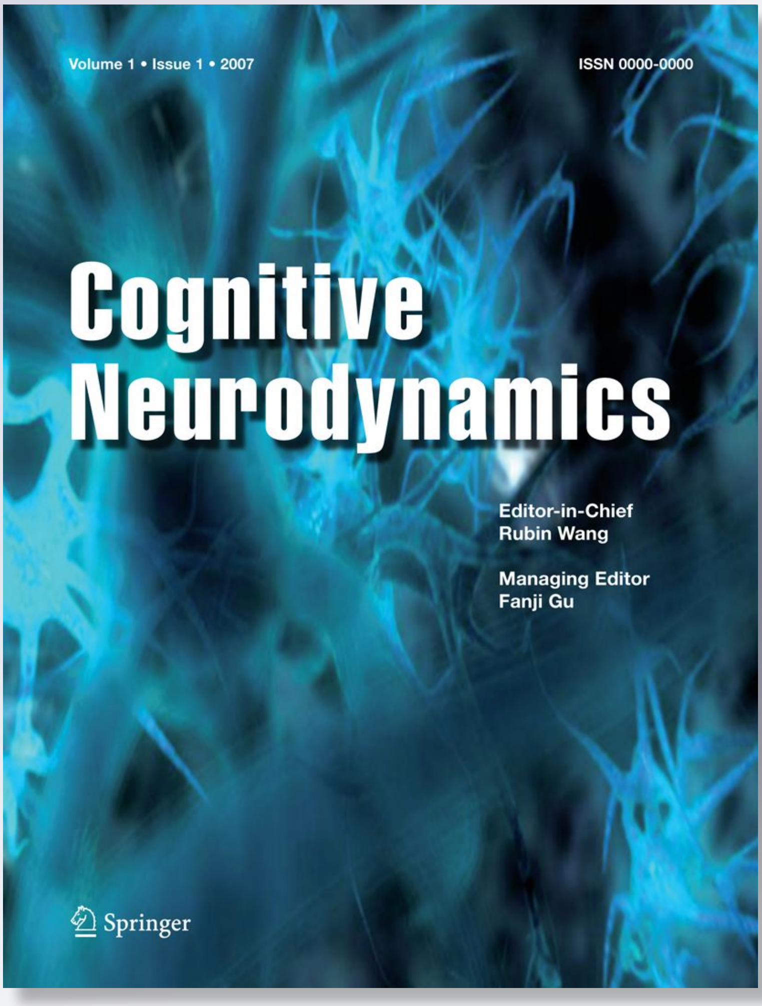
Visual one-shot learning as an ‘anti-camouflage device’: a novel morphing paradigm

Tetsuo Ishikawa & Ken Mogi

Cognitive Neurodynamics

ISSN 1871-4080

Cogn Neurodyn
DOI 10.1007/s11571-011-9171-z



Springer

Cogn Neurodyn
DOI 10.1007/s11571-011-9171-z

RESEARCH ARTICLE

Visual one-shot learning as an ‘anti-camouflage device’: a novel morphing paradigm

Tetsuo Ishikawa · Ken Mogi

Received: 10 March 2011 / Revised: 13 August 2011 / Accepted: 19 August 2011

Abstract Once people perceive what is in the hidden figure such as Dallenbach’s cow and Dalmatian, they seldom seem to come back to the previous state when they were ignorant of the answer. This special type of learning process can be accomplished in a short time, with the effect of learning lasting for a long time (visual one-shot learning). Although it is an intriguing cognitive phenomenon, the lack of the control of difficulty of stimuli presented has been a problem in research. Here we propose a novel paradigm to create new hidden figures systematically by using a morphing technique. Through gradual changes from a blurred and binarized two-tone image to a blurred grayscale image of the original photograph including objects in a natural scene, spontaneous one-shot learning can occur at a certain stage of morphing when a sufficient amount of information is restored to the degraded image. A negative correlation between confidence levels and reaction times is observed, giving support to the fluency theory of one-shot learning. The correlation between confidence ratings and correct recognition rates indicates that participants had an accurate introspective ability (metacognition). The learning effect could be tested later by verifying whether or not the target object was recognized quicker in the second exposure. The present method opens a way for a systematic production of “good” hidden figures, which can be used to demystify the nature of visual one-shot learning.

Keywords Visual one-shot learning · Hidden figure · Fluency theory · Metacognition · Morphing

Introduction

Studying various modes of visual perception provides a salient tool for clarifying certain aspects of awareness and consciousness. Considerable cognitive efforts are needed to perceive surroundings in dim light by using scotopic vision, as color information is useless and spatial resolution is much lower than usual. To segregate the figure from its ground is difficult in these impoverished contexts. Mammals, birds and also insects need to judge shapes of objects by, for example, perceiving illusory contour (Nieder 2002). The ability to perceive illusory contour in partial occlusion or in dimly lit situation such as under the moonlight is biologically adaptive in natural surroundings, as it is advantageous to be able to detect species which mimic their environmental patterns (e.g., felid’s camouflage patterns, Allen et al. 2011) as quickly as possible to flee from predators or to target prey. A visual system with such ability to uncover concealed objects is called an “anti-camouflage device” (Ramachandran 1987). Originally the term “anti-camouflage device” referred only to illusory contour perception. The concept is also applicable to the situation of seeing hidden figures such as the grayscale picture of a cow (Dallenbach 1951) and the two-tone image of a Dalmatian dog (Gregory 1970). When people have an initial look at these hidden figures, they can see nothing but black and white meaningless patterns. But once they realize what is in the figure, a rapid perceptual learning occurs

T. Ishikawa (✉)
Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, G3-46, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8502, Japan
e-mail: tetsuoishikaha@gmail.com

