

A normative set of 98 pairs of nonsensical pictures (doodles)

TAKEHIKO NISHIMOTO, TAKASHI UEDA, KAORI MIYAWAKI, AND YUKO UNE
Waseda University, Tokyo, Japan

AND

MASARU TAKAHASHI
Saitama Institute of Technology, Saitama, Japan

Our purpose in the present study is to provide a normative set of nonsensical pictures known as *doodles* and to demonstrate the role of semantic comprehension in facilitating recall of pictorial stimuli. The set consists of 98 pairs of doodles. Experiment 1 standardized these pictorial stimuli with respect to several variables, such as appropriateness of verbal labels, relationship between two doodles, and correct recall. Appropriateness of verbal labels was rated higher for pictures presented in pairs than for pictures presented singly. Experiment 2 used the standardized set of doodles in a recall experiment similar to those of Bower, Karlin, and Dueck (1975) and others. As we expected, semantic interpretation can strongly facilitate recall. Multiple regression analysis showed that several measures had significant power of explanation for recall performance. The full set of norms and pictures from this article may be downloaded from <http://brm.psychonomic-journals.org/content/supplemental>.

Doodles are nonsensical pictures that are very difficult to understand without being given a thematic clue or label. The pictures themselves supply very few interpretive clues; therefore, people rarely report a coherent understanding when they must rely on the standard conventions typically used in viewing illustrations. An example of a pair of doodles is shown in Figure 1. One appropriate interpretation that links these two doodles is the following: “Piano keys and the hair of Beethoven while he is playing the piano.” This phrase or caption makes both pictures and their pairing a meaningful sequence and, in turn, can facilitate associative recall. Evidence for the latter comes from Bower, Karlin, and Dueck (1975), who used doodles in their memory experiments. They used both a free recall task and a cued recall task and found a strong interaction between visual and verbal encoding. That research was among the first to call attention to the potential of doodles as a useful tool for the study of memory and cognition.

Doodles have been used in three ways in psychological experiments: They have served as stimuli for experiments designed to examine interactions between visual and verbal encoding (e.g., Bower et al., 1975; Hayashi & Une, 2004; Kitagami, 2000; Klatzky & Rafnel, 1976; McAninch & Austin, 1993; Rafnel & Klatzky, 1978; Takahashi & Inoue, 2009). Bower et al. found that participants’ recall of individual doodles (i.e., presented individually) was much better when they received interpretations of these pictures during a preceding study phase. Also, participants

who studied pairs of doodles with a linking interpretation (as in Figure 1) showed greater associative recall and matching (recognition) than did participants who received no interpretation.

Second, nonsensical pictures have been used in neuropsychological studies of memory. In a picture memory task, Iidaka et al. (2001) used doodles (adopted from Nishimoto & Takahashi, 1996) in addition to line drawings of concrete objects (adopted from Snodgrass & Vanderwart, 1980) in an fMRI study using young and older adult participants. Doodles functioned as meaningless abstract line drawings, whereas other drawings represented concrete objects; all stimuli occurred in pairs and lacked any additional interpretation. Three conditions for a picture encoding task were involved: pairs of abstract stimuli (two doodles), pairs of concrete-related stimuli (two common objects, such as “ashtray and cigarette”), and pairs of concrete-unrelated stimuli (two common objects, such as “alligator and piano”). Iidaka et al. found that older adults showed reduced activation in the bilateral parieto-temporo-occipital areas with doodles (i.e., in the abstract condition) and that these participants showed reduced activation in the right temporo-occipital area extending to the fusiform gyrus in the concrete-unrelated condition. The young participants had a positive correlation between correct recognition and cortical signal change in the left parahippocampal gyrus and in the right middle frontal gyrus under the concrete-unrelated condition, but the older participants did not. Also, the young participants

