論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	What humans learn from the visual world : insights from task-irrelevant processes of visual stimuli
著者(和文)	星野英一
Author(English)	Eiichi Hoshino
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第10431号, 授与年月日:2017年3月26日, 学位の種別:課程博士, 審査員:中村 清彦,新田 克己,三宅 美博,青西 亨,長谷川 修
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第10431号, Conferred date:2017/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

What humans learn from the visual world:

insights from task-irrelevant processes of visual stimuli

Department of Computational Intelligence and Systems Science

Eiichi Hoshino

Abstract

Visual input from the world is abundant in information. We extract relevant information from visual input depending on task demands. Together with this task-relevant process, task-irrelevant processes occur automatically (e.g. Hollingworth, 2007; Fiser & Aslin, 2001). Without goals, task-irrelevant processes contain ill-posed problems regarding absence of what to learn definitely. It has yet to be elucidated that what kinds of memory representations are available from visual knowledge learned through task-irrelevant manners. Based on previous studies, I hypothesize that only the view-specific representation is available in task-irrelevant processes of visual information. The overall aim of this thesis is to investigate what kind of memory representation is available in task-irrelevant processes of visual information.

Here, I conducted two original studies. In Study 1, I investigated memory representation of objects in scenes when scenes were combined into comprehensive spatial information of an individual space. Study 1 consisted of a study phase and a test phase. In the study phase, the subjects were instructed to learn objects during scene integration or on a monochromatic background with attention control. In the test phase, the subjects were instructed to judge whether objects with rotation in depth were presented in the learning phase. Based on the result, it was suggested that memory representations of objects in scenes without attention may be processed into a two-dimensional representation bounded to the scene as a texture.

In Study 2, I investigated memory representation of scenes in implicit learning. Two-dimensional scenes, which were generated from an artificial grammar, were presented to the subjects. The results indicate that the subject could learn rules of the

artificial grammar in two-dimensional patterns. It is suggested that knowledge of a finite experience of visual scenes could be crystalized in different levels of relations among visual objects in each individual scene representation.

Together with results of Study 1 & 2 and previous studies, I conclude that only the view-specific representation is available in task-irrelevant processes of visual information.

Contents

Chapter 1. Introduction

1.1 Task-irrelevant processes of visual information

The world is three-dimensional although the visual input is two-dimensional. We rely much on visual information to understand how the world consists. We collect a set of information in flow of visual input from the environment instinctively without conscious effort or formal instruction (Cosmides & Tooby, 1994) and construct knowledge of the visual world. The knowledge involves not only task-relevant but also task-irrelevant processes. In task-relevant processes, relevant information is selected from visual input which is potentially abundant in information. The relevant information is intensively processed to achieve task goals. At the same time, it has been known that task-irrelevant processes occur automatically (Hollingworth, 2007; Janzen & Turennout, 2004; Fiser & Aslin, 2001, 2005; Reber, 1989; Meegan, 2005), as described the details in Chapter 2.

We often utilize task-irrelevant processes. For instance, when you lose your way on the way back home from your friend's house, you might happen to have feeling that a scene lying before you is somewhat familiar to you and relying on that feeling may lead to successful navigation. Consider that your primary task was to get to your friend's house by using some landmarks which had been taught by your friend. The task-relevant process was recognition of the landmarks. However, in the situation where explicit knowledge of the landmarks did not work, other knowledge such as the feeling of familiarity to a scene would help, which might be acquired through task-irrelevant processes. In fact, a metacognitive state that you feel you know it, termed the feeling of

knowing, is helpful to judge as accurately as above chance levels (Nelson, Gerler & Narens, 1984). Nevertheless its validity, little is known about task-irrelevant processes of visual information. To date, much of social demands and hence academic studies focus on how preset goals are achieved with lucid teaching evidences in explicit manners (Pothos et al., 2011). Behind the scenes, task-irrelevant processes occurs by aggregating associative information that is regardless of goals but is possibly worth to individuals for the future use (Hutchinson & Turk-Browne, 2012).

The distinction between task-relevant and task-irrelevant processes is an objective distinction of how we process visual information. Closely related conceptions viewed from subjective perspective are explicit learning and implicit learning, which are a fundamental and ubiquitous process of cognition. Explicit learning is defined, in contrast to implicit learning, as learning with conscious awareness. For example, learning objects and letter sequences are explicit learning in a task requiring to remember them whereas learning associated information of the objects and the rule underlying sequences are implicit learning (Janzen & Turennout, 2004; Reber, 1989). Consequently learning something other than primary task demands is considered as implicit learning. Implicit learning seems to be of the highest utility especially in situations that require making decision about unfamiliar events and that cannot be solely solved with knowledge from explicit learning (Reber, 1989). This thesis discusses studies of implicit learning altogether.

Before move onto the next section, I will describe definitions of terms used in this thesis. An *Object* is defined as that it has size, occupies a space and is perceived as a single unit of perception. If multiple *objects* are joined together, the resultant is also called an *object* (Spelke, 1990). Behaviorally, an *object* is defined by answers to the

question “what is there?”. A *scene* is defined as two-dimensional information that consists of multiple objects. *Representation* is defined as what form memory has (Marr, 1982).

1.2 Issues

The task-irrelevant processes have utility in coping with uncertainty of the world (Hutchinson & Turk-Browne, 2012). However, without goals, task-irrelevant processes contain ill-posed problems regarding absence of what to learn definitely and bring about "curse of dimensionality" with indefinite information to be processed. Evidences suggest that task-irrelevant processes of visual information are rather limited within a range. For example, it has been shown that memory of objects in a picture of a scene involves relative position to the scene but not absolute position to the frame of the picture (Hollingworth, 2007). The range of task-irrelevant processes of visual information, however, has yet to be explored. Unveiling the nature of visual task-irrelevant processes leads to understanding of human cognition in which implicit learning might play important roles. By doing so, we could pursue concepts as well as sense or feeling beyond perception because they might be learned implicitly. In addition, it offers opportunities to investigate broad knowledge that fuels human creativity without limits of what to learn.

It is assumed that there are several kinds of memory representations that could be encoded during visual task-irrelevant processes. Visual input is always a scene. First, as intact information, view-specific representation can be encoded. Next, by segmenting objects from the scene and calculating information about the objects, such as what they are and where they are, object-based representation can be encoded. Both of these representations can be helpful in recognition of scenes, or objects embedded in scenes, in in later occasion. However, view-specific representation is advantageous to recognition of scenes, but disadvantageous to recognition of objects because recognition of objects requires object segmentation from view-specific representation. In contrast,

object-based representation is advantageous to recognition of objects, but disadvantageous to recognition of scenes due to additional load of integrating object representations into scenes. These disadvantaged cases would result in slow reaction time or poor accuracy. In addition to these two representations, abstract representation can be encoded, if multiple scenes are seen. Abstract representation is encoded through extracting common information from scenes and has typical features of them. This representation has an advantage that it can reduce memory usage. If scenes contain depth information spatial representations could be encoded. There are two spatial representations, namely, egocentric and allocentric representations. Egocentric representation is viewpoint dependent representation of space and is equivalent to the view-specific representation. On the other hand, allocentric representation is view independent representation of space and is map-like representation that helps to navigate the environment. As visual input is always viewed from a specific viewpoint, to encode allocentric representation needs additional process on egocentric input. Broadly speaking, memory representations of visual information are view-specific as intact and object-based and abstract as processed.

The overall aim of this thesis is to investigate what kind of memory representation is available in task-irrelevant processes of visual information. Specifically, I focus on task-irrelevant processes of visual information, which is not intended to remember explicitly. It will provide the human nature of incidental judgment in visual cognition according to a previous experience under task-irrelevant processes. Together with previous studies, I hypothesize that only the view-specific representation is available in task-irrelevant processes in visual information.

Here, I conducted two original studies. In Study 1, I investigated memory

representation of objects in scenes when scenes were combined into comprehensive spatial information of an individual space. Study 1 investigated whether task-irrelevant processes of scenes construct object-based representations or view-specific representations. Study 1 consisted of a study phase and a test phase. In the study phase, the subjects were instructed to remember a particular kind of objects while not instructed to remember the rest of objects. The objects instructed to remember were task-relevant objects while the rest were task-irrelevant objects. One group learned objects during combining scenes. Another group learned objects on a monochromatic background. In the test phase, both groups of subjects were instructed to judge whether objects, with/without rotation in depth from viewpoints used in the learning phase, were presented in the learning phase. The results showed that rate of correct judgment for task-irrelevant objects viewed from novel viewpoints were significantly lower than that from familiar viewpoints, in the condition of scene integration. Based on the result, it was suggested that memory representations of objects in scenes with little attention may be processed into a two-dimensional representation bounded to the scene as a texture.

In Study 2, I investigated memory representation of scenes in implicit learning. Study 2 investigated whether task-irrelevant processes of scenes construct an abstract representation from multiple scenes or multiple view-specific representations. The subjects were instructed to learn two-dimensional scenes (exemplars), which were generated from an artificial grammar, in the learning phase. After reaching a criterion of learning, the subjects viewed novel scenes (probes), which consisted of scenes generated from the same artificial grammar and scenes not generated from the same artificial grammar, in the judgment phase. The subjects were instructed to judge whether the scenes were the same rule of scenes presented in the learning phase. The

analysis was conducted using dissimilarities among patterns, which are defined with n-gram probabilities and the Levenshtein distance. The results showed that subjects were able to learn rules of two-dimensional visual patterns (exemplars) and made categorical judgment of probes based on knowledge of exemplar-based representation. My analysis revealed that subjects' judgment distinguishes exemplars, which are similar to probes in configural relations of visual elements, suggesting the existence of configural processing in exemplar-based representations. In addition, the subjects' judgment distinguishes exemplars which are little similar to probes in element-based processing, implicating the elimination of dissimilar exemplars. Exemplar representation was preferred to prototypical representation through tasks requiring discrimination, recognition and working memory. Relations of the studied judgment processes to the neural basis are discussed. I conclude that knowledge of a finite experience of two-dimensional visual patterns can be crystalized in different levels of relations among visual elements.

Together with results of Study 1 & 2 and previous studies, I conclude that only the view-specific representation is available in task-irrelevant processes of visual information.

1.3 Contents of this paper

This thesis is composed of five chapters. I described the significance and the aim of the thesis in Chapter 1. Related studies were given in Chapter 2. Chapter 3 and 4 provided my original studies Study 1 and Study 2 in detail, respectively. In Chapter 5, I discussed the results of Study 1 and 2 and made a conclusion. Figures, references and my publications are listed at the end of the document.

Chapter 2. Related studies

2.1 Task-irrelevant processes of visual stimuli

Humans automatically process visual information other than task demands. A study demonstrating object-position binding to scenes indicates that humans process objects' relative position to scenes (Hollingworth, 2007). Hollingworth presented his subjects to scenes containing multiple objects. The subjects were instructed to remember objects and objects' orientation. They were also informed that they would be asked if objects were left-right mirror reflected in a later recognition task regardless of positions of test objects. Thus, the task-relevant processes were to remember objects and their orientation. After the presentation of learning scenes, the subjects were instructed to answer whether test objects were mirror reflected. The test objects were in either of four conditions: test objects presented in/without scenes of the learning phase and positions of test objects were same/different position compared with the learning phase. The results showed that, only in the scene present cases, the correct rate of test objects presented in the same position were significantly higher than that of test objects presented in different position. No significant difference was observed in the scene absent cases. This indicates that the subject automatically processed task-irrelevant information, positions of objects relative to scenes. This task-irrelevant process was observed both in natural scenes drawn in a perspective view and in two-dimensional scenes placing objects in a grid.

In addition, when comparing performances of object recognition for objects presented at the same position as the leaning phase, objects presented within scenes of the learning

phase were better remembered than objects presented in isolation. Actually, he demonstrated it one year prior to this study and named it the object-to-scene binding effect (Hollingworth 2006).

Fiser & Aslin presented multiple scenes that containing sequences of objects to their subjects over time. The subjects passively viewed the scenes and showed sensitive to conditional probability (i.e. $P(A|B)$) between objects (i.e. A and B) of the sequences (Fiser & Aslin, 2001, 2005). Thus, irrelevance to task demands, the subjects automatically calculated the conditional probability over the temporal dimension. It is suggested that conditional probability provides representation of potential chunk in scenes.

Janzen and Turennout demonstrated that the parahippocampal region collects spatial information in relation to various objects, in a joint encoding of space and objects (Janzen & Turennout, 2004). They designed excellent but simple experiment in which subjects were passively explored a virtual reality museum on a computer screen with special attention to objects of a specific group. The museum was modeled with objects placed on tables that were located at either diverging points where subjects needed to decide which way to go, or just single turning point where no decision would be demanded. After 25 minutes of study phase, functional magnetic resonance images were obtained while subjects performed a simple object recognition test in which isolated objects were presented from a canonical view on a white background, in a two-alternative forced-choice manner. The result was that the parahippocampal gyrus was more activated for objects located at decision points compared to those at non-decision points. It was suggested that neural activations of the parahippocampal region reflects the navigational relevance of landmark objects, which process spatial

information and possibly involve in encoding of the navigational relevance. To produce comprehensive spatial information, it is necessary that the egocentric representations which primary obtained from perceptual information are combined together. In a case of absence of self-motion, the production of comprehensive spatial information is crucially relied on the integration of different scenes. If scenes are episodically dispersed, it is strongly required that there be navigational objects in the scenes to combine these scenes into coherent spatial representations.

2.2 Visual memory

Problems in recognition of visual objects and scenes arise from visual memory retention. Early studies on retention of photographs showed that visual memory was highly robust and held large capacity (Nickerson, 1965; Shepard, 1967). However these studies used distractors highly different from target stimuli (Hollingworth, 2005). According to the Atkinson and Shiffrin memory model (Atkinson & Shiffrin, 1968), long-term memory is transferred through short-term memory. Thus, knowing the nature of the visual short-term memory (VSTM) would help to understand the nature of the visual long-term memory (VLTM).

Hoshino and Mogi investigated that decay of VSTM using three retention periods (Hoshino & Mogi, 2010). Their subjects were instructed to remember identities and locations of nonsensical objects. It was hypothesized that if object memory contained information of identity and location, decay of object memory would result in deletion of both identity and location. However, if information of identity and location were separable in processing, then memory decay would first bring about false memory with deletion of either identity or location. The result showed that there was false memory

before deletion, supporting the latter hypothesis. In addition, VSTM was investigated in the context of change blindness paradigm. According to the studies memory representations of the visual details were actually poor (O'Regan, 1992; O'Regan& Noe, 2001; Rensink, O'Regan & Clark, 1997). However, when the attention was administered to a target object in a natural scene, participants showed relatively accurate discrimination performance on subsequent long-term memory tests (Hollingworth & Henderson, 2002). The change detection performance on the objects after a short-term and even 24-hour delay was significantly high, indicating object-based representation of VSTM and VLTM is robust (Hollingworth, 2005). Although VSTM has a restricted capacity of objects representations (e.g. at most 3 or 4 objects (Luck & Vogel, 1997)), the representations are gradually consolidated into VLTM which has considerably large capacity (e.g. modest forgetting of presented objects with 402 intervening objects and no forgetting with 10) (Hollingworth, 2004). A subsequent eye movement recording study indicates that humans are naturally shifting attention to objects in scenes (Henderson, 2010) and forming object-based representation.

In contrast to object-based representations of scenes, view-specific representation is encoded when layouts or contexts of scene are focused. In such cases, response time was retarded when subjects viewed the same scenes with unfamiliar viewpoints (Diwadkar & McNamara, 1997) and accuracy of recognition test decreased when subjects viewed fragmented scenes (Biederman, 1972) and objects in isolation (Hollingworth, 2006). Hollingworth's study illustrated strong object-to-scene binding in VSTM and VLTM (Hollingworth, 2006). View-specific representation can have depth information, particularly when views depth cues, such as perspective view or binocular disparity (Diwadkar & McNamara, 1997; Chua & Chun, 2003). Benefit from

view-specific representation is faster detection time when spatial configurations (context) around targets are the same (Chun & Jiang, 1998). This effect is termed contextual cueing. Contextual cueing occurs task-irrelevantly without consciously recognizing contexts. Contextual cueing is diminished when scenes are viewed from unfamiliar viewpoints in three-dimensional space (Chua & Chun, 2003).

