UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Dual Processes in the Acquisition of Categorical Concepts

Permalink
https://escholarship.org/uc/item/5m2808nk

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 25(25)

ISSN
1069-7977

Author
Yamauchi, Takashi

Publication Date
2003

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Dual Processes in the Acquisition of Categorical Concepts

Takashi Yamauchi (tya@psyc.tamu.edu)
Department of Psychology; Mail Stop 4235
Texas A&M University, College Station, TX 77840 USA

Abstract

Two experiments and computational simulations investigated the way people make classifications and inferences when the information about category membership was available to participants. On a classification question, participants were asked to predict the category to which a stimulus belongs, and on an inference question, participants were asked to predict the feature value of a stimulus given the category membership of the stimulus. Given classification questions, participants ' performance was influenced greatly by the concrete appearance of individual stimuli, but such an influenced was not present in participants answering inference questions.

Classification and inference constitute two of the most important aspects of concept acquisition (Smith, 1994). In this article, I will examine whether or not a similarity-based process, such as the one formalized in the Generalized Context Model (Nosofsky & Zaki, 2002), can account for judgment processes involved in classification and in inference.

Despite the formidable successes of the similarity-based account of concept formation (Medin & Schaffer, 1978), recent findings suggest that forming categorical knowledge involves multiple routes, which include a similarity-based associative process as well as a rule-based abstract process (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Kruschke, 1998). The validity of this hybrid view has been, however, questioned recently for at least two reasons. First, most of the findings that support the hybrid account are also consistent with exemplar-based models (Nosofsky & Johansen, 2000). Second, studies have shown that depending on the way people interact with a category, they extract different types of feature information (Whittlesea, Brooks, & Westcott, 1994). For this reason, it is difficult to probe the mechanism underlying category formation by simply analyzing the effect of category learning.

In this article, I will investigate how people arrive at classification judgments and inference judgments when the membership about categories is readily available to participants at the time of judgment (see Yamauchi & Markman, 2000 for a similar procedure). In so doing, I intend to demonstrate that classification and inference – two of the most important functions of categories – involves two separable processes.

Overview of the Experiments In one experiment, participants were given a sample sheet depicting 10 members of two categories, and were asked to answer 60 classification questions or 60 inference questions on the basis of the 10 samples (Figure 1). Each category consisted of schematic figures of imaginary bugs that possess 5 features with binary values and a category label (e.g., "monek" or "plaple").

The two categories have a family-resemblance structure, which was derived from prototypes (M0 & P0) (Table 1). On the classification questions, participants predicted the category label of a stimulus given 5 dimensions of feature values (1 1 1 1 0 ?). On the inference questions, participants predicted the value of one of the 5 features, given 4 dimensions of feature values and its category label (? 1 1 1 0 1). Thus, the two types of questions were formally equivalent if category labels and category features are the same thing (Table 1).

Two key variables – feature-matching and feature-manifestation – measured the extent to which participants adopt a similarity-based associative process. I manipulated the number of matching features of the test stimulus to the prototype of the corresponding sample category, and devised two levels of matching features (i.e., high- and medium-levels and see Table 1).

In the second key variable – feature-manifestation, I controlled the appearance of individual stimuli. One set of test stimuli was composed of the same instances as used for the sample stimuli (i.e., Same-manifestation and Set A in Figure 2). The other sets of test stimuli consisted of new instances that were different from the sample stimuli (Different-manifestation; and Sets B/C/D/E in Figure 2). These new instances, however, had some abstract characteristics in common with the features of the sample stimuli (e.g., having eight legs). Thus, answering these questions required awareness of commonalties that go beyond specific appearance of individual items.

The dependent variable of the experiment was the number of responses made in accordance with the prototype of the corresponding category (i.e., category-accordance responses, and see Table 2 for the definition). For example, given the classification question of stimulus M1, responding with the value 1 (i.e., selecting "Monek") was considered a category-accordance response, and given the inference question of stimulus M1, responding with the value 1 (i.e., selecting the long horns) was considered a category-accordance response.

