Emergence of Color Constancy Illusion through Reinforcement Learning with a Neural Network

Katsunari Shibata, *Member, IEEE,* Shunsuke Kurizaki

***Abstract*—Our parallel and flexible brain that must be the origin of our flexibility processes visual signals without being noticed, and due to the unawareness, the contradiction between our perception after the process and original visual property is exposed as "Optical Illusion". The authors form the hypothesis that optical illusion can be acquired through or supported by the learning so as that we behave more appropriately in everyday life. In this paper, "color constancy" is focused on and the authors try to explain its emergence through the learning of a simple "colored-object guidance" task by reinforcement learning with a neural network whose inputs are raw image signals. In the task, it is required to move an object whose color is chosen randomly to the proper location that differs depending on the object color. Half of the field is covered by a translucent filter whose color and angle are chosen randomly at each episode. It was observed that the hidden neurons came to represent the object color mainly not depending on the filter color after reinforcement learning. In the subsequent supervised learning and test, the neural network with new output neurons was trained to output the object color only under the condition of no filter, but, when images covered by colored filter were the input as test patterns after learning, the network outputs were very close to the original object color.**

***Index Terms*—color constancy, optical illusion, function emergence, reinforcement learning, neural network, unconscious process.**

I. Introduction

Although the robot intelligence has been growing, it must not be too much to say that their flexibility is far inferior to humans'. It is obvious that the way of processing is quite different between humans and robots. The processing in our brain is massively parallel and flexible in harmony, and that enables to generate appropriate behaviors by considering many things comprehensively. On the other hand, the robot process is usually consisted of a series of inflexible functional modules developed by humans, and so-called "Frame Problem"[1] still remains as an unresolved issue. Parallel module allocation named "Subsumption Architecture" has been proposed[2], but the difficulty in the interface design among modules seems to disturb the development of more than simple robots.

Since the brain is a massively parallel processing system, while in contrast, our consciousness seems sequential, it is difficult for us to be conscious of all of what the brain is doing exactly even for our own brain. We often misunderstand that what we are conscious of are all or most of the processing in our brain. However, that is obviously wrong even from the fact that we cannot perceive the process of well-known "orientation-selective cells". Probably, so many unconscious processes that we cannot perceive or understand through the sequential consciousness must occupy our brain process actually and support our flexibility behind the scenes. Nevertheless, when we develop an intelligent robot, we deeply rely on the understanding of our brain function by the sequential consciousness. That causes the insurmountable wall in flexibility between humans and robots, the authors think.

By going through the unconscious process, the perceived property is transformed from the physical or original one. Usually in our daily life, the process must be very helpful for our flexible recognition, but the change without notice makes us feel as a contradiction. That must be "illusion".

There are so many amazing optical illusions such as shown in [3][4]. Fig. 1 is one of them and became the trigger of this research. It is incredible that the color of her right eye is actually gray, isn't it? It seems that the illusion can be explained by "Retinex theory"[5] or other compensation by the global color tendency, but we understand the mechanism is not so simple when we see the "Cornsweet edge illusion"[6][4]. Many researchers have been tackling the optical illusions not only by finding a new one, but also by modeling them[7][8] or sometimes by imaging the brain itself[9], and mechanism elucidation of underlying flexible human visual systems has been their target. The authors are interested in the mechanism of how such systems emerge rather than the mechanism of how such flexible visual systems work. That is because the parallel brain considers much wider variety of things in a complicated way than we expected, and "emergence" must be a more promising strategy than "manual design" to realize such flexible functions although that seems a longer way round, the authors think.

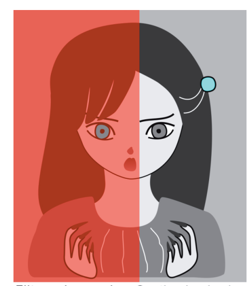


Fig. 1. One of the amazing optical illusions[3]. The eye color is actually gray. Copy permission from Prof. Kitaoka.