T. Nishimoto, nishi@waseda.jp



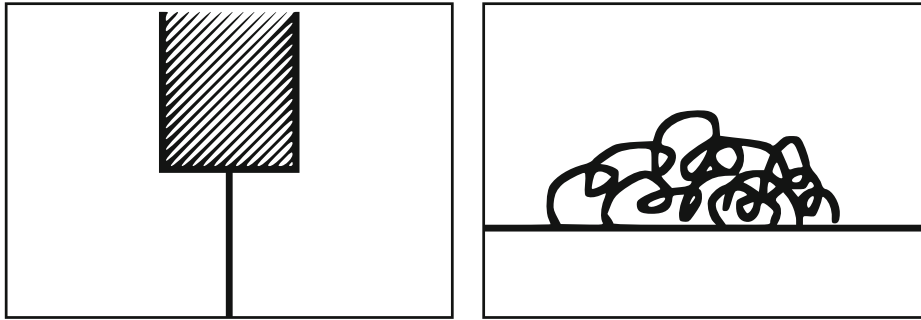


Figure 1. A pair of doodles with the label “Piano keys and the hair of Beethoven while he is playing the piano.”

showed a positive correlation in the right fusiform gyrus, right middle frontal gyrus, and left inferior temporal gyrus with doodles, whereas the older participants did not. That study illustrated the value of doodles as meaningless and abstract stimuli in investigating age-related changes in the neural mechanisms of picture encoding.

Third, in developmental psychology, doodles are often used in false-belief tasks (Barquero, Robinson, & Thomas, 2003; Carpendale & Chandler, 1996; Chandler & Helm, 1984; Doherty & Wimmer, 2005; Lalonde & Chandler, 2002; Perner & Davies, 1991; Saltzman, Strauss, Hunter, & Archibald, 2000). Figure 2 illustrates a sample of pictures shown to children 5–7 years of age (Lalonde & Chandler, 2002). They were asked to describe the full content of a picture on the basis of the label “A ship arriving too late to save a drowning witch.” The original doodle, represented in the upper right drawing in Figure 2, depicts only the bow of a ship and the tip of a witch’s hat.¹ After being presented the full picture, children were allowed to see only

the original doodle and then were asked “What will *X* (i.e., a named classmate) think this is?” This task requires children to bracket their own understanding that the triangles represent the bow and the tip and then to inhibit this dominant information, in order to appreciate that someone could arrive at a different conclusion. It is assumed that this kind of task can yield results that are in line with those produced by other measures of false belief.

Thus, doodles have proven to be useful in various fields of research. The present study is the first attempt to standardize doodles. It is expected that standardized doodles may serve as a useful normative resource for researchers in the fields of memory and cognition.

First, we collected data on the basis of the appropriateness of labels. Although the original labels, such as “piano keys” and “the hair of Beethoven while he is playing the piano,” shown in Figure 1 (labels of left and right doodles, respectively) are humorous and creative, their appropriateness remains subjective. Therefore, the present study was

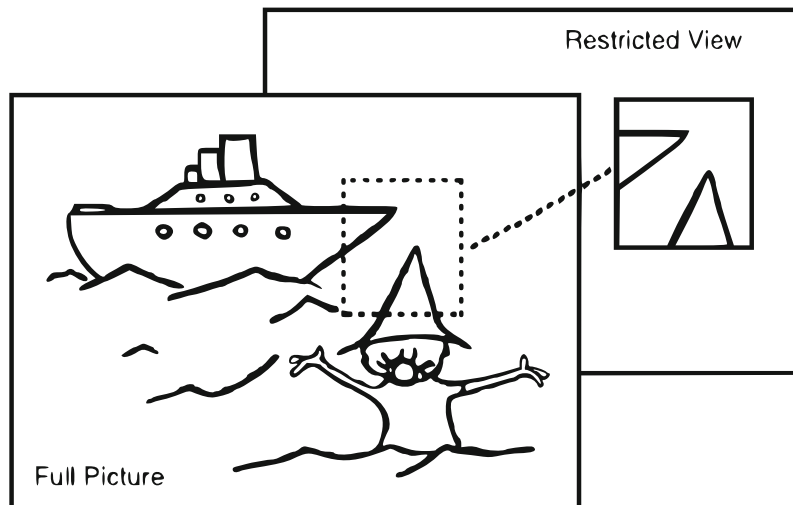


Figure 2. A typical example of doodles (the upper right picture, named “restricted view”) with the label “A ship arriving too late to save a drowning witch.” The lower left picture, named “full picture,” is the interpretative content of the label. From “Children’s Understanding of Interpretation” by C. E. Lalonde and M. J. Chandler, 2002, *New Ideas in Psychology*, 20, p. 169. Copyright 2002 by Elsevier Science Publishers. Reproduced with permission.

designed to collect data that offered a normative measure of the appropriateness of labels for a specific set of doodles.