T. Ishikawa · K. Mogi
Sony Computer Science Laboratories, Inc., Takanawa Muse Bldg, 3-14-13 Shinagawa-ku, Tokyo 141-0022, Japan

and is completed in a very short time. The dramatic transition from an unconscious impasse to a conscious epiphany is thought to be a special type of learning called visual one-shot learning (Mogi et al. 2005; Mogi and Tamori 2006, 2007; Giovannelli et al. 2010), or “Aha!” experience (Gick and Lockhart 1995). During such an experience of insight (Bowden et al. 2005), it is thought that synaptic connectivities are changed rapidly to form a new association (Hebb 1949). Therefore, one-shot learning associates meaningless patterns to meaningful understanding to gain a new concept unknown before the learning. It is differentiated from mere memory formation: encoding, retaining, and retrieval of well-known objects or things. This learning effect is long lasting. It is also called the Eureka effect (Ahissar and Hochstein 1997). Two-tone degraded images of human face are named Mooney faces (Mooney 1957). In the insightful moment when subjects perceive Mooney faces, neural synchronization spreads all over the brain, which lasts for about 100 ms (Rodriguez et al. 1999). Such synchronization is thought to be related to some of the Gestalt rules and feature binding (Singer 2009) or a mechanism for transient functional neurocognitive connectivity (Werner 2009). In general, when “Mooney” objects (i.e., bi-level quantized images of various objects) and their original grayscale photographs are presented alternately, activities of inferior temporal and parietal regions are enhanced (Dolan et al. 1997). Activities in the early retinotopic cortex and foveal confluence are modulated by top-down interpretation as well as the ventral visual stream and the lateral occipital complex (Hsieh et al. 2010). Repetitive transcranial magnetic stimulations (rTMS) over the parietal cortex during presentation of the undegraded images disrupt the identification of the degraded counterparts 30 min later (Giovannelli et al. 2010). The activation of left amygdala predicts memory performance 1 week later in a similar paradigm (Ludmer et al. 2011), suggesting the importance of emotional aspects of one-shot learning.

The abrupt realization of the hidden figure provides a robust experimental tool to investigate certain aspects of conscious visual perception in its systematic and temporal richness. Historically, several psychological tests using the aforementioned types of hidden figures have been developed to study Gestalt perception such as completion or closure: Street Gestalt completion test (Street 1931), a new closure test (Mooney and Ferguson 1951), Gestalt completion task and Snowy pictures task (Ekstrom et al. 1976), and Waterloo Gestalt closure task (Bowers et al. 1990). In many cases, however, there is a lack of specific description about how to create these figures. In the case of Waterloo Gestalt closure task, there is a short statement that artists drew a series of stimuli based on their experience. The famous hidden figures of cow (Dallenbach 1951) and Dalmatian (Gregory 1970) are actually photographs: the former was a collection of Leo Potishman and the latter was taken by the photographer, R. C. James. The conditions for shooting these photos and other details are not known. A simple quantisation of photographic images almost always results in either too obvious or difficult to perceive images and do not induce insight (Mogi and Tamori 2006, 2007). It is still unknown how to make “good” hidden figures systematically without manual retouch based on heuristics. Repetitive usage of the same hidden figure is essentially invalid, as each hidden figure can be only effective to a naïve subject. Thus, there is a shortage of a battery of controlled stimuli to be used in an experiment. Note that the stimuli used in almost all of the preceding imaging studies were too difficult for subjects to recognize by themselves. An often used practice, instead of waiting for realization on their own, is to present pairs of the problem (the degraded two-tone picture) and the solution (the original grayscale or color picture) alternately to provide an answer directly and immediately (“alternate presentation paradigm”, Dolan et al. 1997; Hsieh et al. 2010; Giovannelli et al. 2010; Ludmer et al. 2011). However, learning processes with cognitive effort to try to solve a problem without seeing the hint or answer are impaired in the alternate presentation paradigm. These studies therefore focused not on spontaneous (unsupervised) learning but on forced (supervised) learning. An induced insight is different from a spontaneous one (Ludmer et al. 2011). The unaided perception of “good” hidden figures like Dallenbach’s cow or Gregory’s Dalmatian is concerned with the latter, where learning process takes some time, from a few seconds to a few minutes or more. In some cases the realization takes place after quite a long time, e.g., hours to days. Practically finite experimental time makes it difficult to adopt the free exploration paradigm without time limit.

Here we present a novel procedure which enables production in quantity of hidden figures to clarify the behavioral characteristics of unsupervised visual one-shot learning. By morphing “Mooney” objects with the original grayscale images, figures of varied perceptual difficulties were produced. As a “happy medium” of previous paradigms, i.e., the alternate presentation (the answer is presented immediately) and the static presentation (presentation of a hidden figure as still image for a prolonged time), the technique presented here sets out to morph a hidden figure and its solution with intermediate images between them. The morphed sequence of images facilitate the subject’s quasi-spontaneous one-shot learning in a short time.

We expected that the percentage of perceived responses indicated by button press before the end of the movie would be high enough in the present morphing method. The correct rate was predicted to be considerably larger than that of in the conventional static paradigm. Features of individual stimuli (reaction time, morphing levels necessary for perceiving objects, and correct rates) and their interrelationship were investigated. When the correct rates are on a comparable level, reaction time is usable as an index of the degree of difficulty. The spectrum of reaction time for individual movies is supposed to be broader, if the multifariousness of the visual world is reflected in the stimuli.

The fluency theory (Oppenheimer 2008) has been applied to insight (Topolinski and Reber 2010). It predicts that when a certain cognitive process is executed fluently, the confidence and belief of truth about the process become high, independently of the objective truth. The fluency theory applied to visual one-shot learning would predict that the reaction time as an objective index of fluency must be correlated with confidence levels in a negative manner. We also investigated the effect of repetitive presentations of the same stimuli to confirm the accomplishment of visual one-shot learning. Once one-shot learning has occurred, the learning effect is found to be long-lasting. The comparison between the first and the second exposures would confirm the basic "once and for all" nature of "one-shot" learning. If the one-shot learning has an adaptive function, reaction times should be shorter in the second time compared to the first. The confidence levels in the second exposure should be larger than that for the first time. Finally, we examined the relationship that bridges the objective and subjective aspects of one-shot learning. A positive correlation between the objective correct rates and the subjective confidence levels would suggest that the subjects could judge their internal states accurately by introspection, suggesting proper metacognitive abilities.