Lotto R. B. et al. pointed out as follows[4]. "What we perceive accords not with the features of the retinal stimulus or the properties of the underlying objects, but with what the same or similar stimuli have typically signified in both the experience of the species over the eons and the experience of individuals over their lifetimes", and importance of the statistics of past experience is emphasized. For example, if two line segments are located in alignment, it is highly possible that a part of the line is hidden behind something. When

K. Shibata is with the Department of Electrical and Electronic Engineering, Oita University, Oita, 870-1192 JAPAN e-mail: shibata@oita-u.ac.jp

S. Kurizaki was with Oita University, but now with Tokyo Electron Kyushu, JAPAN.

illuminated by the red evening sun, gray object is highly possible to be cyan originally. They seem to suggest that illusions are formed through experiences in humans, and the process can be modeled simply by using statistical method such as Bayesian inference.

However, a serious question arises when we see it in the aspect of "emergence from scratch". Before applying Bayesian inference, a target hypothesis should be formed at first, but who tells that the event whose probability should be estimated is "the eye color is originally cyan when the color is gray under the red illumination"? Forming an appropriate hypothesis actually needs high intelligence supported by enough knowledge. Just as can be seen in "Frame Problem"[1], there are many possible hypotheses, and if no knowledge are presupposed and just on the empirical basis, it takes a huge time even to reject the trivial ones. Therefore, a model is strongly desired in which functions emerge without any direct design or intention in a system with high degree of freedom.

Then, let us begin with catching the optical illusions as one of the parallel and flexible unconscious functions. The hypothesis: "the functions can be acquired through or supported by the learning to behave more appropriately in everyday life" is formed. As a model of our learning, reinforcement learning with a neural network that has no direct intension to develop an illusion but only has the criteria to get more reward is used. In this paper, "color constancy illusion" is focused on, and the authors try to explain its emergence through learning from the necessity of object recognition not depending on lighting conditions by introducing a "colored-object guidance" task as a very simple model of our daily life. Through that, the origin of our intelligence wants to be explored.

II. Function Emergence through Reinforcement Learning[10]

In the general approach when using reinforcement learning in a robot or agent, the entire system is modularized into some functional modules such as recognition, planning and control, and reinforcement learning is used only for mapping from state space to action space. Aiming to autonomous function acquisition, the authors have employed a very simple but unique approach as shown in Fig. 2. The system is consisted of just one layered neural network(NN) whose inputs are raw sensor signals and whose outputs are motion commands, and other pre-installed functions are excluded as much as possible. The NN is trained by the training signals derived from reinforcement learning algorithm at each time step, that is, reinforcement learning trains the entire process from sensors to motors. It may seem to be inefficient at the first impression, and actually the learning is slow. However, our approach enables purposive function emergence including recognition, memory and so on in the NN. That is expected because the system cannot be optimized to get more reward and less punishment without acquiring necessary functions. The functions works in parallel, flexibly and in harmony without designing the interface manually. The approach is also analogous to the fact that in the real lives, a nerve system connects from sensors to actuators. For the case of this paper,

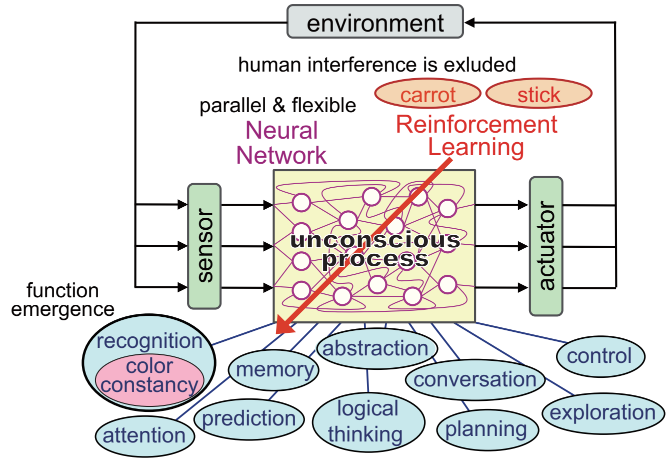


Fig. 2. Function emergence through reinforcement learning using a neural network as a parallel and flexible learning system.

it is expected that "color constancy" ability can be acquired through the learning of "colored-object guidance" task.