Second, we collected data on the appropriateness of the labels that were generated or the extent to which a doodle was meaningful. When doodles are presented without labels, we regard them as being “meaningless pictures.” Indeed, when people are asked to interpret doodles without labels, they tend to elicit a diversity of labels, sometimes resulting in response failures, because no ideas come to the respondent’s mind. The more meaningless a doodle is, the more difficult it is for people to comprehend it; therefore, the diversity of labels that are generated is likely to be greater. To evaluate the extent to which a single doodle is meaningful, we presented doodles without labels and asked the participants to name them and to rate the appropriateness of the label that they made. We named this *generated-label appropriateness* (GLA). To evaluate the variability of the labels that were produced, we calculated *H* (described below), reflecting the response disagreements among participants (Snodgrass & Vanderwart, 1980).

Third, we measured perceptions of relationships between right and left doodles of a pair. Many people may find it easy to link two doodles and comprehend them, if they are perceived to be very similar to each other. However, such associations may be spurious, in that they may reflect only the different surface features of a picture or may depend on some fragment of the drawing, which may improve the memory for the picture. Bower et al. (1975) originally emphasized this point. Accordingly, it is important to collect ratings of perceived picture similarity or relationships by presenting pairs of doodles without labels. From a practical perspective, this procedure also provides a baseline measure of recall performance.

In summary, in the present research, we attempt to standardize five basic measures: label appropriateness (LA), GLA, variability of generated labels (*H*), picture relationship (PR) between doodle pairs, and correct recall. We expect that the present doodle set will be useful in the same manner as the normative set of common and concrete objects (Nishimoto, Miyawaki, Ueda, Une, & Takahashi, 2005). The present set consists of 50 pairs of original doodles newly drawn for this study and 48 pairs adopted from Nishimoto and Takahashi (1996).

In Experiment 1, we outline the methodology and present results of these measurements. In Experiment 2, we replicated the experiments of Bower et al. (1975) and demonstrated how doodles can be used advantageously in experiments investigating memory and cognition. The stimuli used by Bower et al. were adopted from Price (1972, which is also accessible on the Web; Vision Office, 1999);² our replication expands the generality of the Bower et al. findings to a different set of doodles.

EXPERIMENT 1

Normative Data for Doodles

Method

Participants. A total of 333 undergraduate and graduate students of Waseda University voluntarily served in Experiment 1 in return

for course credit. Each was assigned to one of four tasks, which are outlined below.

Materials. We collected 359 doodles and corresponding labels that putatively afford an interpretive meaning for the doodle. These pictures were drawn and labeled by undergraduate students taking an introductory psychology course. Among these doodles, 98 pairs were selected according to three criteria: The shape of the doodle is neither too simple nor too complex, the shape of the doodle is unlike other doodles, and the label creator seems to understand and follow the instructions. For each doodle, see Appendix B in the online supplemental materials.

Procedure. The procedure and instructions varied with each of the four tasks.

Task 1: LA rating. The LA of each doodle was rated in two conditions (single, pair). In the single condition, pairs were divided into two single (i.e., left and right) doodle subsets. On these trials, a doodle was presented with its label. Participants were instructed to rate how appropriate the doodle and its corresponding label were. Using a 7-point scale, where 1 = *not appropriate at all* and 7 = *quite appropriate*, 33 participants rated the left-subset doodles (LA–single–L group) and 29 rated the right-subset doodles (LA–single–R group).

In the paired condition, a pair of doodles was presented together with their respective labels. Participants were asked to rate the left and right labels separately, each on the 7-point scale. Doodles were divided into two subsets that each comprised 49 pairs. Thirty-one participants rated Set 1 (paired-1 group), and 35 rated Set 2 (paired-2 group).

Doodles were presented in a booklet that featured a single doodle (or one doodle pair) on each page. Over 98 trials in the single condition or over 49 trials in the paired condition, each participant rated 98 doodles, which were randomly ordered for each booklet. Participants were instructed to write their judgments regarding the appropriateness of the label for each doodle on the pages and were under no time restriction during this task. The task took approximately 30 min.