Through the variation of morphing and temporal transition parameters, we constructed an external means to control the perception of figures in the conscious domain. Morphing provides a means of dynamically probing into the cognitive processes of one-shot learning, as opposed to the typical static approaches of the conventional hidden figure research. The analysis of results shed light on the interaction of the search and memory recall processes involved.

Methods

Subjects

Eight healthy adult volunteers (4 males, aged 25–31 years; mean age 28.8 years) took part in the experiment. Participants were all right-handed, and had normal or corrected-to-normal vision. They were all native Japanese speakers. Instructions by the experimenter and verbal reports by participants were provided in Japanese.

Materials

Thirty-two grayscale photographs (300 × 300 pixels) of commonly familiar objects were blurred with Gaussian filter (radius = 3 pixels) and binarized to make ambiguous two-tone images. Movies were made by the computer software MorphX 2.9.5 (Norrkross Software) to display the whole morphing sequence in 1% morphing transitions, from the degraded image (e.g., Fig. 1 leftmost column) to the blurred grayscale original (e.g., Fig. 1 rightmost column). The frame rate was 5 fps (1% morphing every 200 ms). The movie sequence consisted of 101 frames, with a total duration of 20.2 s. The subjects sat in a comfortable position at a viewing distance of 60 cm from the computer display (Apple 13-inch MacBook). Stimuli (10° × 10°) were presented against a black background. We confirmed through enquiries after the sessions that all subjects saw all stimuli for the first time at the first exposure in this experiment.

The objects were selected from the list of "A Standardized Set of 260 pictures" (Snodgrass and Vanderwart 1980) which contained some familiar categories: insects, musical instruments, vegetables, fruits, animals, vehicles, carpenter's tools, etc.

Procedure

Subjects self-pacedly started watching movies by clicking the mouse to extinguish a fixation cross and tried to perceive what was in the movie. They were instructed to stop the movie by clicking the mouse button and to shut their eyes when they recognized the object in it or when the movie was finished without recognition. The subjects were asked to close their eyes at the moment of realization, in order to prevent the inspection of the freeze-frame which remained on screen until the experimenter recorded the frame number of the movie and refreshed the display for the next trial. The subjects then verbally reported the name of the object and their "sureness" in a 11-point (0–10) scale. With the experimenter's verbal cue ("ready"), they opened their eyes and proceeded to the next trial (Fig. 2).

The experiment was conducted twice after the practice session (two trials): In the initial exposure, thirty movies were played in a pseudorandom order. After a few minutes break, the same set of stimuli were presented in the same order in the second exposure. The interval for the presentation of the same stimulus was more than 15 min.

Statistical analysis

When no response occurs before the end of the movie, "reaction time" cannot be defined. In statistics, such a time-out trial is called a censored observation. To deal with the censored data properly, a survival (time to event) analysis

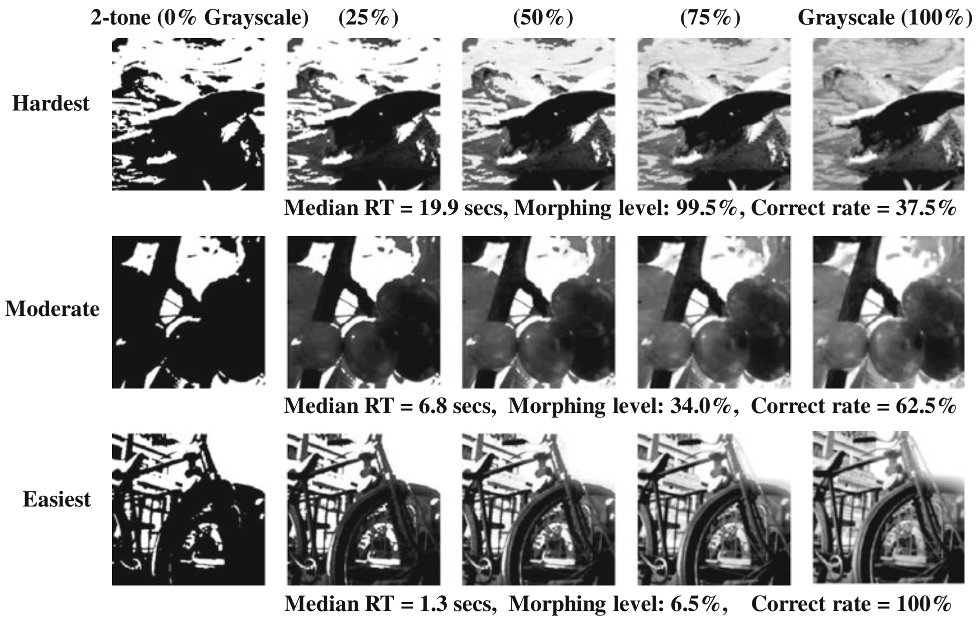


Fig. 1 Typical frame examples extracted from the morphing movies. *Top row*: the hardest to perceive movie ("Alligator"), *middle row*: moderate difficulty movie ("Cherry"), and *bottom row*: the easiest to perceive movie ("Bicycle"). The median reaction time, median morphing level and mean correct rate of these movies are set down underneath each figures. Difficulties of movies are assessed by the median reaction time, or the median morphing level. Time required for a half of participants to perceive is usable as index of difficulty unless more than a half of them time out (which was not the case for all stimuli in the present study). Therefore, the hardest movie meant the longest/highest RT/morphing level. The easiest one would be the stimulus with the shortest/lowest RT/morphing level. Moderate stimulus had an intermediate difficulty, i.e., the 15th RT/morphing level among all 30 stimuli