The concrete learning algorithm is as follows. Based on reinforcement learning algorithm, training signals are generated, and supervised learning is performed. This eliminates the need for supplying training signals from outside, and autonomous learning can be realized. In this paper, for continuous inputs and outputs, actor-critic[11] is used as a reinforcement learning method. Therefore, the output of the NN is divided into a critic output, which evaluates the state, and actor outputs, which generate motions. At first, TD-error is represented as

$$\hat{r}_{t-1} = r_t + \gamma P(\mathbf{s}_t) - P(\mathbf{s}_{t-1}) \tag{1}$$

where r_t is the reward given at time t, γ is a discount factor, $\mathbf{s}_t$ is the sensor signal vector at time t, and $P(\mathbf{s}_t)$ is the critic output when $\mathbf{s}_t$ is the input of the network. Here, the sigmoid function whose value ranges from -0.5 to 0.5 is used as a non-linear output function. When transforming between the NN output and critic value, 0.5 is added or subtracted to adjust the value range. The training signal for the critic output is computed as

$$P_{d,t-1} = P(\mathbf{s}_{t-1}) + \hat{r}_{t-1} = r_t + \gamma P(\mathbf{s}_t), \tag{2}$$

and the training signals (vector) for the actor outputs are computed as

$$\mathbf{a}_{d,t-1} = \mathbf{a}(\mathbf{s}_{t-1}) + \hat{r}_{t-1}\mathbf{rnd}_{t-1} \tag{3}$$

where $\mathbf{a}(\mathbf{s}_{t-1})$ is the actor output vector when $\mathbf{s}_{t-1}$ is the input vector of the network, and $\mathbf{rnd}_{t-1}$ is the random number vector that was added to $\mathbf{a}(\mathbf{s}_{t-1})$ as exploration factors. Then $P_{d,t-1}$ (actually 0.5 is subtracted to adjust it to the value range of the network output) and $\mathbf{a}_{d,t-1}$ are used as training signals, and the NN with the input $\mathbf{s}_{t-1}$ is trained once according to BP (Error Back Propagation)[12]. What the readers are asked is to bear in mind that the learning is very simple and general, and no special learning for color constancy is applied.

III. Simulation

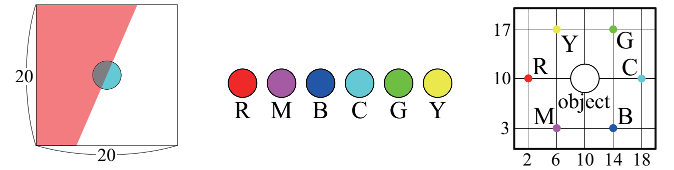
At first, reinforcement learning of "colored-object guidance" task is performed, and it is observed whether color constancy function emerges or not in the hidden neurons. After that, to see whether the color constancy illusion occurs or not actually, additional supervised learning and test are performed.

A. Learning of "Colored-Object Guidance" Task

The task is very simple. As shown in Fig. 3(a), there is a 20×20 square field, and an object whose shape is circle with radius 2 is located at the center at each episode. The object color is chosen randomly at each episode among 6 (Red, Green, Blue, Cyan, Magenta, Yellow) as shown in Fig. 3(b). The object moves according to the sum of the two-dimensional actor output vector $\mathbf{a}(\mathbf{s}_t)$ of the neural network and the random number vector $\mathbf{rnd}_t$ for exploration. The location of the goal differs depending on the object color as shown in Fig. 3(c), but the goal cannot be seen actually. When the object touches the goal point, a reward is given and the episode is terminated. No penalty is imposed.

The important point is that a half of the field is covered by a translucent colored filter. The boundary of the filtered area always passes the center of the field. The angle of the boundary is chosen randomly between 0 and 360 degree. The state of 0 degree means that the right half of the field is covered by the colored filter, while the other half is not covered. The state of 90 degree means that the upper half is covered. The filter color is also chosen randomly at each episode among the 6 colors and no filter. The 6 filter colors are the same as the 6 object colors as mentioned. The color of the filtered area is calculated by the average of the original color and filter color. Therefore, in Fig. 3(a), since the object color is cyan and the filter color is red, the sensor signal from each pixel where the object is covered by the filter is gray. That means that the three color signals are $(R, G, B) = (127, 127, 127)$. That is the average of Red $(255, 0, 0)$ and Cyan $(0, 255, 255)$. To move the object to the goal, it is required to recognize the object color correctly by eliminating the effect of the filter color.