Experiment 1

Previous research has shown that classification requires comparison of matching features derived from specific exemplars. For this reason, classification judgments should be prone to the manipulations introduced to the concrete appearance of the test stimuli. Specifically, the number of category-accordance responses should decline as the similarity level shifts from the high-level to the medium-level of feature-match, and when the appearance of test stimuli is different from that of sample stimuli (i.e., stimuli with different feature manifestation). If participants given inference questions employ an equivalent process as predicted in classification questions, similar response patterns should appear. However, a previous study employing an incremental inference-learning task revealed that participants tend to attend to underlying abstract commonalties rather than concrete stimulus appearance to make a judgment (Yamauchi & Markman, 1998). For this reason, the effect of the feature-matching and feature-manifestation would be less conspicuous in participants making inference judgments than in participants making classification judgments. I tested this hypothesis in Experiment 1.

Participants & Materials Participants were 223 undergraduate students at Texas A&M University, who were randomly assigned to either the classification condition (N=106) or the inference condition (N=117). The stimulus materials were schematic illustrations of cartoon bugs, which were produced from 5 sets (A, B, C, D, and E) of prototypes (Figure 2). The stimuli obtained from Set A depicted 10 samples and 20 test questions. The remaining sets, B, C, D, and E, were employed to produce two versions of test stimuli for counterbalancing. All the test stimuli were divided into two levels of feature-match – high- and medium-levels, and two types of feature manifestation. Altogether, the 60 test stimuli consisted of 20 stimuli with the same manifestation (Set A) and 40 stimuli with different manifestations (Set B/D & C/E).

Procedure & Design The procedure of Experiment 1 involved answering 60 classification questions or 60 inference questions shown on a computer screen. For each question, the computer showed the sample stimuli on the left and the question stimulus on the right side of the screen. Participants indicated their responses by clicking one of the two buttons. The design of the experiment was a 2x2x2 factorial – (Question-type: classification vs. inference – between-subjects factor) x (Feature-match: high vs. medium – within-subjects factor) x (Feature-manifestation: same vs. different – within-subjects factor).

Results The results from this experiment clearly suggest that participants in the classification condition were more sensitive to the concrete appearance of individual stimuli. There was a significant interaction between feature-match and question-type; $F(1, 219)=64.7$, $MSE=0.013$, $p<0.001$. Planned comparisons indicated that the response scores obtained from participants in the inference condition were significantly higher than those from participants in the classification condition at the medium-level of feature match but not at the high-level of feature; $t(221)=4.47$, $p<0.001$.

	High	Medium	Same	Different
classification	0.81	0.64	0.76	0.68
inference	0.78	0.74	0.78	0.75

Note: Average scores for classification and inference questions as a function of feature-matching and feature-manifestation

The results from feature-manifestation also indicated that participants answering classification questions were influenced by the specific appearance of examples more often than participants answering inference questions. There was a statistically significant interaction between feature-manifestation and question-type; $F(1, 219)=23.2$, $MSE=0.015$, $p<0.001$. Planned comparisons indicated that participants in the inference condition made significantly more category-accordance responses than participants in the classification condition given the stimuli composed of different feature-manifestation; $t(221)=2.96$, $p<0.01$ (Bonferroni adjustment). Participants in the two conditions were not statistically distinguishable given the stimuli composed of the same feature manifestation; $t(221)=0.64$, $p>0.10$. The main effect of feature-manifestation was also significant; $F(1, 219)= 47.6$, $MSE=0.015$, $p<0.001$. All the other effects, including the three-way interactions between feature-match, feature-manifestation and question-type, as well as a main effect of question-type, did not reach a significant level; the three-way interaction, $F(1, 219)=1.19$, $MSE=0.01$, $p>0.10$; the main effect of question-type, $F(1, 219)=1.89$, $MSE=0.096$, $p>0.10$. Clearly, classification judgments make use of concrete exemplar information to a larger extent than inference judgments require.