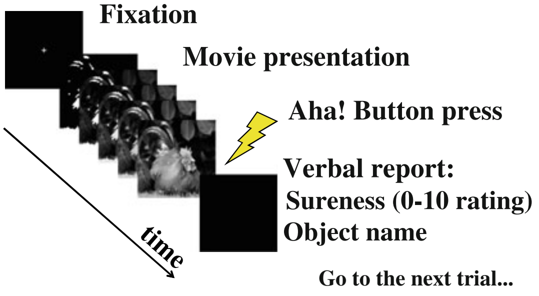


Fig. 2 The time course of a single trial is depicted in illustrative manner. Until the moment the subject presses the button, a morphing movie is played in an ascending fashion, from 0% of *grayscale* (two-tone) frame to 100% of *grayscale* one. The replay speed is 5 frames per second. All movies are constructed from 101 frames, with a maximum duration of 20.2 s. After the key press or at the end of the movie, the participant were asked to report verbally the degree of confidence and what was perceived. For the movie in this figure, the correct answer was "Chicken"

with the Kaplan–Meier method (Bland and Altman 1998), which has been applied to the analysis of an another domain of insight, i.e., matchstick problem (Chi and Snyder 2011), was conducted. A logrank test was used to compare the time to event curve. Regardless of whether participants responded or not in each trial, the answer for the movie (object name) and confidence rating were available. Hence the two-tailed paired/one-sample *t*-test, Pearson's correlation analysis, or two-way analysis of variance (ANOVA) was applied to the correctness of the answer (correct rate) and confidence data. The significance level was set to 0.05 for all statistical tests. Multiple comparisons are corrected by the Bonferroni method.

Results

The earlier frames of morphing movies were degraded and ambiguous so that the subjects found it hard to perceive what was in the frame. However, as the movie frames gradually got close to the original grayscale picture in the movie sequence, they could perceive something (either correctly or incorrectly) in the frame. Key press responses before reaching the end of morphing movies were observed in 92.0 ± 5.5% (mean ± SD) of trials in the first presentation. Significantly more reactions (95.4% ± 5.5% of trials) occurred in the second presentation ($t(7) = 2.65$, $P = 0.03$). Participants answered faster in the second exposure than the first exposure (first quantile: 15.8%, median: 35.5%, the third quantile: 69.3% for the first time and first quantile: 9.0%, median: 18.0%, the third quantile: 42.3% for the second time (logrank test: $\chi^2(1) = 20.5$, $P = 5.9\text{e–}6$). Verbal reports of the object name were judged by the experimenter to be either correct (correct