Fig. 4 shows the learning system and signal flow of this task. The inputs of the neural network are the color signals from 20×20 grid points in the field after normalization between 0 and 1. The values are inverted between 0 and 1, and so all the RGB inputs for the grid with white color are zero. At the initial state, 12 grid points catch the object among total 400 grid points. The outputs of the network consist of one critic and two actor outputs. The output function of each neuron in the neural network is sigmoid function with the value range from -0.5 to 0.5 as mentioned. Each training signal is also limited between -0.4 and 0.4 to avoid the saturation area of the sigmoid function. Each of the two actor outputs corresponds to the lateral or vertical move, and the object moves by the vector $2.5\sqrt{2}(\mathbf{a}(\mathbf{s}_t) + \mathbf{rnd}_t)$ within the length limit $\sqrt{2}$. The size of the training signal vector is also limited to 0.4 that is identical to the object move $\sqrt{2}$. The neural network has five layers; 1200(400×3)-100-40-12-3 from input to output. All the initial connection weights except for the output neurons are set randomly between -1.0 and 1.0, and those for the output neurons are all 0.0. The reward 0.9 is given when the object reaches the goal, that is, when the distance between



(a) field, object and translucent filter (b) 6 colors for moving object or filter (c) goal for each colored object

Fig. 3. Colored-object guidance task. Each object or filter color is chosen randomly among 6 colors at each episode. The goal location is different depending on the object color, but not depending on the filter color.

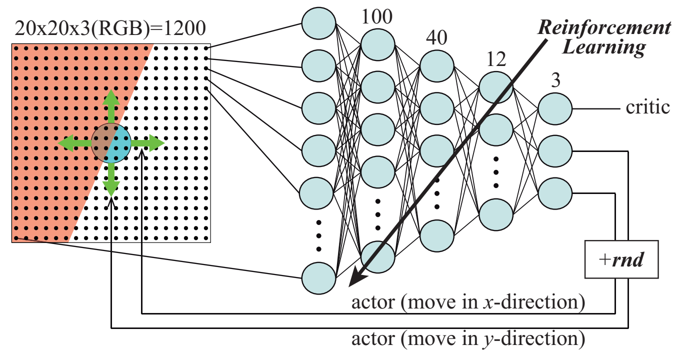
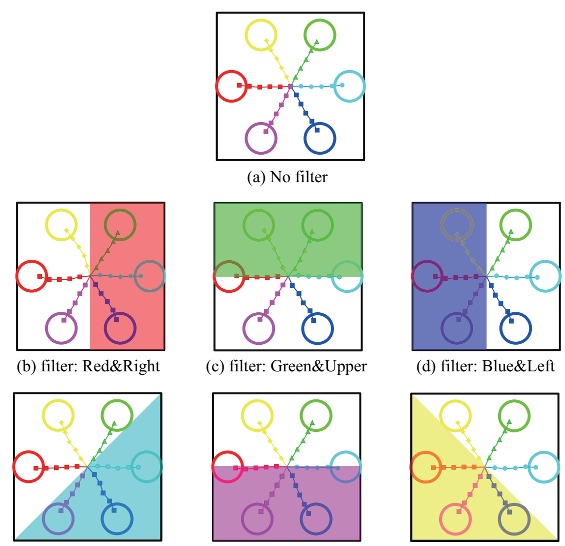


Fig. 4. Learning of "colored-object guidance" task by reinforcement learning using a neural network. The object motion vector is decided by the two actor outputs, but is normalized so as that the maximum size of the vector is $\sqrt{2}$.

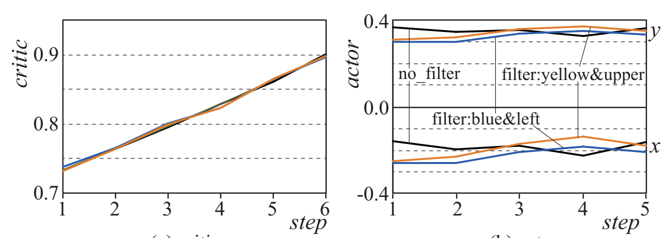


(a) No filter

(b) filter: Red&Right (c) filter: Green&Upper (d) filter: Blue&Left

(e) filter: Cyan&LowerRight (f) filter: Magenta&Lower (g) filter: Yellow&LowerLeft

Fig. 5. Object trajectories for some sample filter conditions after learning.



(a) critic (b) actor

Fig. 6. Change of Critic and actor outputs in one episode for the yellow object after learning.

