Comparing Categorization Models — A psychological experiment

Pei Wang
Center for Research on Concepts and Cognition
Indiana University

May 14, 1993

1 Background

In this paper, a psychological experiment is proposed to compare different theories on the internal structure of categories.

There are three well known psychological theories on this issue (Medin and Ross 92):

Classical theory: According to this theory, concepts have defining features that act as criteria for determining category membership. As a result, the "membership" relation between a concept and given examples is binary, that is, an example either is or isn't a member of the concept, according to whether the example have the features.

Prototype theory: According to this theory, a concept is characterized by properties that shared by most of its members. As a result, category boundaries become fuzzy in the sense that some examples are more "typical" members than others since they share more properties with the "prototype" of the concept, that is, the statistical description of the concept's members.

Exemplar theory: According to this theory, a concept is determined by given examples, and membership of a new example is measured by its similarity to particular examples that exemplify the concept. In such a case, the membership is fuzzy, too, but the reason is different from that of prototype theory. Here, the fuzziness comes from the fact that the similarity relation usually is a matter of degree.

I'm working on an intelligent reasoning system, NARS (Wang 93). This system's memory organization also suggest a theory about category structure in an intelligent system (for both human and computer).

The basic ideas of this theory about categorization are:

1. A concept is characterized by its inheritance relations with other concepts in the system;

2. Each inheritance relation is either extensional (by giving an exemplar) or intensional (by giving a property) to the concept;

3. Each relation have a truth value indicating to what extent the inheritance is successful, according to the system's experience;

4. Each time a concept is used by the system, only some of its relations are considered, and the choice is based on a priority distribution defined among the relations;

5. All concepts are dynamic in the sense that as the system running, new relations are added, some old relations are deleted, and the priority distribution is adjusted by the system itself to adapt to its environment.

NARS can support such a category model, since:

1. It uses a term-oriented language, in which every judgment represents an inheritance relation between two terms (or concepts, categories);
2. It can do revision, deduction, induction, and abduction in an unified manner;
3. The memory is divided into chunks, and each chunk store knowledge that directly related to the same concept;
4. There is a priority distribution maintained among the judgments within a chunk;
5. Each time a concept is addressed to solve a problem, only some of the judgments are used in the corresponding chunk, and the probability of a judgment's being used is determined by its priority.

The internal structure of a concept in NARS is dynamic, but not random. When the system is experienced enough on a concept, some of its relations become stable, and have a higher probability to be addressed when the system use the concept. Let's call them the "dominant relations" of the concept. If these relations happen to be intensional, they become properties that an example should have to be a member of the concept. When these properties are binary, the concept will fit in with the classical theory, otherwise it will meet the prediction of the prototype theory. On the other hand, if the dominant relations happen to be extensional, they provide exemplars for the concept, and the membership of new examples will be determined by how similar they are to the exemplars.

Therefore, I claim here:

1. Usually, both intensional and extensional considerations are involved when the membership of an example is determined for a concept, and the truth values of the concept's dominant relations may be close to binary or far from binary;
2. Therefore, the classical theory, prototype theory, and exemplar theory all correctly describe special cases of a more general categorization mechanism;
3. Since the internal structure of a concept changes according to the system's experience, the concept will show different features at different stage, so no previous theory can provide a satisfactory explanation for all the phenomena;
4. In summary, NARS provide a better model for categorization than the three theories mentioned above.

2 Method

To test the category model of NARS, and to compare it with the other theories, I designed a psychological experiment of "concept learning from examples".

The experiment works like this: subjects are asked to learn a concept on figures. There are three sets of figures: P1 – P6 are positive examples of the concept, N1 – N6 are negative examples of the concept, and X1 – X6 are used to test subjects' criterion in categorization.

The experiment consists of the following stages:

1. Only Figure 1 is given to a subject. The subject is told that P1 and P2 are instance of a concept C, but N1 and N2 are not. Then Figure 2 is given to the subject. With Figure 1 still available, the subject is asked to evaluate the membership of X1 – X6 to C in a 0 – 10 scale, where 10 means "Yes", 8 means "almost is", ..., 5 means "unsure", ..., 0 means "No".

2. Figure 3 is given to a subject. The subject is told that P3 and P4 are also instance of C, but N3 and N4 are not. With Figure 1, 2, and 3 available, the subject is asked to re-evaluate the membership of X1 – X6 to C in the 0 – 10 scale.