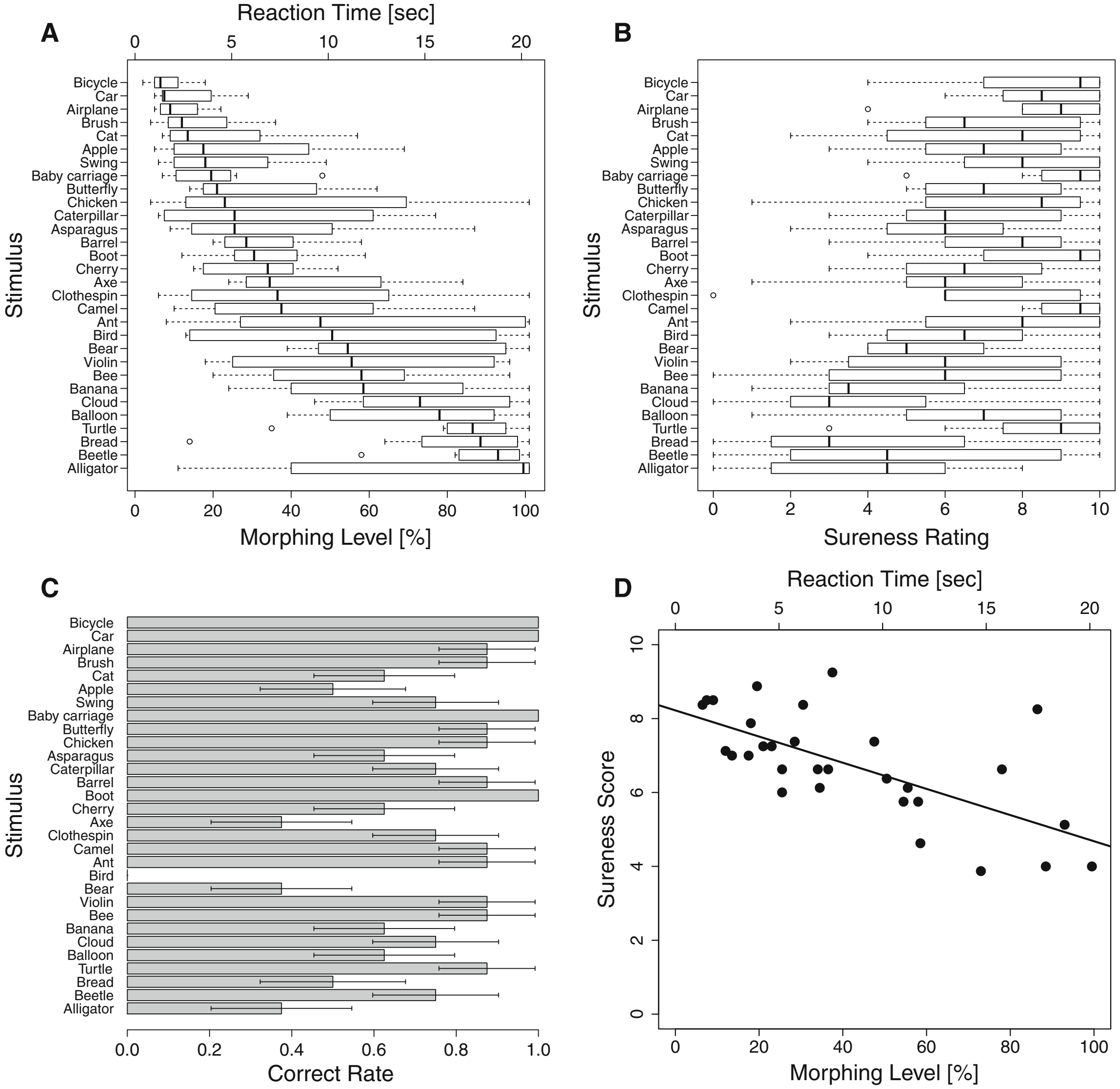


Fig. 3 **a** The reaction times (*upper abscissa*) and morphing levels (*lower abscissa*) of individual movies in the first exposure are indicated in a *box-whisker plot*. The *vertical bold line* in a *box* indicates the median (i.e., the second quartile, Q2). The *left* and *right* sides of the *box* show the first quartile (Q1) and the third quartile(Q3), respectively. The *left* and *right* end of the whisker correspond to the minimum and maximum of the data in 1.5 times of interquartile range (IQR = Q3 − Q1) from Q1 or Q3, respectively. *Blank circles* show the data point outside the range of [Q1 − 1.5 IQR, Q3 + 1.5 IQR]. The movies are sorted by the ascending order of the median of response time from the *top* to the *bottom* of the graph. **b** The sureness scores of individual movies are shown in a *box-whisker diagram*. The alignment sequence of stimuli is the same as **a**. **c** The correct rate of answer object name is plotted for each stimulus in a *barplot*. The *error bars* indicate standard error of the mean. **d** The correlation between median morphing levels and mean sureness scores was significant (Pearson's correlation coefficient $r = -0.67$, $P = 4.9\text{e}{-}05$)

answer, CA) or incorrect (incorrect answer, IA, including a non-perceived trial). Taking into account of whether the answer was correct or not, 72.5 ± 9.7% of the movies were correctly perceived in the first exposure and significantly more 81.7 ± 6.2% of the movies were recognized aright in the second exposure ($t(7) = 3.67$, $P = 0.008$).

To elucidate the features in a variety of movies, the following analysis was conducted on individual stimuli using the data of the first exposure. Distributions of the morphing levels defined as percentages of containing grayscale picture (or RTs) at the time when participants perceived the hidden object were calculated for each movie and sorted by median RTs (Fig. 3a). In many cases, the distribution was not symmetric and had a fat tail in the right side. The sureness score (Fig. 3b) and the correct rate (Fig. 3c) were plotted in the same order as in Fig. 3a. The

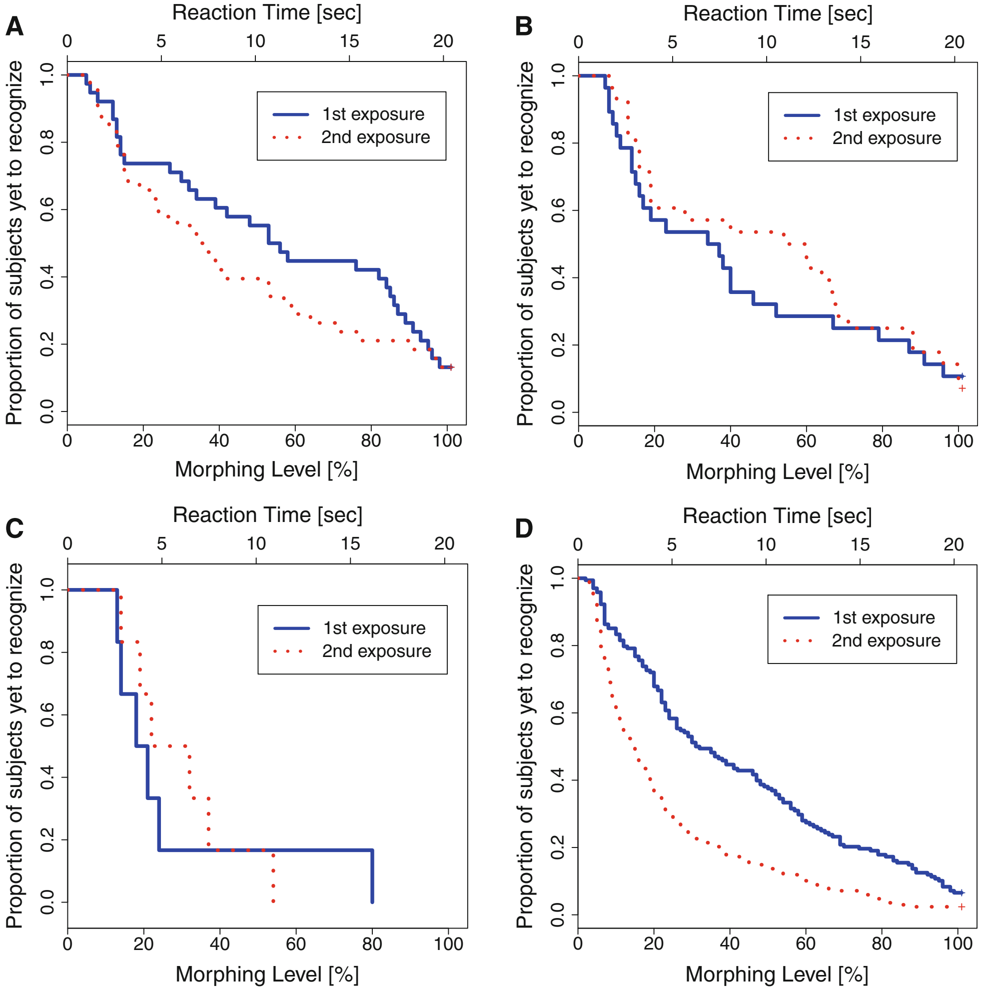


Fig. 4 The time to *event curves* using Kaplan–Meier (KM) method are compared on exposure time. There are four conditions because of the combination of correctness (IA or CA) for each exposure time (1st or 2nd exposure). **a** IA–IA condition: when answers were incorrect in the first and the second exposure, the *KM-curves* are not significantly different (logrank test: $\chi^2(1) = 0.7$, $P = 0.399$). **b** IA–CA condition: when a correct perception occurred not for the first time but for the second time, a logrank test did not reveal significant difference between the first and the second exposure ($\chi^2(1) = 0.3$, $P = 0.563$). **c** CA–IA condition: when the answer changed to incorrect in the second exposure although the answer was correct in the first exposure, the *time to event curves* did not differ from each other ($\chi^2(1) = 0.1$, $P = 0.741$). **d** CA–CA condition: only in this condition, or when the answer in the second presentation was the same as the first correct one, the *KM-curves* of the second exposure dipped faster than that of the first one ($\chi^2(1) = 35.2$, $P = 2.94\text{e}{-}09$)

longer the subjects took to perceive, the less were confidence levels reported by them. (Fig. 3d). There was a significant negative correlation between median RTs (or morphing levels) and the mean sureness score ($r = -0.67$, $P = 4.9\text{e-}05$). The mean confidence ratings were correlated positively with mean correct rates ($r = 0.56$, $P = 0.0012$). Median RTs (or morphing levels), however, did not correlate with mean correct rates ($r = -0.36$, $P = 0.051$).

