

А. С. Карбалевич,
преподаватель кафедры методологии
и методов психологических исследований БГПУ

ФОРМИРОВАНИЕ КАТЕГОРИЙ С НЕЛИНЕЙНЫМ РАСПРЕДЕЛЕНИЕМ ПРИЗНАКОВ

Центральным в когнитивной психологии является вопрос об осмыслении полученного человеком опыта и ментальной организации действительности, единицами которой являются категории — ментальные модели. Психика раскладывает реальность на категории, формируя понятия. Категоризация является основополагающим механизмом как для человеческого, так и для искусственного интеллекта. Процесс категоризации важен не только для классификации объектов и явлений окружающего мира, но и для поддержания таких функций, как прогнозирование, принятие решений, планирование, интерпретация, коммуникация.

Существует предположение о том, что сами категории могут иметь различную внутреннюю структуру. Исследования L. W. Barsalou [1] выявили наличие тенденции у испытуемых формировать такие нетаксономические категории, как «вещи, которые необходимо вынести первыми при пожаре». Эксперимент D. I. Medin [2] показал, что ландшафтные дизайнеры склонны категоризовать растения в соответствии с функциями, которые они выполняют в проекте (деревья, дающие тень; деревья, служащие для создания орнаментного ансамбля и др.). В то же время дизайнеры с легкостью могли переключаться на схему построения категорий по таксономическому принципу. Так, G. Lakoff [3] выделяет четыре типа категорий: пропозициональные, образно-схематические, метафорические, метонимические. R. N. Shepard [4] в своем исследовании использовал шесть типов категорий.

Существует предположение о том, что организация знаний посредством формирования различных типов категорий происходит в рамках нескольких систем [5]. Однако не все авторы придерживаются данного мнения. На современном этапе существует два подхода к моделированию интеллектуальных систем.

Первый подход опирается на построение единого, универсального механизма. Примером может служить классическая теория категоризации, теория прототипов либо теория

экземпляров. Классическая теория категоризации предполагает для каждой категории наличие обязательных существенных признаков. Данную теорию и ее современные трактовки часто называют моделями, основанными на правилах (rule-based models). Выведение правила из ряда наблюдаемых закономерностей предполагает процедуру проверки гипотез относительно существенности признаков [6–8].

Возникновение теории прототипов (prototype-based model) связывают с именем E. Rosch. Автор определяет прототип как «член категории, наиболее полно характеризующий структуру данной категории как целого». Это своеобразная сумма признаков всех ее членов. Математически это можно назвать средним или центральной тенденцией. Новый объект относится к той категории, с прототипом которой он имеет наибольшее подобие [9].

Тенденция ухода от абстракций (правил, схем, прототипов) довольно ярко прослеживается в современной когнитивной психологии (exemplar-based model). Исследователи, занимающиеся изучением памяти и перцептивной категоризации, все чаще говорят о запечатлении конкретных образцов (следов, экземпляров). В 1978 г. появилась одна из первых наиболее успешных моделей данного типа — контекстная модель (context model — CM). Ее авторами стали D. L. Medin и M. M. Schaffer [10]. Позже модель была дополнена R. M. Nosofsky и переименована в обобщенную (generalized) контекстную модель (GCM) [11].

Второй подход в моделировании интеллектуальных систем опирается на идею о том, что сознание человека использует различные механизмы для решения одних и тех же задач. Наличие разнообразного инструментария для решения одной задачи содействует быстрой и успешной адаптации к изменениям во внешней среде. Например, модель COVIS предполагает соревнование двух механизмов, один из которых основан на сходстве, а второй — на формировании правил [12]. В то же время модель ATRIUM предполагает кооперацию двух механизмов: механизма, основанного на пра-

вилах, и механизма, основанного на сохранении отдельных экземпляров в памяти [13].

