IIAI Open Conference Publication Series
IIAI Letters on Informatics and Interdisciplinary Research
Vol. 003, LIIR060
DOI: https://doi.org/10.52731/liir.v003.060

Simulating Reaction Time for Eureka Effect in Visual Object Recognition Using Artificial Neural Network

Kazufumi Hosoda *, Shigeto Seno †, Tsutomu Murata *

Abstract

The human brain can recognize objects hidden in even severely degraded images after observing them for a while, which is known as a type of Eureka effect, possibly associated with human creativity. A previous psychological study suggests that the basis of this ”Eureka recognition” is neural processes of coincidence between input-dependent firing and stochastic-spontaneous firing. Here we constructed an artificial-neural-network-based model that simulated the characteristics of the human Eureka recognition.

Keywords: Eureka effect, Visual object recognition, Artificial neural network.

1 Introduction

Understanding and reconstructing human creativity in terms of neural modeling is important for constructing creativity support systems [1]. Eureka effect, also known as enlightenment, illumination, or Aha! effect, is thought to be associated with human creativity [2]. As a phenomenon in a visual Eureka effect, humans can recognize objects hidden in even severely degraded images (such as Mooney images [3]) suddenly after observing them for a while. Once recognized, the objects remain clear in the conscious awareness.

In a previous psychological study of this “Eureka Recognition (ER)” [4], reaction time (RT) of human subjects to recognize hidden objects in degraded binary (black-and-white) images followed a certain equation in which the RT was determined by both “individual ability” and “image difficulty that varies across images.” This quantitative relation was explained by assuming that the human brain stochastically activated missing components of a hidden object in a degraded image and ER occurred when multiple missing components were coincidentally activated. A coarse-grained mathematical model that repeats stochastic activation until the coincidence occur, using psychologically-measured image difficulty, explained the following characteristics of the psychological results: (I) RTs over subjects for a degraded image followed a normal distribution on a logarithmic time scale, (II) means and standard deviations of the above lognormal distributions over various degraded images

* Center for Information and Neural Networks (CiNet), NICT, Osaka, Japan
† Graduate School of Information Science and Technology (IST), Osaka University, Osaka, Japan

showed a linear relation, and (III) image difficulty showed a discrete distribution characterized by natural numbers. However, this coarse-grained model does not include specific neural mechanisms of ER and how ER can be implemented by artificial neural network (ANN) models. Note that the Ratcliff drift diffusion model [5], a well-studied standard model for the task of perceptual decision making, can not explain the above three characteristics.

In this study, we constructed an ANN-based model to simulate the ER characteristics observed in the previous human study [4]. Being applied to the same degraded images used in the human experiment (e.g., shown in Figure 1), our model successfully reproduced the above three characteristics of ER. Our model can show differences and similarities between the recognition of humans and a certain ANN, and it will contribute to the elucidation of the mechanism of Eureka effect and human creativity.

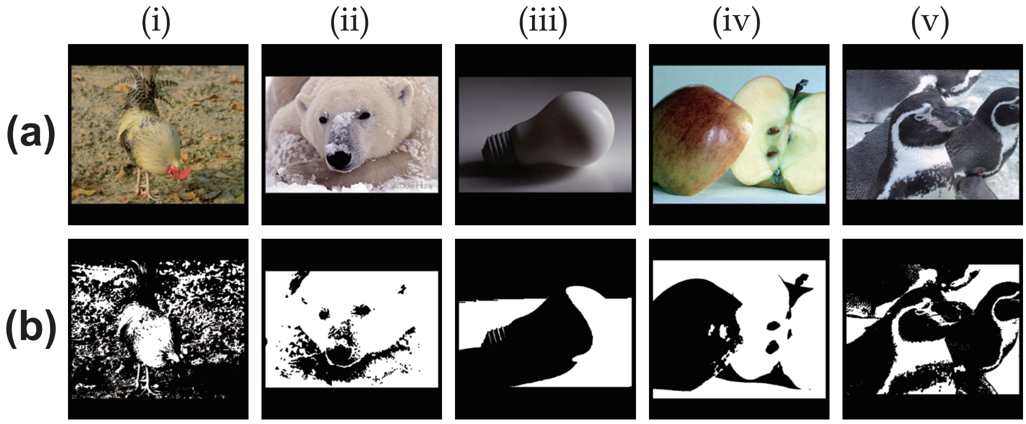

Figure 1: Examples of images used in this study. Five of the 90 images are shown as examples for each of the color (a) and binary (b) images. The same images were used in the previous psychological experiments.

