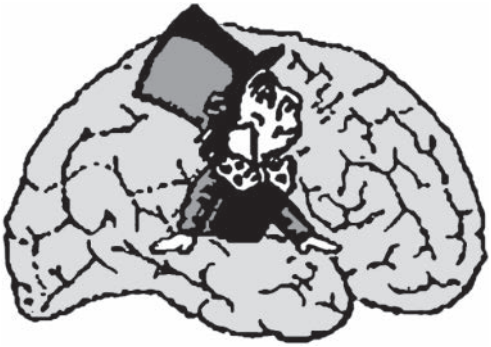
КОГНИТИВНАЯ НАУКА В МОСКВЕ

НОВЫЕ ИССЛЕДОВАНИЯ



МАТЕРИАЛЫ КОНФЕРЕНЦИИ 2017

ПОД РЕД. Е.В. ПЕЧЕНКОВОЙ, М.В. ФАЛИКМАН

УДК 159.9

ББК 81.002

К57

К57 Коллективный

Когнитивная наука в Москве: новые исследования. Материалы конференции 15 июня 2017 г.

Под ред. Е.В. Печенковой, М.В. Фаликман. – М.: ООО «Буки Веди», ИППиП. 2017 г. – 596 стр.

Электронная версия

ISBN 978-5-4465-1509-7

УДК 159.9

ББК 81.002

ISBN 978-5-4465-1509-7

ВЛИЯНИЕ ВЕРОЯТНОСТНЫХ СИГНАЛОВ НАГРАДЫ И НАКАЗАНИЯ НА ВЫБОР ИЗ ДВУХ АЛЬТЕРНАТИВ

Г. Л. Козунова*, Н. А. Воронин, В. В. Венидиктов, Т. А. Строганова
chukhutova@gmail.com
Московский государственный психолого-педагогический университет, Москва

Аннотация. Поведение человека в условиях частичной неопределенности исхода характеризуется тенденцией к подведению относительной частоты поведения к вероятности его подкрепления — закон вероятностного соответствия. Мы исследовали роль вероятностных сигналов награды и наказания в формировании этого поведения. 29 здоровых взрослых испытуемых выполняли серию из 4 тестов по 40 проб каждый, в которых выбор одного из пары стимулов подкреплялся в 70% случаев, а второй — в 30%. Результаты показали, что лишь у половины испытуемых выработалось предпочтение выгодного стимула в первом тесте, в то время как оставшиеся испытуемые демонстрировали накопительный эффект обучения от первого теста к последнему. Поведение испытуемых в тех тестах, где обучения не произошло, характеризовалось слабой тенденцией повторять свой предыдущий выбор после получения награды и переключаться на другой стимул — после наказания. Интересно, что в периоды, непосредственно предшествовавшие успешному обучению, знак обратной связи не оказывал прямого влияния на последующее решение. В эти периоды испытуемые демонстрировали парадоксальную восприимчивость только к редким, нетипичным сигналам обратной связи, которые приводили к выбору заведомо невыгодного стимула. Мы предполагаем, что формированию адаптивного навыка предшествовала имплицитная прагматическая оценка обоих стимулов на основе регулярных сигналов награды и наказания, однако редкие рассогласования с ожидаемым результатом провоцировали эксплицитное поисковое поведение. Таким образом, активизация поискового поведения в результате ошибки предсказания может лежать в основе феномена вероятностного соответствия.

Ключевые слова: обучение с подкреплением, положительная и отрицательная обратная связь, неопределенность, ошибка предсказания, поисковое поведение

Финансирование: Исследование выполнено при поддержке базового финансирования МЭГ-центра.

Одним из механизмов адаптивного поведения является обучение с подкреплением — тенденция повторять действия, за которыми последовала награда, и прекращать поведение, сопряженное с наказаниями. Вопрос о том, какое подкрепление — положительное или отрицательное — эффективнее для формирования навыка, до настоящего времени остается дискуссионным. В целом исследования обучения с подкреплением не позволяют сделать обоснованного

и непротиворечивого вывода о большей эффективности того или другого типа обратной связи (ОС) (Frank et al., 2004).

С другой стороны, существуют устойчивые индивидуально-возрастные различия в восприимчивости к положительной и отрицательной ОС. Более того, при определенных неврологических расстройствах может избирательно нарушаться обучаемость на основе только наград или только наказаний (для обзора см. Козунова, 2016). Действительно, функциональная организация системы кодирования подкрепления включает два относительно независимых пути для обработки положительной и отрицательной ОС. Так при получении неожиданной награды регистрируется мощный фазический ответ нейронов стриатума, чувствительных к повышению концентрации дофамина, а при отсутствии ожидаемой награды удлиняются временны́е интервалы между импульсами в их тонической активности (Schultz, Dickinson, 2000).