Предположительно, категории, основанные на легко вербализуемых правилах, формируются посредством эксплицитной системы за счет активности префронтальных отделов коры головного мозга. Все члены категории обладают, как минимум, одним общим признаком и правило может звучать как «все черные объекты – это категория А, все белые объекты – это категория Б». Но бывают категории, где члены объединены по принципу семейного сходства (обычно это перцептивное сходство). Тогда правило может звучать как «большинство черных объектов – это категория А, а большинство белых – это категория Б». Такое правило нельзя с легкостью вербализовать и при категоризации новых объектов оно не позволяет избежать ошибок. Всегда есть вероятность, что новый объект, хоть и является черным, но при этом с определенной долей вероятности может принадлежать к категории Б, где большинство членов имеют белую окраску. Логично было бы предположить, что такие категории формируются посредством имплицитной системы за счет активности подкорковых структур под действием дофамина. Процесс обучения заключается в постепенном выстраивании ассоциаций между перцептивными характеристиками объектов и верными ответными реакциями.

В данной работе представлены результаты исследования влияния смены закономерности ряда признаков на формирование категорий, обозначаемых R. N. Shepard как «Тип 3» (рисунок 1). Это категории, с распределенными нелинейно признаками экземпляров и имеющие три размерности (три признака), из которых только две являются релевантными. Категории данного типа могут быть обозначены как «правило+исключение», то есть три экземпляра объединены в категорию по общему признаку, а четвертый экземпляр является исключением.

Если предположить, что работает одна система, генерирующая и проверяющая гипотезы (правила категоризации), то при обучении на таком стимульном материале процент правильных ответов испытуемых должен сначала колебаться около 50 %, а затем, при обнаружении релевантного признака, подняться примерно до 70–80 %, так как четвертая часть стимульного материала не подходит под общее правило. Если правило дает большое количество правильных ответов, а отказаться от выдвинутой гипотезы психологически трудно [14; 15], то, возможно, на этом этапе испытуемый будет пытаться модифицировать найденное правило, делая его более сложным с целью

учета стимула-исключения. При достаточном уровне когнитивных способностей и/или длительном времени обучения испытуемый может выучить категории и перестать делать ошибки.

При предположении, что работает имплицитная система, ожидается довольно медленное научение, при котором уровень ответов будет возрастать от 50 % (случайное угадывание) до какой-то величины, близкой к 100 %. Тем не менее, работа такой системы всегда будет давать ошибки категоризации, так как поочередно, хотя и в разной степени, усиливаются то одни, то другие ассоциативные связи (например «белый цвет – категория А» и «белый цвет – категория Б»).

Если предположить, что в процессе распознавания объектов будут активно участвовать обе системы, одна из которых (эксплицитная) проверяет гипотезы относительно существенности признаков, а вторая (имплицитная) выстраивает ассоциативные связи между перцептивными формами и номинативным содержанием, то результаты их работы будут совпадать. Найденное правило не будет противоречить постепенно формирующемуся прототипу, а исключение будет давать ошибки категоризации и по одной, и по второй системе. Кривая научения практически не будет отличаться от кривой научения при работе только одной имплицитной системы.

Таким образом, задание просто выучить две контрастные категории не дает возможности выяснить, сколько систем задействованы в процессе научения. Поэтому в предлагаемом эксперименте использовалась полная смена закономерности (reversal shift), когда после некоторого обучения то, что называлось категорией «А», переименовывается в категорию «Б» и наоборот. Смена происходила по релевантному для категоризации признаку (intra-dimensional shift).

Если предположить, что работает одна система, генерирующая и проверяющая гипотезы (правила категоризации), то при смене правила процент правильных ответов сначала должен упасть до 50 % или даже ниже (предположительно до 25 % из-за ошибок, связанных с категоризацией стимула-исключения), а затем опять подняться до прежнего уровня (например, 70–80 %), так как смена простого правила на противоположное происходит довольно быстро [8].