Task 2: Generated label and its appropriateness rating. On each trial, single unlabeled doodles were presented in random order. Participants were asked to create a label to the presented doodle. Next, they rated its appropriateness (GLA) using the same 7-point scale as was used for Task 1. If the participant could not generate a label, the response “nothing comes to mind” was recorded.

Thirty participants were assigned to the left doodles (GLA–single–L) group, and 30 were assigned to the right doodles (GLA–single–R) group.

Task 3: PR rating. Thirty-five participants took part in the PR rating. Paired unlabeled doodles were presented to the participants, who rated their perceived relationship between left and right doodles on a 7-point scale, where 1 = *no relationship* and 7 = *very related*. There was no time restriction on this rating task. Five patterns of randomized order were prepared and randomly assigned to each participant. A booklet was used to present the stimuli and to record the score.

Task 4: Recall. A recall task was executed to standardize recall performance of the doodles. Doodle pairs were divided into four subsets: two subsets comprised 25 pairs, and two subsets comprised 24 pairs. Thirty participants for each subset took part in the experiment. For each subset, 10 randomized orders were arranged and randomly assigned to the participants.

In a preliminary encoding phase, a pair of the doodles with labels was projected using a personal computer on a screen for 10 sec. After the encoding phase and a 3-min retention phase, a cued recall test was conducted. The left doodle was presented as a cue, and participants were asked to draw the doodle that corresponded to the presented cue on a blank page in the booklet. After 20 sec, they were asked to cease drawing and to turn the page of the booklet.

The doodles that participants drew were rated on a 5-point scale by 5 other participants, and the average scores were computed for

each doodle. Raters were instructed to score the doodles using the following judging criteria: 4 = a drawn picture is identical to the target doodle; 3 = almost the same as the target, but small parts are slightly different in shape or number; 2 = important elements are drawn, but some small parts are not correctly drawn, missing, drawn in wrong position, or horizontally flipped; 1 = not quite totally different, but important elements are missing; and 0 = totally different, or no drawing was done at all.

Raters were also instructed to judge the drawings by their similarities to the targets and to ignore individualities in drawing (e.g., line thickness and object size). Finally, the 5 raters' judged scores of a participant's drawing were averaged and were used as the recall score.

Results and Discussion

The normative data on doodles are shown in Appendix A in the online supplemental materials, and pictures of all the doodles are presented in Appendix B. Appendix A presents the following information for each pair of doodles: labels for left/right pictures (label-L/R), the means for LA for left/right pictures in single presentations (LA-single-L/R) and for LA for left/right pictures in paired presentations (LA-pair-L/R), the measure of generated label disagreement for left/right pictures (H-L/R), the means for GLA for left/right pictures (GLA-L/R), the proportion of labeling failure (LF, "nothing comes to mind") for left/right pictures (LF-L/R), the means for PR between left and right pictures (PR), and the cued recall score for right doodles (recall).

We computed the *H* metric (Snodgrass & Vanderwart, 1980), an index of generated label disagreement, as follows:

$$H = \sum_{i=1}^k P_i \log_2(1/P_i),$$

where *k* is the number of different labels for each doodle and *P_i* is the proportion of participants producing the *i*th label. The generated label for each doodle was judged as being the same if it satisfied any of the following criteria: (1) It was the same as another generated label; (2) it was the same as another generated label, but the label was written in a different Japanese script (e.g., kana or Chinese

Table 1
Summary Statistics for the Measured Variables

	<i>M</i>	<i>SD</i>	Max	Min
LA-single-L	4.01	1.03	6.36	2.09
LA-single-R	4.10	0.81	6.36	2.31
LA-pair-L	4.43	0.93	6.61	2.65
LA-pair-R	4.27	0.86	6.34	2.17
<i>H</i> -L	3.06	0.98	4.51	0.63
<i>H</i> -R	3.39	0.85	4.62	0.77
GLA-L	4.75	0.78	6.40	2.72
GLA-R	4.75	0.66	6.30	3.18
LF-L(%)	6.26	7.71	40.00	0.00
LF-R(%)	4.73	6.42	43.33	0.00
PR	3.45	0.93	6.26	2.06
Recall	2.28	0.55	3.49	1.04

Note—LA-single-L/R, label appropriateness for left/right picture in single presentation; LA-pair-L/R, label appropriateness for left/right picture in pair presentation; *H*-L/R, statistics of label disagreement for left/right picture; GLA-L/R, generated label appropriateness for left/right picture; LF-L/R, percentage of labeling failure for left/right picture; PR, picture relationship; Recall, cued recall score for right doodle.

characters; "glass and straw" or "glass/straw"); (3) the label used abbreviations of words used in another generated label; or (4) it was an idiomatic name that subsumed the other generated label. We eliminated LFs when computing *H*. The total number of LFs for single doodles was 323, which was 5.5% of the total number of responses.

