. Perhaps the most well researched aspect of surface level is our understanding of an observer's perceived viewpoint of objects. The impetus for research on this topic probably came from the strong claims of viewpoint invariance in the early RBC model. Hayward and Williams (2000), Tarr and Bülthoff (1995), and Tarr and Pinkert (1989) all provided evidence for recognition costs (i.e., decreased reaction times) associated with rotating the viewpoint of an object from its original presentation, casting doubt upon the invariance built into the RBC model. The more an object is rotated from its original studied view, the longer recognition takes. There are also models of object recognition that make explicit use of surface features. For example, Poggio and Edelman (1990) created a computer model of a neural network that learned to recognize 3-dimensional images in different orientations using a view-based matching algorithm (i.e., geons were not included in the model).

The 90's debate surrounding interpretations of viewpoint was largely a matter of degree. Structuralists first argued for invariance, later conceding that viewpoint could aid object recognition (under very specific conditions). Those exploring edge + surface explanations documented elements of recognition that could not be accommodated in a structuralist framework. The role of color in object recognition remains an open question, but it appears to be following the same research trajectory as viewpoint.

3. Contributions of color research

3.1 Color information is ancillary to object recognition

Beiderman and Ju (1988) first argued that structural (edge-based) properties of objects are theoretically preferred over viewpoint, texture, and color information. It is not the case that these features can't be used, but that they are only useful in certain circumstances when object shape is compromised or extremely variable (e.g., sorting laundry, Biederman & Ju, 1988). Beiderman and Ju (1988) assessed color contribution by measuring participants' naming times of simple line drawings of objects compared to the fully-detailed color pictures of those objects. Beiderman and Ju (1988) failed to obtain any significant differences between the naming times of the two versions of the objects. If surface and color information contributed to recognition, then the fully detailed color versions of the pictures should have been named more quickly. Beiderman and Ju concluded that color and texture were not the primary means to object recognition.

Similarly, Ostergard and Davidoff (1985) examined the contribution of color to object recognition. They provided evidence that color pictures elicited faster naming times, but that presenting the objects in their correct color didn't matter. They explained this result indirectly as a function of shape. That is, color provided extra luminance or contrast that aided in shape extraction. In a follow-up experiment, Davidoff and Ostergard (1988) produced evidence that color did not impact reaction time (in a semantic classification task). They concluded that color is not part of the semantic (i.e., meaningful) representation of objects. They left open that there may be some other representation of objects that includes color information (e.g., ancillary verbal information). Cave, Bost, and Cobb (1996) explored color and pattern manipulations of pictures in repetition priming. They demonstrated that changes in color did not influence repetition priming; whereas, shape did. Cave et al. concluded that repetition priming is insensitive to physical attributes that are not attended (i.e., color or size).

3.2 Color information is an inherent property of objects

In contrast to these results presented above, evidence for the importance of color information has been compounding. Price and Humphreys (1989), Tanaka and Presnell (1999) and Wurm et al. (1993) all had participants engage in some form of an object classification task (i.e., does a picture match a previously presented word). They found that color information facilitated the recognition of objects, but only those with very strong color associations. For example, an orange colored carrot (i.e., high color diagnostic HCD object) was named more quickly than its grayscale compliment; but there were no differences in reaction time between color and grayscale versions of a sports car (i.e., low color diagnostic LCD object). These studies provide evidence that color is an important component in object recognition, but *only* for highly color diagnostic objects. Naor-Raz et al. (2003) also explored color diagnosticity in a Stroop task where participants named objects or words that were matched or mismatched with their appropriate color. They found that response times were significantly faster for objects in their typical color (e.g., a yellow banana) than atypical (e.g., a purple banana). This pattern was reversed when colored words were used to describe the objects (i.e., seeing the word banana in either yellow or purple ink). Naor-Raz et al. (2003) concluded that their results provide evidence that color is encoded in object representation at different levels (i.e., perceptual, conceptual, and linguistic).

Evidence also implicates color processing in recognition of everyday objects that are not color diagnostic. Rossion and Pourtois (2004) revisited the naming times of the Snodgrass and Vanderwart object picture set (260 objects) in which they created three conditions: line drawings (the original set), gray-level detailed drawings, or color detailed drawings. They found that color aided recognition, and that while this was more pronounced for color diagnostic items, color also aided the recognition of low color diagnostic or variable colored items (e.g., man-made objects).