Если мы предполагаем, что работает одна имплицитная система, то процент правильных ответов после смены правила должен резко упасть, а затем медленно и постепенно повышаться, причем скорость обучения на этот раз должна быть существенно ниже скорости первоначального обучения, так как сформиро-

ванные ассоциативные связи будут находиться в противоречии с образованием новых. Тренды, характеризующие скорость обучения, должны оказаться непараллельными.

Если предположить, что работают сразу две системы, то первой на смену правила должна отреагировать эксплицитная система, что даст довольно ощутимую прибавку правильных ответов вскоре после смены правила. Однако из-за того, что на этот раз имплицитная система будет работать не согласованно с эксплицитной, а, наоборот, противоречить ей довольно длительное время, требующееся на затухание следов старых связей и образование новых, процент правильных ответов будет гораздо ниже первоначально достигнутой величины.

Испытуемые. В эксперименте принял участие 31 студент 2–3 курсов факультета психологии БГПУ (7 юношей, 24 девушки).

Стимульный материал. 8 геометрических фигур, различающихся по цвету (серый/зеленый), форме (квадрат/треугольник), размеру (маленький/большой). Каждая категория содержала по 4 экземпляра. Признаки экземпляров распределены нелинейно. Категории данного типа могут быть обозначены как «правило+исключение», то есть три экземпляра объединены в категорию по общему признаку, а четвертый экземпляр является исключением. Наглядное представление фигур показано на рисунке 1.

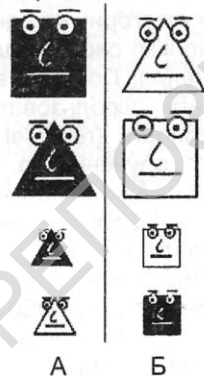


Рисунок 1 – Категории типа 3 по R. N. Shepard

Процедура. Первая серия эксперимента состояла из десяти блоков, в каждом из которых было по четыре объекта, относящихся к двум категориям, то есть всего было восемь объектов для категоризации. Объекты внутри блока предъявлялись в случайном порядке. Каждый испытуемый для формирования понятий выполнял стандартное задание классификации объектов и получал обратную связь. Предъявление объектов и фиксация ответов испытуемого происходили с помощью программы SuperLab. Каждая серия начиналась с демонстрации знака «+» на белом фоне на

1 с. Потом на экране появлялось изображение объекта на 10 с. В течение этого времени испытуемый должен был дать ответ, к какой из двух категорий относится объект, нажимая одну из двух клавиш на стандартной клавиатуре. Сразу после его ответа на экране на 3 с появлялась обратная связь – правильное название категории. Вторая серия предъявлялась по тому же алгоритму за исключением того, что правильные названия категорий менялись местами.

Фиксировалось количество правильных ответов (от 0 до 8), а также номера объектов, правильно и неправильно отнесенных к категории.

Результаты и обсуждение. На рисунке 2 представлена полная кривая научения после усреднения данных, выстроенная на основании количества правильных ответов. Несмотря на небольшое количество экземпляров в категориях (по четыре в каждой), процесс научения занимает достаточно продолжительный период: испытуемые смогли правильно категоризировать только около 5–6 стимулов из 8 к концу периода обучения. На графике видно, что предположение относительно более низкого темпа обучения после смены закономерности ряда признаков оказалось верным. При сравнении средних показателей обучения до и после смены правила была выявлена статистическая значимость различий, что подтверждает наглядные данные ($t(30)=6,81$; $p<0,0001$).