Experiment 2

Experiment 2 was designed to minimize external differences between the two tasks. In this experiment, participants were asked to make classification judgments or inference judgments on the basis of a single sample stimulus (prototypes of each category and see Figure 3a). In the classification question, participants were asked to indicate the probability that a question stimulus belongs to the same type as the corresponding sample stimulus. In the inference question, participants were asked to indicate the probability that a question stimulus has the same feature as the corresponding sample stimulus. Participants indicated their estimated probability in a 0-100 scale. In this manner, participants in the two conditions received the same stimuli and were asked to compare each test stimulus directly to a sample stimulus.

If inference and classification diverge in their fundamental decision processes, then the discrepancy observed in Experiment 1 should be replicated in this simplified setting as well.

Participants & Materials Participants were 86 undergraduate students at Texas A&M University, who were randomly assigned to the classification condition (N=44) or to the inference condition (N=42). The materials employed in Experiment 2 were analogous to those used in Experiment 1. In this experiment, I adopted the stimulus sets A and B only (Figure 2).

The sample stimuli were prototypes from the two categories – monek and plaple. All test stimuli were composed of new feature instances that were different from those depicted the sample stimuli (New manifestation). Altogether, there were 40 test stimuli. Among them, 30 stimuli were divided into three levels of feature-match (10 stimuli each for high, medium, and low level of feature-match) in a similar manner described in Experiment 1.

Along with these three levels of feature-match, I also devised 10 "contradictory" test stimuli (i.e., inconsistent questions). Table 2 shows the configuration of these inconsistent questions. In these questions, participants were asked to predict the probability that the test stimulus has the inconsistent value (e.g., (1 1 1 1 ?/0 1) or (1 1 1 1 1 ?/0) and see Figure 3b).

These inconsistent questions help probe the extent to which participants perceive the equivalence of stimuli. If a sample stimulus and a test stimulus are treated as perceptually equivalent, then participants will be likely to give the same label (or the same feature value) to these stimuli. In other words, the more participants acknowledge abstract commonalties of underlying features, the less likely that they endorse inconsistent values. In this manner, these inconsistent questions would measure the extent to which participants perceive commonalties across different feature instances.

Procedure & Design The procedure of this experiment was identical to that described in Experiment 1 except that participants were asked to indicate their responses with a 0-100 scale on the basis of a single sample stimulus (Figure 3a). The figures that were shown in the two conditions were identical. The design of the experiment was a 2 x 4 factorial – 2 (Question-type: classification vs. inference – between-subjects factor) x 4 (Feature-match: high, medium, low, and inconsistent questions – within-subjects factor). The dependent variable of this experiment was the probability scores that participants indicated to each question.

Results The results from Experiment 2 were in accord with the view that participants answering classification questions and participants answering inference questions interpret individual features in a different manner. As in Experiment 1, there was a significant interaction between question-type and feature-match; $F(1, 82)=6.52$, $MSE=169.5$, $p<0.05$. Participants in the two conditions differed both at the medium-level of feature-match and at the low-level feature match, but not at the high-level of feature-match; the high-level of feature-match, $t(84)=0.12$, $p>0.10$; the medium-level of feature-match, $t(84)=2.76$, $p<0.05$ (Bonferroni); the low-level of feature-match, $t(84)=5.18$, $p<0.001$. Clearly, participants in the classification condition were influenced by the level of matching features, but such an influence was less noticeable in participants in the inference condition.

Given inconsistent questions, participants in the inference condition were much less likely to endorse inconsistent features ($M=28.8$), as compared to participants in the classification condition ($M=41.5$); $t(84)=3.42$, $p<0.01$. This result indicates that participants in the inference condition were aware of abstract commonalties across different instances to a larger extent than participants in the classification condition were.