3. Figure 4 is given to a subject. The subject is told that P5 and P6 are instance of C, too, but N5 and N6 are not. With Figure 1, 2, 3, and 4 available, the subject is asked to re-evaluate the membership of X1 – X6 to C in the 0 – 10 scale.

After all the subjects are tested, their evaluations on X1 – X6 at the same stage are averaged to get an average membership for each of the 6 testing examples. Totally there are $6 \times 3 = 18$ values, and we will refer Xi's average evaluation at stage j as A_{ij}.

The experiment is designed according to the following consideration:

At stage 1, all the subject know about concept C is the two positive examples (P1 and P2), and the two negative examples (N1 and N2). These examples are designed in such a way that P1 and P2 are quite similar to each other, and at the same time very different from N1 and N2. As a result, it is hard to build dominant intensional relations for C, that is, to find remarkable properties that shared by P1 and P2, but not by N1 and N2, since there are too many candidates. Under such a situation, the subject's internal representation of C is expected to meet the description of the exemplar theory, that is, the subject's evaluation on X1 – X6 is mainly determined by how similar they are to the given examples. Therefore, I predict that A_{11} and A_{21} will have the highest values (since they are more similar to P1 and P2), while A_{51} and A_{61} will have the lowest values (since they are more similar to N1 and N2).

At stage 2, new examples are introduced, and they are designed to make the "exemplar model" hard to work. P3 and P4 looks different from P1 and P2 in many ways, and N3 and N4 are designed to make trouble for the similarity evaluations: X1 and X2 are also similar to N3, though they are still similar to P1 and P2; it is hard to say whether X3 is more similar to P4 or to N4. On the other hand, there is a obvious property shared by all positive examples, but by none negative examples: three upright rectangles. Under such a situation, the subject's internal representation of C is expected to meet the description of the classical theory, that is, the subject's evaluation on X1 – X6 is mainly determined by whether there are three upright rectangles, which is a binary property. Therefore, I predict that A_{32} and A_{42} will be close to 1, while A_{12}, A_{22}, A_{52}, and A_{62} will be close to 0.

At stage 3, new examples refuse the previous "necessary and sufficient condition" model, since P5 and P6 don't have three upright rectangles, but N5 and N6 have. Putting all examples together, it is easy to see that what distinguish P1 – P6 from N1 – N6 is that all positive examples tend to have their components in a line. Under such a situation, the subject's internal representation of C is expected to meet the description of the prototype theory, that is, the subject's evaluation on X1 – X6 is mainly determined by whether they have the property that all the components tend to be aligned. This property is there from the very beginning, but it is not dominate until stage 3, at that time all other competing properties are refused and more evidence is accumulated. As a result, I predict that A_{53} and A_{63} will have the highest values at this stage, while A_{33} and A_{43} will have the lowest values.

As a complement of the experiment, we can ask the subject his/her membership standards at each stage after the whole experiment is finished, and compare it with our expectation.

The subjects of the experiment should be adults, since children may have problem to understand the task.

At each stage, the subject should be given enough time to study the given examples and to evaluate the testing examples, such as 5 minutes or until the subject make up his/her mind.

To eliminate the effect of memory and attention, at stage 2 and 3 the previous examples are still shown to the subject.

To eliminate the effect of order in testing examples, Figure 2 should be re-ordered randomly for each subject.

3 Significance

If we can obtain the expected results from the experiment, it will has the following implications:

1. We get evidence in favor of NARS's category model;

2. The result can be extended from artificial categories (as currently used) to natural categories (Rosch *et al.* 76), and new experiments can be similarly designed to test NARS's category model.

3. We can re-evaluate the previous theories on category structures, re-explain their results, and looking for their limitations and counter examples. For instance, we can design further experiments to determine the conditions under which subjects will set up a classical (or prototype, exemplar) model for a category, and the conditions under which subjects will change their model from one type to another (not necessarily in the "exemplar to classical to prototype" order).

4. NARS shows that similarity relationship can be evaluated in a dynamic and context-dependent way, therefore, it can provide a basis for determining categories. The questions about this issue (Medin and Ross 92, p377) come from traditional opinions that categories are static and context-independent.