Полученные данные дают основание говорить о наличии тренда лишь до смены закономерности (рисунок 3), причем тренд проявляется при построении его как по всем блокам, так и начиная с любого блока первой серии обучения. Наличие тренда для результатов первых десяти блоков (до смены правила) говорит о том, что прослеживается динамика в процессе обучения, а это значит, что испытуемые действовали не наугад. Полученные данные согласуются с предположениями о наличии либо одной имплицитной системы, либо двух взаимодействующих систем категоризации. С большой вероятностью можно предположить, что при изолированной работе эксплицитной системы процесс научения шел бы быстрее, так как проверить гипотезы относительно весомости каждого из признаков довольно просто из-за малого их количества (всего три признака). Количество исключений также невелико (по одному исключению в каждой категории). Однако резкого скачка в количестве правильных ответов не прослеживается, а довольно плавная кривая обучения не может быть объяснена работой только одной эксплицитной системы.

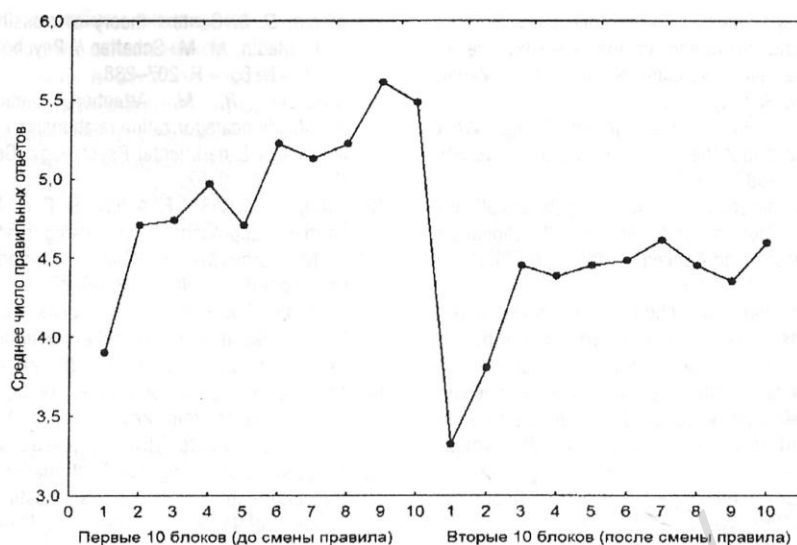


Рисунок 2 – Кривая обучения после усреднения данных по выборке

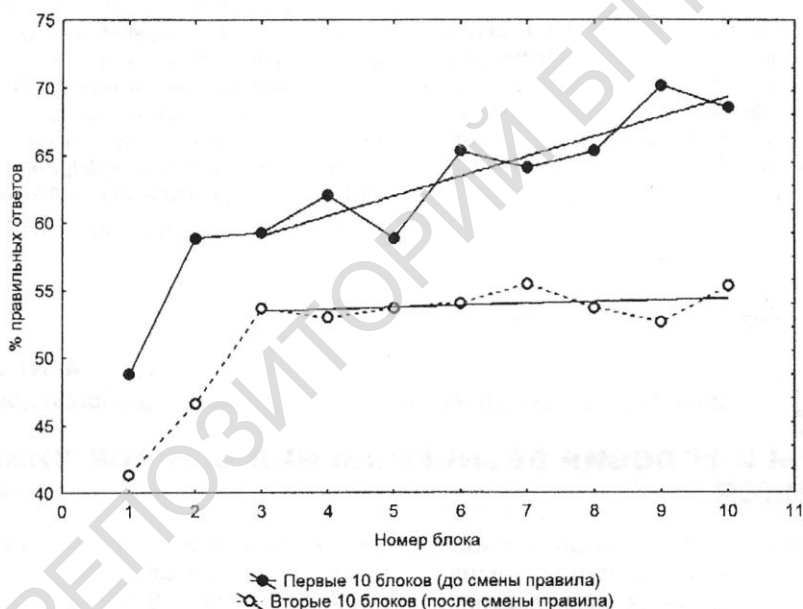


Рисунок 3 – Сравнительная схема проявления тренда (начиная с третьего блока) до и после смены закономерности ряда признаков

После смены правила процент правильных ответов падает, что предсказывается всеми моделями, однако затем довольно быстро возрастает, и, начиная с третьего блока, исчезает тренд, то есть обучение перестает происходить. Полученные данные лучше всего согласуются с предположением о наличии работающих вместе двух систем категоризации. Очевидно, что процесс переучивания является еще более трудоемким, что может объясняться нейробиологической основой функционирования семантической памяти: образование и распад нейронных связей требуют времени, поэтому имплицитная система

является более медленной, что и проявляется в отсутствии тренда после третьего блока. Наличие довольно резкого скачка в количестве правильных ответов между первым и третьим блоками можно интерпретировать как смену эксплицитного правила.