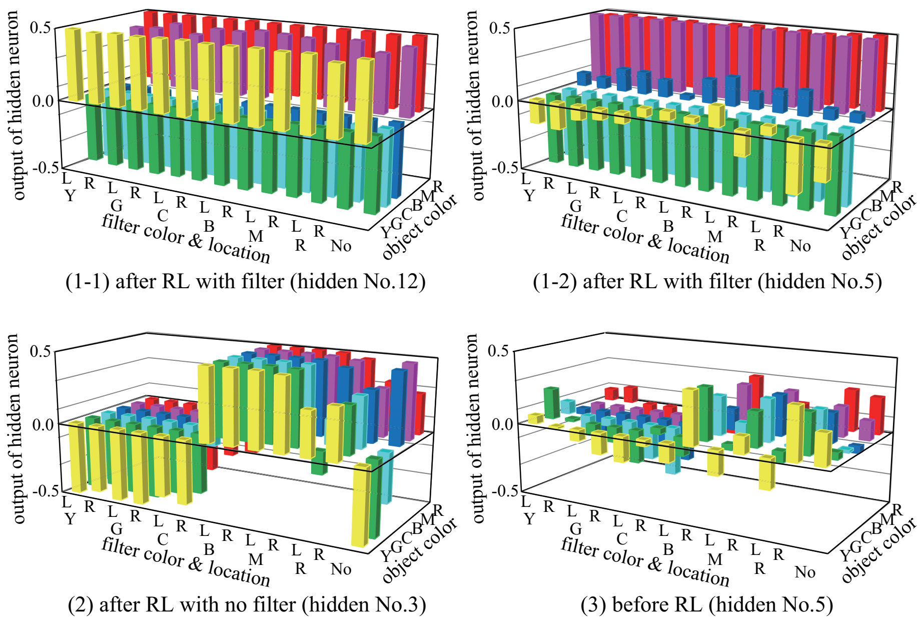


Fig. 7. Comparison of the output difference of one or two typical top hidden neurons depending on the object color, filter color and angle among the three cases: (1) after reinforcement learning(RL) of "colored-object guidance" task with filter, (2) after RL of the same task with no filter and (3) before RL (initial weights). The letters 'R', 'M', 'B', 'C', 'G', 'Y' indicate colors: red, magenta, blue, cyan, green, and yellow respectively. The letters 'R' and 'L' indicate the filter angle (location): 0° (right) and 180° (left) respectively.

object center and goal is less than 2.0. The discount factor γ in Eq. (2) is set to 0.96. The range of the random number that is added to the actor outputs for exploration is as large as ± 2.0 at first, and linearly decreased until 0.0.

Fig. 5 shows the object trajectories for some sample filter conditions after 200,000 episodes of learning. It can be seen that the object moves differently depending on the object color and reaches the correct goal in 5 steps that is the optimal. The trajectory is slightly different depending on the filter color, but very similar to each other.

Fig. 6 shows the change of critic and actor in one episode for some cases of the yellow object. In one case, no filter is applied. In other two cases, the filter is blue that is the opposite of the object color and located left half, or the filter is yellow that is the same as the object color and located upper half. It is seen that in either case, the critic is smoothly increasing, and cannot be distinguished from the ideal curve that is computed from the reward $r = 0.9$ and the discount factor $\gamma = 0.96$. As for the actor outputs, to move the object to the upper left direction, the actor output for the x-element is negative, while that for the y-element is positive even though there are some difference among the three cases.

B. What the Hidden Neurons Came to Represent

The outputs of 12 top hidden neurons that are in the hidden layer closest to the output layer are observed when the object is located at the center of the field with various combinations of object and filter colors. Fig. 7 shows some typical hidden neuron's output for 78 combinations of object and filter colors: (6 object color $\times$ (6 filter color $\times$ 2 filter angle(location) + no filter)). For comparison, other than "after RL with filter" case, hidden neurons are also observed in the cases of "after