2.3 Spatial representations

The nature of spatial representations underlying human cognitive system has long been discussed (Tolman, 1948). Evidences suggest that both egocentric and allocentric representations (Klatzky, 1998) exist in parallel and support successful navigation. Burgess's research indicates complementary roles for these representations, with increasing dependence on allocentric representations with the amount of movement between presentation and retrieval, the number of objects remembered, and the size, familiarity and intrinsic structure of the environment (Burgess, 2006). There are likely to be several differences in the nature of memory involved in both representations. The allocentric representations are thought to be primarily stored in long term memory integrated from several independent egocentric representations, whereas egocentric representations is often invoked as working memory. Although these two representations are modality general, supposing, in case of visual domain, the allocentric and the egocentric representations are comparable to abstract and view-specific representations, respectively. It has not been shown for certain how and where the allocentric long-term-memory (LTM) is finally encoded. Unlike animals, which primarily use path integration system (which is based on idiopathic or motor sensory information) for spatial update, human is able to construct a spatial

representation from sequential views on computer screen without self-motion (Ekstrom et al., 2003). Their study showed a neural network of human spatial navigation based on cells that responds (1) to place or location primarily in the hippocampus, (2) to views of landmarks in the parahippocampal region and to the subject's navigational goals and (3) to conjunctions of place, view and goal in the frontal and temporal lobes. Involvement of the parahippocampal area in encoding of scenes (Epstein et al., 1999; Bar & Aminoff, 2003) and object locations (Maguire, 1998) has been observed in several studies.

2.4 Object recognition

Understanding the mechanism underlying object recognition is one of the subjects of intense debate. There are two major theories of how object recognition is achieved: viewpoint-specific process is affected by objects orientation comparing with previously stored visual memory (Tarr & Bülthoff, 1995), whereas viewpoint-invariant process extracts objects characterizations termed "geon structural descriptions" (GSD) which consists of a fundamental parts of object and their relationships (Biederman, 1987, 2000). Human is capable to infer the novel orientations of an object (Biederman & Bar, 1999), possibly using either GSD (Spetch & Friedman, 2003) or mental rotation (Shepard & Metzler, 1971). If a presented object is somewhat familiar, it would be easier to infer the different orientations from one-shot view. Several studies demonstrated that extraction or formation of orientation-invariant representations from even a novel object is highly automatic and quick, in repetition blindness (RB) paradigm (Harris & Dux, 2005; Coltheart, Mondy & Coltheart; 2005). In RB, the second occurrence of a repeated stimulus is less likely to be reported, compared with the

occurrence of other stimuli. In experiment of Coltheart et al., repeated stimulus was inserted with depth rotation and significant repetition blindness was found for all orientation differences (Coltheart, Mondy & Coltheart; 2005).

Understanding of visual object embraces recognition, identification and categorization, which imply distinction or judgment in reference to prior knowledge (Palmeri & Gauthier, 2004). Such knowledge involves memory of perception and motion concerning objects and more abstract concept (Barsalou, 1999; Quinn & Eimas, 2000, Tyler & Moss, 2001; Sloutsky, 2010). A current view of neural representation posits that perceptual and motor memories are stored in distributed brain regions that overlap or are connected with regions active during learning (Martin, 2007; Damasio, 1989; Barsalou, 1999; Patterson, Nestor & Rogers, 2007). However, it is controversial as to how concept is represented. One view suggests that conceptual processing arises from simultaneous activation of relevant areas for tasks (Damasio, 1989; Martin, 2007; Gainotti, 2011) or the other postulates an amodal site for integration of multimodal information (Patterson, Nestor & Rogers, 2007). Evidence suggests that category selectivity is represented in the temporal lobe. As well as strong category selective areas, the fusiform face area (FFA) (Kanwisher, McDermott & Chun, 1997) and the parahippocampal place area (PPA) (Epstein & Kanwisher, 1998), remaining area of the inferior temporal cortex (IT) may serve a function of categorical representation (Kriegeskorte et al. 2008). Specifically anterior part of IT serves integration of task relevant features (Sigala & Logothetis, 2002). The fusiform gyrus shows across-category sensitivity concerning objects' feature statistics in a way that objects with more shared features or attributes with others (i.e. animals share many features: "has four legs", "has eyes", etc. than tools) are represented more laterally than those

with less shared features according to a feature-based model (Tyler et al., 2013). In fact, the lateral fusiform gyrus is activated during viewing and matching biological objects (face and animals) whereas the medial fusiform gyrus is activated by non-biological objects (tools and houses). Moreover, faces engage more focally to the lateral portion of the fusiform than animals and tools engage more laterally than houses (Chao, Haxby & Martin, 1999). In contrast to the fusiform, the anterior temporal lobe (ATL) exhibits within-category sensitivity in fine-grained recognition rather than basic-level (Rosch et al., 1976) discrimination. Studies of brain lesion show that semantic dementia (SD) patients, who have damage to ATL, represent loss of semantic abilities, specifically in the category of living things, and become unable to make fine distinction between highly similar objects, suggesting that ATL plays a role of amodal hub-site for semantic knowledge (Patterson, Nestor & Rogers, 2007). In addition, ATL shows sensitivity for familiarity (Jefferies & Ralph, 2006) and feature co-occurrence (Tyler et al., 2013). Taken together, similarity and familiarity are the important issues in neural representation of knowledge. I explored how similarity affects judgment with controlling familiarity of exemplars in Study 2.

2.5 Unsupervised category learning

Animals have ability of unsupervised category learning. They instinctively adopt arrangements of environmental cues to their understanding of the world in a way that they get advantages to survive. For example of the visual domain, indigo buntings learn layout of stars to navigate themselves (Emlen, 1975). Honey bees released at a site new to them can go back to their hives by selecting the best matching snapshot of known scenes and comparing it with their current retinal image (Cartwright & Collette, 1987).

Bees use certain visual cues, such as edges and colors of scenes (Collette & Collette, 2002) and arrangement of landmarks. Those abilities of learning arrangements in certain biased ways seem to be genetically implemented in evolution (Marcus, 2004). As the highest of organisms, humans have a flair for several kinds of patterns. Humans categorize complex patterns with extensive training (Reber, 1967; Ashby & Maddox, 1992), possibly including aforementioned stellar cues and scenes as well if taught. In addition, humans tend to conjecture and to attribute rules to information even it is random, which sometimes results rationally fallacious reasoning like clustering illusion (Kahneman & Tversky, 1972; Gilovich, Vallone & Tversky, 1985) although these irrational short cuts of reasoning may rather help category construction under uncertainty (Gigerenzer & Goldstein, 1996). It is important to explore what subjective judgment reflects in relation to subjective experience of exemplars, as equally as studies of how humans extract rules.

Chapter 3. Memory representation of objects in scenes when scenes are combined into comprehensive spatial information of an individual space (Study 1)

3.1 Introduction

We are able to recognize objects from novel viewpoints (Biederman, 1987). Especially, it is easy if they are objects used in daily life, such as chairs. In that situation, object representations are segmented from a scene. However, it is suggested that object representation is bound to a scene, as demonstrated by Hollingworth's results that accuracy of object recognition is higher when objects are presented in learned scenes than presented in isolation (Hollingworth, 2006). This "object-to-scene binding effect" (Hollingworth, 2006) is a task-irrelevant process of visual scenes. In addition, Janzen & Turennout demonstrated that the parahippocampal region processes objects as well as spatial information, in a joint encoding of objects and scenes (Janzen & Turennout, 2004). This indicates that memory representation of an object includes information of its surrounding.

Taken together, it is suggested that memory representation of objects is bound to memory representation of scenes, regardless of still pictures (Hollingworth, 2007) or a movie of virtual reality environment (Janzen & Turennout, 2004). The previous studies, however, did not address whether memory representation of objects contains three-dimensional information. In the previous studies, subjects viewed objects from familiar viewpoints during recognition. If memory representation of objects contains

three-dimensional information, subjects must be able to recognize objects from novel viewpoints. In that case, objects representation are segmented from scenes. In addition, it was ambiguous how much the subjects felt depth of scenes although it is important to recognize depth of scenes to investigate whether memory representation of objects contains three-dimensional information.

The interpretation of scenes is an important factor in contextual behaviours and the formation of episodic memory. The classic key idea of spatial information processing has been analysed by O'Keefe and Nadel (O'Keefe & Nadel, 1978). Together with discovery of "place cells" and concept of "cognitive map" (O'Keefe & Dostrosky, 1971), they proposed the cognitive map theory in which place cells, dead reckoning system and landmark navigation are combined into allocentric map-based representation in hippocampal formation. In addition, Yamaguchi et al. have proposed a mechanism of hippocampal memory encoding of episodic events from novel temporal inputs caused by theta phase precession (Yamaguchi, 2003). To produce allocentric long-term memory (LTM), it is necessary that the egocentric representations primarily obtained from perceptual information are combined together. In a case of absence of self-motion, the production of allocentric LTM crucially relies on the integration of different scenes. If scenes are episodically dispersed, it is required that there be navigational landmarks in the scenes to combine these scenes into coherent allocentric representations.

The current study focuses on the observation of object recognition underlying human cognition after episodically dispersed views are combined into comprehensive spatial information of an individual space. Additionally, attention enhances the visual LTM (VLTM) of previously attended objects embedded in a natural scene (Hollingworth & Henderson, 2002), which is supported by a dynamic evolution model on attention and

memory (Wang & Yu, 2006), suggesting that object representations in LTM may also be affected by attention.

The aim of this study is to reveal whether object representation has three-dimensional information when objects are learnt during scene integration. I conducted two experiments in which the subjects were instructed to remember particular objects. In Experiment 1, objects were presented three-dimensional perspective scenes and the subjects were required scene integration whereas in Experiment 2 just two objects were presented.

3.2 Methods

Two groups of subjects were participated in either of two experiments. Both experiments consisted of a learning phase and a test phase. The procedures of the test phase were the same in both experiments. To investigate the nature of object representations within human cognition in LTM that is learnt during scene-integration, Experiment 1 was designed especially with regard to viewpoints with attention control, in the context of integration of the spatial information. In the leaning phase, participants were, in turn, viewed two dispersed views in which the several objects were located. They were instructed to remember objects on green bases and their position whereas those on blue bases were distractors at the moment. Likewise, Experiment 2 was carried out as the same as Experiment1 but the background in learning phase was changed to monochromatic. After the learning phase, they were required to conduct a two-alternative forced-choice recognition test. The objects presented in the test phase were divided into three types; i.e. objects viewed from same viewpoint as the learning phase, those from different viewpoint and novel objects. Chairs were used as objects

which were chosen from everyday use objects in same basic-level category to make recognition from depth-rotated viewpoints easy. The experiments were designed to focus on studying the nature of object representations in LTM, rather than in short-term memory (STM) (Hayward et al. 2006) in which the multi-angle object representations would be more easily established than those in LTM.

Experiment 1

Participants: Nine subjects (two females and seven males) aged from 22 to 60 years old) participated in Experiment 1. All subjects reported normal or corrected-to-normal vision.

Learning phase

Stimuli and apparatus: A simulated exhibition was set up with virtual reality computer graphics. A space was constructed of randomly chosen junctions (e.g. right- or left-turn corner or crossroads) with randomly placed four chairs chosen out of a pool of 62. Two of four chairs were put on green bases and the two others were on blue bases as indicated in Figure 1 (a) and (b). Scenes from two perspectives were taken with a resolution of 640 x 480 pixels and a focus of view of 45 degree.

Procedure: The first scene depicted the area around a junction (’decision points’), while the second scene was taken from the same junction. In each trial, two scenes were presented for seven seconds each in turn, with an interval of black background of 1.2 seconds (Figure 1 (a), (b)). One of the two chairs on a green base was placed in the junction area in the first scene (a), which also appeared in the second scene (b) in front, serving as an ”anchor” to the second scene. Another chair on a green base was

presented only in the first scene. The subjects were instructed to remember chairs on green bases and their locations. Thus, chairs on green bases were task-relevant objects and those on blue bases were task-irrelevant objects. After 1.2 seconds presentations of the two scenes, participants were asked to point out the position of the chair by moving the red pointer placed at the centre of the scene with the mouse (c), similar to the task in egocentric pointing (Wang & Spelke, 2000). In the pointing task, the first scene in the leaning phase was overlaid with a grid and at the bottom of this scene; a scene of the chairs on green base in second scene was presented. There were ten trials (two scene presentations + a pointing task) in which a total of forty chairs were presented.

Figure 1. Experiment 1- Learning phase.
A sample stimulus set in a trial. An object learning task was performed during scene-integration. Participants were required to remember the objects on green bases and its location. (a),(b) The learning snapshots. The objects on the green bases are attended objects while those on blue bases are unattended. (c) The location pointing task. The object at the bottom, which was only visible in the second scene (i.e, (b)), was required the subjects to point out the location within the egocentric view of the first scene.

Test phase

Stimuli and apparatus: Stimuli used this phase were images from two different viewpoints for each of forty chairs presented in the learning phase. One of images was chosen from one of presented viewpoints in the learning phase. Another image was

chosen from one of un-presented viewpoints (i.e. 90, 180 or 270 degrees rotated in depth from viewpoints in the learning phase). In addition, 44 images, which consisted of 22 novel chairs with two viewpoints, were prepared. The total of 124 images then individually presented in a random sequence.

Procedure: Thirty minutes after the set of learning phase, test phase was performed (Figure 2). The subjects were instructed to answer whether chairs had been presented in the learning phase regardless of viewpoints, as quickly as possible by pressing a button (”yes” or ”no”) in a two-alternative forced-choice manner. The chairs were presented either in the ‘same viewpoint’ as in the learning phase or in a ‘different viewpoint’. Thus, the subjects were asked to conduct a recognition task for chairs viewed from the same or different view points, which were attended (presented on green bases) or unattended (presented on blue bases) during the learning phase.

Figure 2. Test phase.
A sample stimulus set in a trial in the test phase. The object recognition test in which learnt objects were viewed from familiar or novel viewpoints was performed in a two-alternative forced-choice paradigm. The objects learnt in learning phase were displayed with a rotation (0, 90, 180 or 270 degrees) in depth.

Experiment 2

Participates: Three subjects (1 female and 2 males aged from 20 to 60 years old) participated in experiment 2. All subjects reported normal or corrected-to-normal vision.

Learning phase

Stimuli and apparatus: Scenes with two chairs was presented in a monochromatic background. Each chair was presented above either a green bar which indicated subjects required to remember, or a blue bar as a distractor. The properties of the three-dimensional perspective of scenes were set up the same as that of Experiment 1 and the chairs used were the same as Experiment 1. The total of sixty scenes were presented to each subject. Sixty scenes consist of twenty chairs viewed from two different viewpoints and twenty chairs viewed from one viewpoint. Viewpoints were randomly chosen from four viewpoints (a canonical view and 90, 180 or 270 degrees rotated in depth from the canonical view). Half of them were attended in the learning phase.

Procedure: Participants were instructed to remember chairs above green bars in a two-object-scene presented on a computer screen. Thus, chairs above green bases were task-relevant objects and those above blue bases were task-irrelevant objects. After 300 ms fixation, scenes were presented for 2000 ms followed by a blank until participants pressed the key to continue to the next trial. Total of 30 trials were performed by each participant.

Test phase

Stimuli and apparatus: The settings were the same as Experiment 1 except for number of novel chairs. 20 novel chairs viewed from two viewpoints, instead of 22, were used and therefore the total of 120 images were presented.

Procedure: Thirty minutes after the learning phase, the rotated object recognition test was performed. The experimental design was the same as the test phase in Experiment 1 other than the presented duration of a stimulus is substituted by 1000 ms.

3.3 Results

In Experiment 1, the rate of correct judgment (old or new) was significantly higher for the task-relevant objects (chairs on green bases) compared to the task-irrelevant objects (chairs on blue bases). A 2 by 2 repeated-measures analysis of variance (attention by viewpoint) on the rate of correct judgment showed a significant effect of attention ($F(1, 8) = 7.977$, $P = 0.022 < 0.05$), whereas no significant effect of viewpoints and no interaction between same and different viewpoints ($F(1, 8) = 0.397$, $P = 0.546$) (Figure 3 (a), (b)). Planned contrasts revealed that the rate of correct judgment for task-irrelevant objects was significantly lower than that for task-relevant objects when viewed from different viewpoints ($p < .05$) whereas there was no significant difference when viewed from the same viewpoints. The rate of correct judgment for task-irrelevant objects was higher when viewed from the same viewpoints as the learning phase than viewed from different viewpoints, although the difference was not statistically significant (Figure 3 (a)). In addition, the rate of correct judgment for objects located at a certain spatial configuration, such as right- or left-turn corner or crossroads, was not consistent in conditions with any particular viewpoints. A few participants reported that

remembering the objects and their location was too difficult and could not confidentially discriminate old or new for the most of the objects. However the tendency of less rate of correct judgment for the task-irrelevant objects when viewed from different viewpoints was observed among participants. A motivated participant showed that good performances in recognition for task-relevant objects viewed from both the same and different viewpoints, but again his performances for task-irrelevant objects viewed from different viewpoints was significantly poor.

In Experiment 2, no statistically significant difference was observed among the four conditions (Figure 3 (b)).