Критическим условием для кодирования сигналов ОС как подкрепления является фактор их неожиданности. Характерные изменения активности дофамин-чувствительных нейронов исчезают, как только награды становятся полностью предсказуемыми, а наказания — контролируемыми (Schultz, Dickinson, 2000). Человек во многих повседневных ситуациях получает неполную и непостоянную ОС. Можно предполагать, что в таких условиях повышенная чувствительность к случайным, нетипичным сигналам ОС в связи с их неожиданностью может препятствовать поведенческой адаптации. Мы исследовали, как влияют регулярные и редкие сигналы награды и наказания на поведение здоровых взрослых испытуемых при обучении в условиях частичной неопределенности исхода.

Метод

Мы использовали модифицированную методику вероятностного обучения при выборе из двух альтернатив. 29 взрослым испытуемым предлагались на выбор два абстрактных стимула, причем выбор одного из них приводил к выигрышу, а другого — к проигрышу в 70% случаев (частая ОС). В оставшихся 30% случаев испытуемые получали нерепрезентативную ОС: за выбором выгодного стимула следовал штраф, а за альтернативным решением — награда (редкая ОС). Мы провели серию из 4 таких тестов по 40 проб каждая, в которых использовались разные пары стимулов. Схема эксперимента приведена на рис. 1.

Результаты

Результаты показали, что за редким исключением (3 человека из 29) у испытуемых сформировался навык предпочтения более выгодного стимула. Однако мы наблюдали значительные индивидуальные различия в количестве опытов, необходимых для появления этого навыка. Около половины выборки вырабатывали его уже в первом тесте (14 из 29), а у оставшихся испытуемых он формировался постепенно в последующих тестах по мере тренировки. К последнему тесту различия в поведении этих двух подгрупп сглаживались до уровня статистической тенденции (рис. 2).

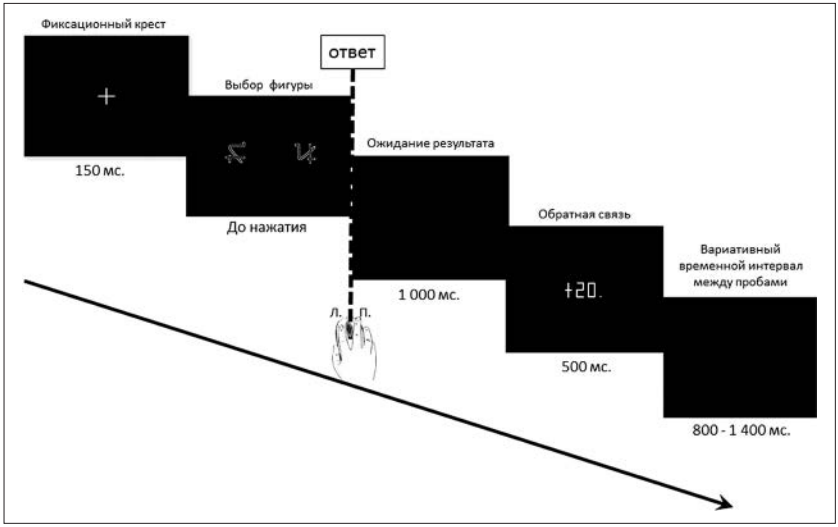


Рисунок 1. Испытуемому предлагалось выбрать одну из двух фигур, которые случайным образом менялись местами. Формой ответа служило нажатие на левую или правую кнопку пульта соответственно положению выбираемой фигуры. Через 1 секунду после нажатия кнопки выбора на экране появлялась обратная связь о результате сделанного выбора: выигрыш или штраф. По окончании серии из 40 таких проб испытуемому показывали общую сумму баллов, заработанных за всю игру.

