Когнитивная наука в Москве: новые исследования. Материалы конференции 23–24 июня 2021 г. Под ред. Е.В. Печенковой, М.В. Фаликман, А.Я. Койфман. М.: «Буки Веди», ИППиП. 2021 г. – 556 стр.

БЫСТРОЕ МНОГОСТУПЕНЧАТОЕ ПОСЛЕДОВАТЕЛЬНОЕ ОБУЧЕНИЕ С ОТЛОЖЕННОЙ ОБРАТНОЙ СВЯЗЬЮ У КРЫС В ТЕСТЕ «ЭКСТРЯПОЛЯЦИОННОЕ ИЗБАВЛЕНИЕ»

Н.А. Бондаренко*
Ninabonda52@gmail.com
ООО «НПК «Открытая наука», Московская обл.

В повседневной жизни (например, при посещении банкомата) нам приходится быстро запоминать цепочку из нескольких последовательных действий, приводящую к желаемому результату (подкреплению). Данный тип обучения квалифицируют как «быстрое многоступенчатое последовательное обучение с отложенной обратной связью» (Gläscher et al., 2010; Daw et al., 2011; Walsh and Anderson, 2011; Tartaglia et al., 2017).

Аналогичное многоступенчатое обучение с отложенной обратной связью встречается и у животных. Например, в тесте «Экстраполяционное избавление» (ТЭИ) крысу помещают внутрь ограничительного цилиндра, нижним концом опущенного в емкость с водой. Для того чтобы покинуть столь стрессогенную ситуацию, животное должно совершить цепочку действий: выйти из ограничительного цилиндра, поднырнув под его нижним краем, затем найти трап, подвешенный на бортик емкости, и уже по нему выбраться из воды. Ранее мы показали, что: a) 90% крыс с первой попытки проходят этот «квест»; b) вылезание из воды является единственным подкреплением (обратной связью), обеспечивающим запоминание животным всей цепочки действий; c) при повторных помещениях крыс в установку ТЭИ животные быстро оптимизируют способ избавления из цилиндра.

У людей быстрое многоступенчатое последовательное обучение с отложенной обратной связью может быть описано на основе алгоритма эпизодического контроля (Lehmann et al., 2019). Однако у животных существование эпизодической памяти многие подвергают сомнению. Альтернативой является гипотеза об использовании животными классического алгоритма временной разницы (например, TD-0, Sutton, 1988), где обучающим сигналом является предсказание будущего. Обучение в соответствии с этим алгоритмом, в отличие от алгоритма «эпизодического контроля», имеет свою специфику: наибольшее количество ошибок субъекты совершают в точках выбора, наиболее удаленных от вознаграждения (феномен «градиента подкрепления» Халла). На основе алгоритма TD обучения можно сделать прогноз для обучения крыс в ТЭИ: обучение подныриванию в цилиндре будет происходить максимально быстро, если вынуть животное из воды сразу после подныривания. Если же после подныривания животное должно сначала найти трап, да еще и научиться вылезать по нему из воды, обучение подныриванию должно замедлиться. Целью настоящей работы была экспериментальная проверка данной гипотезы.

В эксперименте использовали 3 группы половозрелых крыс-самцов линии Вистар, Животных из группы №1 сразу после завершения подныривания экспериментатор извлекал из воды. Животные из группы №2 после подныривания сначала плавали, затем находили сетчатый трап и по нему вылезали из воды. Животные из группы №3 после

подныривания также находили трап и вылезали по нему из воды, но этот трап был сделан из гибкой спирали, поэтому для вылезания требовалась значительная ловкость. Обнаружено, что при трехкратной экспозиции к ТЭИ крысы групп 2 и 3 обучились безошибочно пользоваться трапом. В то же время, межгрупповых различий динамики обучения подныриванию выявлено не было. Это означает отсутствие эффекта «градиента подкрепления», поэтому гипотеза об использовании животными алгоритма TD обучения в ТЭИ не подтвердилась. Можно предположить, что быстрое многоступенчатое последовательное обучение с отложенной обратной связью у крыс в ТЭИ происходит с участием некоего аналога функции «эпизодического контроля» человека. Дальнейшие эксперименты будут направлены на проверку данного предположения.

Daw ND, Gershman SJ, Seymour B, Dayan P & Dolan RJ Model-based influences on humans' choices and striatal prediction errors. Neuron 69, 1204–15. DOI: 10.1016/j.neuron.2011.02.027 (2011).

Lehmann MP, Xu HA, Liakoni V, Herzog MH, Gerstner W, Preuschoff K. One-shot learning and behavioral eligibility traces in sequential decsion making. Elife. 2019. 8:e47463. doi: 10.7554/eLife.47463.

Walsh MM, Anderson JR. Learning from delayed feedback: neural responses in temporal credit assignment. Cogn Affect Behav Neurosci. 2011. 11(2):131-43. doi: 10.3758/s13415-011-0027-0.

Gläscher J, Daw N, Dayan P, O'Doherty JP. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron. 2010. 66(4):585-95. doi: 10.1016/j.neuron.2010.04.016.

Tartaglia EM, Clarke AM, Herzog MH. What to Choose Next? A Paradigm for Testing Human Sequential Decision Making. Front Psychol. 2017. 8:312. doi: 10.3389/fpsyg.2017.00312.

Sutton, 1988 Sutton, R. S. (1988). Learning to predict by the method of temporal differences. Machine Learning, 3:9-44.