Taken together, the results from Experiment 2 clearly suggest that classification and inference make use of concrete exemplar information in different degrees.

	High	Medium	Low	Inconsistent
Classification	47.7	37.5	28.2	42.3
Inference	47.3	47.0	44.9	28.8

Note: Average estimation scores for the classification and inference questions as a function of feature-matching

Computational Simulations Experiments 1 and 2 indicated that classification judgments were more likely to rely on specific exemplar appearance, while inference judgments tend to focus on abstract commonalities of features. To corroborate this suggestion, I investigated whether or not the latest version of Nosofsky's Generalized Context Model (GCM, Nosofsky & Zaki, 2002) can account for the data obtained in the classification condition and in the inference condition.

The formula (1) is an extension of the GCM introduced in the Nosofsky & Zaki study (2002). In this formula, the probability that a probe item *i* is classified into Category A is expressed as a function of the number of matching features between all items in Categories A and B, and item *i*.

$$P(A \mid i) = \frac{[\sum_{j \in Ca} sim(j,i)]^{\gamma}}{[\sum_{j \in Ca} sim(j,i)]^{\gamma} + [\sum_{j \in Cb} sim(j,i)]^{\gamma}} \quad \text{-- (1)}$$

$$sim_{ij} = \exp(-c \times d_{ij}{}^{r}),\ d_{ij} = \sum_{m=1}^{M} w_m \mid h \times x_{im} - x_{jm} \mid$$

x_{im} and x_{jm} denote the values of exemplars *i* and *j* on dimension *m,* respectively, and w_m is the attention weight given to dimension *m* ($0 \le w_m \le 1$ and $\sum w_m = 1$). *h* is the parameter that adjusts the appearance of individual exemplars. For the items that are constructed with the same feature instances, *h* is set to 1. For the items that are composed of different feature instances, *h* is set to vary ($1 \le h$). This parameter is introduced to accommodate the different feature manifestation adopted in the current experiments. *c* is an overall sensitivity parameter ($0 \le c < \infty$). γ is a response scaling parameter ($0 \le \gamma < \infty$). With $\gamma = 1$, participants are supposed to respond probabilistically, and with $\gamma > 1$, participants respond more deterministically. *r* is a parameter associated with similarity metric. In this simulation, *r* is set to 1.

For the inference questions, I tested whether or not participants would employ the same feature-matching process, as shown in (1). In this case, the probability that participants choose a category-accordance response A given a probe item *i* is characterized in the same manner specified in (1). In (1), the similarity distance between item *i* and *j* is obtained by examining the disparity between individual feature values. Given an inference question asking the value of horns, for example, the similarity distance between two items was measured along 5 dimensions – head, body, legs, tail and labels, but excluding horns.

To simulate the setting in Experiment 2, I modified (1) slightly (see (2)). In Experiment 2, participants received a single sample, and were asked to estimate the probability that the probe item has a particular category label or a feature. Because the probe item can belong to any category, (1) is expanded to accommodate this situation by introducing a new parameter *mc*.

$$P(A \mid i) = \frac{[\sum_{j \in Ca} sim(j,i)]^{\gamma}}{[\sum_{j \in Ca} sim(j,i)]^{\gamma} + [\sum_{j \in Cb} sim(j,i)]^{\gamma} + mc} \quad \text{--(2)}$$

where $mc = \sum_{k=1}^{K} [\sum_{j \in Ck,\, j \notin Ca, Cb} sim(j,i)]^{\gamma}$

Because a single prototype stimulus was shown as a sample representing a category, the similarity distance between a probe item and a single prototype is computed for each category. Experiment 2 also has "inconsistent" questions. For example, an inconsistent classification question asks the probability that an item has the label "plaple" while the sample shows the label "monek" (e.g., P(B|i)) These inconsistent questions were handled by calculating the complement of the consistent features (e.g., 1-P(A|i)). In this setting, it is assumed that participants first estimate the probability that a test stimulus has a value consistent with the sample stimulus (P(A|i)) and then estimate the probability that the test item does not have the consistent value (1-P(A|i)). For all simulations, the best parameter values were sought by an iterative search routine that minimized the sum of squared errors (SSE).