3.3 Explaining the conflicting findings

There are several explanations for conflicting results with respect to color. Probably the most pronounced is the fact that researchers have disagreed on the nature of color diagnosticity (and which items are most appropriate). For example, color diagnostic items tend to be vegetables, fruits, animals, and man-made objects. Studies emphasizing shape often use only man-made items, while those emphasizing color include more natural objects. Nonetheless, the distinction of category has recently been excluded as the predominant reason for conflicting findings, as suggested by Nagai and Yokosawa (2003) and Therriault et al. (2009). Of greater concern is that studies that argue that color is not important in object recognition often do so from a null result. That is, these studies report an *absence* of evidence as evidence that color is not utilized (Biederman & Ju, 1988). Simply put, it is problematic to accept the null hypothesis; it does not provide a solid base to build theory.

4. Our contribution to understanding the role of color in object recognition

4.1 On developing color object stimuli

In a recent article, Therriault, Yaxley, and Zwaan (2009) explored a range of recognition and object representation tasks using color stimuli. We made use of highly detailed photographs of objects. There are several important points to note about our selection of stimuli and their development. First, we only selected high color diagnostic items, most were concepts adapted from Naor-Raz et al. (2003). As noted by Tanaka and Presnell (1999), color diagnostic items used in earlier studies were later found to be problematic (e.g., camera or flowerpot). Consequently, we excluded any objects that were identified as problematic from earlier studies. Once we obtained quality photos, the pictures went through a washing process where we removed all color information (i.e., we transformed them to grayscale using Adobe Photoshop). This insured that once we re-colored the objects they would only contain one color and that we could directly control this color (i.e., all red object colors used the exact same red).

Three different color versions of the objects were created: grayscale, appropriately colored (congruent), and inappropriately colored (incongruent). This departs from previous studies that typically employ two conditions (a grayscale image compared to the appropriate colored version or studies that pit an appropriate colored object against an inappropriately colored version). Experimentally, our design allows comparison of the relative contribution of color (appropriate and inappropriate) to a control (the grayscale image).

Each picture occupied a 3 inch square space (72 pixels per inch) presented on a white computer background controlled using the software program E-Prime (Schneider et al., 2002). Also included in our design were 72 filler items that were not color diagnostic and were randomly colored. The filler items were incorporated to de-emphasise the likelihood that participants would become aware of the color diagnostic nature of our experimental items. The final 24 experimental objects were created in one the following range of colors: brown, green, red, and orange and were repainted with the appropriate translucent color (using the standard RGB code values for each of our colors).

Figure 2 presents two example stimuli in each of the three conditions (for demonstration simplicity, I only included red items). One potential criticism against using color diagnostic items as stimuli is that they are all either food items or animals, and that these could be treated differently than man-made objects. In our study, more than a third of our experimental pictures were man-made objects (see figure 3 for two example man-made items).

4.2 Experimental tasks and results

Therriault et al. (2009) created a set of 4 experiments using the stimuli described above. In Experiment 1, participants were asked to name objects and their time to respond was measured. Experiment 1b used the same stimuli but queried participants if a presented word matched a subsequent picture (while measuring reaction time). Experiment 2 used a rebus paradigm (i.e., participants read sentences with inserted pictures). A critical noun in a sentence was replaced by its picture and reading time was recorded (Potter, et al., 1986). Experiment 3 mirrored Experiment 2 but used an earlier contextual sentence in an attempt to override the congruent color of the object (e.g., a pumpkin is described as painted green in the sentence prior to the presentation of the target sentence with the pictured object).

Fig. 2. Example natural stimuli demonstrating color conditions: incongruent, black and white, and congruent; respectively (From Therriault, et al., 2009).

Fig. 3. Example man-made stimuli demonstrating color conditions: incongruent, black and white, and congruent; respectively (From Therriault, et al., 2009).

Experiment 1 provided a measure of pure recognition. Our results indicated that images presented in congruent color facilitated naming time, whereas incongruent color information actually interfered with naming time (when compare with the control gray-scale image). Experiment 1b provides information on the conceptualization/visualization of the object, as participants had to verify if a presented word matched its picture. Again, congruent color facilitated verification decisions, whereas incongruent color information interfered with verification. Experiments 2 and 3 provided a test of object recognition in which the task was to use the information in the context of comprehending a sentence. In both cases, the same pattern emerged: congruent stimuli aided recognition processes and incongruent stimuli harmed recognition processes. The consistency in color processing across different methods is striking. Below, Figure 4 presents the reaction time data for all of our experiments (error bars depict standard error).