To compare the naïve observation (before learning occurred) with the second time one (when a learning might or might not have already occurred), the performance for a particular image was classified into four categories: IA–IA, IA–CA, CA–IA, and CA–CA conditions, based on the correctness of the subject's answer in the first and the second exposures. The probabilities of stimuli categorized for each conditions were IA–IA: 15.8 ± 4.9%, IA–CA: 11.7 ± 8.3%, CA–IA: 2.5 ± 3.2% and CA–CA: 70.0 ± 12.0% of stimuli (mean ± SD). Time to event (IA or CA) curve analysis was conducted for the four conditions (Fig. 4). No significant difference was found between the curves of the first presentation and the second one in the IA–IA, IA–CA and CA–IA condition (logrank test : all *P*s >0.05, Fig. 4a to c). On the other hand, only when the condition was the CA–CA one, the reaction time (or morphing level) was smaller in the second exposure than in the first exposure ($P < 0.001$, Fig. 4d).

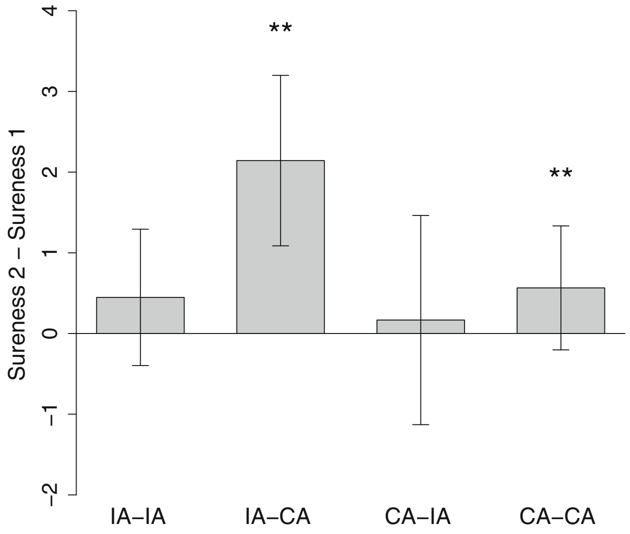


Fig. 5 Difference of sureness ratings between the first presentation and the second one. In the IA–CA condition (two-tailed one-sample t test: $t(27) = 3.55$, $P = 0.0014$) and the CA–CA condition ($t(167) = 3.16$, $P = 0.0019$), the differences were significantly different from zero. In other words, sureness ratings were larger in the second exposure than in the first exposure. In the IA–IA condition ($t(37) = 1.08$, $P = 0.287$) and CA–IA condition ($t(5) = 0.10$, $P = 0.921$), a t-test did not reveal significant difference of sureness scores between the exposure time. *Error bars* show the standard error of the mean. *Asterisks* show the significance of P values adjusted in multiple comparison: ** $P < 0.01$

Differences of sureness scores (i.e., $\Delta S = \text{Sureness } 2 - \text{Sureness } 1$) indexed the change of confidence in the second exposure from the first time. ΔS was plotted for four experimental conditions (Fig. 5). If there is no change of sureness ratings, ΔS must be zero. To test the null hypothesis, one-sample t-tests were carried out. Significant differences from zero were yielded in the IA–CA and CA–CA conditions (Ps < 0.01). For the rest, i.e., in the IA–IA and CA–IA conditions, the confidence levels were statistically unchanged (Ps > 0.05).

There are two parameters which characterize the performance: subjective confidence and objective accuracy. The relationship between them is described in a line plot (Fig. 6). The sureness data were summed up into six levels, because of the numbers of lower sureness scores were too small. To clarify the effect of sureness and the number of exposures on the correct rate, a two-way ANOVA was conducted. It showed the significant main effect of sureness ($P < 0.001$), whereas the effect of exposure ($P > 0.05$) and interaction ($P > 0.05$) were not significant. A correlation analysis was performed to examine the linear relationship between sureness and correct rate. Correct rate was positively correlated with sureness in both of the exposure times: Pearson's correlation coefficient $r = 0.93$ ($P = 0.007$) for the first exposure and $r = 0.96$ ($P = 0.003$) for the second exposure.

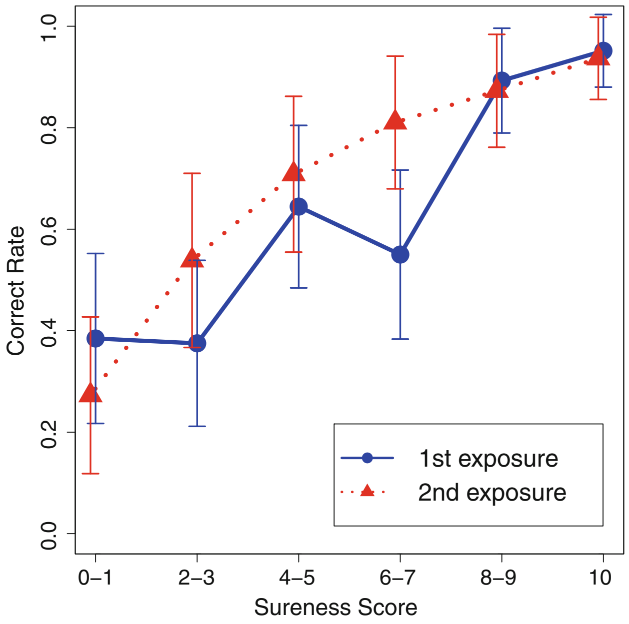


Fig. 6 Correct rate as a function of sureness scores and number of exposures is shown in a *line graph*. The *error bars* represent the standard error of the mean. Two-way ANOVA (Sureness × Exposure) revealed that the main effect of sureness was significant ($F(5, 468) = 20.88$, $P = 7.47\text{e}{-19}$). The effect of exposure ($F(1, 468) = 0.398$, $P = 0.097$) and the interaction ($F(5, 468) = 0.32$, $P = 0.051$) were not significant. Correct rate and sureness rating is positively correlated in both the first exposure (Pearson's correlation coefficient $r = 0.93$, $P = 0.007$) and the second exposure ($r = 0.96$, $P = 0.003$)