5. NARS also make the idea of "knowledge-based organization of concepts" and the idea of "similarity-based organization of concepts" (Medin and Ross 92) to be its special cases, since in it "similarity judgment" is one type of inference the system used to determine membership. On the other hand, the system can use other types of inference to carry out the task, and usually the system can combine results from different types of inference to get a final decision.

6. Since NARS is not a model specially designed for categorization, but a model for high-level cognition in general, it suggests many interesting explanations of human cognition to cognitive psychologists, such as about the relations between categorization and reasoning, memory, learning, attention, and so on. Actually, in NARS they are treated as different aspects of a unified process in a unified system.

7. Since NARS's category model is applicable both to human minds and to computers, it suggest a possible way to achieve artificial intelligence. We can even design experiments that can be used to test human subjects and computer subjects in the same way. (Since the current version of the experiment use figures, it cannot be used on NARS yet.)

References

[1] Medin, D. and Ross, B. (1992). *Cognitive Psychology*, Chapter 12. Harcount Brace Jovanovich, Inc.

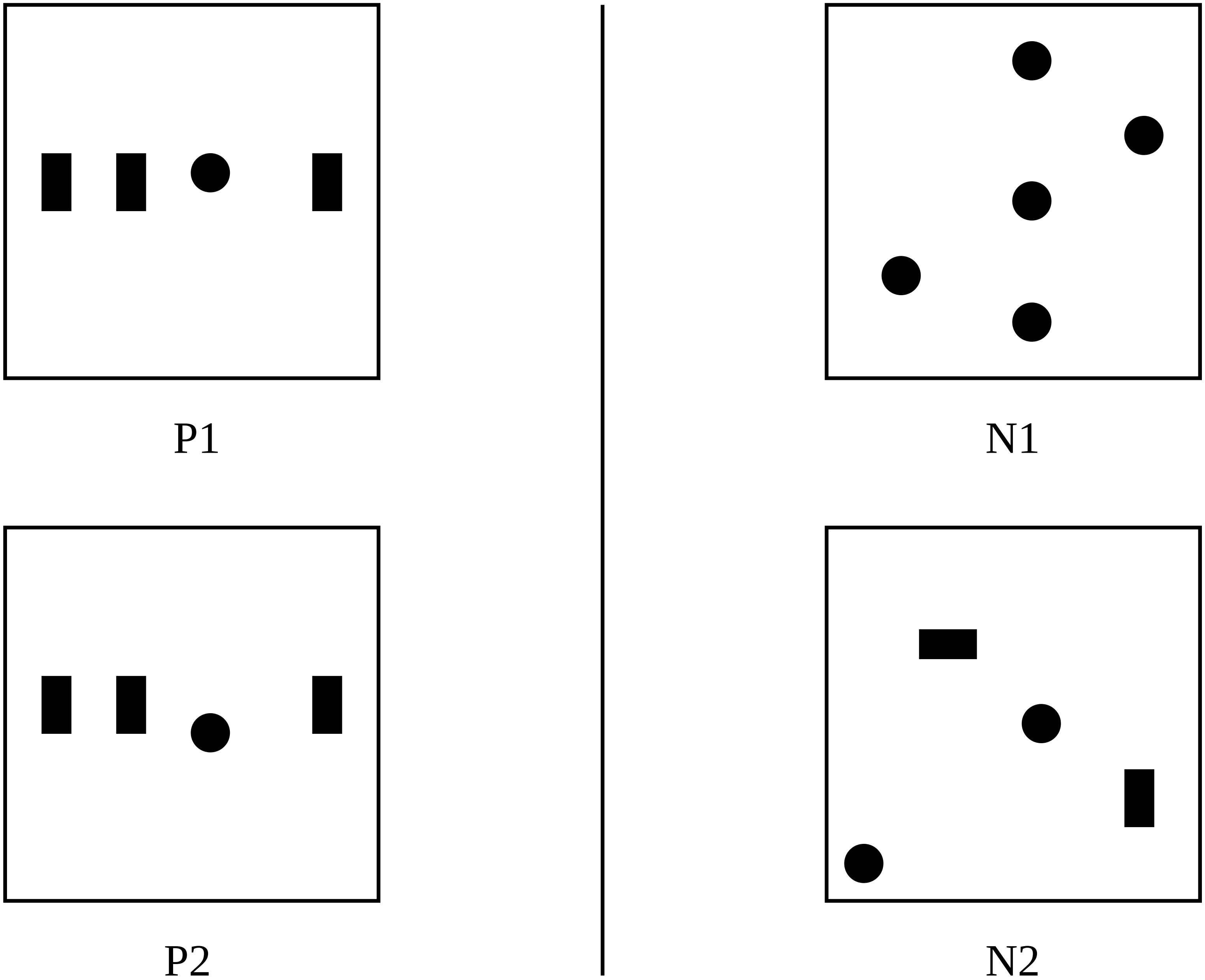
[2] Rosch, E., Mervis, C. B., Gray, W., Johnson, D., and Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology 7*, 573 – 605.