Table 1 presents summary statistics for the measured variables. According to the minimum and maximum values, each measure has some variability among pictures. LA for paired pictures was higher (*M* = 4.43 and 4.27 for left and right pictures, respectively) than for single pictures [*M* = 4.01, *t*(97) = 9.67, *p* < .01, for left pictures; *M* = 4.10, *t*(97) = 3.49, *p* < .01, for right pictures]. That is, appropriateness was rated as being higher when pictures were presented in pairs than when they were presented separately. These results indicate that recognition for paired doodles had a mutual effect on the recognition of individual doodles in each pair.

Next, we calculated correlations among the measures. Table 2 presents this correlation matrix. As we expected,

Table 2
Correlations Among the Measures

	LA-Single-L	LA-Single-R	LA-Pair-L	LA-Pair-R	<i>H</i> -L	<i>H</i> -R	GLA-L	GLA-R	LF-L	LF-R	PR	Recall
LA-single-L	1.00											
LA-single-R	-.01	1.00										
LA-pair-L	.91**	.04	1.00									
LA-pair-R	.10	.84**	.15	1.00								
<i>H</i> -L	-.58**	.01	-.59**	-.01	1.00							
<i>H</i> -R	-.06	-.24**	-.06	-.35**	.26**	1.00						
GLA-L	.62**	-.05	.63**	-.04	-.73**	-.08	1.00					
GLA-R	.01	.27**	.01	.37**	-.17	-.63**	.13	1.00				
LF-L	-.40**	.18	-.40**	.07	.46**	.08	-.68**	-.14	1.00			
LF-R	-.11	-.16	-.06	-.26**	.09	.39**	-.06	-.60**	.02	1.00		
PR	.59**	.20*	.59**	.38**	-.48**	-.27**	.45**	.18	-.32**	-.10	1.00	
Recall	.15	.08	.15	.17	-.05	-.13	-.03	.12	.02	-.01	.21*	1.00

Note—LA-single-L/R, label appropriateness for left/right picture in single presentation; LA-pair-L/R, label appropriateness for left/right picture in pair presentation; *H*-L/R, statistics of label disagreement for left/right picture; GLA-L/R, generated label appropriateness for left/right picture; LF-L/R, proportion of labeling failure for left/right picture; PR, picture relationship; Recall, cued recall score for right doodle. **p* < .05. ***p* < .01.

significant correlations emerged between LA and GLA, and between LA for single and paired pictures. Moreover, LA and GLA were negatively correlated with *H*, indicating that, when pictures have highly appropriate labels, participants tend to give common concepts for those pictures even if they are presented without labels. LF was negatively correlated with indexes of label appropriateness (LA and GLA). These results show that doodles with high appropriateness are easy to interpret (i.e., less nonsensical), whereas those with low appropriateness are more nonsensical if they are presented without labels. PR correlated with most appropriateness variables; that is, pairs including pictures with high appropriateness were rated as being well associated to each other. This result indicates that participants rated the relationship on the basis of the concepts for pictures that come to mind spontaneously. Appropriateness indexes for left pictures correlated with the relationship more highly than did those for right pictures, which suggests that the interpretation of the left pictures has a more significant impact on the whole impression of pairs. Recall was significantly correlated only with PR.