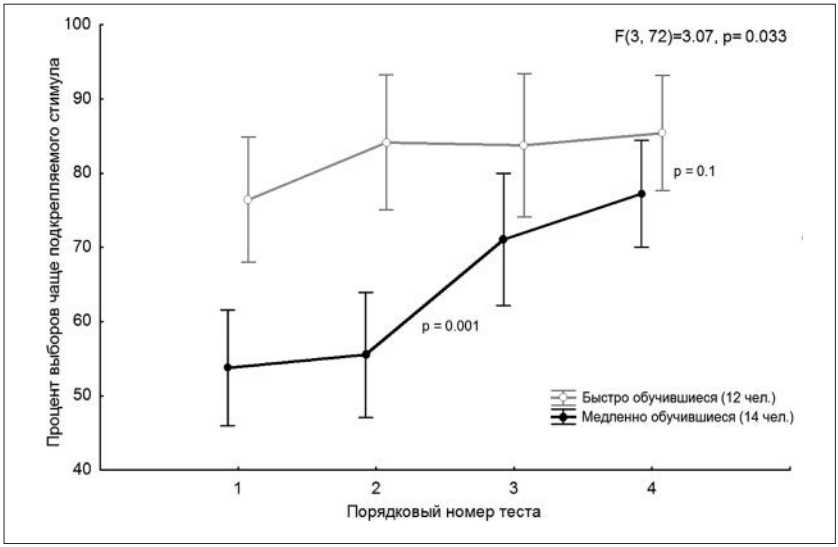


Рисунок 2. Динамика адаптивного поведения в зависимости от индивидуально-типологических особенностей темпа обучения

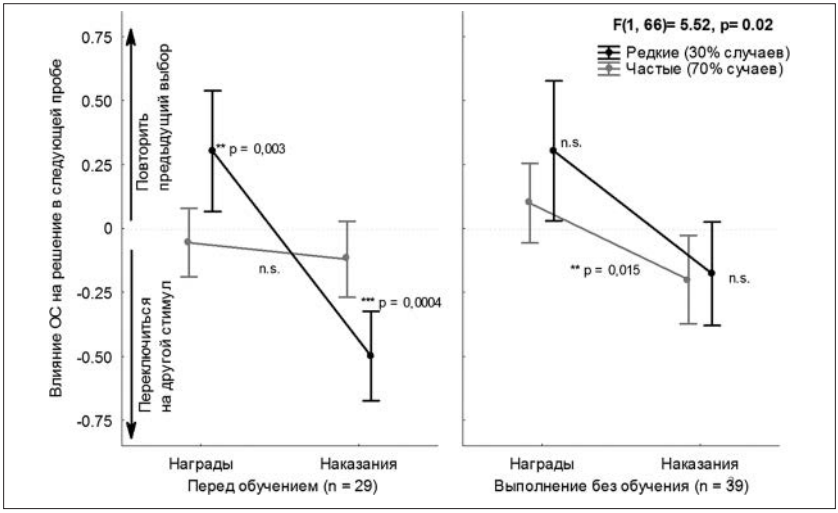


Рисунок 3. Чувствительность к редким сигналам награды и наказания как предиктор успешности обучения

Чтобы проверить гипотезу о роли неожиданных сигналов награды и наказания в формировании адаптивного навыка в условиях частичной неопределенности исхода, мы проанализировали, как они влияют на последующий выбор испытуемого. Мы сравнили случаи неуспешного обучения в отдельных тестах (50±15% выборов стимула) и начальные периоды обучения, предшествовавшие появлению предпочтения выгодного стимула.

Оказалось, что избирательная чувствительность к редким и нетипичным сигналам ОС была характерна только для тех испытуемых, которые впоследствии обучились успешно.

Парадоксально, но в этот период частые и легко предсказуемые сигналы ОС не оказывали на поведение испытуемых прямого влияния. В то же время случайные и нетипичные награды и наказания стабильно провоцировали их на выбор заведомо невыгодного стимула в следующей пробе. В отличие от них, испытуемые, которые так и не смогли обучиться, одинаково реагировали на высоковероятные и низковероятные сигналы ОС. На протяжении всего теста они демонстрировали слабую тенденцию повторять свой выбор после наград и переключаться на другой стимул после наказаний (рис. 3).

Обсуждение результатов

Результаты нашего эксперимента на уровне групповой статистики в точности соотносятся с данными нейроэкономических исследований (см. Shanks et al., 2002), в которых показано, что в условиях непостоянной ОС частота вырабатываемого поведения соответствует вероятности его подкрепления (см. рис. 2). Закон вероятностного соответствия представляет собой классический пример

неадаптивного поведения человека, так как объективно наиболее выгодной стратегией в таких условиях является постоянный выбор только чаще подкрепляемого стимула.

Большинство исследователей сходятся во мнении, что феномен вероятностного соответствия обусловлен иллюзорным восприятием испытуемыми нестабильности ОС как скрытой закономерности (Wolford et al., 2004). Однако эта интерпретация не раскрывает механизма принятия решения в зависимости от вероятностных сигналов ОС. Также без объяснения остается тот факт, что большинство людей по мере тренировки и накопления опыта постепенно преодолевают закон вероятностного соответствия (Shanks et al., 2002). Полученные нами данные дополняют эти наблюдения тем, что степень предпочтения выгодного стимула не только усиливается по мере выполнения задачи, но и переносится на последующие задачи с новыми стимулами (рис. 2).