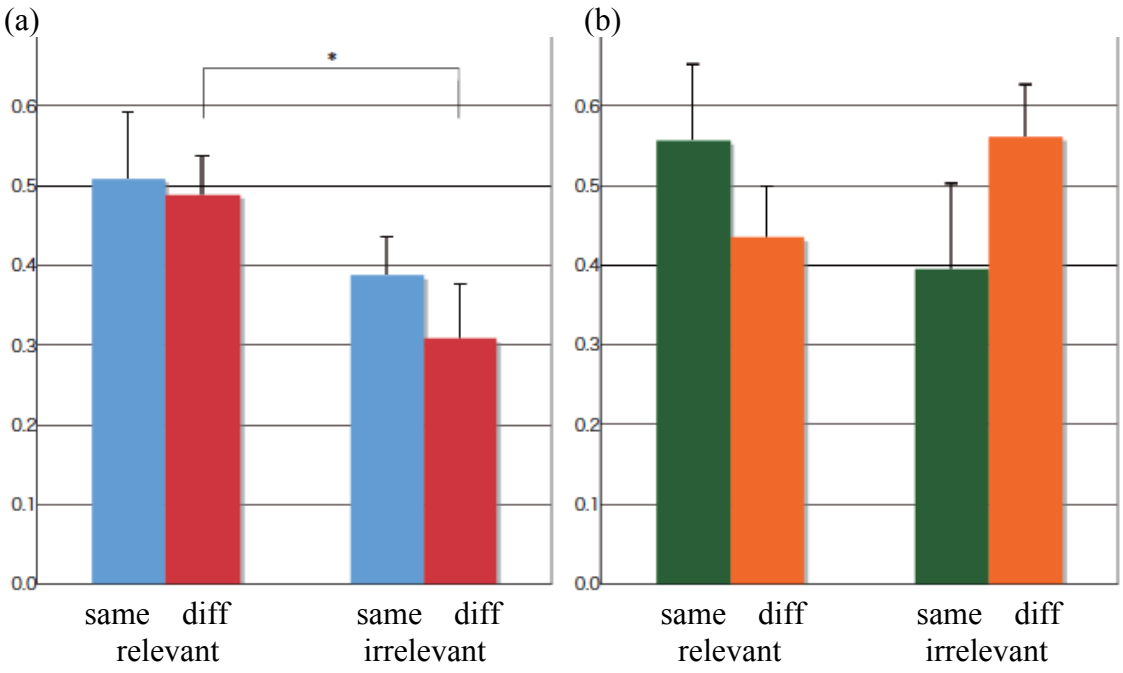

Figure 3. The rates of correct judgment for the objects. (a, b) Rate of correct judgment in Experiment 1and 2, respectively. The bottom labels indicates types of test objects, i.e. viewpoints and task relevancies. Error bars are standard errors across the subjects.

3.4 Discussion

The results of Experiment 1 showed that memory representations of task-irrelevant

objects did not have three-dimensional structural information. However, no effect of attention and viewpoints were observed in Experiment 2. The results indicate that memory representations of task-irrelevant objects are view-specific and two-dimensional like a texture in VLTM, when scenes are combined into comprehensive spatial information of an individual space.

Why the representations of objects are two-dimensional? In Experiment 1, the subjects were required to integrate scenes as well as to remember chairs. It is possible that there were little spare cognitive resources to process task-irrelevant chairs due to the high cognitive load tasks. It is suggested that implicit learning requires a certain minimal attention (Seger, 1994). A role of attention is extracting focal information and therefore a cognitive stage of object representations may be produced by attention. Thus, in the current study, attention may support to obtain the view-invariant or three-dimensional representations of objects from scenes, but in absence of attention, such representations were never obtained. Rather the object-to-scene binding effect (Hollingworth, 2006) gave rise to a perception of objects as a texture in the scene. In fact, the current result showed that overall memory of task-relevant objects was relatively well established compared with that of task-irrelevant objects, consistent with a previous work (Hollingworth & Henderson, 2002) which supported that attention to objects in scenes enhanced consolidation of memory representations of objects. In contrast, Experiment 2 indicates possibility that task-irrelevant as well as task-relevant objects were attended enough to process object segmentation from scenes in a form of three-dimensional representation because chairs were familiar three-dimensional objects in daily life. It is possible that object segmentation from scenes were easily processed in two-object condition because of no other objects except for two chars and no

requirement of scene integration. Therefore, it is possible that little attention may lead to memory representations of objects to be view-specific. Further experimental data collection is necessary to answer how much operation of scene integration itself might affect representations of objects.

In addition, it is possible that memory representations of scenes are also view-specific in task irrelevant manner. From the current result, it is considered that memory representations of task-irrelevant objects are bound to memory representations of scenes, consistent with previous studies (Hollingworth, 2006; Janzen & Turennout, 2004). Because memory representations of task-irrelevant objects are view-specific as discussed above, memory representations of scenes that bounded with that of objects might be view-specific too. The problem is what kind of information this view-specific representation contains. Evidences show that visual short-term memory does not contain detail of scenes (Rensink, O'Regan & Clark, 1997). However, in consideration of previous studies on visual scenes, I hypothesize the view-specific representation of scenes contain rules of objects' arrangement. I will investigate this issue in Study 2.

One might argue that information processing in Experiment 1 involved cognitive map or allocentric representation. Rigorously speaking, it is necessary to investigate if subjects construct such representation within the same protocol. However, evidences indicate that there would be egocentric representation or view-specific representation rather than allocentric representation in the situation of Experiment 1. Diwadkar & McNamara showed that memory representations of scenes viewed from multiple viewpoints are view-specific (Diwadkar & McNamara, 1997). Actually, it seems that active viewpoint control is necessary to form allocentric representation (Ekstrom et al. 2014). However, in Experiment 1, the subject passively viewed presentation of two

scenes. Thus, allocentric representation might not be formed. In addition, contextual cuing effect (Chun & Jiang, 1997), which involves task-irrelevant processes, diminishes when scenes are viewed from unfamiliar viewpoints (Chua & Chun, 2003). This indicates that memory representation concerning contextual cueing effect is view dependent. Boundary extension indicates that memory representations from viewing scenes contain larger space than the scenes (Intraub & Richardson, 1989). Altogether, it is possible that memory representation constructed in Experiment 1 was view-specific representation which might contains depth and larger 3-dimensional spatial information.

In Experiment 1, the accuracy on recognition of attended objects viewed from rotated viewpoints indicates that the role of attention is producing the view-invariant object representations. Moreover, orienting attention could predict that the nature of object representations would reflect objects' intrinsic representations but not three-dimensional spatial configuration around the objects (But see Hollingworth, 2006). This idea is supported by Mallot and Gillner's study that the local views and objects are recognised individually and not recognised as configurations among objects when navigating in a large-scale environment (Mallot & Gillner, 2000). The current results further indicate that selective activation for navigational objects in the parahippocampal gyrus (Janzen & Turennout, 2004) may not concern processes of three-dimensional spatial configurations. Because the difference between scenes of decision points and non-decision points were the openness of the scenes, it is possible that the activation may reflect types of scenes. There were no differences on recognition performances between the navigational objects and objects used in the pointing task. The manipulation of pointing object location might have given rise to the equivalence of recognition performance between the navigational objects and those used in the pointing

task.

Humans understand the spatial relationship using visual information, constructing egocentric and allocentric representations. Together with these representations, there is a feeling of spatial extent or ”presence”, during the exploration or observation of visual scene and interpreting it. The presence can be either in visible if it is online in which targets are within a view, or invisible if it is offline in which targets are within the integration of prior views. Here, I term the former online space as view dependent space (VDS) and the latter as view independent space (VIS). The comprehensive visual space is integrated through the encoding of individual views. A sensory data of individual views is accounted for the VDS while the mental state of integrated comprehensive space is accounted for the VIS. VIS is only possible to realize explicitly through a collection of VDS. Several researchers have been proposed concepts similar to VIS, e.g. the internal representation (O’Regan & Noe, 2001) and the view-independent three-dimensional descriptions (Mulligan & Daniilidis, 2000).

In the basic idea of philosophy, there are two types of reality which applied variety of academic fields (Weiss, 1996); one is the physical system of reality and another is the mental state of reality. In the visual spatial perception, view dependent reality such as stereopsis has been considered as the former physical reality underlying bottom up stream in the brain. This online reality is a VDS component. On the other hand, VIS is knowledge or mental state based on the involvement in the environment. In terms of visual reality, the act of viewing, which facilitates VDS, with VIS evoke more real reality than just viewing without VIS. Additionally, it is suggested that VIS is the common component of the comprehensive spatial information in the hippocampus, which is important in human spatial cognition. VIS, in the form of implicit memory, is

integrated within spatial exposure, often realized with ”presence”, resulting in involvement in the environment. There is another explanation on the reinforcement of novel view recognition for attended objects. In the Experiment 1, the subjects were instructed to perform the pointing task in which spatial manipulations were required. This spatial manipulation in presented scenes gave rise to involvement in the environment. The involvement may allow the spatial imagery more flexible to access VIS. As a result, the rate of correct judgment from different viewpoints was relatively high. Generally speaking, involvement in the environment is important factor to understand spatial configurations (Wexler & Boxtel, 2005). The involvement in the environment also explains why passive observers, like children accompanying adults are easy to lose their way.

In the current study, there would be four stages in which the involvement in the environment is considered to be generated: (1) just looking at the scene of three-dimensional perspective, (2) the scene-integration, (3) a mental process of inferring the location of a object in the pointing task and (4) the pointing action itself. Further study on involvement in the environment will be needed, including how it is generated and how much it affects on spatial cognition, as well as what makes differences between the nature of objects and scenes in terms of spatial memory formation.

Finally, the link between spatial memory and episodic memory will be discussed. Buzsáki suggested an analogy between the formation of context independent semantic memories from multiple overlapping episodes with common junctions among the episodes and that of time-independent comprehensive spatial information from overlapping multiple junctions of different routes with dead reckoning (Buzsáki, 2005).

Although the activities of place cells in rats are strongly based on path integration system (O'Keefe & Dostrovsky, 1971; O'Keefe & Nadel, 1978; Hafting et al., 2005), the overall mechanism including place cells which underlie mental process of the formation of context independent memories, commonly used in semantic and spatial representations, may be the same as that underlying the scene-integration and object recognition. If the inhibition priming observed in the current study is context dependent, a similar effect may be found in episodic and semantic memory. For instance, an exposure to a contextual sentence with attention control may inhibit the recognition from different aspects of the presented words such as meaning. Buzsáki also mentioned the contribution of theta rhythm oscillation to spatial memory and semantic memory formation (Buzsáki, 2005). The future direction would be analysis with the electroencephalogram (EEG) data, during integration of spatial memory or VIS, compared with the consolidation of semantic memory.

Chapter 4. Memory representation of scenes in implicit learning (Study 2)

4.1 Introduction

Humans receive information flow from the environment, often without conscious efforts or formal instructions (Cosmides & Tooby, 1994). In that process, humans construct knowledge of categories, whereby we can make judgment of a novel event as to whether it is a member of a group defined by previous experience (Ashby& Maddox, 2005). The arrangement of elements is one of the features that define a category, often found in music, language and design of textiles, architecture, landscapes and so on, which all seem to be elaborative and creative works by humans often accompanying an impression or a feeling. . In the visual domain, objects in the two dimensional visual field consist of components that make potentially infinite combinations. Knowledge about which parts of scenes are likely to be together (Biederman, 1972; Graef, Christiansen & Ydewalle, 1990) and which individual scenes are classified together (Friedman, 1979; Oliva, 2005) facilitate our understanding of natural scenes. However, the exact nature of the process in which humans organize initially nonsensical visual scenes into meaningful representations is not known. A key question is whether humans can construct categorical knowledge from two-dimensional visual arrangement alone.

The process of the learning arrangement patterns of stimuli has been studied under sequential exposures to auditory (Saffran, Aslin & Newport, 1996; Saffran, Johnson, Aslin & Newport, 1999; Marcus et al., 1999) and visual (Fiser & Aslin, 2001, 2002; Kirkham, Slemmer & Johnson, 2002; Stobbe, Westphal-Fitch, Aust & Fitch, 2012)

stimuli. Those stimuli are abstract and initially nonsensical for subjects by the exclusion of prior knowledge. Fiser and Aslin presented multiple scenes that contain sequences of elements over time. Subjects exhibited sensitivity to conditional probability (i.e. $P(A|B)$) between elements (i.e. A and B) of the sequences (Fiser & Aslin, 2001, 2005), where the conditional probability was calculated over the temporal dimension. These studies typically focus on the temporal frequency of multiple stimuli over time, and study subjects' sensitivity to this type of information. In terms of cortical processing, such analysis possibly involves the medial temporal lobe (Turk-Browne et al., 2009), where, computationally, information embedded in the mutual relations between elements are processed.

Contextual information is important in the categorical judgment of visual scenes consisting of a variety of elements projected to the retina, A scene containing some cars, lines in-between, and streetlamps probably depicts a car park: A scene consisting of a house and one or a few cars is likely to be a residential area. This kind of category judgment of scenes can be done almost irrespectively of feature complexity, as scene judgment occurs before the identification of features in the scenes (Oliva, 2005).

Category representation has been modeled in two lines of theories. The prototype theory posits that categorization is accomplished by referencing to a common representation or an averaged prototype from multiple exemplars (Rosch, 1973, 1975; Biederman, 1987; Smith & Minda, 2002). This common representation or an averaged prototype is also referred to abstract representation (Bruce & Young, 1986). On the other hand, the exemplar theory relies on the references to exemplars themselves (Kahneman & Tversky, 1972; Hintzman, 1986; Nosofsky, 1986; Poggio & Edelman, 1990; Tarr, 1995; Storms et al., 2000; Mack, Preston & Love, 2013). In addition, there

may be combined representations depending on the two approaches according to task demands (Ross & Makin, 2000; Smith, 2014; McMenamin et al., 2015). In the visual domain, the nature of category representation has well been studied in the recognition of objects (Palmeri & Gauthier, 2004). Evidences suggest that both abstract categorical knowledge of objects and exemplar-specific knowledge coexist in the left and right hemispheres, respectively (Marsolek, 1999), particularly in the fusiform cortices (Garoff et al., 2005). There is a controversy about view-dependency of object representation (Biederman, 1987; Tarr, 1995). In contrast to object-based and feature-based representations, less is known about the representation of visual arrangement. It has been shown that humans are able to generalize one-dimensional visual sequence with feedback but birds were not (Stobbe, Westphal-Fitch, Aust & Fitch, 2012), temporal visual sequence from passive viewing (Fiser & Aslin, 2002) and spatial configurations (Chun & Jiang, 1999; Fiser & Aslin, 2001). Westphal-Fitch et al. demonstrated human ability of detecting groups of elements and preference over well-ordered patterns such as grouped elements in two-dimensional visual patterns (Westphal-Fitch, Huber, Gomez & Fitch, 2012). It remains to be elucidated relation between exemplars of visual arrangements and categorical knowledge of them. Given that there are several evidences that humans can extract categorical regularities through statistical learning (Marcus et al., 1999; Brady & Oliva, 2008), it is hypothesized that humans learn visual arrangement in spatial statistical manners.

Effects of spatial frequency within each exemplar and collective information of multiple exemplars may arise because it is possible that exemplar-based information influences categorization, as suggested in the exemplar theory of category. Similarity of a probe to exemplars influences category judgment, automatically and mandatorily

(Hahn et al., 2010). Thus, it is important to consider how learning exemplars affects judgment of a probe regarding similarity. Exemplar-based knowledge of visual arrangement would enable the subjects to voluntarily find out rules within the presented elements and attribute them to individual events (Barsalou, Huttenlocher & Lamberts, 1998). A recent computational study suggests that humans may acquire such knowledge by learning parts of exemplars as well as relations between them (Lake, Salakhutdinov & Tnenbaum, 2015).

In the actual environment, humans seldom see objects or sequences of objects in isolation. Ensembles of objects constitute a scene, with various conditional probabilities between them. Humans are sensitive to conditional probabilities of sequences in the scene (Fiser & Aslin, 2001, 2005), which reflect rules that generate them. Rules are embedded in the collection of sequences and contain multiple elements with several conditional probabilities, which could be generated from a formal grammar. To the best of my knowledge, there haven't been sufficient experiments which show how humans learn rules within and across scenes, and how they use the acquired knowledge in later judgments of novel scenes. Such cognitive processes share properties with language acquisition, as both consist of elements (i.e. letters or words) with various conditional probabilities between them. The artificial grammar (AG) learning (Reber, 1967) is a useful paradigm to control such information and to study implicit learning. Patterns generated from AG are composed of distinct elements, which can be quantified in terms of the occurrence of frequencies known as n-gram probabilities (also known as transitional probabilities), and the Levenshtein distance (Gomez & Gerken, 1999; Levenshtein, 1966). Studies using AG have shown that humans are able to learn rules of visual sequences along a single (spatial or temporal) dimension (Stobbe et al., 2012,

Conway & Christiansen, 2009). It has been suggested that vision is better at extracting spatial order statistics than temporal order statistics (Conway & Christiansen, 2009). Visual sequence learning was affected by element positions in sequences (Conway & Christiansen, 2009).