2 Methods

2.1 Dataset and ANN image classifier

For input images, we used 90 binary images and their corresponding 90 original color images which were also used in the previous human experiment [4] (e.g., shown in Figure 1). For an ANN image classifier, we used ViT-B/32 (a Vision Transformer [6]) pretrained with ImageNet 1000 classes. Some of the 90 images were not included in the 1000 classes, and we added 15 new classes to the ANN by DONE (Direct ONE-shot learning [7]), using three Creative-Commons-license images on the internet as the training data for each of the 15 classes. We determined "correct class" for each binary image as a coarse-categorically-correct class [8] that were top-1 or top-2 output obtained when inputting the corresponding color image.

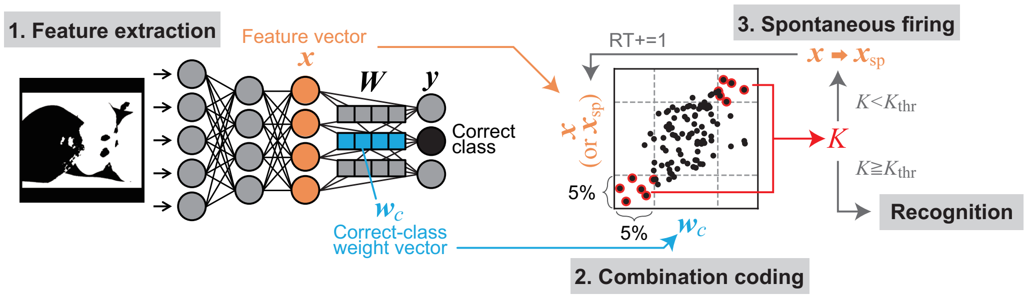

Figure 2: Scheme of the model. The model consists of three parts: an ANN image classifier for "1. Feature extraction", a recognition criterion based on multiple neural factors for "2. Combination coding", and a stochastic process for "3. Spontaneous firing".

2.2 ANN-based Model for Eureka Recognition

Our model is based on the previous coarse-grained model [4], which was inspired by two neurophysiological ideas: "combination coding" and "spontaneous firings." Combination coding proposes that an object is recognized by a combination of simultaneously activated multiple neural clusters, each of which represents a visual feature component of the object [9]. When the neural cluster components necessary for recognizing an object are not enough, the object is not recognized, which is considered to be the case when the visual input is a severely degraded image and is not recognized instantly. Here, it is known that cortical neurons spontaneously fire even without sensory input [10], and their spontaneous firings can be approximated by a stochastic process [11]. If these spontaneous firings coincidentally fill the missing feature, the severely degraded image would be recognized. This is the mechanism of the previous coarse-grained model.

To integrate the mechanism of the previous coarse-grained model into an ANN model, we used a node (neuron) of the input feature vector of the final-dense layer of ANN as a brain neural cluster that represents a visual feature component of the object (Figure 2). When the image was not immediately recognized, we added a stochastic process for the spontaneous firings that randomly changes the feature vector, as an ER process. We repeated this stochastic process until the active neural cluster components necessary for recognizing an object in an image became enough, i.e., the input image was recognized. We defined RT as the number of the repetitions of the stochastic process until the recognition.

Specifically, to construct an ANN-based model that simulates RT for ER, we integrated the following three parts: an ANN image classifier (ViT-B/32), a recognition criterion based on multiple neural factors, and a stochastic process. Note that our model was not for classifying a binary image but for obtaining RT, and the model considered only the correct class because in the human experiment the wrong answer was ignored. The ANN classifier has a final-dense layer, in which the input is a feature vector $\boldsymbol{x}$ with M neurons ($M = 768$) and the output is a vector $\boldsymbol{y}$ for N classes ($N = 1000 + 15$, see Methods) that represents the fitness rating to each class of the output by means of a weight matrix $\boldsymbol{W}$ ($N \times M$). The rating for i-th class y_i ($i = 1, 2, \cdots, N$) is proportional to the cosine similarity between the feature vector $\boldsymbol{x}$ (obtained from image) and the corresponding weight vector $\boldsymbol{w}_i$ (i-th row vector of $\boldsymbol{W}$). Therefore, the weight vector for the correct class $\boldsymbol{w}_c$ indicates neurons to fire or rest for the class.