EXPERIMENT 2

A Replication of Bower et al. (1975) and Others

In this experiment, we applied the metrics associated with the task and stimuli of Experiment 1 to describe recall performance in order to demonstrate the use of doodles in memory experiments. In particular, we replicated the findings of studies by Bower et al. (1975), Klatzky and Rafnel (1976), and Rafnel and Klatzky (1978). In this replication, we examined the respective roles of verbal labels or descriptions in memorizing pictorial stimuli. Our results suggest that appropriate labels help memory for visual objects. Bower et al. asked participants to study pairs of doodles and later tested the memory of these participants by cuing one member of each doodle pair for cued recall. In our study, we followed this procedure and scored the recall sketches of the participants. However, in this experiment, we also considered the usefulness of the appropriate (original) labels in recalling the single doodles by contrasting performance in appropriate and irrelevant label conditions.

Method

Participants. None of 106 college students (47 male and 59 female, mean age = 20.6 years) who took part in the experimental session for course credits had served in Experiment 1. They were assigned randomly to one of the experimental conditions. The experiment was conducted in several sessions with small groups of participants. In addition, 177 students who did not participate in the experimental session served as raters for the recall responses.

Materials. We randomly chose 25 single doodles from the standardized picture set as experimental stimuli. All were selected from the right side of the pair sets. We also randomly chose 25 irrelevant labels from the rest of the standardized set. We made three randomized stimuli blocks, Patterns A, B, and C. Each block differed in the presentation order of the 25 pictures. With these stimuli blocks and the labeling conditions, we had six experimental blocks in total (e.g., A-appropriate, B-irrelevant, etc.). Stimuli were displayed to each participant group with an LED projector, and booklets were used for drawing the recall responses.

Procedure. A study (encoding) phase was followed by a test (recall) phase. In the encoding phase, participants were instructed to memorize the pictures and the corresponding labels (either appropriate or irrelevant). Each participant group saw 25 single doodles from one of the six experimental blocks. Each stimulus was shown on the screen for 10 sec, during which time participants were allowed to remember or rehearse the stimuli freely in mind, but were instructed not to try drawing the pictures they had seen. A 3-min rest was inserted before the recall phase. Following this, in the recall test, only labels (cue labels) were shown to the participants as cues. Participants were told to draw the picture attached to a cue label in 20-sec presentation time. They were informed that each recall response should be drawn in a 5 × 7 cm frame printed on the response booklet, which was handed in during the rest time, after which the cue label would appear on the screen. At this point, they were to cease drawing the current item (even if it was unfinished) and move to the next stimulus item. Then they were to draw a remembered doodle associated by a cue label as accurately as possible and to avoid adding artistic or creative features. After the recall phase, participants were allowed to finish the unfinished pictures or erase unnecessary details as needed.

Results and Discussion

The participants' recall performances were judged by 177 raters who were unfamiliar with the memory experiment according to the judging criteria as mentioned above. One rater judged 3 different participants' 75 response drawings; that is, 5 raters judged 1 participant's recall responses.³

First, we checked the effect of the presentation order on the participants' memory performances. The average recall scores in three stimuli blocks (i.e., A, B, and C) were not significantly different [$F(2,103) = 1.94, p > .15$]. Therefore, we pooled these data in the following analyses.

Labeling effectiveness. In Experiment 2, our interest focused on the effect of the labels as recall cues. The mean recall score was 2.62 ($SD = 0.33$) in the original, appropriate label condition, whereas in the irrelevant label condition, the mean recall score was 1.71 ($SD = 0.45$). The result of a three-way ANOVA (labeling × doodles × participants; the latter two were within-block factors) showed a significant and robust difference in recall performance due to labeling [$F(1,104) = 40.65, p < .001$]. Not surprisingly, proper labels could help recall performance.

Predictability of the standardized measures. In Experiment 2, we applied several indexes developed in Experiment 1 to describe memory performance. We addressed the question of whether these indexes can predict the participants' memory performance by using stepwise multiple regression analyses with the indexes as explanatory variables and by using the recall scores in the appropriate/irrelevant label conditions as the response variable.

In the irrelevant label condition, the best-fit model with respect to the Akaike information criterion (AIC) was defined as

$$\text{Recall performance} = (\text{Intercept}) + [\text{LF}].$$

In this condition, the adjusted R^2 value was .09, suggesting that the memory performance was not well predictable by the measures.