Learning categorical knowledge is achieved through two types of manners, supervised and unsupervised. In supervised category learning, subjects are given feedback of predefined categorical rules and thus performance of learning can be objectively measured (e.g. Minda & Smith, 2001). On the other hand, there is no feedback in unsupervised category learning, and in laboratory based studies subjects are often not required to make correct judgment but subjective judgment on later occasions because underlying structures of events are not always apparent (e.g. Handel & Imai, 1972). In everyday life situation, supervised category learning arises, in prescriptive manner, from teaching rules or customs through language, social convention and education (Pothos et al., 2011) while unsupervised category learning produces self-organized knowledge of category through flow of input without instructions.

Unsupervised category learning occurs automatically or spontaneously during exposure to various kinds of stimuli, such as visual objects (Colreavy & Lewandowsky, 2008; Pothos, Edwards & Perlman, 2011) and vocal sounds (Marcus et al., 1999). Moreover, humans can learn language and read the facial expression and the mind of others without formal training (Davidoff, 2001; Baron-Cohen, et al., 2001; Call & Tomasello, 2008) although there seems to lack of sufficient information to construct rules (Pinker, 2002). The facts that infants selectively attend to language-like and face-like patterns (Saffran, Aslin & Newport, 1996; Simion & Giorgio, 2015) indicate that humans have innate ability of learning patterns selectively. It is suggested that the

ability of learning patterns, such as colors, faces and music, as it were understanding of the world, is concerned with language ability (Kay & Kempton, 1984; Davidoff, 2001; Koelsch et al., 2002) and that language ability itself is innate subserved by the Language acquisition device (Chomsky, 1965) that fills the explanatory gap of poverty of the stimulus (Chomsky, 1980), although there is controversy over this issue (Pullum & Scholz, 2002). Interestingly, humans are able to master rules of language only with positive evidence, which consists of correct rules but not wrong rules, and to judge whether a given sentence abides by the rules. Accordingly, it is hypothesized that humans can also construct categorical knowledge of visual patterns with positive evidence alone.

Study of learning knowledge including visual category learning always faces difficulty of controlling prior knowledge and learning knowledge. It is reported that visual short-term memory (VSTM) is sparse and abstract and its strength is affected by familiarity, attention and context (Rensink, O'Regan & Clark, 1996). Thus familiarity is controlled using, for instance, pre-examined data of familiarity (such as Snodgrass & Vanderwart, 1980) and adopting artificial thus nonsensical stimuli (Hoshino & Mogi, 2010)). The artificial grammar learning (AGL) is a useful paradigm to tackle this difficulty and to investigate rule learning as well, typically in language acquisition. The artificial grammar creates rules based on a finite state grammar which has some states or nodes connected with arrows that indicate transition from one state to another. Each arrow represents a letter or a word so that finite sequences of transition from start state to end state are produced. Resulting sequences, words or sentences, can be quantified in terms of occurrence of frequency known as n-gram probabilities (also known as transitional probabilities), the Levenshtein distance, and so on (Gomez & Gerken, 1999;

Levenshtein, 1966) because they are composed of distinct elements and definable in terms of transitions. In addition, because it is artificially created, subjects do not experienced it before and therefore it allows us to study how they can achieve word segmentation and find word order and word abstract pattern (Saffran, Aslin & Newport, 1996, Gomez & Gerken, 2000).

The current study aimed to investigate whether humans can learn rules of two-dimensional abstract patterns (exemplars) consisting of shapes, which is implicitly learned without explicit instructions, and, if so, how they use the acquired knowledge to judge new patterns (probes) in relation to their finite experience of the exemplars. I set an experimental procedure in which subjects were familiarized with patterns through within-category discrimination, that is unsupervised category learning using positive evidence, and later they were asked to judge whether patterns were members of the familiarized category, that is judgment based on categorical knowledge. The current experiment consisted of two phases, a learning (familiarizing) phase and a judgment phase. The learning phase required subjects to perform implicit learning by familiarizing with visual patterns within a context of a working memory task. Given that Atkinson and Shiffrin's model (Atkinson & Shiffrin, 1968) is taken into account, visual short-term memory, as primary input for familiarizing long-term memory, must be carefully controlled. To focus on the nature of knowledge of visual arrangement, the current study use artificial arrangements of common shapes, so that no single shape would capture subjects' particular attention and shapes are readily encoded, but arrangements needs familiarization where implicit learning would be hypothesized. The paradigm used in the current experiment was mimicry of Reber's AGL (Reber, 1989). The current experiment, however, is different from Reber's in following respects. 1) I

investigated implicit learning while subjects performed discrimination whereas Reber investigated implicit learning while subjects performed reproduction. Reber introduced this method in the visual domain presenting generated letters to subjects as mimicry of written language. This approach has been extended in visual shapes (Stobbe, Westphal-Fitch, Aust & Fitch, 2012, Westphal-Fitch, Huber, Gomez & Fitch, 2012). 2) To focus on features of arrangement itself rather than features associated with positional information I introduced stimuli that have more ambiguous edges by tiling up. In the judgment phase, subjects were required to judge if rules of a pattern matched with the rules of familiarized patterns. What, if any, information could explain judgment of choice? I analyzed relations among patterns using mathematically definable similarities, n-gram probabilities and the Levenshtein distance. In the context of visual learning, explicit conscious report may not reflect legitimate representation of stored information about a visual stimulus seen on earlier occasion (Johansson, Sikström, & Olsson, 2005). Even after familiarization subjects fail to tell exact rules behind it explicitly but they certainly acquire categorical knowledge implicitly during identification (Reber, 1967). Therefore I investigated objective relations between the exemplars and the probes rather than subjective similarity. Successful construction of categorical knowledge will give rise to that subjects' judgment marks out different similarities between probes and knowledge regarding exemplars.

4.2 Methods

Participants: Seventeen subjects (10 females and 7 males age 18-34, with an average of 22.1 and standard deviation of 5.8) participated in this experiment. The number of subjects seems adequate to test the current hypothesis, in reference to artificial grammar

studies that revealed human abilities of rule learning (Marcus et al., 1999; Turk-Browne et al., 2009; Reber, 1967; Gomez & Gerken, 1999). All subjects had normal or corrected-to-normal vision. They were remunerated for participating. They gave written informed consent after being explained about the purpose and nature of the experiments. The experimental protocol was approved by the Brain and Cognitive Sciences Ethics Committee of Sony Computer Science Laboratories. The stimuli were presented on a computer screen. The subjects responded by key pressing.

Stimuli and apparatus: A finite state grammar with five letters, which was described in the previous study (Reber, 1989), was used to generate ruled strings in this experiment. The letters TXVPS in the original study were substituted by shapes, i.e. T to a square, X to a plus, V to a star, P to a circle, S to a triangle as illustrated in Figure 4. Each shape was drawn with white antialiasing lines, within approximately 32 x 32 pixels at center of a 60 x 60 pixels black background patch. Those strings of shapes were diagonally expanded to make units of tiles so that the resulting patterns were symmetric with respect to the pi/4 and 3pi/4 lines. These units were recursively tiled to cover the computer display with a resolution of 1680 x 1050 (Figure 4) to eliminate information regarding apparent tile edges or element positions. Note that, in this protocol, different size of seed sequence of shapes never met the same pattern. The current stimulus design enabled us to examine human understanding of relationship among visual elements.

The finite state grammar generated total of 43 possible strings with lengths up to eight, with corresponding visual patterns. For each subject, 25 patterns were randomly chosen to represent all paths through the grammar for the learning phase while the remaining 18 were reserved for the judgment phase. As a control to the finite state grammar, 43

strings, were randomly generated for teach subject using the same shapes, matching the grammar strings in length. They were converted to visual patterns in the same procedure as the grammar generated patterns. I refer the learning patterns and the judgment patterns, which were presented in the learning phase and the judgment phase, as exemplars and probes, respectively, in a context of categorical judgment.

The patterns were presented on a computer display, which was placed at a distance of approximately 60 cm from the subjects and the visual angle of an element was approximately 1˚. Responses were recorded through key press.

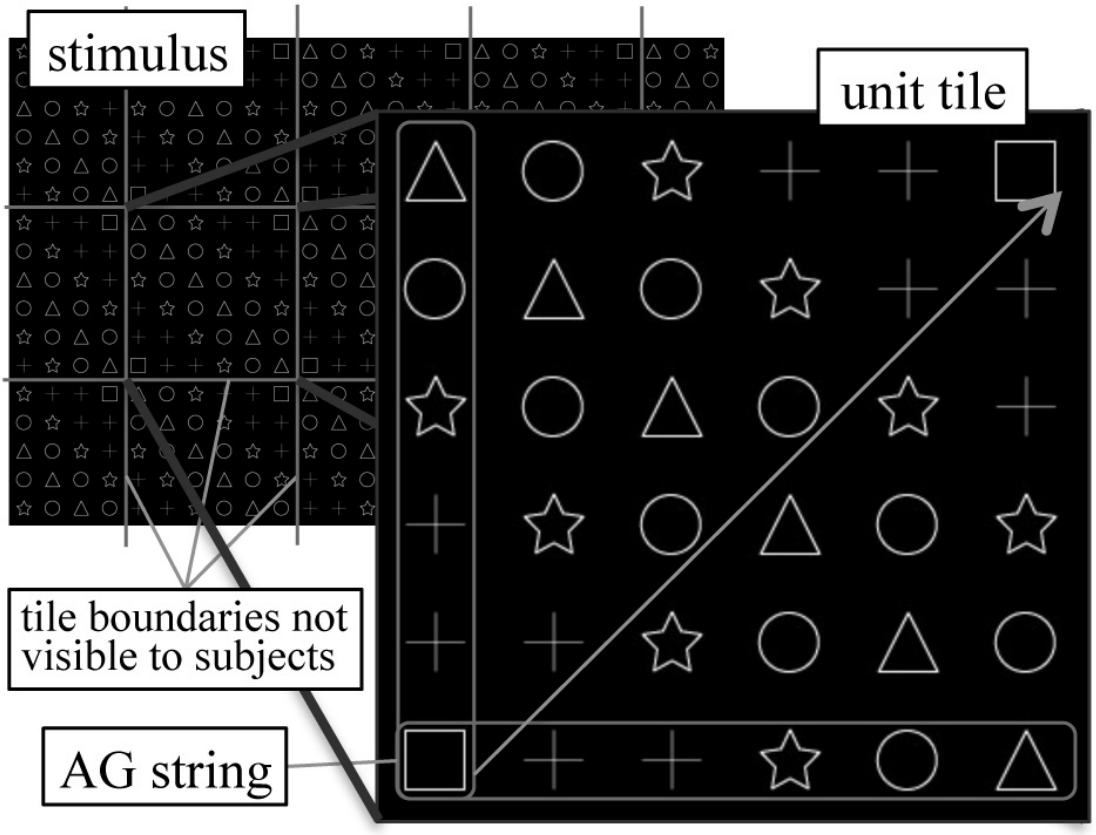

Figure 4. Pattern generation and an example stimulus.

Procedure: The experiment consisted of a learning phase followed by a judgment phase. For the learning phase, 25 learning patterns were randomly divided into 5 sets, each containing 5 patterns for each subject. The number of patterns was set to be 5, based on a pilot study which indicated that most subjects completed the learning phase in the first two consecutive sessions with four patterns, but did not with five patterns, possibly due to limited capacity of visual short-term memory storage depending on stimulus complexity (Alvarez & Cavanagh, 2004).

Much of the procedure in the learning phase was adapted from Reber's work on the artificial grammar (Reber, 1967), which required the subjects to reproduce nonsensical word, with two consecutive correct reproductions required to proceed to the next set. In current experiment, after the presentation of each pattern in a set, the subjects were instructed to answer the order of the presentation, instead of drawing up the patterns. This procedure was designed to control familiarity of patterns at the same level (Reber, 1967), as well as to keep the subjects' attention to the patterns (Conway & Christiansen, 2009). The control of familiarities was particularly important for exemplar-based analysis, assuming familiarities of exemplars were the same.

In the learning phase, the subjects viewed an instruction message saying, "remember the order of presentation of patterns", on the computer display. After pressing a key to proceed, they viewed a blank for 100 ms, a number indicating the order of the presentation for 1 s, and 1 out of 5 patterns for 5 s. This procedure was repeated 5 times without breaks to complete the 5 patterns of a set. Then, the subjects viewed a message saying, "answer the order of presentation by a key press". After pressing a key to proceed, they viewed one of the 5 patterns. The subjects were instructed to answer the

order of the presentation by pressing a number key from 1 to 5. After answering, the subjects viewed the next pattern following a 100 ms blank, until the completion of 5 patterns. These procedures constituted a single trial. No feedback of correct/wrong was given. Trials with the same stimulus set with shuffled orders for presentations and questions were repeated until the criterion of two consecutive correct answers for all 5 orders was reached. When one set was completed, a new set of 5 patterns was learned until all the 5 sets were finished. Except for during 5 consecutive pattern presentations and during answering the order, the experiment was designed not to proceed without a key press so that subjects could have a rest and proceeded the task in their own pace in order to keep their concentration. This information was given to the subjects.

For the judgment phase, a set consisting of 79 patterns was prepared, which included 18 the grammar generated patterns (not used in the learning phase) twice each, and 43 control patterns. The 79 patterns with shuffle orders were presented one by one until the subjects responded. The subject's task was to answer whether the rules of a pattern presented was "same" or "different" compared with the rules for previously learned 25 patterns in the learning phase, in a two-alternative forced-choice procedure. The subjects were specifically instructed as follows: "The 25 patterns you have seen were based on a rule. From now on, patterns will appear on the screen one by one. Please answer by pressing a key whether the pattern is based on the same rule or a different one." No explicit remark about construction of the rules was given, in order not to interfere with the subjects' own conception about the nature of patterns. The subjects were instructed to place their index fingers on the "f" and "j" key as home positions. Half of the subjects were instructed to press the "f" key if they felt a pattern presented was "same" and "j" key if "different", while the other half was instructed vice versa in a

counterbalance.

The subjects were initially informed only about the learning phase and not about the judgment phase, to avoid explicit categorization or rule searching when they tackled the learning phase. After finishing the learning phase, they were given instruction about the judgment phase. After completing computer-based tasks, they answered a written questionnaire about the experiment.

Analysis:

In the categorical judgment of visual arrangements, similarities among exemplars and probes have been a particularly interesting subject for research (Hahn et al., 2010). To analyze similarities among patterns, I introduced measures of dissimilarities concerning relations of elements, namely the Levenshtein distance (here after LD) and n-gram probabilities (also known as trasitional probabilities). LD is defined as minimum number of operations, namely deletions, insertions and substitutions required to convert one sequence into the other (Levenshtein, 1966). I applied LD to analyze relations of elements in seed sequences of the patterns, which were in the bottommost row and the leftmost column of tiles. In addition, n-gram probabilities were used to measure the two-dimensional relations of elements in tiles. An n-gram probability represents a probability of occurrence of an item conditioned on its n-1 contiguous items (i.e. $P(xi \mid x_{i-(n-1)}, \ldots, x_{i-1})$) (Gomez & Gerken, 1999). I defined an n-gram dissimilarity of pattern A compared with pattern B as follows. First, I picked up any n-grams from the unit tile of pattern A. Next, I calculated n-gram probabilities for each of theses n-grams in the unit tile of pattern B. Finally, I obtained an n-gram dissimilarity as 1 minus the mean of the n-gram probabilities (i.e. $1 - \Sigma\, P_{\mathrm{A}}(x_{i,xi-1,\ldots,x_{i-(n-1)}}) \times P_{\mathrm{B}}(x_i \mid x_{i-(n-1)}, \ldots, x_{i-1})$ ($x \in$ A, B). I

thus calculated dissimilarities for 1-, 2- and 3-gram. A 1-gram probability had no conditional probability (i.e. $P(x_i)$) and was equal to the mean frequency of five elements. In 2- and 3- grams, every contiguous sequence was taken from a unit tile to calculate possible combination of n-grams. Note that the resulting n-grams are the same regardless of whether the contiguous sequence was taken horizontally or vertically, because of the symmetry of patterns along the pi/4 and 3pi/4 lines. If an interest element is near an edge of a unit tile and its conditional elements are outside of the tile, the conditional probability was defined with outside elements as the neighbor tiles. The 3-gram was derived from averaging two ways of calculations depending on how conditional elements were assigned, i.e., conditional-conditional-target and conditional-target-conditional. In the current experimental setting with limited number of stimuli, the mean 3-gram of low ranks for particular subjects reached probability zero, which means any of 3 contiguous sequences in their probes did not appear in their low-ranked exemplars, leading to a floor effect and I did not perform statistical test on them. Accordingly, dissimilarities for n-grams with $n > 3$ were not calculated because most of n-grams in a pattern of more than 3 sequences are not found in another pattern when $n > 3$.