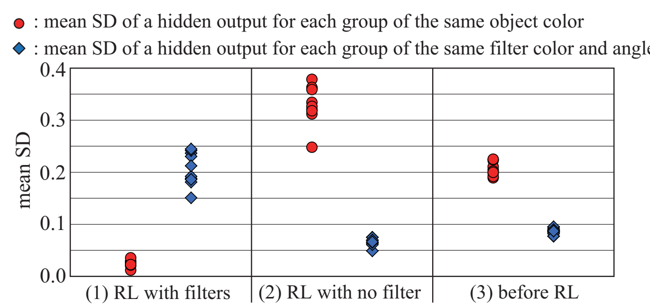


Fig. 8. Comparison of mean standard deviation(SD) of an top hidden neuron output between for each group of the same object color (red circle plot) and for each group of the same filter color and angle. One plot indicates the average over 12 top hidden neurons, and 10 plots are for different simulation runs with a different random number sequence. Small value of the red circle plot indicates that the variation of hidden outputs is small due to the filter color and angle if the object color is the same, and that can be seen after RL with filter condition as shown in Fig. 7(1).

reinforcement learning(RL) of object guidance task without applying any filter" and "before RL". In the case of "after RL with filter", half of the hidden neurons do not change its value so much or only change its value in the case of no filter, but the other hidden neurons change its output mainly according to the object color as shown in Fig. 7(1-1)(1-2). No neurons that change its value mainly according to the filter color are observed. On the other hand, in the case of "after RL with no filter", the output is deeply influenced by the filter color as shown in Fig. 7(2) except for small number of neurons that do not change or change irregularly. In the case of "before RL", since the output of each hidden neuron is decided only by the initial connection weights, the absolute value of the output is smaller in total, and the distribution is more irregular than the others as shown in Fig. 7(3).

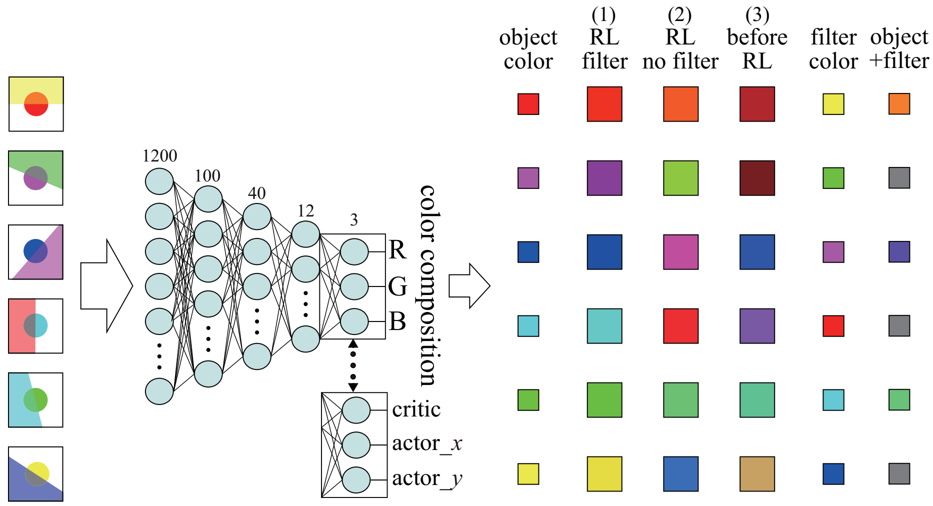


Fig. 9. Comparison of the color composed from the network outputs after supervised learning of object color with no filter among three cases; (1) after reinforcement learning(RL) of "colored-object guidance" task with filter, (2) after RL of the same task with no filter and (3) before RL. Please notice that as the input images at the left side, small size of images around the center are shown, but actually, 20×20 image is inputted to the neural network.

In order to see the tendency quantitatively, Fig. 8 shows the variation of hidden neuron's output in the same object color and also the variation in the same filter color and angle. To show the former property, standard deviation of a top hidden neuron output when the object color is the same, but the filter color and angle are varied is calculated, and then its average over the 6 object colors is calculated. Furthermore, the average of the value over all hidden neurons are calculated and one red circle is plotted in the figure. On the other hand, standard deviation of a top hidden neuron output when the filter color and angle is the same, but the object color is varied is calculated and its average is plotted as one blue diamond. It is quite apparent that the hidden neurons after RL with filter mainly represent the object color without being influenced by the filter color, while on the contrary, those after RL with no filter mainly represent filter color without being influenced by the object color.