[3] Wang, P. (1993). *Non-Axiomatic Reasoning System (Version 2.2).* Technical Report No. 75 of the Center for Research on Concepts and Cognition, Indiana University.

Appendices

Figure 1 – 4

Figure 1: Teaching Examples I



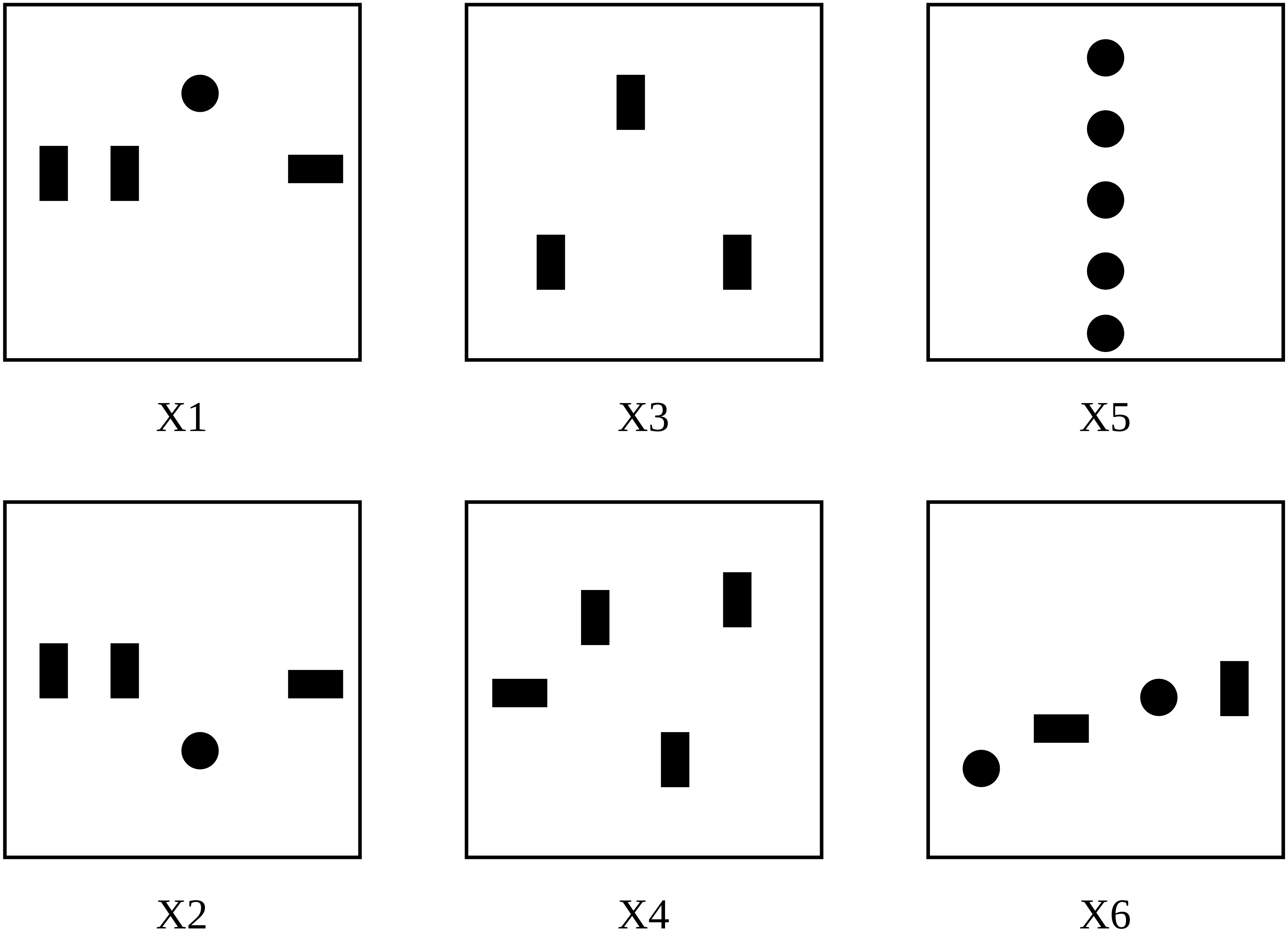
P1

N1

P2

N2

Figure 2: Testing Examples



X1

X3

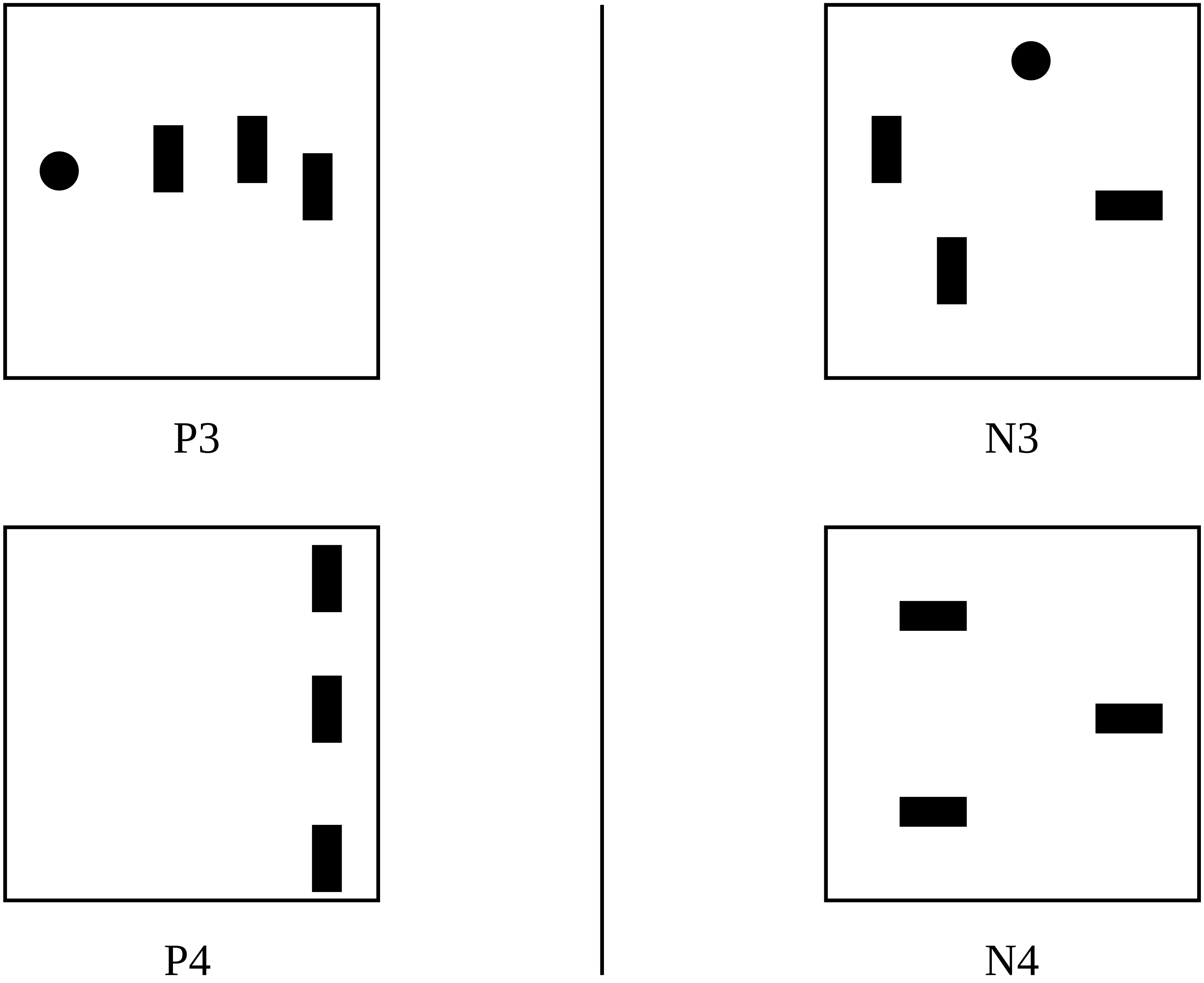