Dissimilarities measure context-dependency or levels of relations among elements (Rentschler et al., 1994). 1-gram is a context-independent measure within patterns and thus is an element-based processing of the patterns, defined as the frequency of each element regardless of its spatial configuration. A larger size of n-ram implicates more context-dependency. Thus, 2- and 3-grams are based on configural processing, reflecting spatial configural relations among multiple elements within patterns. LD is also considered to be configural processing, because editing an element requires

positional information that is defined relative to non-target elements. For example, consider a case in which the string PPXS is converted to the string TXS. To calculate LD, one would first need to compare these two strings, and arrive at a single sequence, allowing for the possibilities of insertions or deletions. In this case, the last two letters XS are the same. Next, one needs to know where different letters are located relative to XS. In this case, it is on the left of X. After replacing P with T or deleting P, one would still need to handle one more P located the leftmost, to be deleted or replaced with T. These manipulations would necessarily involve relations among multiple elements.

To investigate what aspects of relations between exemplars (patterns in the learning patterns) and probes (patterns in the judgment patterns) affect subjects' judgment, I analyzed several kinds of dissimilarities (i.e. LD and n-grams). The analysis was based on objective dissimilarities rather than predefined rules (grammar), in order to focus on subjective experience, while the rules were not the only solution for this unsupervised category learning (Gigerenzer & Goldstein, 1996). I calculated three alternatives in each type of dissimilarity. 1) As exemplar-based analysis, dissimilarities for each probe compared with each of 25 exemplars were calculated. Those dissimilarities were sorted in an ascending order for each probe so that 25 exemplars were ranked in order of similarity. For instance, a rank 1 exemplar of LD was the most similar to, or the least distant from a particular probe in measure of LD. By ranking, it was possible to examine which ranks of exemplar were informative. 2) The mean dissimilarities of all exemplars were calculated, which were equal to the average of dissimilarities for all ranks. The mean dissimilarities are thought to represent collective information or some sort of abstract information of the exemplars. The calculations were irrespective of the interaction between ranks and judgments, conveying alternative information about

knowledge regarding the average distance strategy (Reed, 1972) (See Discussion). 3) Dissimilarities for each probe compared with the most prototypical exemplar were calculated. The most prototypical exemplar had the least dissimilarity among exemplars, which was considered to share the most attributes and the most typical member of exemplars (Rosch, 1975). I call it dissimilarity for the prototypical. There is possibility that each of these three alternative calculations contributes to convey information about dissimilarity. To reduce type I errors, I took a conservative approach in which 25 ranks, the mean and the prototypical were corrected for multiple comparisons all together (i.e. n = 27).

Although the levels of dissimilarities grasp different aspects of patterns, they would share common features as long as they measure dissimilarities. The mean dissimilarities were all significantly correlated with each other (Figure 5). In addition, I separately performed statistical tests on dissimilarities to capture various aspects of the stimuli in each level of relation among elements. Because the weights of dissimilarities for judgment were not known, I conducted the analysis focused on each dissimilarity rather than one between dissimilarities as multiple regressions.

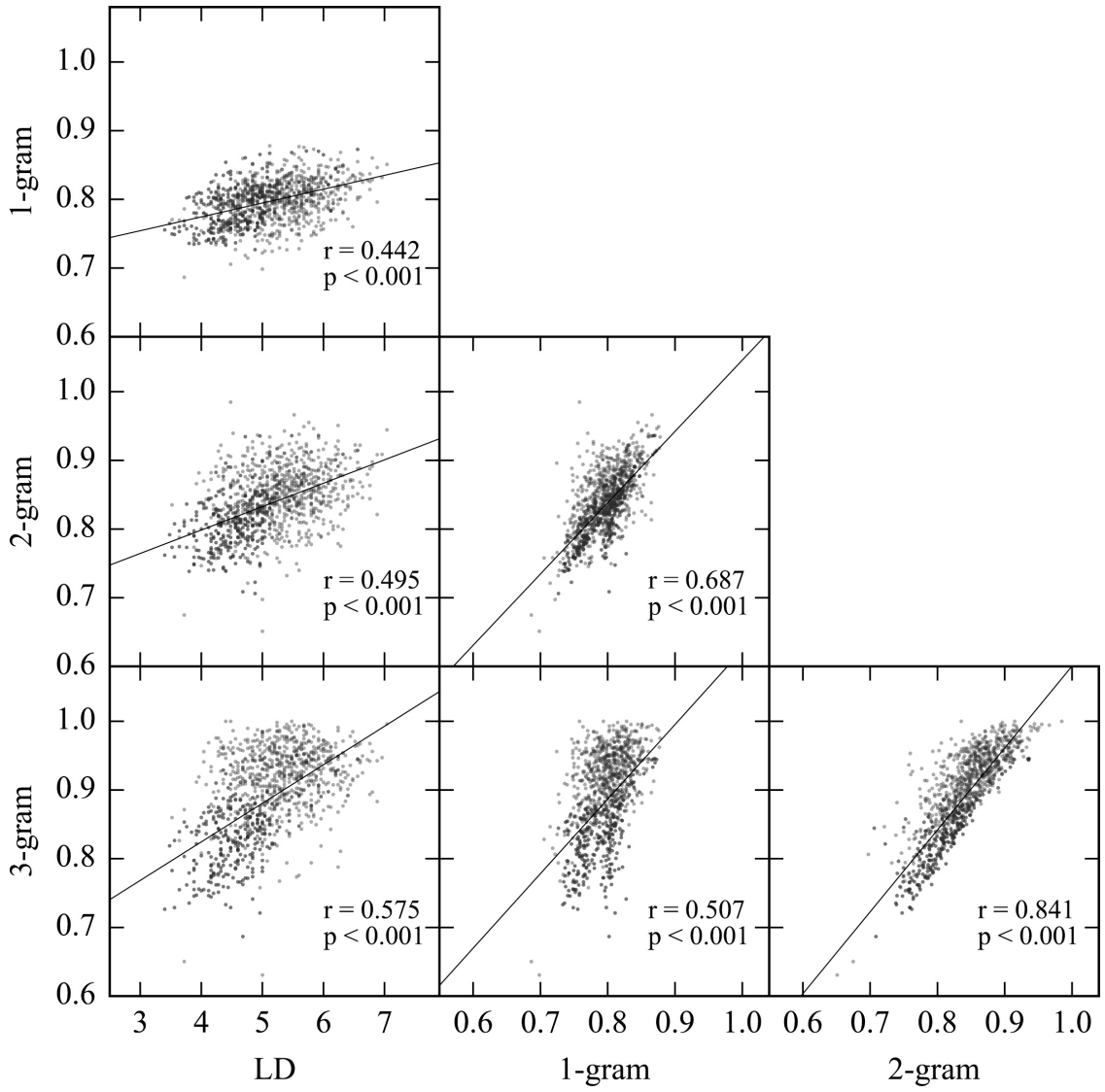

Figure 5. Pearson's correlation coefficients among four types of the mean dissimilarities.

4.3 Results

One subject was excluded from further participation in the experiment for not finishing the learning phase within one and half an hour. Sixteen subjects completed the tasks. They spent forty minutes on average and no more than one hour for the learning phase including intervening breaks. None of the subjects reported that they remembered any single patterns or unit tiles precisely. They were not aware of unit tiles and the purpose

of the judgment phase during the learning phase.

To investigate whether the subjects learned the learning patterns more effectively as sets progressed I first calculated the learning effect of the learning phase. A Page's L test (Page, 1963) revealed that the mean number of errors in a set had a statistically significant descending trend in proportion to number of sets ($p = .000307 < .05$) (Figure 6). (Alternatively, a repeated measures ANOVA determined that the mean errors in a set differed statistically significantly between sets ($F(4, 60) = 4.456$, $p < .01$). Number of trials taken to reach the criterion ranged from 2 to 19 with 6.35 on average. Post hoc tests using the Bonferroni correction revealed that number of sets elicited a slight reduction in the mean errors from Set 1 (1.50 ± .58) to Set 3 (.93 ± .61, p =.041) and 5 (.69 ± .69, p =.052). These effects were possibly due to getting used to the task and finding strategic ways, learning of rules, or mixture of them. This result is consistent with Reber's experiment in which he visually presented letters of AGL (Reber, 1967).

The subjects were marginally able to discriminate the grammar generated and control patterns in the judgment phase although they answered almost equal number of "same" and "different" judgments (mean ± std = same: 38.8 ± 11.3, different: 40.2 ± 11.3). I calculated the sensitivity index d-prime of the signal detection theory with a "hit" defined as a "same" judgment on the grammar generated patterns and a false alarm as a "same" judgment on control patterns. A one-sample t-test across the subjects revealed that the mean d-prime between the grammar generated and control judgment was significantly above zero (Mean = .45, $T(15) = 4.84$, $p = .0002 < .001$, Cohen's $d = 1.21$). The results indicate that the subjects successfully learned aspects of the rules under the two-dimensional patterns implicitly.

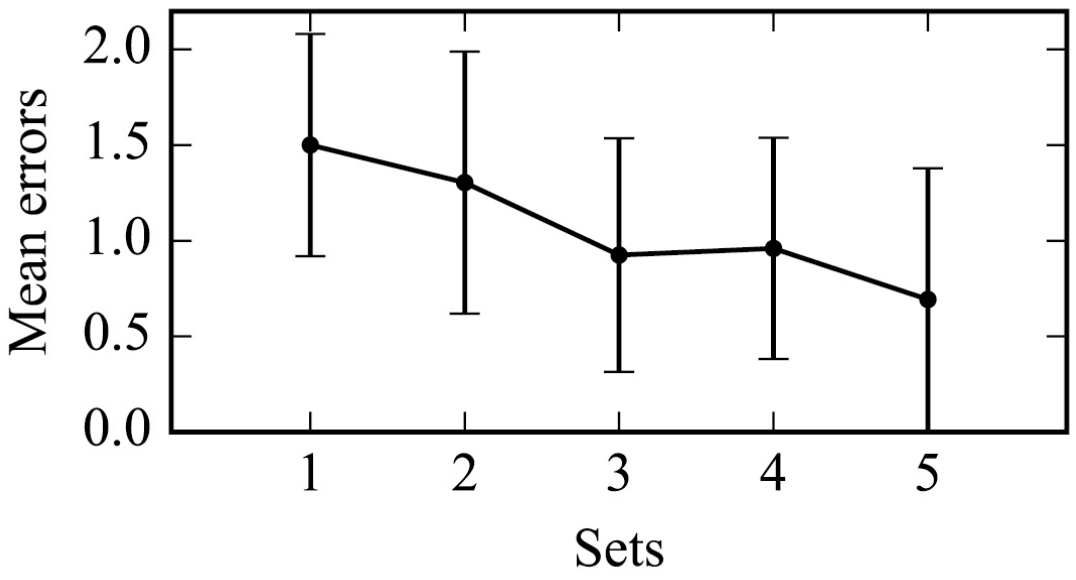

Figure 6. Mean number of errors in the study phase. The error bars indicate standard deviations.

Next, I looked at nature of the acquired knowledge in regard to similarity between exemplars and probes (Figure 7). The response times (RTs) of judgment varied between the subjects, from 1.19 to 5.97 seconds on median. From the RTs and observation by the experimenter, no obvious outliers, such as inadvertent button press or subjects' inattention, were found (minimum .5 to maximum 20 seconds) and all judgments were included to the following analysis.

In order to investigate what information of the exemplars reflects subjects' judgment, I compared dissimilarities between each probe (judgment pattern) and each learning pattern (see Analysis).

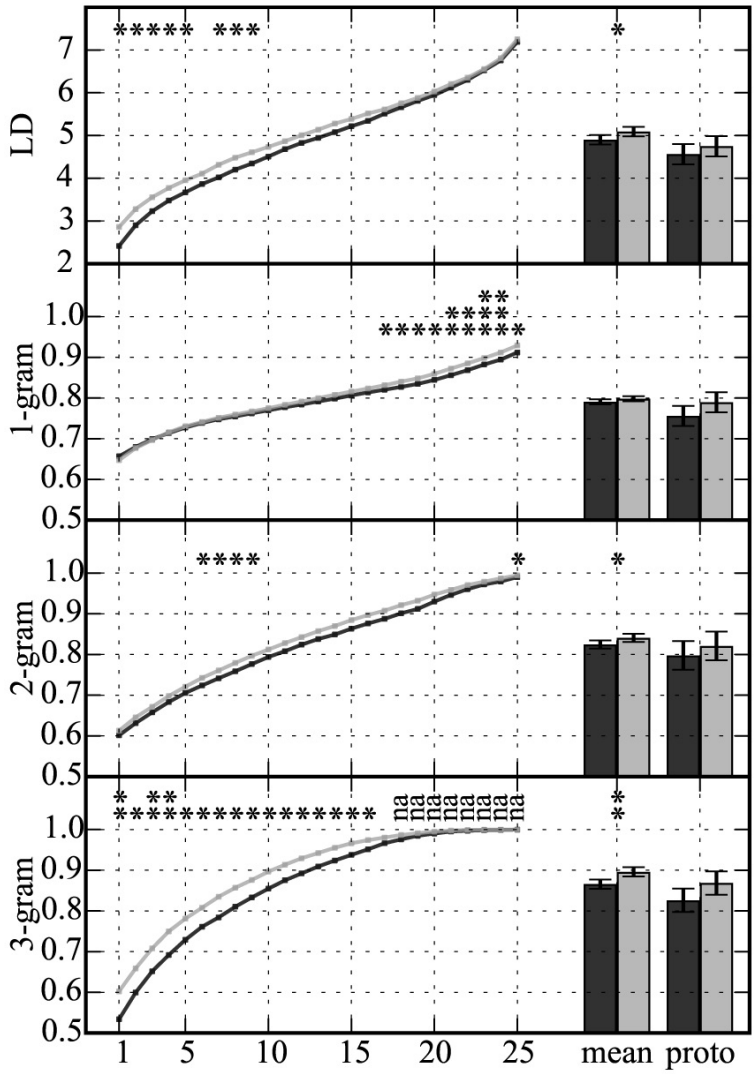

Figure 7. Difference of dissimilarities between judgments: The x-axis indicates ranks, the mean of 25 ranks and the prototypical. The y-axis indicates the mean LD and n-gram dissimilarities along the subjects. The black and gray line/bars are for the "same" and "different" judgments, respectively. The asterisks indicate statistically significant differences between judgments in multiple paired t-tests with Bonferroni correction (where * $p < .05$, ** $p < .01$, *** $p < .001$). The error bars indicate the standard deviations. No statistical tests were performed on ranks labeled "na" due to the ceiling effect.

LDs for each rank, the mean of 25 ranks and the prototypical are shown in Figure 7. A two-way repeated measures ANOVA revealed that the interaction between ranks and judgments was significant ($F(24, 360) = 9.457$, $p < .0001$). Multiple paired t-tests between judgments with Bonferroni correction of n = 27 showed that LDs for rank 1, 2, 3, 4, 5, 7, 8, 9 and the mean of 25 ranks had significant difference (all tests showed "same" < "different" and $p < .05$) whereas LDs for lower ranks and the prototypical did not. This indicates that judgment of a probe was based on exemplars which were highly similar, in terms of LD, to the probes but not on the prototypical within the exemplars.

In three of two-way repeated measures ANOVA for 1-, 2- and 3-gram dissimilarities showed that all the interactions between ranks and judgments were significant ($F(24, 360) = 8.943, 1.627, 15.26$; $p < .0001$, $p = .03 < .05$, $p < .0001$, respectively). Multiple paired t-tests in each dissimilarity between judgments with Bonferroni correction of n = 27 were performed. 3-gram showed a similar tendency to LD. Both 3-gram and LD had significant difference in dissimilarities for higher ranks and the mean (all tests showed "same" < "different" and $p < .05$, Cohen's $d > 0.8$). I also performed the Kolmogorov-Smirnov tests for normality for each distribution and found no violation of normality. Additionally, in the 3-gram there were significant differences in dissimilarities for middle ranks higher than rank 17. I did not perform t-tests on rank 18 to 25 because of the ceiling effects, in which some subjects had dissimilarity = 1.0, meaning no 3 contiguous sequence was found in given learning patterns and the 3-gram probability resulted in zero. Therefore details in lower ranks are not shown. In contrast, in 1-grams dissimilarities for lower ranks from 17 to 25 had significant differences (all tests showed "same" < "different", p values as shown in the legend of Figure 7.), while those for any higher ranks than rank 17, the mean and the prototypical did not, due to

different effect sizes. The results for 2-gram showed that dissimilarities for rank 6 to 9, 25, the mean had significant differences (all tests showed “same” < “different”, $p < .05$.)