Results from computational simulations suggest that the modified GCM was able to account for participants' classification performance very well. More than 88% of the variation was accounted for by the exemplar model given classification questions in Experiments 1 and 2. However, no more than 27% of the variation was accounted for by the same model given the inference data obtained in Experiments 1 and 2 (Table 3).

In order to validate the disparity in the GCM's ability to handle classification questions and inference questions, I fitted the model to the data obtained from each individual participant in Experiment 2. In this analysis, the average SSE score in the inference condition was nearly twice larger than the average SSE score in the classification condition; SSE in the inference condition (M=1.52), SSE in the classification condition (M=3.09), $t(84)>100$, $p<0.001$. A similar result was obtained for the accountability score (percentage of explained variation); Inference (M=0.12), Classification (M=0.39), $t(84) = 4.3$, $p<0.001$. Clearly, the results from the computational simulations suggest that inference involves more than simple feature-matching processes.

Conclusion

The two experiments and the computational simulations indicate that people employ different decision strategies to answer classification questions and inference questions. Specifically, participants exhibit a strong tendency to assess concrete exemplars to obtain classification judgments, while such a tendency is in general absent in participants answering inference questions. Unlike classification, inference seems to guide people to extract abstract commonalties among different instances. I suggest that this disparity arises from the fact that classification and inference are reliant on two different cognitive processes in different degrees. Given the fact that inference and classification constitute two fundamental functions of categories, I suggest that the formation of a concept is intertwined with these two separable processes.

Acknowledgements

The author is indebted to Art Markman for his thoughtful suggestions and advice.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 3*, 442-481.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127*, 107-140.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification. *Psychological Review, 85*, 207-238.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review, 7,* 375-402.

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisisted: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*, 924-940.

Smith, E. E. (1994). Concepts and categorization. In E. E. Smith & D. N. Osherson (Eds.), *An invitation to cognitive science* (Vol. 3, pp. 3-33). Cambridge, MA: MIT Press.

Whittlesea, B. W. A., Brooks, L. R., & Westcott, C. (1994). After the Learning is over: Factors controlling the selective application of general and particular knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 259-274.

Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26(3),* 776-795.

Figure 1: The Sample Stimuli and a Classification Question and an Inference Question

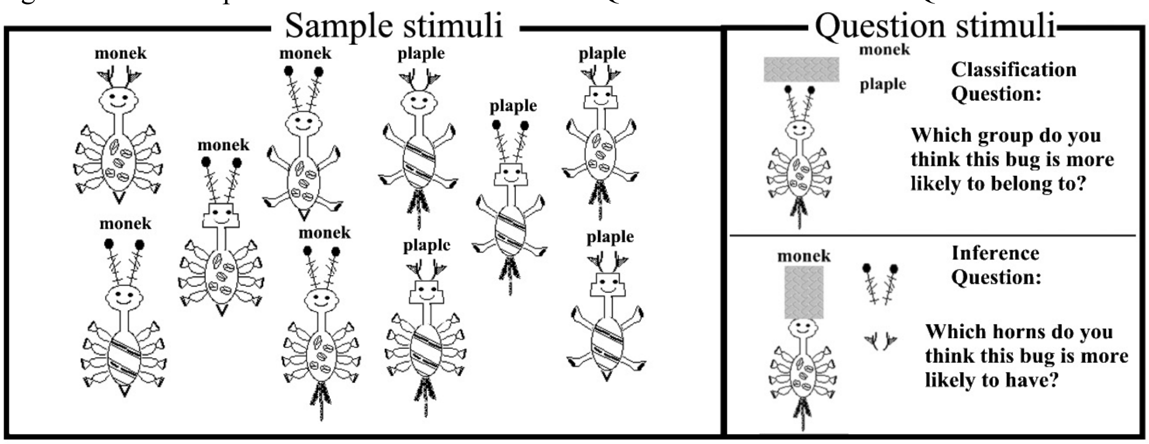

Table 1: the Structure of the Sample Stimuli and of the Question Stimuli in Experiment 1