For the recognition criterion, we defined "specific-neurons-to-fire" and "specific-neurons-to-rest" of the feature vector $\boldsymbol{x}$ as neurons whose corresponding weight values were respectively top 5% and bottom 5% (i.e., fraction of feature elements contributing to recognition, $\theta = 0.05$) in the weight vector $\boldsymbol{w}_c$. Then, we obtained the sum of the numbers of firing specific-neurons-to-fire (top 5%) and resting specific-neurons-to-rest (bottom 5%) in the feature vector $\boldsymbol{x}$, which we denote as K. Setting a threshold number K_{thr}, we regarded the input image as being recognized when $K \geq K_{\mathrm{thr}}$. The threshold value was obtained in common for all 90 binary images as the minimum value of K among 90 color images ($K_{\mathrm{thr}} = 10$), because human subjects recognized all color images instantly and ER did not occur in the human experiment.

When K was less than K_{thr}, the input image was not yet recognized, and the following stochastic process was introduced as the spontaneous firing. Each of M neurons in $\boldsymbol{x}$ randomly fired, rested, or retained the value obtained from the input image, with a probability of p, p, or $1-2p$, respectively. The resultant feature vector with spontaneous firing was defined as $\boldsymbol{x}_{\mathrm{sp}}$. At each time step, K was evaluated for $\boldsymbol{x}_{\mathrm{sp}}$, and this stochastic process was repeated while K was less than K_{thr} (the feature vector was returned back to the original $\boldsymbol{x}$ each step). When K satisfied the threshold, i.e., one or multiple missing components were coincidentally activated, the stochastic process was terminated, and the RT for ER was obtained as the number of time steps of this stochastic process.

The model parameters and variables are summarized in Table 1. There are no hyperparameters because we did not use a learning process by optimization to build the model but just used a pretrained model.

Table 1: The model parameters and variables. See text for details.

Parameters	Description
$\boldsymbol{W}$	$N \times M$ weight matrix of the final-dense layer of ANN (pretrained).
$\boldsymbol{w}_c$	Correct-class weight vector with M elements as a row vector of weight matrix $\boldsymbol{W}$ of ANN (image-class dependent).
θ	Fraction of feature elements contributing to recognition in the feature vector (=0.05).
p	Probability of the spontaneous firing (trial/individuality dependent); the value was randomly assigned to each trial, following common lognormal with mean -3.5 (≈ 0.0003 in linear) and SD 1.
K_{thr}	Number of feature elements necessary for recognition (= 10, as the minimum value among 90 color images).
Variables	**Description**
$\boldsymbol{x}$	Feature vector with M elements obtained as the input of the final-dense layer of ANN (image dependent).
$\boldsymbol{x}_{\mathrm{sp}}$	Feature vector with the spontaneous firing added to $\boldsymbol{x}$.
K	Sum of the number of firing specific-neurons-to-fire and resting specific-neurons-to-rest in the feature vector $\boldsymbol{x}$ or $\boldsymbol{x}_{\mathrm{sp}}$.

3 Results and Discussion

The ANN-based model was applied to obtain RT for the 90 binary images, which were used in the previous human study [4]. To validate the model in comparison with the human results, we investigated if the model reproduced the characteristics of human data described as (I) to (III) in Introduction section.

First, since humans can answer color images instantly and ER does not occur, we set the minimum value of K among the 90 color images as the threshold value ($K_{\text{thr}} = 10$). Therefore, RT was zero when any color image was input.

Next, 90 binary images were input to the model and RTs were obtained. In this case, 20 out of 90 binary images satisfied the threshold ($K_{\text{thr}} = 10$) before the stochastic process of ER, and the ER stochastic process occurred for the other 70 images.

The simulation of ER stochastic process was performed using 100 different spontaneous-firing probabilities p (corresponding to the individuality of 100 human subjects). We made the spontaneous-firing probability p to distribute lognormally (according to the human results), and RT was obtained by simulations for each p value.