Таким образом, исходя из результатов приведенного исследования можно сделать вывод в пользу наличия, как минимум, двух систем категориального обучения.

ЛИТЕРАТУРА

1. Barsalou, L. W. Ad hoc categories / L. W. Barsalou // Memory & Cognition. – 1983. – № 11. – P. 211–227.

2. Atran, S. Thinking about biology. Modular constraints on categorization and reasoning in the everyday life of Americans, Maya, and scientists / S. Atran, D. I. Medin, N. Ross // *Mind & Society*. – 2002. – № 3: 31.
3. Lakoff, G. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. – Chicago : University of Chicago Press, 1987. – 632 p.
4. Shepard, R. N. Learning and memorization of classifications / R. N. Shepard, C. I. Hovland, H. M. Jenkins // *Psychological Monographs: General and Applied*. – 1961. – № 75(13). – P. 1–41.
5. Herzog, S. M. Blending and choosing within one mind: Should judgments be based on exemplars, rules or both? / S. M. Herzog, B. von Helversen // *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, Austin, July 31–August 3, 2013 / TX: Cognitive Science Society; M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.). – Austin. – 2013. – P. 2536–2541.
6. Bruner, J. S. *A study of thinking* / J. S. Bruner, J. Goodnow, G. Austin. – New York: Wiley. – 1956. – 350 p.
7. Levine, M. A. *A cognitive theory of learning: Research on hypothesis testing* / M. A. Levine. – Hillsdale, NJ: Erlbaum. – 1975. – 309 p.
8. Bower, G. Concept identification / G. Bower, T. Trabasso // In R.C. Atkinson (ed.) *Studies in Mathematical Psychology*. – Stanford University Press, 1964. – 418 p.
9. Rosch E. H. Principles of categorization. Cognition and categorization / Eds. by E. H. Rosch, B. B. Lloyd. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978. P. 27–48.
10. Medin, D. L. Context theory of classification learning / D. L. Medin, M. M. Schaffer // *Psychological Review*. – 1978. – № 85. – P. 207–238.
11. Nosofsky, R. M. Attention, similarity, and the identification/categorization relationship / R. M. Nosofsky // *Journal of Experimental Psychology: General*. – 1986. – № 115. – P. 39–57.
12. Ashby, F. COVIS / F. Ashby, E. Paul, W. T. Maddox // *Formal approaches in categorization* / Eds. by E. M. Pothos & A. J. Wills. – Cambridge: Cambridge University Press, 2011. – P. 65–87.
13. Erickson, M. A. Rules and exemplars in category learning / M. A. Erickson, J. K. Kruschke // *Journal of Experimental Psychology: General*. – 1998. – № 127. – P. 107–140.
14. Аллахвердов, В. М. Сознание как парадокс (Экспериментальная психология, т. 1) / В. М. Аллахвердов. – СПб. : Издательство ДНК, 2000. – 528 с.
15. Klayman, J., Ha, Young-won Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, Vol 94(2), – Apr 1987. – P. 211–228.

SUMMARY

The research data of category learning process in the situation of non-linear distribution of characteristics is presented at the article. The author analyzes the data received using a reversal shift method. It is shown that categories of this type are formed by means of two systems: one of which is explicit and is based on the hypothesis testing, and another is implicit and based on perceptual similarity.

Поступила в редакцию 30.07.2015 г.