It was possible that significant differences and effect sizes across levels were not due to the nature of the stimuli used here, such that lower ranked exemplars in 1-gram would appear as higher ranked exemplars in 3-gram. To verify this point, I further calculated correlation coefficients of rank numbers, which each exemplar scored, among dissimilarities. All pairs of dissimilarities showed significant positive correlations of ranks (Figure 8), indicating that different tendencies of ranks among dissimilarities arose from the nature of subjects' judgments (Inselberg & Dimsdale, 1990).

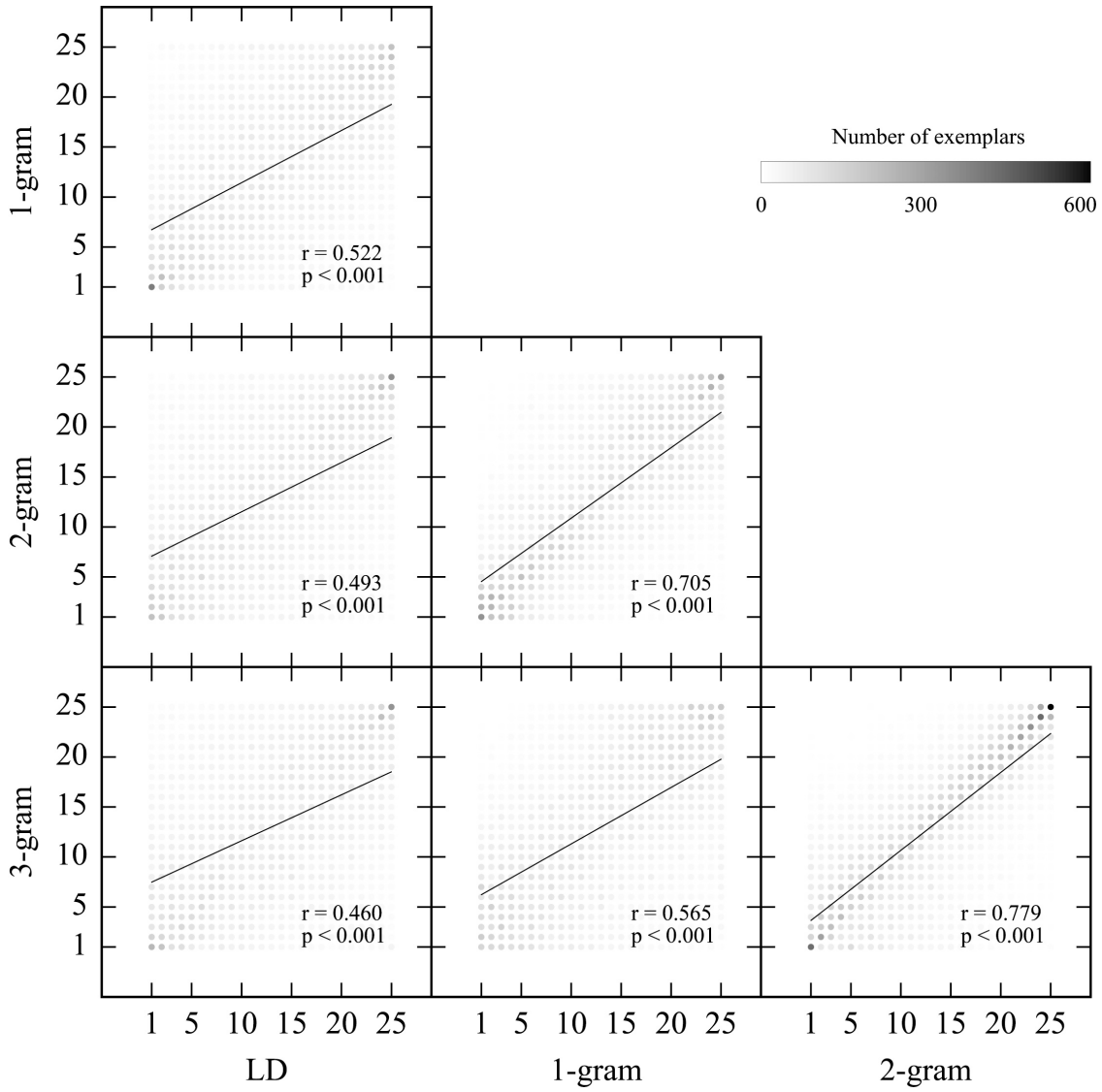

Figure 8. Spearman's rank correlation coefficients among ranks in four types of dissimilarities. A dot color indicates a number of exemplars. The x and y axes are ranks of each dissimilarity.

4.4 Discussion

In Study 2, I used a visual version of AGL paradigm (Reber, 1967) to investigate whether human subjects can extract statistical regularity of two-dimensional patterns.

AGL approach to implicit rule extraction has been validated in previous studies, which showed that statistical learning occurs in familiarization (Saffran, Aslin & Newport, 1996; Marcus et al., 1999; Gomez & Gerken, 1999), including one-dimensional visual patterns (Reber, 1967; Stobbe, Westphal-Fitch, Aust & Fitch, 2012; Conway & Christiansen, 2009), visual temporal orders (Fiser & Aslin, 2002; Kirkham, Slemmer & Johnson, 2002), visual spatial configurations (Chun & Jiang, 1999; Fiser & Aslin, 2001, 2005). Neural correlates of visual statistical learning indicate that statistical learning occurs implicitly with little exposure to stimuli, independent of subsequent explicit familiarity (Turk-Browne, Scholl, Chun, & Johnson, 2009). These studies would suggest that humans could extract statistical regularity under two-dimensional patterns. Consistent with the previous studies (Saffran, Aslin & Newport, 1996; Saffran et al., 1999; Marcus et al, 1999; Fiser & Aslin, 2001, 2002, 2005; Kirkham, Slemmer & Johnson, 2002; Sttobe et al., 2012; Turk-Browne et al, 2009; Reber, 1967; Gomez & Gerken, 1999; Conway & Christiansen, 2009), the current study showed rule extraction over two-dimensional patterns in human subjects as demonstrated by the result of significant learning effects and the signal detection sensitivity.

Regarding the AGL of the visual-spatial format, the current experiment is an extension of a previous study, which showed that humans could learn rules from horizontally displayed visual sequences that were generated by an artificial grammar (Conway & Christiansen, 2009). The Conway and Christiansen study used horizontally displayed one-dimensional sequences and found that the rule learning was affected by elements at the left end of the sequences. This effect was excluded in the current experiment by using tiled patterns in which only the element spatial relations were relevant. Thus, it is suggested that humans are able to learn rules with element spatial

relations without element-position relations, at least in the case of two-dimensional arrangement.

The marginal value of the sensitivity index between the grammar generated and control patterns observed in the experiment indicates that there is no fine-grained categorization according to predefined rules: Categorization might proceed by subjects' individual definition based on their own experience (Gigerenzer & Goldstein, 1996). The predefined rules, however, are not derivable precisely from the limited number of exemplars, as argument concerning poverty of stimulus often suggest in formal language theory (Chomsky, 1965, 1980). In the current experiment, the learning was unsupervised, where the rule extraction occurred in implicit learning in a task in which discrimination recognition and working memory are required. No instruction about rule extraction was given. Taken together with previous findings (Fiser & Aslin, 2001, 2005), it is suggested humans are sensitive not only to isolated sequences of elements embedded in scenes, but also novel scenes that consist of such sequences. In addition to the fact that unsupervised category learning occurs automatically or spontaneously during exposure to visual objects (Colreavy & Lewandowsky, 2008; Pothos, Edwards & Perlman, 2011), the current experiment demonstrated that unsupervised category learning occurs during discrimination of two-dimensonal visual arrangement. Unsupervised category learning typically involves ill-posed problems and demands conjecture or instinct to learn meaningful categorical knowledge (Hume, 1739). It has been suggested that instinctive learning or reasoning has validity (Gigerenzer & Goldsetin, 1996). The nature of subjects' judgments observed in the current study is in line with its validity, shedding light on the nature of human perception of the visual arrangement.

Accordingly, I investigated how humans make judgments based on previous knowledge of similar patterns, using dissimilarity measures. Importantly, the result that the subjects' judgment distinguished dissimilarities (Figure 7) indicates that subjects successfully made categorical judgment based on dissimilarities of familiarized knowledge and that categorical knowledge is acquired through positive evidence alone. The detailed analysis of the data showed that different levels and ranks of dissimilarities differently exerted subjects' judgment. LD and 3-gram showed significantly different dissimilarities in high ranks according to judgment. In relatively high context-dependent measures such as LD and 3-gram, dissimilarities of more similar rank exemplars and the mean dissimilarities were effectively reflected on judgment. On the other hand, in context-independent measures, such as 1-gram, dissimilarities of less similar rank exemplars were reflected on judgment. These results were consistent with a previous study in which single elements and sequence of elements were extracted from scenes (Fiser & Aslin, 2001). The result across levels of processing indicates that element-based and configural processing coexist, which is consistent with the notion that local and global processing are separable (Tanaka & Farah, 1993, Rentschler et al., 1994). The former possibly recruits the fusiform area (Kanwisher, McDermott & Chun, 1997, Gauthier, Skudlarski, Gore & Anderson, 2000). I propose that processing of visual arrangement can be conducted implicitily, with gradual arrangements regarding how many elements are taken into account at once.

Effects of dissimilarity on judgment were larger in 3-gram than 2-gram. This discrepancy was possibly due to the fact that knowledge of rules was more heavily represented in 3-gram than 2-gram, while the memory of embedded sequences in larger spatial configurations was inhibited (Fiser & Aslin, 2005). There were few common

n-grams for n>=4 between patterns in the current experiment. Thus, 3-gram was the largest informative sequence in this context, whereas 2-gram was less informative. 2-gram not embedded in 3-gram might have contributed to the significant difference observed in the result of 2-gram. Fiser & Aslin asked their subjects to judge familiarity of a single sequence of elements embedded in scenes (Fiser & Aslin, 2001, 2005). The subjects were able to remember element sequences with perfect conditional probability $p = 1.0$, but not with non-perfect conditional probability $p = 0.5$ or 0.66, when those two types of sequences were presented equal times. Results in this study indicate that humans are sensitive to various conditional probabilities between elements of spatial sequence.

Reed's categorization strategies (Reed, 1972) explain the characteristics of element-based and configural processing. He documented four important strategies of subjective categorization, namely prototype, proximity algorithm, cue validity and average distance. The average distance strategy entails judgment based on the mean distances between a probe and all exemplars. In his study, the average distance strategy, as well as the prototype strategy, explained subjects' behavior in categorizing multidimensional faces. The mean dissimilarity in the current analysis is equivalent to the average distance strategy. The proximity algorithm is a sort of the exemplar theory. It predicts that judgment is based on the most similar exemplar to a probe, which is equivalent to the k-nearest neighbor (k-NN) method. The K-NN method is one of useful computational model of pattern recognition (Altman, 1992), where k in k-NN represents the number exemplars taken into account for a given classification. In the current analysis, the shared tendency between LD and 3-gram helps to explain the characteristics of highly context-dependent or informative configural processing. The

judgment may primarily based on the proximity or the k-NN algorithm strategy in configural processing. On the other hand, the tendency in element-based processing is likened to distal algorithm. Therefore it is possible that judgment is based on elimination of highly dissimilar exemplars regarding element-based processing. These possibilities can be proposed on the premise that each dissimilarity analysis is separately discussed. Exemplars in high ranks of LD and 3-gram would be different from those in low ranks of 1-gram. Nevertheless, all the types of dissimilarities positively correlated (Figure 8). Further studies will be necessary to elucidate issues concerning which strategy most contributes to judgment. In contrast to many studies in statistical learning which have focused on temporal frequencies, the current study investigated spatial frequency of element sequences within patterns. As a result, I were able to analyze categorical judgment based on relations between probes and exemplars, keeping knowledge of individual exemplars (Barsalou, Huttenlocher &Lamberts, 1998). The analysis was extended to comparison and accumulation of exemplars, reflected in ranking and mean dissimilarity, respectively.

The subjects repeatedly learned each example through within-category discrimination, until they reached a certain learning criteria. Accordingly, I could assume that the subjects were familiarized to the exemplars equally. The results suggest no prototypical representation was constructed, where the prototypical exemplars were not marked out at any levels of dissimilarities depending on judgment. It is possible that the prototypical exemplar, which is the least dissimilar exemplar among 25 exemplars, does not represent the actual prototype (Rosch, 1975). The prototype approach assumes that generalized knowledge is formed in category learning, whereas exemplar approach requires memory of individual exemplars. Both approaches have advantages depending

on the nature of task (Ross & Makin, 2000; Smith, 2014). Briscoe & Feldman showed that humans perform a middle point of both extreme approaches in a supervised category learning with multiple feature dimensions (Briscoe & Feldman, 2011). They claimed that prototype and exemplar models are in trade-off relationship, and are too biased to fit complex and too variance to fit any predefined rules, respectively. The current result is in favor of exemplar-based representation as shown in several ranks of dissimilarities (Figure 7). These findings are consistent with natural language categories (Storms et al., 2000) and evidence from neural data (Mack, Preston & Alison, 2013).

When contrasted with the results for the prototypical in the analysis, it is possible that the mean dissimilarity (average distance) contributes to a more abstract category representation as collective information of exemplars (Reed, 1972). A study investigated neural correlates with a visual identification task demonstrated that abstract category is represented in the left occipital cortex and IT, while specific exemplar is represented in the right occipital cortex and IT (McMenamin et al., 2015). Likewise, Marsolek's previous study, in which stimuli were presented on left or right visual field, reported essentially the same result (Marsolek, 1999). More specifically, the core areas of abstract categorical representation and exemplar representation may be left and right fusiform gyri, respectively (Garoff et al., 2005). Garoff et al. showed that specific minus non-specific recognition and non-specific recognition minus forgetting are associated with activity in right and left fusiform gyri during encoding, respectively. In their study, subjects viewed and judged presented visual objects, choosing from three alternatives, "same", "similar" or "new", with respect to knowledge in a prior learning phase, which was conducted in a very similar manner to the current experiment. They designated a "same" response to a "same" object as specific recognition, a "same" to a "similar"

object or a "similar" response to a "same" object as non-specific recognition. The present results are consistent with the view that exemplars contributing to strong exemplar-based knowledge lead to specific recognition accompanied by the right fusiform activation while exemplars contributing to abstract knowledge such as average distance lead to non-specific recognition accompanied by the left fusiform activation. In addition, the characteristics of exemplars would be already determined by the fusiform cortices in the learning phase (Garoff et al., 2005).

Previous studies for visual category learning have dedicated much attention to analysis on multiple features of objects (Ashby & Maddox1992, Sigala & Logothetis, 2002, Briscoe & Feldman, 2011) but little on arrangement of visual elements. Instead, arrangement has been examined in the context of statistical learning, in which humans have excelled over animals (Stobbe et al., 2012, Westphal-Fitch et al., 2012). Studies of animals revealed that animals have ability to learn statistical regularity depending on number of elements and complexity of grammar determined by linearity (Scharff &Nottebohm, 1991) as a sophisticated illustration by Wilson et al. (Wilson et al., 2013, Figure 1 A) but prefer different processes from humans. They tend to process local configural relations (Cerella, 1980; Wilson, Smith & Petkov, 2015) and do not transfer knowledge in abstract level (Seki, Suzuki, Osawa & Okanoya, 2013).

On the other hand, humans understand global as well as local relations (Navon, 1977; Wilson, Smith & Petkov, 2015). Infants have ability to learn abstract sequences (Marcus et al., 1999; Gomez & Gerken, 1999) and adults can transfer abstract knowledge across modalities (Altmann, Dienes & Goode, 1995). Attention alters these statistical learning processes (Ravignani, Westphal-Fitch, Aust, Schlumpp & Fitch, 2015). Many studies of statistical learning use auditory stimuli or in fewer cases visual

and tactile (Conway & Christiansen, 2005), all which focused on temporal frequency. This temporal statistics may be processed in the hippocampus, an area thought to be associated with episodic memory (Turk-Browne et al., 2010). It is, however, not known whether spatial statistical arrangement is processed in a similar manner to temporal one, engaging brain areas that is associated with episodic memory (Turk-Browne et al, 2009). The global and local processing of visual input shows some similarity to temporal statistics regarding how animals tend to process (Westphal-Fitch et al., 2012; Wilson et al., 2015), and thus some shared mechanisms would be involved. One region possibly involved in spatial arrangement processing is the fusiform cortex, which shows sensitivity for feature statistics (Tyler et al., 2013; Wright et al., 2015). Extensive familiarization facilitates categorical selectivity in the fusiform. Not only faces but also objects of visual expertise activate the lateral side of the fusiform, also known as fusiform face area (Gauthier, Skudlarski, Gore & Anderson, 2000; Xu, 2005). Extensive training of tool-like novel objects elicits focal activation of the medial fusiform gyrus, a region known to be tool-selective (Weisberg, Turennout & Martin, 2007). These familiarization effects indicate that the fusiform may aggregate information of objects and categorize according to their statistics of features. On the other hand, the perirhinal cortex (PRC) in ATL plays a prominent role of discrimination between semantically similar objects (Wright, Randall, Clarke & Tyler, 2015) and between objects in the context with high degree of feature ambiguity (Saksida, Bussey, Buckmaster & Murray, 2007). Thus, it is possible that the fusiform cortex is involved during familiarization, such as the learning phase in the current study, whereas PRC is involved when decision is required, such as the judgment phase.