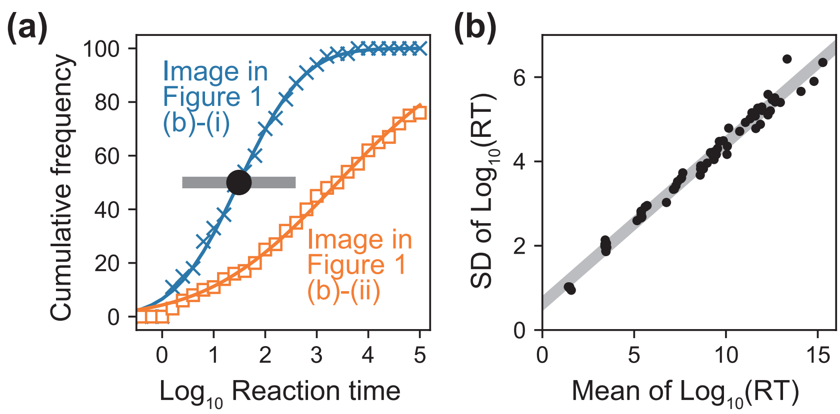

Figure 3: Statistical characteristics of the reaction time. (a) Cumulative frequency of RT. Results for two example binary images are shown. (b) The relationship between the mean and standard deviation of log RT in the 70 binary images with the ER stochastic process. The gray line shows the regression line.

3.1 Statistical Characteristics

Figure 3(a) shows cumulative distributions of the common logarithm of RT with the one hundred p values for inputs of two example binary images (blue crosses and orange squares). These distributions were well fitted by normal distributions (blue and orange lines) as the property (I) of human study (see Introduction), providing values of the logarithmic means and SDs (as indicated by the black circle and gray bar for blue example).

Figure 3(b) shows the relationship between the means and SDs obtained by the model using the 70 binary images with the ER stochastic process. The result clearly shows a linear relationship between the means and SDs over the 70 images ($R = 0.99$), successfully reproducing the property (II) of human study.

The linear relationship between mean and SD indicates that these lognormal distributions in Figure 3(a) just have different horizontal scales. The linear relationship is reproduced by a mechanism based on not accumulating processes but simultaneous processes, because the mean should have been proportional to the variance (not SD) in the case of accumulation processes. Note that typical models for perceptual decision making, such as drift diffusion model [5] and Bayesian attractor model [12], can not reproduce these results.

3.2 Differences between Images

Figure 4(a) shows a frequency distribution of the mean log of RT of the 90 binary images in our model. Twenty images with log RT as ”0” were recognized immediately without the ER stochastic process. This result successfully provided the corresponding property (III) of human study.

The histogram is not continuous, but has periodic peaks. In this model, the periodic peaks are reproduced because the difficulty of recognizing binary image is discretely expressed by the number of missing components necessary for recognition (as bars with integers). Roughly speaking, when the number of missing components necessary for recognition is n, the binary image is not recognized until n components spontaneously fire at the same time, and thus RT is proportional to p^{-n} on average.

Figure 4(b) quantitatively shows the relationship between the previous human experiment and our model simulation about the mean log of RTs (i.e., image difficulty) of 90 binary images. It shows large differences with a small significant correlation ($R = 0.28, P = 0.007; \alpha = 0.05$) between them, possibly illustrating both differences and similarities in ways of coding visual objects to neural activities, which is still under elucidation across research fields. In order to investigate the differences and similarities, more detailed analyzes such as behavior of ANN states and parameter/model dependencies are required in the future.

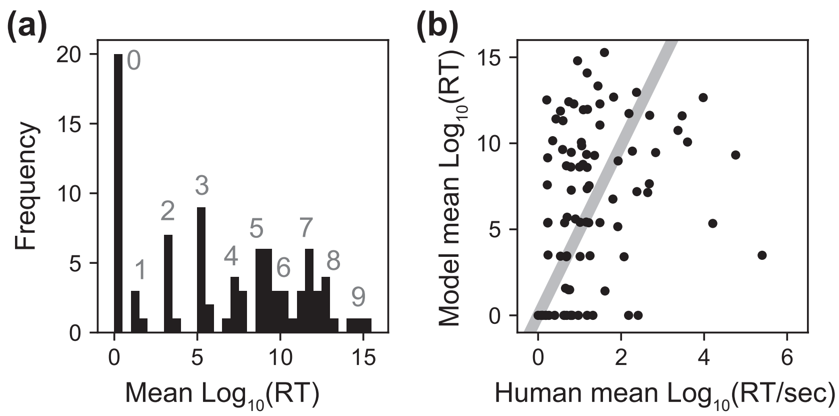

Figure 4: Difference between images. (a) Histogram of log RT (i.e., image difficulty) of 90 binary images. (b) Relationship between the human experiment and our model simulation about the mean log of RTs of 90 binary images. The gray line shows a line passing through both the origin and the centroid.