C. Test of "Color Constancy Illusion"

In order to test the "color constancy illusion", additional supervised learning is applied to the neural network. The output neurons are replaced to 3 new neurons with 0 connection weights from all the top hidden neurons. At each time, an object with a randomly chosen color is located at the center of the field with no filter, and each of the RGB values of the object color is given as training signal for the corresponding output after linearly transformed into -0.4 to 0.4. Learning was done for 10,000 presentations. All the connection weights were modified including those modified in RL, but the hidden representations did not change so much by the supervised learning. After learning, the object is located at the center but with a colored filter as a test, and the three outputs are observed. This time, a half of the field and also a half of the object are covered by a colored filter, and the other halves are not covered by it. If "color constancy" has been formed through the learning of the object guidance task, it is expected that the output will be close to the object color by eliminating

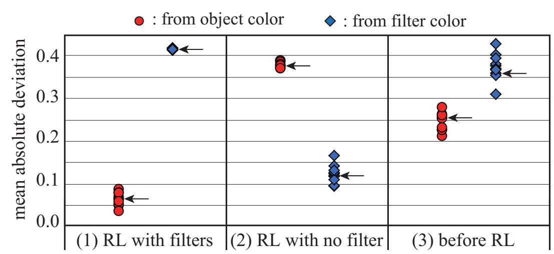


Fig. 10. Comparison of mean absolute deviation from the composite output to the object color or filter color among three cases. 10 simulation runs with different random number sequence are done for each case. The small arrows indicate the data for the neural network used in Fig. 7 and Fig. 9.

the filter effect because the hidden neurons represent the object color not depending on the filter color. Fig. 9 shows comparison of the object color composed from the network outputs for some object color and filter conditions. It can be seen that the composite color is more similar to the object color in the case of "after reinforcement learning(RL) with filter" than in the other two cases. It is interesting that in the case of "after RL with no filter", the color is influenced more from the filter color. That should be because recognition of object color was learned all over the field, and discrimination between object color and filter color was not learned since no filter appeared during RL. In the case of "before RL", the input signals from the field other than around the center is always 0 during the supervised learning, and so in the test phase, the remaining initial connection weights affect the output.

In order to see the tendency quantitatively, Fig. 10 shows the average of mean absolute deviation from the outputs of the network to each of the object color and filter color for the test patterns. The error for the 6 learning patterns with no filter became sufficiently small through learning in all the 3 cases. For the test patterns in which an object appears with a colored filter, the average of absolute deviation over 6 object colors, 6 filter colors, and 4 filter angles is calculated. In the case of "after RL with filter", the output is significantly closer to the

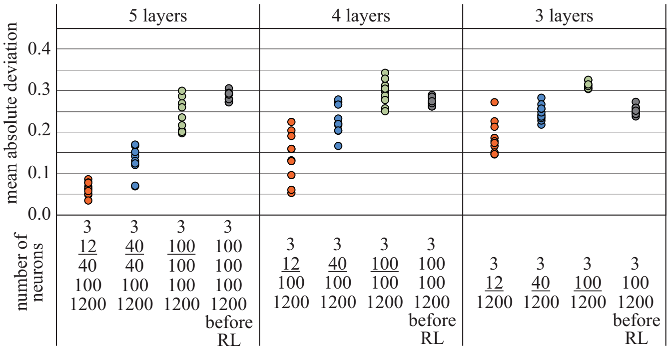


Fig. 11. Difference of mean absolute deviation from the composite output to the object color depending on the network structure. 10 simulation runs with different random number sequence are done for each case as in Fig. 10.

object color than to the filter color. In the case of "before RL", the outputs are closer to the object color, but not so close as the case of "after RL with filter". As mentioned above, the output in the case of "after RL with no filter" is closer to the filter color compared with the other cases. These results suggest the possibility that "color constancy", in other words, "filter color compensation" emerges through learning in daily life situation such as eating a banana in the reddish sunset, the authors think.

Figure 11 shows how the network structure influences the mean absolute deviation between the composite color and the real object color. It is seen that the error decreases as the number of the top hidden neurons that are the closest to the output layer is getting smaller. It also becomes smaller as the number of layers increases. Here, the color constancy emerges in the process to transform the higher dimensional visual information by the demand of generating appropriate actor (motion) and critic (state value). It is thought that the small number of top hidden neurons and many layers help to generate representations closer to the output. Therefore, it is thought that the network structure deeply influences to the emergence. A similar effect can be seen in [13].