Discussion

The participants "perceived" some figures (irrespective of correctness) and responded in 92.0 ± 5.5% of stimuli in the first exposure before the end of the movie. Considering only the CA cases, 72.5 ± 9.7% of stimuli were perceived correctly for up to 20.2 s of presentation in the first trial. It is significantly larger than that of the previous alternate presentation paradigm: for example, the spontaneous recognition rate was 27 ± 3% for up to 10 s in experiment 2 of Ludmer et al. (2011). Moreover, in the conventional static presentation paradigm, where participants kept on searching a stationary hidden figure (Dallenbach 1951; Gregory 1970; Mogi et al. 2005; Mogi and Tamori 2006, 2007), the spontaneous recognition rate was relatively low. For instance, the correct rates of the famous hidden figures were 6.2% for "Cow" and 12.0% for "Dalmatian" after 300 s in a series of mega-lab experiments with the number of subjects $N = 113$ (Mogi, Sekine and Tamori, unpublished data). In comparison, more than 70% of the stimuli were perceived correctly within a few tens of seconds in our study. Hence, it is suggested that the present study should provide a suitable method to investigate spontaneous one-shot learning rather than induced insight.

In the first exposures, the morphing levels necessary for cognition and the ratings for sureness for individual stimuli were negatively correlated. It is possible that the decline of confidence level as RTs (or morphing levels) became larger due to the decrease in the correct rate. However, this was not the case, because the RTs (or morphing levels) and correct rate were not significantly correlated in the first presentation. Therefore, the results support the fluency hypothesis of insight (Topolinski and Reber 2010), which predicts that a more fluent processing (i.e., the shorter RT) induces a more confident feeling in the subjects.

The morphing levels, or response time lengths can be thought of as indicators of the degree of difficulty for recognition of the objects. A higher sureness score would indicate a lower degree of difficulty, and hence a higher fluency. Individual stimuli used in the present study indicated a wide spectrum of RTs or morphing levels (Fig. 3). In turn, the battery of stimuli used reflected a wide range in the degree of difficulty. Neither too difficult nor too easy problems, but problems with a right degree of difficulty, are proposed to be one of the prerequisites for insight (Hebb 1949). A robust set of hidden figures with appropriate difficulty levels, (i.e. “good” hidden figures) is needed to demystify the nature of one-shot learning, because a hidden figure can be used only once per a subject in an experiment. By applying the present method, new hidden figures with various difficulty levels can be made in a systematic manner.

The reaction time (or morphing level) of recognition at the second trials was smaller (or lower) than the first trial only in the CA–CA cases, suggesting that one-shot learning leads to an appropriate prior knowledge to be used in the next trial when participants perceived correctly in the naïve state. Once learning is accomplished in the right direction, it should be adaptive to save time in grasping the situation instantaneously using the “anti-camouflage” device to detect hidden objects without additional exploration.

Familiarity of objects’ names were ensured by using the standardized stimuli set (Snodgrass and Vanderwart 1980). A complete list of all objects used in the experiment is provided in Fig. 3. Note that the initial letter of most words was A, B or C, apart from a few exception. In Japanese, however, names of objects started with varied phonemes. Therefore there was no chance that a certain type of phonological priming effect occurred accidentally.

The same stimuli order used in the first and second exposure might have led to the possibility of contextual effect or trial-to-trial dependencies. If such an effect occurred generally, time to event curves would have been different between the first and second exposures. However, this was not always the case. The change of time to event curve was only observed in the CA–CA condition. Although this analysis alone does not necessarily signify that there was no trial-to-trial dependencies of any kind, it does suggest that contextual effects, if any, had a limited impact on the main results.

The average rating of sureness in the second trial was higher than that in the first one in the IA–CA and the CA–CA conditions. For the former case, it is possible that the participants were aware of the incorrectness of their answers at the first exposure, reflected in the change to the correct answer at the second trial. For the latter case, it is possible that they became more confident because of the repetition of subjective feeling of certainty, or fluency.

There was a positive correlation between the rating of sureness and the correct rate. Thus, the subjective evaluations of confidence were consistent with the objective performance, providing a measure for metacognition. The scene consistency between the perceived figure and ground that ensures the accuracy of object recognition (Davenport and Potter 2004) may subserve the appropriate metacognition.

Humans can learn to recognize difficult Mooney faces after a brief (about 5 s) exposure of unambiguous counterparts. Neurons in the inferior temporal cortex of primates exhibit an enhancement reflecting the neural substrate of the rapid perceptual learning (Tovee et al. 1996). A patient with anterograde amnesia (Korsakoff’s syndrome) possibly caused by a medial temporal lobe damage showed no evidence of perceptual learning (Ramachandran 1995). Although his response time for the recognition of hidden figures was relatively normal, there was no reduction in latency when the same figures were exposed more than once. These evidences suggest that the inferior temporal visual cortex is involved in visual one-shot learning.