By contrast, in the appropriate label condition, the best-fit model was defined as

$$\text{Recall performance} = (\text{Intercept}) + [\text{Recall}] + [H-R] \\ + [\text{GLA-R}] + [\text{LF}] + [\text{LA-single-R}] + [\text{LA-pair-R}].$$

In this condition, the adjusted R^2 value was .49. It is not surprising that the index concerning the memory (recall) contributes significantly to this explanation. More interesting and relevant is the finding that the appropriateness of the label (H , GLA , LA) also has explanatory power. The predictability of H suggests that the easiness of the verbal encoding would affect the to-be-stored pictures. In summary, as was the case in experiments reported by Bower, Klatzky, and their colleagues (Bower et al., 1975; Klatzky & Rafnel, 1976; Rafnel & Klatzky, 1978), the present experiment demonstrates that appropriately labeled pictorial stimuli are easier to memorize.

GENERAL DISCUSSION

The purpose of the present study was to develop a normative set of nonsensical pictures (i.e., doodles) and to demonstrate the role of semantic comprehension in facilitating recall of pictures. A total of 98 pairs of doodles were adopted for standardization.

In Experiment 1, we presented the method and the results of this standardization. In Experiment 2, we followed the recall design of Bower et al. (1975) and others. As we expected, the effect of the labels as recall cues was highly significant, which indicates that semantic interpretation can strongly facilitate free recall. Multiple regression analysis showed that LA has an explanatory power for recall performance. Overall, these findings are consistent with the common concept that memory is aided whenever contextual cues can elicit appropriate schemata into which the material to be learned fits (Bower et al., 1975).

There are two issues for further research. The first is concerned with the originality of the doodle labels. Doodles often evoke an "aha" reaction, reflective of the originality of a label. Originality may affect memory and cognition. However, the effects of originality on recall performance were not assessed in this study. Further studies should provide data on the effect of originality.

The second is concerned with the cultural specificity of the doodles and with the generality of the results of the present study to other cultures. Indeed, some pictures and labels used in this study are specific to the Japanese culture and are likely to be unfamiliar to individuals from other cultures: Item 12 (Label-L: minced and steamed fish), Item 17 (Label-L: Mount Fuji), Item 20 (Label-L: drumstick of wooden fish; Label-R: wooden fish), Item 32 (Label-L: Japanese sword from the front; Label-R: blade of Japanese sword), Item 50 (Label-L: top of Sazaesan's head; Label-R: top of Namihei's head), and Item 70 (Label-L: belly button; Label-R: horns of Thunder God) are examples of these. Although approximately 10% of the pictures used in the present study may be culture specific, such specificity cannot be entirely avoided in this kind of a database. Therefore, it is expected that the doodles and

labels developed here would be partially revised when they are used with other cultures. Although several doodles are specific to the Japanese culture, nevertheless, Experiment 2 replicated the result of Bower et al. (1975). Moreover, several variables (H , LA , and GLA) made a significant contribution to recall performance. Both of these facts are indicative of the validity of the results of this study.

The set of doodles developed in this study can be applied usefully in different fields of experimental psychology and would also have various uses as experimental material. For example, they can be presented separately as well as in pairs. Moreover, because they are meaningless pictures that, by themselves, do not give sufficient interpretive clues, they can also be used as abstract material. We expect that these standardized doodles would be used in a variety of psychological studies.

AUTHOR NOTE

The present work was supported in part by Waseda University under Research Grants 2005B-049 and 2006B-041. The authors gratefully acknowledge Editor Gregory Francis for his encouraging advice, and we thank anonymous reviewers for their constructive and helpful suggestions for the first version of the manuscript. Address correspondence concerning this article to T. Nishimoto, Department of Psychology, Faculty of Letters, Arts, and Sciences, Waseda University, 1-24-1 Toyama, Shinjuku-ku, 162-8644 Tokyo, Japan (e-mail: nishi@waseda.jp).