Finally, the results of the current study suggest the existence of element-based and

configural processing in visual arrangement in humans, which is consistent with a computational study (Lake, Salakhutdinov & Tnenbaum, 2015). The co-existence of them in element-based representation suggests that visual representation would be distributed along two axes, spatial relations within exemplars and multiple individual exemplars. The spatial axis is responsible for levels of processing, from the element-based to the configural of multiple elements within each exemplar. This axis reflects the online analysis of spatial and perceptual information. The axis of multiple individual exemplars is for categorical knowledge, and is subserved by a single exemplar to conjoint representation of multiple exemplars or prototypes. Knowledge of exemplars involves memory system, and serves as the basis for judgment of forthcoming events, possibly engaging the fusiform cortices. Although it is possible that there are other axes or measurements that capture better aspects of visual representation, this objective analysis sheds light on human judgment regarding exemplar-based knowledge. Specifically, the current analysis provides the evidence of both axes within a single experiment.

In conclusion, the current study provides several import theoretical implications about the nature of visual representation. Humans are able to learn rules of two-dimensional arrangement in statistical manner. The rules contain categorical knowledge that is dominated by exemplar-based representation, and is used in later judgment of new patterns. The exemplar-based representation possibly involves the fusiform cortex and embraces configural and element-based processing concurrently. The configural processing tends to process in k-NN algorithm whereas the element-based processing is useful in elimination approach. The ability of processing visual arrangement may responsible for human creativity in which infinite potentialities are promised with

arrangement.

Chapter 5. General discussion

This thesis investigated task-irrelevant processes of visual information. Specifically, it is assumed that the candidates for memory representations of scenes are view-specific, object-based and abstract representations or the combination of them. Compared to view-specific representation, object-based representation is gone through additional segmentation process from scenes and abstract representation is gone through additional abstraction process of scenes. The latter two additional processes were investigated in Study 1 and 2, respectively.

In Study 1, it is indicated that objects in scenes are likely to be encoded in view-specific representation rather than object-based representation, when the scenes are combined into comprehensive spatial information of an individual space. Memory representations of task-irrelevant objects are bound to memory representations of scenes. Furthermore, it is possible that scenes are also implicitly encoded in view-specific representations. Allocentric representation might not be formed (Diwadkar & McNamara, 1997; Ekstrom, 2014; Chua & Chun, 2003). It was yet to be investigated what kind of information this view-specific representation contains.

In Study 2, it is indicated that humans are able to learn rules of two-dimensional arrangement implicitly. The rules are spatial statistics of objects in scenes. Thus, I suggest that view-specific representation learned in task-irrelevant processes contains visual spatial statistics of objects. Furthermore, the results indicates that memory representations of scenes are view-specific representation based on each experienced scene but not abstract representation. Humans are able to discriminate scenes

immediately (Oliva, 2005). The cognitive mechanism related with this ability is therefore possible to discriminates scenes by comparing scenes with specific scenes which are previously experienced.

Altogether, this thesis demonstrates that available information in task-irrelevant processes is view-specific representation and it is possible that no additional process would be assumed. Because learning visual arrangement of objects seems automatic, it is considered that this automatic process may prevent memory representation from segmenting objects from scenes and, as a result, memory representation of task-irrelevant objects lacks three-dimensional structural information. This consideration is consistent with studies proposing that discrimination of scenes is faster than identification of objects (e.g. Oliva, 2005). On the other hand, it is contrary to the notion of a computational approach that visual scenes are processed from low level information, such as edges and surface, to high level information, such as objects and scenes.

This study has several limitations. First, task-relevant processes contained multiple processes, which were not broken down to a single process. It is needed to be broken down more precisely to determine which task-relevant process affects the results, especially in Study 1. Next, potential task-irrelevant processes are infinite. Therefore, other processes than the current investigation are assumed. Finally, subjects' prior knowledge affected the result. In consideration of these limitations, the future direction of this study would be to investigate task-irrelevant visual learning using more precise contrast of conditions in adults and to investigate task-irrelevant visual learning in infants to reveal effects of experiences.

In summary, this thesis propose that it is necessary to take into account the effect of

information regarding object arrangement in visual experiment, which is encoded in task-irrelevant processes. This notion is consistent with evidence from neural data (Janzen & Turennout, 2004; Meegan & Honsberger, 2005). Humans are masters of making sense of patterns even they are nonsense at first, as represented by learning ability of language. Considering that, I believe that certain biased ways, or natural constraints in task-irrelevant processes provide insight into human understanding of the world, visual and otherwise. Moreover, task-irrelevant processes of visual information may give rise to richness of subjective experience of visual events by adding associative information to primary meaning of stimuli and consequently bring about concepts of events with impression or feeling, such as emotion (Sharot, Delgado & Phelps, 2004) and aesthetic pleasure (Reber, Schwarz & Winkielman, 2004).

References.

Altman, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. Am. Stat. 46, 175–185 (1992).

Altmann, G. T. M., Dienes, Z. & Goode, A. Modality independence of implicitly learned grammatical knowledge. J. Exp. Psychol. Learn. Mem. Cogn. 21, 899–912 (1995).

Alvarez, G. A. & Cavanagh, P. The capacity of visual short-term memory is set both by visual information load and by number of objects. Psychol. Sci. 15, 106–11 (2004).

Ashby, F. G. & Maddox, W. T. Complex decision rules in categorization: Contrasting novice and experienced performance. J. Exp. Psychol. Hum. Percept. Perform. 18, 50–71 (1992).

Ashby, F. G. & Maddox, W. T. Human category learning. Annu. Rev. Psychol. 56, 149–178 (2005).

Atkinson, R. C. & Shiffrin, R. M. in Psychology of Learning and Motivation 2, 89–195 (1968).

Aust, U. & Huber, L. The role of item- and category-specific information in the discrimination of people versus nonpeople images by pigeons. Anim. Learn. Behav. 29, 107–119 (2001).

Bar, M. & Aminoff, E. Cortical Analysis of Visual Context. Neuron 38, 347–358 (2003).

Bar, M. VISUAL OBJECTS IN CONTEXT. (2004). doi:10.1038/nrn1476

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y. & Plumb, I. The 'Reading the Mind in the Eyes' Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. J. Child Psychol. Psychiatry 42, 241–251 (2001).

Barsalou, L. W. Perceptual symbol systems. Behav. Brain Sci. 22, 577-609–60 (1999).

Barsalou, L. W., Huttenlocher, J. & Lamberts, K. Basing categorization on individuals and events. Cogn. Psychol. 36, 203–72 (1998).

Biederman, I. Perceiving real-world scenes. Science 177, 77–80 (1972).

Biederman, I. Recognizing depth-rotated objects: a review of recent research and theory. Spat. Vis. 13, 241–53 (2000).

Biederman, I. Recognition-by-components: A theory of human image understanding. Psychol. Rev. 94, 115–117 (1987).

Biederman, I. & Bar, M. One-shot viewpoint invariance in matching novel objects. Vision Res. 39, 2885–99 (1999).

Brady, T. F. & Oliva, A. Statistical Learning Using Real-World Scenes: Extracting Categorical Regularities Without Conscious Intent. Psychol. Sci. 19, 678–685 (2008).

Briscoe, E. & Feldman, J. Conceptual complexity and the bias/variance tradeoff. Cognition 118, 2–16 (2011).

Bruce, V. & Young, A. Understanding face recognition. Br. J. Psychol. 77 (Pt 3), 305–27 (1986).

Burgess, N. Spatial memory: how egocentric and allocentric combine. Trends Cogn. Sci. 10, 551–557 (2006).

Buzsáki, G. Theta rhythm of navigation: link between path integration and landmark navigation, episodic and semantic memory. Hippocampus 15, 827–40 (2005).

Call, J. & Tomasello, M. Does the chimpanzee have a theory of mind? 30 years later. Trends Cogn. Sci. 12, 187–192 (2008).

Cartwright, B. A. & Collette, T. S. Landmark maps for honey bees. Biol. Cybern. 57, 85–93 (1987).

Cerella, J. The pigeon's analysis of pictures. Pattern Recognit. 12, 1–6 (1980).

Chao, L. L., Haxby, J. V & Martin, A. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nat. Neurosci. 2, 913–919 (1999).

Chater, N., Tenenbaum, J. B. & Yuille, A. Probabilistic models of cognition: conceptual foundations. Trends Cogn. Sci. 10, 287–91 (2006).

Chomsky N. Aspects of the Theory of Syntax. Aspects of the Theory of Syntax. Cambridge, MA: MIT Press; 1965. 224 p.

Chomsky N. Rules and Representations. New York: Columbia University Press; 1980. 308 p.

Chua, K. & Chun, M. M. Implicit scene learning is viewpoint dependent. Percept. Psychophys. 65, 72–80 (2003).

Chun, M. M. & Jiang, Y. Top-Down Attentional Guidance Based on Implicit Learning of Visual Covariation. Psychol. Sci. 10, 360–365 (1999).

Chun, M. M. Contextual cueing of visual attention. Trends Cogn. Sci. 4, 170–178 (2000).

Collett, T. S. & Collett, M. Memory use in insect visual navigation. Nat. Rev. Neurosci. 3, 542–552 (2002).

Colreavy, E. & Lewandowsky, S. Strategy development and learning differences in supervised and unsupervised categorization. Mem. Cognit. 36, 762–75 (2008).

Coltheart, V., Mondy, S. & Coltheart, M. Repetition blindness for novel objects. Vis. cogn. 12, 519–540 (2005).

Conway, C. M. & Christiansen, M. H. Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences. J. Exp. Psychol. Learn. Mem. Cogn. 31, 24–39 (2005).

Cosmides, L. & Tooby, J. Better than Rational : Evolutionary Psychology and the Invisible Hand. Am. Econ. Rev. 84, 327–332 (1994).

Damasio, A. R. Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. Cognition 33, 25–62 (1989).

Davidoff, J. Language and perceptual categorisation. Trends Cogn. Sci. 5, 382–387 (2001).

de Fockert, J. W. The Role of Working Memory in Visual Selective Attention. Science (80-.). 291, 1803–1806 (2001).

Diwadkar, V. A. & McNamara, T. P. Viewpoint Dependence in Scene Recognition. Psychol. Sci. 8, 302–307 (1997).

Ekstrom, A. D. et al. Cellular networks underlying human spatial navigation. Nature 425, 184–8 (2003).

Emlen, S. T. The Stellar-Orientation System of a Migratory Bird. Sci. Am. 233, 102–111 (1975).

Epstein, R. & Kanwisher, N. A cortical representation of the local visual environment. Nature 392, 598–601 (1998).

Epstein, R., Harris, A., Stanley, D. & Kanwisher, N. The parahippocampal place area: recognition, navigation, or encoding? Neuron 23, 115–25 (1999).

Fiser, J. & Aslin, R. N. Unsupervised statistical learning of higher-order spatial structures from visual scenes. Psychol. Sci. 12, 499–504 (2001).

Fiser, J. & Aslin, R. N. Encoding multielement scenes: statistical learning of visual feature hierarchies. J. Exp. Psychol. Gen. 134, 521–37 (2005).

Fiser, J. & Aslin, R. N. Statistical learning of higher-order temporal structure from visual shape sequences. J. Exp. Psychol. Learn. Mem. Cogn. 28, 458–467 (2002).

Folstein, J. R., Gauthier, I. & Palmeri, T. J. Mere exposure alters category learning of novel objects. Front. Psychol. 1, 40 (2010).

Gainotti, G. The organization and dissolution of semantic-conceptual knowledge: Is the 'amodal hub' the only plausible model? Brain Cogn. 75, 299–309 (2011).

Garoff, R. J., Slotnick, S. D. & Schacter, D. L. The neural origins of specific and general memory: the role of the fusiform cortex. Neuropsychologia 43, 847–859 (2005).

Gauthier, I., Skudlarski, P., Gore, J. C. & Anderson, A. W. Expertise for cars and birds recruits brain areas involved in face recognition. Nat. Neurosci. 3, 191–7 (2000).

Gigerenzer, G. & Goldstein, D. G. Reasoning the fast and frugal way: models of bounded rationality. Psychol. Rev. 103, 650–69 (1996).

Gilovich, T., Vallone, R. & Tversky, A. The hot hand in basketball: On the misperception of random sequences. Cogn. Psychol. 17, 295–314 (1985).

Gomez, R. L. & Gerken, L. Infant artificial language learning and language acquisition. Trends Cogn. Sci. 4, 178–186 (2000).

Gomez, R. L. & Gerken, L. Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. Cognition 70, 109–35 (1999).

Graef, P. De, Christiaens, D. & Ydewalle, G. Perceptual effects of scene context on object identification. Psychol Res 52, 317–329 (1990).

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. Nature 436, 801–806 (2005).

Hahn, U., Prat-Sala, M., Pothos, E. M. & Brumby, D. P. Exemplar similarity and rule application. Cognition 114, 1–18 (2010).

Handel, S. & Imai, S. The free classification of analyzable and unanalyzable stimuli. Percept. Psychophys. 12, 108–116 (1972).

Harris, I. M. & Dux, P. E. Orientation-invariant object recognition: evidence from repetition blindness. Cognition 95, 73–93 (2005).

Hayward, W. G., Kong, H. & Harris, I. M. Dissociating viewpoint costs in mental rotation and object recognition. 820–825 (2006).

Henderson, J. M. Object-based attentional selection in scene viewing Antje Nuthmann. J. Vis. 10, 1–19 (2010).

Hintzman, D. L. 'Schema abstraction' in a multiple-trace memory model. Psychol. Rev. 93, 411–428 (1986).

Hollingworth, A. Scene and Position Specificity in Visual Memory for Objects. J. Exp. Psychol. Learn. Mem. Cogn. 32, 58–69 (2006).

Hollingworth, A. Constructing Visual Representations of Natural Scenes: The Roles of Short- and Long-Term Visual Memory. J. Exp. Psychol. Hum. Percept. Perform. 30, 519–537 (2004).

Hollingworth, A. Object-position binding in visual memory for natural scenes and object arrays. J. Exp. Psychol. Hum. Percept. Perform. 33, 31–47 (2007).

Hollingworth, A. The Relationship Between Online Visual Representation of a Scene and Long-Term Scene Memory. 31, 396–411 (2005).

Hollingworth, A. & Henderson, J. M. Accurate visual memory for previously attended objects in natural scenes. J. Exp. Psychol. Hum. Percept. Perform. 28, 113–136 (2002).

Hoshino, E. & Mogi, K. in In Neural Information Processing. Theory and Algorithms 255–261 (2010). doi:10.1007/978-3-642-17537-4_32

Hume D, A Treatise of Human Nature by David Hume, reprinted from the Original Edition in three volumes and edited, with an analytical index, by L.A. Selby-Bigge, M.A. (Oxford: Clarendon Press, 1896). [cited 2016 Oct 1]. [Internet] available from http://oll.libertyfund.org/titles/342

Hutchinson, J. B. & Turk-Browne, N. B. Memory-guided attention: Control from multiple memory systems. Trends Cogn. Sci. 16, 576–579 (2012).

Inselberg, A. & Dimsdale, B. Parallel coordinates: a tool for visualizing multi-dimensional geometry. in Proceedings of the First IEEE Conference on Visualization: Visualization `90 361–378 (IEEE Comput. Soc. Press, 1990). doi:10.1109/VISUAL.1990.146402

Intraub, H. & Richardson, M. Wide-angle memories of close-up scenes. J. Exp. Psychol. Learn. Mem. Cogn. 15, 179–87 (1989).

Janzen, G. & Turennout, M. Van. Selective neural representation of objects relevant for navigation. 7, 673–677 (2004).

Jefferies, E. & Lambon Ralph, M. a. Semantic impairment in stroke aphasia versus semantic dementia: A case-series comparison. Brain 129, 2132–2147 (2006).

Jiang, Y., Chun, M. M. & Olson, I. R. Perceptual grouping in change detection. Percept. Psychophys. 66, 446–453 (2004).

Johansson, P. Failure to Detect Mismatches Between Intention and Outcome in a Simple Decision Task. Science 310, 116–119 (2005).

Kahneman, D. & Tversky, A. Subjective probability: A judgment of representativeness. Cogn. Psychol. 3, 430–454 (1972).