X5

X2

X4

X6

Figure 3: Teaching Examples Ⅱ



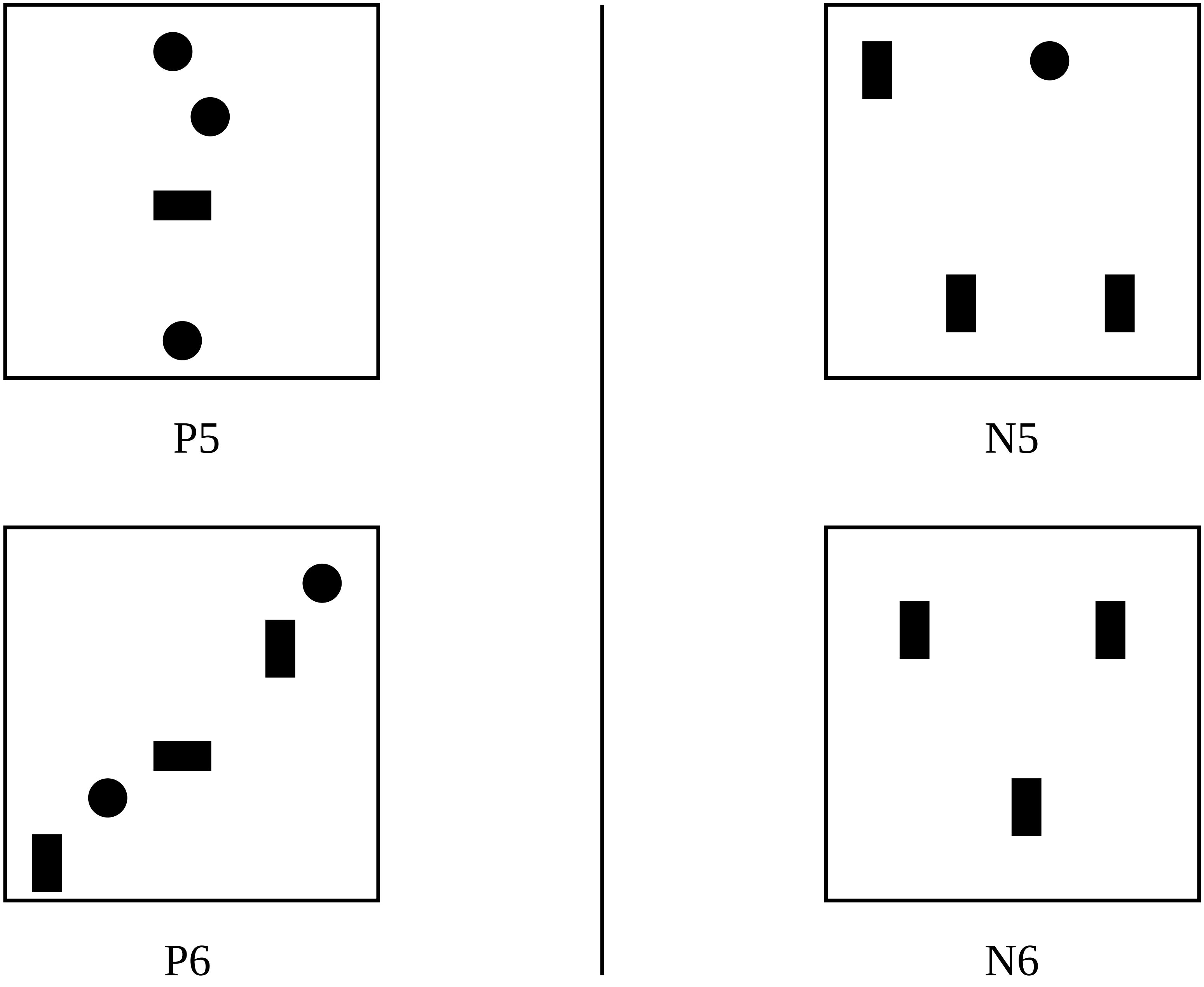
P3

N3

P4

N4

Figure 4: Teaching Examples Ⅲ



P5

N5

P6

N6