Sample Stimuli

	Horns	Head	Body	Legs	Tail	Labels		Horns	Head	Body	Legs	Tail	Labels
M1	1	1	1	1	0	1	P1	0	0	0	0	1	0
M2	1	1	1	0	1	1	P2	0	0	0	1	0	0
M3	1	1	0	1	1	1	P3	0	0	1	0	0	0
M4	1	0	1	1	1	1	P4	0	1	0	0	0	0
M5	0	1	1	1	1	1	P5	1	0	0	0	0	0
M0	1	1	1	1	1	1	P0	0	0	0	0	0	0

Question Stimuli

Horns	Head	Body	Legs	Tail	Labels		Horns	Head	Body	Legs	Tail	Labels
1	1	1	1	0	*1*	**High**	***0***	0	0	0	1	*0*
1	1	1	0	***1***	*1*		0	0	0	1	***0***	*0*
1	1	0	***1***	1	*1*		0	0	1	***0***	0	*0*
1	0	***1***	1	1	*1*		0	1	***0***	0	0	*0*
0	***1***	1	1	1	*1*		1	***0***	0	0	0	*0*
1	1	1	0	0	*1*	**Medium**	***0***	0	0	1	1	*0*
1	1	0	0	***1***	*1*		0	0	1	1	***0***	*0*
1	0	0	***1***	1	*1*		0	1	1	***0***	0	*0*
0	0	***1***	1	1	*1*		1	1	***0***	0	0	*0*

Notes: The values enclosed with the rectangular boxes are those used for the classification questions and the values with bold typeface are those used for the inference questions. The responses consistent with these values are defined as "category-accordance responses."

Figure 2: Five Sets of Prototypes

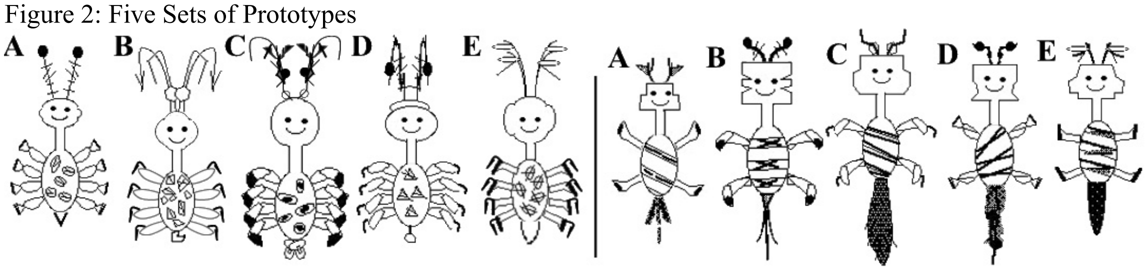

Figure 3a: Samples of a Classification Question – (a) and an Inference Question – (b) Used in Experiment 2