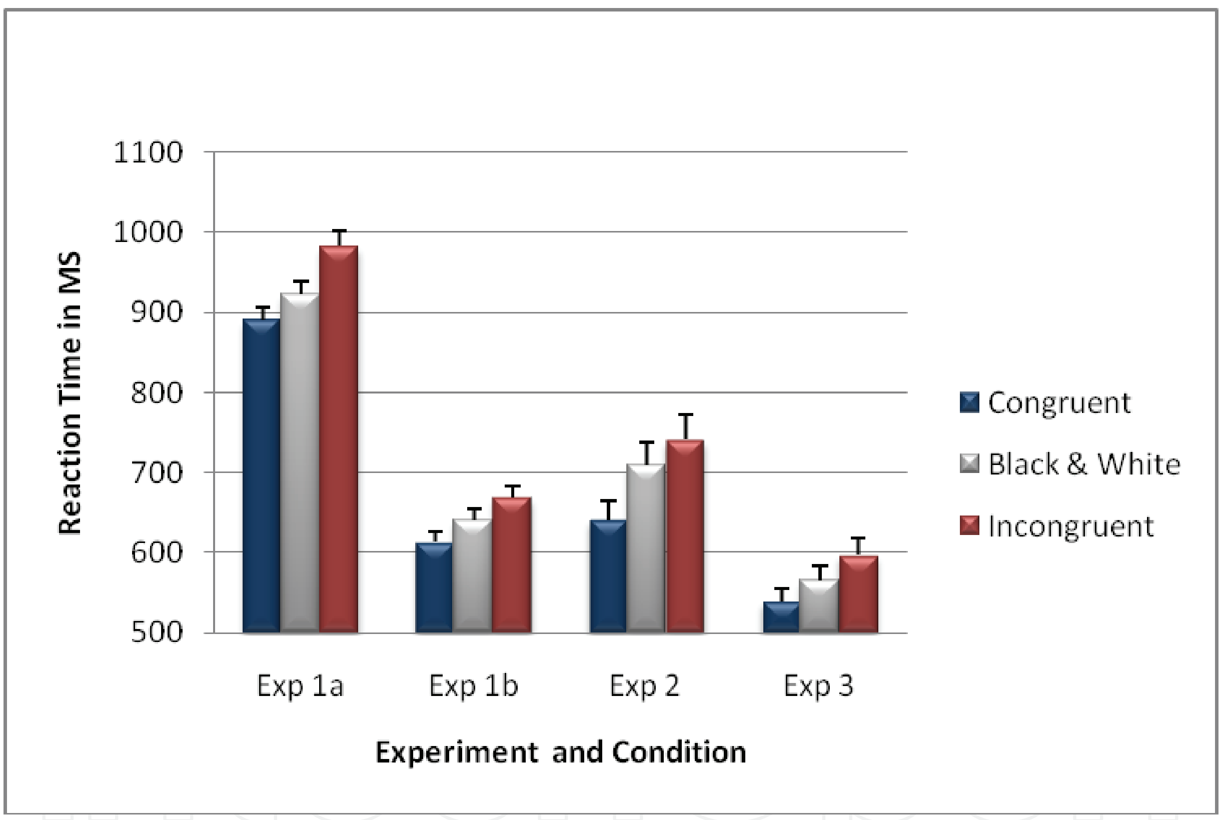


Fig. 4. Reaction time results of all experiments (From Therriault et al., 2009)

We would argue that the experimental bar is set high for our color items. In isolating color we had to present stimuli that were not completely natural. For example, notice that the stems of both the apple and strawberry are incorrectly colored. However, we can be certain that a single color was responsible for differences in reaction time. Results from our experiments consistently demonstrate that object recognition is much more flexible than relying on simple shape extraction from brightness, depth, and color. Knowing that a strawberry is red contributes to recognizing that object in a fundamental way, above and beyond its shape.

5. On finding the right conceptual analogy in object recognition

5.1 The original speech segmentation analogy

Biederman's (1987) article was a landmark paper; to this day it remains a highly cited and informative guide to those interested in object recognition. In that piece, Biederman enlisted research on speech perception. In short, he argued that object recognition is akin to speech segmentation (i.e., the idea that although speech is a continuous sound wave, the listener splits these sounds into primitives in their mind). For example, a novice learning a new language will often complain that it is difficult to tell where one word begins and another stops. Often, comunication at this stage is characterized as gibberish. With skill, the learner begins to make the proper segmentations in the soundwave to distinguish words. In English, all of the words we can create are formed on a small set of primitives or in linguistics called phonemes (there are roughly 46). From these primitives we can form thousands of words and even create new ones. So too, geons are the primitives that we can combine in a multitude of ways to help us recognize and distinguish objects in our environment.