In sum, we proposed a novel paradigm to create new hidden figures with a broad spectrum of difficulty levels in a systematic manner by using a morphing technique. Through gradual changes from a blurred and binarized two-tone image to the blurred grayscale image of the original photograph including objects in a natural scene, spontaneous one-shot learning could occur at a certain stage of morphing when a sufficient amount of information is restored to the degraded image. A negative correlation between confidence levels and the reaction time or morphing level necessary for perceiving objects is consistent with the fluency theory of insight (Topolinski and Reber 2010) in the domain of visual one-shot learning. A strong correlation between the confidence ratings and the correct recognition rate indicates that the subjects had accurate metacognition. The comparison between the first and the second exposures confirms the basic “once and for all” nature of “one-shot” learning. The learning effect could be tested later by checking whether the target object is recognized quicker in the second exposure. The results reported here suggest a potential relationship between the underpinning mechanism of one-shot learning and the neural substrate of short-term (Rutishauser et al. 2010) or long-term (Ludmer et al. 2011) memory formation and retrieval. Thus, we note that memory components of one-shot learning is the result of the learning and not the other way around. Further sets of research are needed to clarify the reason why and how one-shot learning leads to long lasting memories based on just a single trial. The present method paves a new path for the production of "good" hidden figures in large quantities, which can be used to eventually demystify the nature of visual one-shot learning.

Acknowledgments We thank Tomomitsu Herai, Eiichi Hoshino, Shinichi Nozawa, Ayako Onzo, Kei Omata, Takayasu Sekine, Tamami Sudo, Tsugumi Takano, Fumiko Tanabe, Fumihiko Taya, Mayumi Toshima, and Toru Yanagawa, for their helpful discussion. We also thank three anonymous reviewers for their insights.

References

Ahissar M, Hochstein S (1997) Task difficulty and the specificity of perceptual learning. Nature 387:401–406

Allen WL, Cuthill IC, Scott-Samuel NE, Baddeley R (2011) Why the leopard got its spots: relating pattern development to ecology in felids. Proc Biol Sci 278(1710):1373–1380

Bland JM, Altman DG (1998) Survival probabilities (the Kaplan-Meier method). BMJ 317(7172):1572

Bowden EM, Jung-Beeman M, Fleck J, Kounios J (2005) New approaches to demystifying insight. Trends Cogn Sci 9(7):322–328

Bowers K, Regehr G, Balthazard C, Parker K (1990) Intuition in the context of discovery. Cogn Psychol 22(1):72–110

Chi RP, Snyder AW (2011) Facilitate insight by non-invasive brain stimulation. PLoS One 6(2):e16655

Dallenbach KM (1951) A picture puzzle with a new principle of concealment. Am J Psychol 64:431–433

Davenport JL, Potter MC (2004) Scene consistency in object and background perception. Psychol Sci 15(8):559–564

Dolan RJ, Fink GR, Rolls E, Booth M, Holmes A, Frackowiak RS, Friston KJ (1997) How the brain learns to see objects and faces in an impoverished context. Nature 389:596–599

Ekstrom RB, French JW, Harman HH, Dermen D (1976) Manual for kit of factor-referenced cognitive tests. Princeton Educational Testing Service, New Jersey

Gick ML, Lockhart RS (1995) Cognitive and affective components of insight. In: Sternberg RJ, Davidson JE (eds) The nature of insight. MIT Press, Cambridge, pp 197–228

Giovannelli F, Silingardi D, Borgheresi A, Feurra M, Amati G, Pizzorusso T, Viggiano MP, Zaccara G, Berardi N, Cincotta M (2010) Involvement of the parietal cortex in perceptual learning (Eureka effect): an interference approach using rTMS. Neuropsychologia 48(6):1807–1812

Gregory RL (1970) The intelligent eye. Weidenfeld & Nicolson, London

Hebb DO (1949) The organization of behavior: A neuropsychological theory. Wiley, New York

Hsieh PJ, Vul E, Kanwisher N (2010) Recognition alters the spatial pattern of FMRI activation in early retinotopic cortex. J Neurophysiol 103(3):1501–1507

Ludmer R, Dudai Y, Rubin N (2011) Uncovering camouflage: amygdala activation predicts long-term memory of induced perceptual insight. Neuron 69(5):1002–1014

Mogi K, Tamori Y (2006) Making good hidden figures. Perception 35 ECVP Abstract Supplement 35

Mogi K, Tamori Y (2007) Temporal characteristics of visual one-shot learning. Perception 36 ECVP Abstract Supplement 51

Mogi K, Sekine T, Tamori Y (2005) Slow and fast processes in visual 'one-shot' learning. Perception 34 ECVP Abstract Supplement 15–16

Mooney CM (1957) Age in the development of closure ability in children. Can J Psychol 11(4):219–226

Mooney CM, Ferguson GA (1951) A new closure test. Can J Psychol 5(3):129–133

Nieder A (2002) Seeing more than meets the eye: processing of illusory contours in animals. J Comp Physiol A Neuroethol Sens Neural Behav Physiol 188(4):249–260

Oppenheimer DM (2008) The secret life of fluency. Trends Cogn Sci 12(6):237–241

Ramachandran VS (1987) Visual perception of surfaces: a biological theory. In: Petry S, Meyer GE (eds) The perception of illusory contours. Springer, Berlin Heidelberg New York, pp 93–108

Ramachandran VS (1995) 2-D or not 2-D?—that is the question. In: Gregory R, Harris J, Heard P, Rose D (eds) The artful eye. Oxford University Press, Oxford New York, pp 249–267

Rodriguez E, George N, Lachaux JP, Martinerie J, Renault B, Varela FJ (1999) Perception's shadow: long-distance synchronization of human brain activity. Nature 397:430–433

Rutishauser U, Ross IB, Mamelak AN, Schuman EM (2010) Human memory strength is predicted by theta-frequency phase-locking of single neurons. Nature 464:903–907

Singer W (2009) Distributed processing and temporal codes in neuronal networks. Cogn Neurodyn 3(3):189–196

Snodgrass JG, Vanderwart M (1980) A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. J Exp Psychol Hum Learn 6(2):174–215

Street RF (1931) A gestalt completion test. Contributions to education 481. Columbia University Teachers College, New York

Topolinski S, Reber R (2010) Gaining insight into the "Aha" experience. Curr Dir Psychol Sci 19:402–405

Tovee MJ, Rolls ET, Ramachandran VS (1996) Rapid visual learning in neurones of the primate temporal visual cortex. Neuroreport 7(15–17):2757–2760

Werner G (2009) Consciousness related neural events viewed as brain state space transitions. Cogn Neurodyn 3(1):83–95