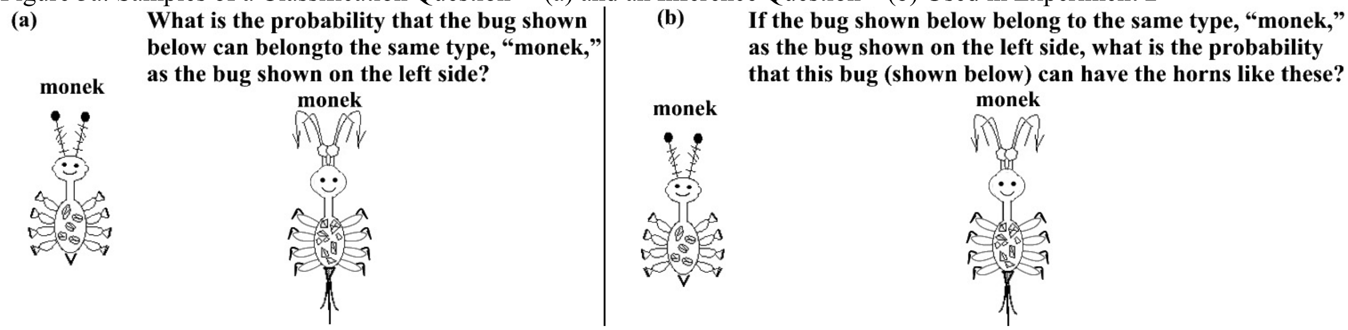

Figure 3b: Samples of Inconsistent Questions; Classification– (a) and Inference– (b)

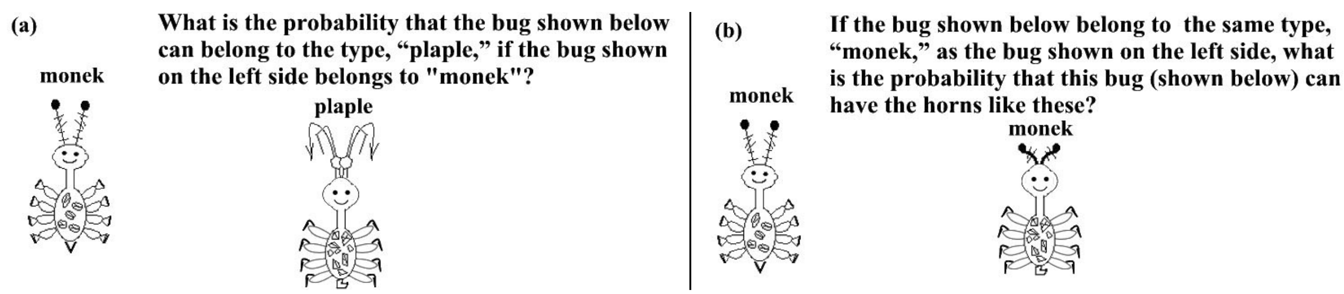

Table 2: The Structure of Inconsistent Questions Used in Experiment 2

	Horns	Head	Body	Legs	Tail	Label	Horns	Head	Body	Legs	Tail	Label
Prototypes	1	1	1	1	1	1	0	0	0	0	0	0
Questions	1	1	1	1	***0/?***	1	0	0	0	0	***1/?***	0
	1	1	1	***0/?***	1	1	0	0	0	***1/?***	0	0
	1	1	***0/?***	1	1	1	0	0	***1/?***	0	0	0
	1	***0/?***	1	1	1	1	0	***1/?***	0	0	0	0
	0/?	1	1	1	1	1	***1/?***	0	0	0	0	0
	1	1	1	1	1	***0/?***	0	0	0	0	0	***1/?***

Table 3: A Summary of the Simulation Results

Experiment 1	w1	w2	w3	w4	w5	w6	c	h	*r*	SSE	% explained
Classification	0.058	0.1061	0.106	0.655	0.075	NA	9.005	1.074	0.615	0.176	88.8
Inference	0.1248	0.021	0.129	0.082	0.082	0.56	2.29	1.04	0.96	0.108	23.9

Experiment 2	w1	w2	w3	w4	w5	w6	c	h	*r*	mc	SSE	% explained
Classification	0	0.141	0.115	0.601	0.143	NA	1.615	1.015	1.01	0.51	0.037	95.0
Inference	0.151	0	0.259	0.14	0.45	0	0.94	1	0.99	0.056	0.799	26.0