Kanwisher, N., McDermott, J. & Chun, M. M. No Title. J. Neurosci. 17, 4302–4311 (1997).

Kanwisher, N., McDermott, J. & Chun, M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17, 4302–4311 (1997).

Kay, P. & Kempton, W. What Is the Sapir-Whorf Hypothesis? Am. Anthropol. 86, 65–79 (1984).

Kirkham, N. Z., Slemmer, J. A. & Johnson, S. P. Visual statistical learning in infancy: evidence for a domain general learning mechanism. Cognition 83, B35-42 (2002).

Klatzky, R. Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. Spat. Cogn. - An Interdiscip. approach to Represent. Process. Spat. Knowl. 1–17 (1998). doi:10.1007/3-540-69342-4_1

Koelsch, S. et al. Bach speaks: a cortical ‘language-network’ serves the processing of music. Neuroimage 17, 956–66 (2002).

Kriegeskorte, N. et al. Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. Neuron 60, 1126–1141 (2008).

Lake, B. M., Salakhutdinov, R. & Tenenbaum, J. B. Human-level concept learning through probabilistic program induction. Science 350, 1332–8 (2015).

Levenshtein, V. I. Binary codes capable of correcting deletions, insertions, and reversals. Sov. Phys. Dokl. 10, 707–710 (1966).

Levine, J. MATERIALISM AND QUALIA : THE EXPLANATORY GAP. 64, 354–361 (1983).

Luck, S. J., Vogel, E. K. & Shapiro, K. L. Word meanings can be accessed but not reported during the attentional blink. Nature 383, 616–618 (1996).

Luck, S. J. & Vogel, E. K. The capacity of visualworking memory for features and conjunctions. Nature 390, 279–281 (1997).

Mack, M. L., Preston, A. R. & Love, B. C. Decoding the brain's algorithm for categorization from its neural implementation. Curr. Biol. 23, 2023–2027 (2013).

Maguire, E. A., Frith, C. D., Burgess, N., Donnett, J. G. & O'Keefe, J. Knowing where things are parahippocampal involvement in encoding object locations in virtual large-scale space. J. Cogn. Neurosci. 10, 61–76 (1998).

Mallot, H. A. & Gillner, S. Route navigating without place recognition: What is recognised in recognition-triggered responses? Perception 29, 43–55 (2000).

Marcus, G. F., Vijayan, S., Bandi-Rao, S. & Vishton, P. M. Rule Learning by Seven-Month-Old Infants. Science 283, 77–80 (1999).

Marcus, G. F. The Birth of the Mind: How a Tiny Number of Genes Creates the Complexities of Human Thought. (Basic Books, 2004). at <http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0465044069>

Marr D. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. New York: Freeman.

Marsolek, C. J. Dissociable Neural Subsystems Underlie Abstract and Specific Object Recognition. Psychol. Sci. 10, 111–118 (1999).

Martin, A. The Representation of Object Concepts in the Brain. Annu. Rev. Psychol. 58, 25–45 (2007).

McMenamin, B. W., Deason, R. G., Steele, V. R., Koutstaal, W. & Marsolek, C. J. Separability of abstract-category and specific-exemplar visual object subsystems: Evidence from fMRI pattern analysis. Brain Cogn. 93, 54–63 (2015).

Meegan, D. V. & Honsberger, M. J. M. Spatial information is processed even when it is task-irrelevant: Implications for neuroimaging task design. Neuroimage 25, 1043–1055 (2005).

Minda, J. P. & Smith, J. D. Prototypes in category learning: the effects of category size, category structure, and stimulus complexity. J. Exp. Psychol. Learn. Mem. Cogn. 27, 775–799 (2001).

Moore, C. M. & Egeth, H. Perception without attention: evidence of grouping under conditions of inattention. J. Exp. Psychol. Hum. Percept. Perform. 23, 339–352 (1997).

Mulligan, J. & Daniilidis, K. View-independent scene acquisition for tele-presence. in Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000) 105–108 (IEEE, 2000). doi:10.1109/ISAR.2000.880933

Navon, D. Forest before trees: The precedence of global features in visual perception. Cogn. Psychol. 9, 353–383 (1977).

Nelson, T. O., Gerler, D. & Narens, L. Accuracy of feeling-of-knowing judgments for predicting perceptual identification and relearning. J. Exp. Psychol. Gen. 113, 282–300 (1984).

Nickerson, R. S. Short-term memory for complex meaningful visual configurations: A demonstration of capacity. Can. J. Psychol. 19, 155–160 (1965).

Nosofsky, R. M. Attention, similarity, and the identification-categorization relationship. J. Exp. Psychol. Gen. 115, 39–57 (1986).

O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map . Preliminary evidence from unit activity in the freely-moving rat. Brain Res. 34, 171–175 (1971).

O'Keefe, J., Nadel, L.: The Hippocampus as a Cognitive Map. Oxford, Clarendon (1978)

O'Regan, J. K. & Noë, a. A sensorimotor account of vision and visual consciousness. Behav. Brain Sci. 24, 939-973-1031 (2001).

O'Regan, J. K. SOLVING THE 'REAL' MYSTERIES OF VISUAL PERCEPTION : THE WORLD AS AN OUTSIDE MEMORY. Canadian 46, 461–488 (1992).

Oliva, A. in Neurobiology of Attention (eds. Itti, L., Rees, G. & Tsotsos, J. K.) 251–257 (Elsevier, 2005). doi:10.1016/B978-012375731-9/50045-8

Oliva, A. & Torralba, A. The role of context in object recognition. 11, (2007).

Palmeri, T. J. & Gauthier, I. Visual object understanding. Nat. Rev. Neurosci. 5, 291–303 (2004).

Patterson, K., Nestor, P. J. & Rogers, T. T. Where do you know what you know? The representation of semantic knowledge in the human brain. Nat. Rev. Neurosci. 8, 976–987 (2007).

Pinker, S. The Blank Slate: The Modern Denial of Human Nature. (Viking Press, 2002).

Poggio, T. & Edelman, S. A network that learns to recognize three-dimensional objects. Nature 343, 263–266 (1990).

Pons, F. & de Rosnay, M. in Swiss Journal of Psychology 59, 215–216 (2000).

Pothos, E. M., Edwards, D. J. & Perlman, A. Supervised versus unsupervised categorization: Two sides of the same coin? Q. J. Exp. Psychol. 64, 1692–1713 (2011).

Pothos, E. M. et al. Measuring category intuitiveness in unconstrained categorization tasks. Cognition 121, 83–100 (2011).

Quinn, P. C. & Eimas, P. D. The Emergence of Category Representations During Infancy: Are Separate Perceptual and Conceptual Processes Required? J. Cogn. Dev. 1, 55–61 (2000).

Ravignani, A., Westphal-Fitch, G., Aust, U., Schlumpp, M. M. & Fitch, W. T. More than one way to see it: Individual heuristics in avian visual computation. Cognition 143, 13–24 (2015).

Reber, A. S. Implicit learning of artificial grammars. J. Verbal Learning Verbal Behav. 6, 855–863 (1967).

Reber, A. S. Implicit learning and tacit knowledge. J. Exp. Psychol. Gen. 118, 219–235 (1989).

Reber, A. S. Implicit learning of artificial grammars. J. Verbal Learning Verbal Behav. 6, 855–863 (1967).

Reber, R., Schwarz, N. & Winkielman, P. Processing Fluency and Aesthetic Pleasure : Is Beauty in the Perceiver ' s Processing Experience ? 8, 364–382 (2004).

Reed, S. K. Pattern Recognition and Categorization. Cogn. Psychol. 3, 382–407 (1972).

Rensink, R. A., O'Regan, J. K. & Clark, J. J. To See or not to See: The Need for Attention to Perceive Changes in Scenes. Psychol. Sci. 8, 368–373 (1997).

Rentschler, I., Treutwein, B. & Landis, T. Dissociation of local and global processing in visual agnosia. Vision Res. 34, 963–71 (1994).

Rosch, E. Cognitive reference points. Cogn. Psychol. 7, 532–547 (1975).

Rosch, E. H. Natural categories. Cogn. Psychol. 4, 328–350 (1973).

Rosch, E. & Mervis, C. B. Family resemblances: Studies in the internal structure of categories. Cogn. Psychol. 7, 573–605 (1975).

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. Basic objects in natural categories. Cogn. Psychol. 8, 382–439 (1976).

Ross, B. H. & Makin, V. S. in The nature of cognition (ed. Sternberg, R. J.) 205–241 (MIT Press, 1999). doi:doi: 10.1145/271125.271148

Saffran, J. R., Aslin, R. N. & Newport, E. L. Statistical learning by 8-month-old infants. Science 274, 1926–8 (1996).

Saffran, J. R., Johnson, E. K., Aslin, R. N. & Newport, E. L. Statistical learning of tone sequences by human infants and adults. Cognition 70, 27–52 (1999).

Saksida, L. M., Bussey, T. J., Buckmaster, C. A. & Murray, E. A. Impairment and Facilitation of Transverse Patterning after Lesions of the Perirhinal Cortex and Hippocampus, Respectively. Cereb. Cortex 17, 108–115 (2006).

Scharff, C. & Nottebohm, F. Study of the behavioral deficits following lesions of various parts of the zebra finch song system: Implications for vocal learning. J. Neurosci. 11, 2896–2913 (1991).

Seger, C. a. Implicit learning. Psychol. Bull. 115, 163–96 (1994).

Seki, Y., Suzuki, K., Osawa, A. M. & Okanoya, K. Songbirds and humans apply different strategies in a sound sequence discrimination task. Front. Psychol. 4, 1–9 (2013).

Sharot, T., Delgado, M. R. & Phelps, E. A. How emotion enhances the feeling of remembering. Nat. Neurosci. 7, 1376–1380 (2004).

Shepard, R. N. & Metzler, J. Mental rotation of three-dimensional objects. Science 171, 701–3 (1971).

Shepard, R. N. Recognition Memory for Words , Sentences , and Pictures. J. Verbal Learning Verbal Behav. 6, 156–163 (1967).

Sigala, N. & Logothetis, N. K. Visual categorization shapes feature selectivity in the primate temporal cortex. Nature 415, 318–320 (2002).

Simion, F. & Giorgio, E. Di. Face perception and processing in early infancy: inborn predispositions and developmental changes. Front. Psychol. 6, 1–11 (2015).

Sloutsky, V. M. From Perceptual Categories to Concepts: What Develops? Cogn. Sci. 34, 1244–1286 (2010).

Smith, J. D. Prototypes, exemplars, and the natural history of categorization. Psychon. Bull. Rev. 21, 312–331 (2014).

Smith, J. D. & Minda, J. P. Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. J. Exp. Psychol. Learn. Mem. Cogn. 28, 800–811 (2002).

Snodgrass, J. G. & Vanderwart, M. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. J. Exp. Psychol. Hum. Learn. 6, 174–215 (1980).

Spelke, E. S. Principles of Object Perception. Cogn. Sci. 14, 29–56 (1990).

Spetch, M. L. & Friedman, A. Recognizing rotated views of objects: interpolation versus generalization by humans and pigeons. Psychon. Bull. Rev. 10, 135–40 (2003).

Stobbe, N., Westphal-Fitch, G., Aust, U. & Fitch, W. T. Visual artificial grammar learning: comparative research on humans, kea (Nestor notabilis) and pigeons (Columba livia). Philos. Trans. R. Soc. B 367, 1995–2006 (2012).

Storms, G., De Boeck, P. & Ruts, W. Prototype and Exemplar-Based Information in Natural Language Categories. J. Mem. Lang. 42, 51–73 (2000).

Tanaka, J. W. & Farah, M. J. Parts and wholes in face recognition. Q. J. Exp. Psychol. Sect. A 46, 225–245 (1993).

Tarr, M. J. Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. Psychon. Bull. Rev. 2, 55–82 (1995).

Tolman, E. C. Cognitive maps in rats and men. Psychol. Rev. 55, 189–208 (1948).

Turk-Browne, N. B., Scholl, B. J., Chun, M. M. & Johnson, M. K. Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities Without Awareness. J. Cogn. Neurosci. 21, 1934–1945 (2009).

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K. & Chun, M. M. Implicit perceptual anticipation triggered by statistical learning. J. Neurosci. 30, 11177–11187 (2010).

Tye, M. The Imagery Debate. The MIT Press (1991). doi:10.2307/2220377

Tyler, L. K. et al. Objects and Categories: Feature Statistics and Object Processing in the Ventral Stream. J. Cogn. Neurosci. 25, 1723–1735 (2013).

Tyler, L. K. & Moss, H. E. Towards a distributed account of conceptual knowledge. Trends Cogn. Sci. 5, 244–252 (2001).

Volpe, B. T., Ledoux, J. E. & Gazzaniga, M. S. Information processing of visual stimuli in an ‘extinguished’ field. Nature 282, 722–724 (1979).

Wang, R. F. & Spelke, E. S. Updating egocentric representations in human navigation. Cognition 77, 215–50 (2000).

Wang, R. & Yu, J. A neural model on cognitive process. Adv. Neural Networks-ISNN 2006 50–59 (2006).

Wang, W. C., Lazzara, M. M., Ranganath, C., Knight, R. T. & Yonelinas, A. P. The Medial Temporal Lobe Supports Conceptual Implicit Memory. Neuron 68, 835–842 (2010).

Weisberg, J., van Turennout, M. & Martin, A. A Neural System for Learning about Object Function. Cereb. Cortex 17, 513–521 (2006).

Weiss G. When is real, real: Computers, the new reality and human consciousness. Acta polytechnica scandinavica. Ci Vol. 105 (9 ref.), pp.89-95 (1996).

Westphal-Fitch, G., Huber, L., Gomez, J. C. & Fitch, W. T. Production and perception rules underlying visual patterns: effects of symmetry and hierarchy. Philos. Trans. R. Soc. B 367, 2007–2022 (2012).

Wexler, M. & van Boxtel, J. J. A. Depth perception by the active observer. Trends Cogn. Sci. 9, 431–8 (2005).

Wilson, B. et al. Auditory Artificial Grammar Learning in Macaque and Marmoset Monkeys. J. Neurosci. 33, 18825–18835 (2013).

Wilson, B., Smith, K. & Petkov, C. I. Mixed-complexity artificial grammar learning in humans and macaque monkeys: evaluating learning strategies. Eur. J. Neurosci. 41, 568–578 (2015).

Wright, P., Randall, B., Clarke, A. & Tyler, L. K. The perirhinal cortex and conceptual processing: Effects of feature-based statistics following damage to the anterior temporal lobes. Neuropsychologia 76, 192–207 (2015).

Xu, Y. Revisiting the role of the fusiform face area in visual expertise. Cereb. cortex 15, 1234–1242 (2005).

Yamaguchi, Y. A theory of hippocampal memory based on theta phase precession. Biol. Cybern. 89, 1–9 (2003).

Publications.

Papers

Eiichi Hoshino, Ken Mogi. Multiple Processes in Two-Dimensional Visual Statistical Learning. PLOS One (In press)

Hoshino, E., Taya, F., and Mogi, K. (2007). Memory formation of object representation : natural scenes. Advances in Cognitive Neurodynamics ICCN 2007 (2008): 457-461.

Hoshino, E. and Mogi, K. (2010). Evidence for False Memory before Deletion in Visual Short-Term Memory Lecture Notes in Computer Science, 2010, Volume 6443/2010, 255-261, DOI: 10.1007/978-3-642-17537-4_32

Conferences

Eiichi Hoshino, Ken Mogi. Trade-off in the effect of attention for visual short term memory. Association for the Scientific Study of Consciousness 2010 (Poster)

Eiichi Hoshino, Ken Mogi. Evidence for false memory before deletion in visual short-term memory. International Conference on Neural Information Processing 2010. (Oral presentation)

Hoshino, E., Taya, F. and Mogi, K. (2008a). Implicit processing of the location and identity information in human. Association Scientific Study of Consciousness 12th. Taipei, Taiwan. (Poster)

Hoshino, E., Taya, F. and Mogi, K. (2008b). Implicit processing of the location and identity information in human. The Society for Neuroscience 38th Annual Meeting. Washington, USA. (Poster)

Eiichi Hoshino, Fumihiko Taya and Ken Mogi. "Contextual memory formation and goal-directed behavior". Society for Neuroscience 2007 (Oral presentation)
Eiichi Hoshino, Fumihiko Taya and Ken Mogi. " Memory formation of object representation : natural scenes". International Conference on Cognitive Neurodynamics 2007 (Oral presentation)

Hoshino E, Taya F, Mogi K. (2006). The extension of egocentric space perception, ASSC10 (Association for the Scientific Study of Consciousness), Oxford U.K., 25th Jun. (Poster)

Hoshino E, Taya F, Mogi K. (2006). Egocentric space and object perception. SfN2006 (Society for Neuroscience), Atlanta US, 18th Oct. (Poster)