REFERENCES

- BARQUERO, B., ROBINSON, E. J., & THOMAS, G. V. (2003). Children's ability to attribute different interpretations of ambiguous drawings to a naive vs. a biased observer. *International Journal of Behavioral Development, 27*, 445-456. doi:10.1080/01650250344000064
- BOWER, G. H., KARLIN, M. B., & DUECK, A. (1975). Comprehension and memory for picture. *Memory & Cognition, 3*, 216-220.
- CARPENDALE, J. I., & CHANDLER, M. J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development, 67*, 1686-1706. doi:10.2307/1131725
- CHANDLER, M. J., & HELM, D. (1984). Developmental changes in the contribution of shared experience to social role-taking competence. *International Journal of Behavioral Development, 7*, 145-156. doi:10.1016/S0163-6383(84)80207-6
- DOHERTY, M. J., & WIMMER, M. C. (2005). Children's understanding of ambiguous figures: Which cognitive developments are necessary to experience reversal? *Cognitive Development, 20*, 407-421. doi:10.1016/j.cogdev.2005.05.003
- HAYASHI, M., & UNE, Y. (2004). Hypermnnesia in recognition and recall tasks using doodles [in Japanese]. *Japanese Journal of Cognitive Psychology, 1*, 13-24.
- IIDAKA, T., SADATO, N., YAMADA, H., MURATA, T., OMORI, M., & YONEKURA, Y. (2001). An fMRI study of the functional neuroanatomy of picture encoding in younger and older adults. *Cognitive Brain Research, 11*, 1-11. doi:10.1016/S0926-6410(00)00058-6
- KITAGAMI, S. (2000). The influence of verbal encoding on the memory of visual information [in Japanese]. *Japanese Journal of Psychology, 71*, 387-394.
- KLATZKY, R. L., & RAFNEL, K. (1976). Labeling effects on memory for nonsense pictures. *Memory & Cognition, 4*, 717-720.
- LALONDE, C. E., & CHANDLER, M. J. (2002). Children's understanding of interpretation. *New Ideas in Psychology, 20*, 163-198. doi:10.1016/S0732-118X(02)00007-7
- MCANINCH, C. B., & AUSTIN, J. L. (1993). Effect of caption meaning on memory for nonsense figures. *Current Psychology, 11*, 315-323. doi:10.1007/BF02686789
- NISHIMOTO, T., MIYAWAKI, K., UEDA, T., UNE, Y., & TAKAHASHI, M. (2005). Japanese normative set of 359 pictures. *Behavior Research Methods, 37*, 398-416.

- NISHIMOTO, T., & TAKAHASHI, M. (1996). A set of nonsensical pictures (doodles) for use in experiments of memory and cognition [in Japanese]. *Waseda Psychological Reports*, **29**, 63-90.
- PERNER, J., & DAVIES, G. (1991). Understanding the mind as an active information processor: Do young children have a "copy theory of mind?" *Cognition*, **39**, 51-69. doi:10.1016/0010-0277(91)90059-D
- PRICE, R. (1972). *Doodles*. Los Angeles: Price/Stern/Sloan.
- RAFNEL, K. J., & KLATZKY, R. L. (1978). Meaningful-interpretation effects on codes of nonsense pictures. *Journal of Experimental Psychology*, **4**, 631-646. doi:10.1037/0278-7393.4.6.631
- SALTZMAN, J., STRAUSS, E., HUNTER, M., & ARCHIBALD, S. (2000). Theory of mind and executive functions in normal human aging and Parkinson's disease. *Journal of the International Neuropsychological Society*, **6**, 781-788. doi:10.1017/S1355617700677056
- SNODGRASS, J. G., & VANDERWART, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 174-215.
- TAKAHASHI, M., & INOUE, T. (2009). The effects of humor on memory for non-sensical pictures. *Acta Psychologica*, **132**, 80-84. doi:10.1016/j.actpsy.2009.06.001
- VISION OFFICE (1999). The doodles homepage. Retrieved October 8, 2009, from www.doodles.com.
- VOOLAI, P. (2003). Constructing digital databases of the periphery of Estonian riddles: Database *Estonian Doodles*. *Folklore*, **25**, 87-92. Retrieved October 8, 2009, from www.folklore.ee/folklore/vol25/doodles.pdf.

NOTES

1. This is one of the most famous examples of a doodle, which the American musician Frank Zappa used as the title and as cover art for an album released in 1982.
2. The database Estonian Doodles is another extensive resource that includes about 650 types of doodles called *riddles* (Voolaid, 2003).
3. The 177th rater judged only 2 participants' responses.

SUPPLEMENTAL MATERIALS

The doodle stimuli and norms from this article may be downloaded from <http://brm.psychonomic-journals.org/content/supplemental>.

(Manuscript received November 19, 2009;
revision accepted for publication March 13, 2010.)