5.2 A proposed analogy: word recognition and the word superiority effect

One could argue that we do not need to stray too far from the visual domain to find an appropriate analogy that captures the nature of both structural and view-based approaches to object recognition. A good candidate would be the recognition processes employed during reading (i.e., word identification). Considerable research in cognitive psychology has documented the contribution of individual letters (bottom-up) and word knowledge (top-down) in word recognition. A fairly well known demonstration is the word superiority effect (Rayner & Pollatsek, 1989; Reicher, 1969). In a typical experiment exploring this effect, participants are presented with a single word, a single letter, or a pseudo-word (on a computer screen) and asked if the display contained a critical letter. For example, given one of the following stimuli (*cork*, *o*, or *lork*), the participant would be asked if the display had an *o* in it. At first blush, one would assume that the letter *o* in isolation would lead to the fastest verification times. This is not the case. Participants were significantly faster to verify the letter *o* in the word *cork* than the *o* in isolation or the pseudo word *lork*. These counter-intuitive results are easily explained as a confluence of bottom-up (i.e., the processing of the individual letters) and top-down processing (i.e., knowledge of the word cork and our experiences with it as a whole unit). Word recognition isn't discriminatory; any activation that helps in the recognition process will be used. In this example case, there are two levels of potential activation with a word that we know (and, incidentally, why we don't see the effect with non-words). In the same fashion, geons represent the parts, bottom-up approach to object recognition; whereas, view-based information and surface features are often better characterized as top-down. Object recognition mirrors word recognition; any activation that helps in the recognition process will be used.

6. Synthesis and concluding remarks

Similar to the word superiority effect, Therriault et al.'s data (2009) can be taken to provide evidence for a *color superiority effect*--the stimuli from our study easily map onto reading (i.e., an incongruent colored object is equivalent to a pseudo-word; a congruent colored object is equivalent to a known word; and a grayscale image is equivalent to a letter in isolation). Our reactions times also mirror the pattern obtained in reading research on the word superiority effect.

Structural accounts of object recognition provide a solid base to ground the shape component of recognition, but they are simply not sufficient to accommodate color. Color is an intrinsic property of many objects and is represented at all levels of the cognitive system as reviewed in this chapter and even in low-level categorization of scenes (e.g., Goffaux et al., 2005; Olivia & Schyns, 2000). Structuralists argued that those who examine surface features (e.g., color) are essentially arguing for a view-based template theory (Biederman, 2007; Hummel, 2000). At the heart of this debate was an either-or-approach, pitting features against templates. Current views on object recognition are much more integrative and pragmatic. Foster and Gilson (2002), Hayward (2003), and Tanaka et al. (2001) all provide examples of how research benefits from the integration of structural and view-based approaches. I would offer that the research presented in this chapter provides an opportunity to build a more complete, albeit less economical, explanation of object recognition.

So, where is the future of color research in object recognition heading? The tent exploring elements of object recognition is large enough to accommodate a more diverse group of disciplines beyond perception (and we would all benifit from it). For example, research in biology suggests that the brain has evolved to separate brightness, depth, color, and movement (Livingston & Hubel, 1987). This begs the question, what ecological advantage does color vision provide? Is it a surprise that color diagnostic items are often natural items (e.g., food or animals)? Primate research provides evidence that vision has optimized to differentiate edible fruits from background colors (Summer & Mollon, 2000). Similarly, Changizia, Zhang, and Shimojo (2006) provide evidence that primate vision has also optimized for colors associated with skin and blood. In the area of cognition, Stanfield and Zwaan (2001), and Zwaan et al. (2002, 2004) all demonstrate rapid interactions between language and visual representations. Connell (2007) and Richter and Zwaan (2009) point out that text color can make use of (interfere) with the representation of object color. There remain challenges with respect to the timing of recognition and its integration (modularity), but research in these varied disciplines will bring us a more complete picture of the role of color in object recognition.

7. References

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115-147.

Biederman, I. (1990). Higher-Level Vision, In *Visual Cognition and Action*, D. N. Osherson et al., The MIT Press, MA.

Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research, 39,* 2885-2899.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for 3D viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 1162-1182.

Biederman, I., & Gerhardstein, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition: Reply to Tarr and Bülthoff (1995). *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1506-1514.

Biederman, I., & Ju, G. (1988). Surface vs. Edge-Based Determinants of Visual Recognition. *Cognitive Psychology, 20,* 38-64.

Cave, C. B., Bost, P. R., & Cobb, R. E. (1996). Effects of color and pattern on implicit and explicit picture memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22 (3),* 639-653

Connell, L. (2007). Representing object color in language comprehension. *Cognition, 102*, 474-485.