IV. Conclusion and Discussion

In this paper, it was pointed out that unconscious process by the brain as a parallel processing system and also the emergence of such process through daily life learning are important, and since we are not conscious of the process, the result of the process is felt as contradiction or illusion. Aiming to show the possibility of the emergence of illusion and also underlying flexible function, a simple "colored-object guidance" task under the condition of a various color filter was learned by reinforcement learning with a neural network. By observing the hidden neurons' outputs and also by additional supervised learning and test, it was confirmed that "color constancy" function emerged through learning from the purpose of moving an object to the corresponding goal whose location is decided by the object color. In this method, no knowledge about the task or no inference target that is required for statistical methods such as Bayesian inference is necessary, and the useful representation emerges in the hidden neurons in the neural network as a parallel processing system as the result of the optimization to get more reward and less punishment. That is a very important aspect to utilize a parallel processing system effectively and to avoid the "Frame Problem".

The "color constancy illusion" shown in this paper can be explained also by Retinex theory[5], but our actual illusion is not so simple but influenced by many factors as mentioned in Introduction. When we examine the illusion using simple color patches with a variety of conditions, we actually notice that it is not easy to say the condition exactly that the color constancy illusion occurs. For example, a small gray square on a large light red square does not cause the same illusion as in Fig. 1, but a reddish gray square is located next to the gray square, we can see the same gray square is closer to cyan. The authors expect that these complicated occurrence mechanism must be formed through the experiences with various conditions, for example, one small gray (not cyan) object is just put on a large light-red paper.

The illusion may not be an acquired ability, but be an inherent ability for humans. If so, similar emergence mechanism may exist in the evolutional process. The authors think it possible that inherent illusion is supported also by the learning in our brain after birth in which illusion has been already acquired through evolutional process before its birth, and that makes the visual recognition more adaptive and flexible.

Acknowledgment

The authors would like to express their gratitude to Prof. Akiyoshi Kitaoka, Ritsumeikan Univ. for triggering them to this research by the amazing optical illusions and giving them some useful comments. This research was supported by JSPS Grant in-Aid for Scientific Research #19300070 & #23500245.

References

[1] Dennett, D. Cognitive Wheels: The Frame Problem of AI, *The Philosophy of Artificial Intelligence*, Margaret A. Boden, Oxford University Press, pp. 147-170 (1984)

[2] Brooks, R. A. Intelligence Without Representation. *Artificial Intelligence*, Vol. 47, pp.139-159 (1991)

[3] Kitaoka, A.: Akiyoshi's illusion pages, http://www.ritsumei.ac.jp /˜akitaoka/index-e.html, or Introduction to Visual Illusions, Asakura Shoten (2010) in Japanese

[4] Lotto, R.B., Purves, D., Nundy S.: Why We See What We Do, *American Scientist*, Vol. 90, No. 3, pp. 236–243 (2002)

[5] Land, E.H.: The Retinex Theory of Color Vision, *Scientific American*, Vol. 237, No.6, pp. 108–128 (1977)

[6] Cornsweet, T.: Visual Perception, Academic Press (1970)

[7] Inui, T., Hongo, S. & Kawato, M., A computational model of brightness illusion and its implementation, *Perception*, Vol. 19, pp. 401 (1990)

[8] Robinson, A. E, Hammon, P. S. & de Sa, V. R.: Explaining brightness illusions using spatial filtering and local response normalization, *Vision Research*, Vol. 47, pp. 1631–1644 (2007)

[9] Kuriki, I., Ashida, H., Murakami, I & Kitaoka, A.: Functional brain imaging of the Rotating Snakes illusion by fMRI, *J. of Vision*, Vol. 8, No. 10, pp. 1–10 (2008)

[10] K. Shibata: Emergence of Intelligence through Reinforcement Learning with a Neural Network, *Advances in Reinforcement Learning*, A. Mellouk (Ed.), InTech, pp. 99–120 (2011)

[11] Barto, A.G., Sutton, R.S., & Anderson, C.: Neuron-like adaptive elements that can solve difficult learning control problems, *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13: 834-846 (1983)

[12] Rumelhart, D.E, Hinton, G.E., and Williams, R.J.: Learning Internal Representations by Error Propagation, *Parallel Distributed Processing*, The MIT Press, pp. 318-362 (1986)

[13] Shibata, K.: Spatial Abstraction and Knowledge Transfer in Reinforcement Learning Using a Multi-Layer Neural Network, *Proc. of ICDL5*, 36 (CD-ROM) (2006)