4 Conclusion

We simulated the reaction time for the Eureka effect in visual object recognition of degraded difficult-to-recognize binary images by using an ANN-based model. Our model well reproduced the characteristics of the previous human experimental results [4] and quantitatively showed the difference between human and the ANN-based model in image difficulty. The coincidence of multiple stochastic neural activities, which plays an essential role in our ANN-basd model, seems to be consistent with a conventional theory that creativity arises from the association of multiple seemingly independent processes [1]. Of course just because our model is able to reproduce human results does not mean that their mechanism is the same, however, at least our model, unlike other models for perceptual decision making, is consistent with the above human results. Our model may contribute to the elucidation of the mechanism of creativity and construction of creativity support systems.

Acknowledgments

We thank Drs. Toshio Yanagida, Kunihiko Kaneko, Izumi Ohzawa, Toshiyuki Kanoh, Hideki Kashioka, Tetsuya Shimokawa, and Mr. Keigo Nishida for helpful advice and discussions. This work was supported in part by MIC under a grant entitled "R&D of ICT Priority Technology (JPMI00316)".

References

[1] K. Wang and J.V. Nickerson, "A literature review on individual creativity support systems," Comput Human Behav 74, pp.139-151, Sep. 2017

[2] M. Botella, F. Zenasni, and T. Lubart, "What Are the Stages of the Creative Process? What Visual Art Students Are Saying.," Frontiers in Psychology, vol. 9, 2018, Accessed: Nov. 01, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02266

[3] T. J. McKeeff and F. Tong, "The Timing of Perceptual Decisions for Ambiguous Face Stimuli in the Human Ventral Visual Cortex," Cerebral Cortex, vol. 17, no. 3, pp. 669–678, Mar. 2006, doi: 10.1093/cercor/bhk015.

[4] T. Murata, T. Hamada, T. Shimokawa, M. Tanifuji, T. Yanagida, "Stochastic Process Underlying Emergent Recognition of Visual Objects Hidden in Degraded Images," PLOS ONE 9(12), e115658, Dec. 2014

[5] R. Ratcliff and J. N. Rouder, "Modeling Response Times for Two-Choice Decisions," Psychol Sci, vol. 9, no. 5, pp. 347-356, Sep. 1998

[6] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," presented at ICLR, Sep. 2020

[7] K. Hosoda, K. Nishida, S. Seno, T. Mashita, H. Kashioka, and I. Ohzawa, "It's DONE: Direct ONE-shot learning with quantile weight imprinting," arXiv:2204.13361, Jun. 2022

[8] N. Eshed, “Novelty Detection and Analysis in Convolutional Neural Networks,” May 2020, doi: 10.7298/7273-6v18.

[9] K. Tsunoda, Y. Yamane, M. Nishizaki, and M. Tanifuji, “Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns,” Nat Neurosci, vol. 4, no. 8, Art. no. 8, Aug. 2001, doi: 10.1038/90547.

[10] M. Tsodyks, T. Kenet, A. Grinvald, and A. Arieli, “Linking Spontaneous Activity of Single Cortical Neurons and the Underlying Functional Architecture,” Science, vol. 286, no. 5446, pp. 1943–1946, Dec. 1999, doi: 10.1126/science.286.5446.1943.

[11] P. Lansky, P. Sanda, and J. He, “The parameters of the stochastic leaky integrate-and-fire neuronal model,” J Comput Neurosci, vol. 21, no. 2, pp. 211–223, Oct. 2006, doi: 10.1007/s10827-006-8527-6.

[12] S. Bitzer, J. Bruineberg, and S. J. Kiebel, “A Bayesian Attractor Model for Perceptual Decision Making,” PLOS Computational Biology, vol. 11, no. 8, p. e1004442, Aug. 2015