Davidoff, J. and Ostergaard, A. (1988) The role of color in categorical judgments. *Quarterly Journal of Experimental Psychology, 40,* 533–544

Foster, D. H., & Gilson, S. J. (2002). Recognizing novel three-dimensional objects by summing signals from parts and views. *Proc. R. Soc. Lond. B. Biol. Sci. 269*, 1939-1947

Goffaux, V., Jacques, C., Mouraux, A. Olivia, A., Schyns, P. G, & Rossion, B. (2005). Diagnostic colors contribute to the early stages of scene categorization: Behavioral and neurophysiological evidence. *Visual Cognition, 12 (6),* 878-892

Grossberg, S. and Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures and neon color spreading. *Psychological Review*, 2 (2), 173-211.

Hayward, W. G. (2003). After the viewpoint debate: Where next in object recognition. *Trends in Cognitive Sciences, 7,* 425-427.

Hayward, W. G., & Williams, P. (2000). Viewpoint costs and object discriminability. *Psychological Science, 11,* 7-12.

Lacey, S., Hall, J., & Sathian, K. (2010). Are surface properties integrated into visuohaptic object representations? *European Journal of Neuroscience, 31*(10), 1882-1888.

Livingston, M. S., & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *Journal of Neuroscience, 7,* 3416-3468.

Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.

Marr, D. & Nishihara, H. K. (1978). Representing and recognition of the spatial organisation of three-dimensional shapes. *Proceedings of the Royal Society, London, B200*, 269-294.

Naor-Raz, G., Tarr, M. J., & Kersten, D. (2003). Is color an intrinsic property of object representation? *Perception, 32,* 667-680.

Oliva, A. and Schyns, P. (2000) Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41,* 176–210

Ostergaard, A. and Davidoff, J. (1985) Some effects of color on naming and recognition of objects. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 11,* 579–587

Potter, M., Kroll, J. F., Yachzel, B., Carpenter, E., & Sherman, J. (1986). Pictures in sentences: Understanding without words. *Journal of Experimental Psychology: General, 115,* 281-294.

Price C. J., Humphreys, G. W. (1989). The effects of surface detail on object categorization and naming. *Quarterly Journal of Experimental Psychology A, 41,* 797-828.

Rayner, K., & Pollatsek, A. (1989). *The Psychology of Reading*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81 (2),* 275–280

Richter, T. & Zwaan, R.A. (2009). Processing of color words activates color representations. *Cognition*, 111, 383-389.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33,* 217-236.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's Guide.* Pittsburg: Psychology Software Tools Inc.

Stanfield, R.A. & Zwaan, R.A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12,* 153-156.

Summer, P., & Mollon, J. D. (2000). Catarrhine photopigments are optimized for detecting targets against foliage background. *Journal of Experimental Biology, 203*, 1963-1986.

Tanaka, J. W., & Presnell, L.M. (1999) Color diagnosticity in object recognition. *Perception & Psychophysics, 61,* 1140–1153

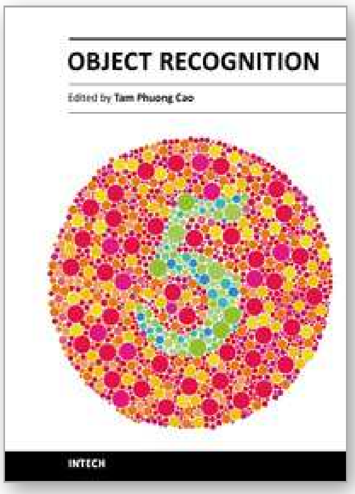
Tanaka, J. W., Weiskopf, D. & Williams, P. (2001). Of color and objects: The role of color in high-level vision. *Trends in Cognitive Sciences, 5,* 211-215.

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon-structural-descriptions or by multiple-views? *Journal of Experimental Psychology: Human Perception and Performance, 21(6),* 1494-1505.

Wurm, L. H., Legge, G. E., Isenberg, L. M., & Luebker, A. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 899-911.

Zwaan, R.A., Madden, C.J., Yaxley, R.H., & Aveyard, M.E. (2004). Moving words: Dynamic mental representations in language comprehension. *Cognitive Science, 28,* 611-619.

Zwaan, R.A., Stanfield, R.A., Yaxley, R.H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13,* 168-171.



Object Recognition

Edited by Dr. Tam Phuong Cao

ISBN 978-953-307-222-7

Hard cover, 350 pages

Publisher InTech

Published online 01, April, 2011

Published in print edition April, 2011

Vision-based object recognition tasks are very familiar in our everyday activities, such as driving our car in the correct lane. We do these tasks effortlessly in real-time. In the last decades, with the advancement of computer technology, researchers and application developers are trying to mimic the humanâ€™s capability of visually recognising. Such capability will allow machine to free human from boring or dangerous jobs.