В нашей работе впервые описано, что кроме адаптивных стратегий вероятностного соответствия и максимизации возможной награды значительную часть поведения испытуемых могут составлять периоды, в которых отсутствует предпочтение того или иного стимула. Именно на них целесообразно исследовать психологические механизмы обучения ввиду максимальной восприимчивости испытуемых к поступающей ОС. Мы показали, что непосредственно перед появлением адаптивного навыка испытуемые демонстрировали парадоксальную восприимчивость к нетипичным сигналам ОС (рис. 3). Это свидетельствует о том, что уже с самых первых опытов они имплицитно присваивали обоим стимулам относительную прагматическую ценность на основе высоковероятной ОС. Однако редкие рассогласования с этой оценкой провоцировали их на эксплицитное поисковое поведение. У тех испытуемых, которые так и не смогли обучиться в текущем тесте, дифференциация между высоковероятными и низковероятными сигналами ОС отсутствовала. По-видимому, формирование прогноза о результате собственного выбора и сличение актуального результата с ним является необходимым этапом поведенческой адаптации к условиям неопределенности исхода. Таким образом, активизация поискового поведения в результате ошибки предсказания может лежать в основе феномена вероятностного соответствия.

Литература

Козунова Г.Л. Обучение в условиях вероятностного подкрепления и его роль в адаптивном и дезадаптивном поведении человека // Современная зарубежная психология. 2016. Т. 5. № 4. С. 85 – 96. doi:10.17759/jmfp.2016050409

Frank M.J., Seeberger L.C., O'Reilly R.C. By carrot or by stick: cognitive reinforcement learning in parkinsonism // Science. 2004. Vol. 306. No. 5703. P. 1940 – 1943. doi:10.1126/science.1102941

Schultz W., Dickinson A. Neuronal coding of prediction errors // Annual Review of Neuroscience. 2000. Vol. 23. No. 1. P. 473 – 500. doi:10.1146/annurev.neuro.23.1.473

Shanks D.R., Tunney R.J., McCarthy J.D. A re-examination of probability matching and rational choice // Journal of Behavioral Decision Making. 2002. Vol. 15. No. 3. P. 233 – 250. doi:10.1002/bdm.413

Wolford G., Newman S. E., Miller M. B., Wig G. S. Searching for patterns in random sequences // Canadian Journal of Experimental Psychology / Révue canadienne de psychologie expérimentale. 2004. Vol. 58. No. 4. P. 221 – 228. doi:10.1037/h0087446

A Role of Probabilistic Rewards and Punishments in Decision Making during 2-Alternative Selection Task

Kozunova G. L.*, Voronin N. A., Venidictov V. V., Stroganova T. A.
chukhutova@gmail.com
Moscow State University of Psycology and Education

Abstract. In a typical probability learning task, participants are presented with a repeated choice between two response alternatives, one of which has a higher payoff probability than the other. Previous research has found that people match their response probabilities to payoff probabilities, thus diverging from the rational strategy of allocating all their responses to the high-payoff alternative. We investigated the role of probabilistic rewards and punishments in this phenomena. A group of 29 neurotypical adults completed four probabilistic selection tasks, each consisting of 40 trails. In the selection task, participants chose one of the two abstract stimuli. One response alternative was correct on 70% of trials and the other on 30% of trials. The results revealed that only half of our participants rapidly developed adaptive preference for an advantageous alternative while performing the first task in the sequence, whereas the other half gradually improved their success rate from the first to the fourth task. More interestingly, right before establishing an adaptive bias in their response probability, our participants were almost insensitive to the regular feedback signal. Instead, they showed paradoxical sensitivity to rare, non-representative feedback signals guiding them toward the disadvantageous choice. We hypothesized that just before switching towards a correct behavioral strategy, implicit representations of the stimuli's pragmatic value has been already formed, but rare deviations from a stimulus' predicted value activated a new behavioral search seeking to achieve a maximum expected reward. This seeking behavior may underlie the "probabilistic matching" phenomenon and explain why participants fail to maximize despite the apparent simplicity of the problem facing them.

Keywords: probabilistic learning, feedback, reward, punishment, uncertainty, prediction error, exploration behavior

Financing: The study was supported by the basic funding of the MEG Center Motion Perception Abnormalities in ASD.