

---

## ПСИХОЛОГИЯ ОБРАЗОВАНИЯ EDUCATIONAL PSYCHOLOGY

---

### Обучение в условиях вероятностного подкрепления и его роль в адаптивном и дезадаптивном поведении человека

*Козунова Г.Л.,*

*кандидат психологических наук, старший научный сотрудник, Центр нейробиологических исследований  
(МЭГ-центр), ФГБОУ ВО МГППУ, Москва, Россия,  
chukhutova@gmail.com*

В статье рассматривается обучение человека в условиях частично неопределенного исхода собственных действий, моделирующее один из механизмов адаптивного поведения в естественной среде. Базовые механизмы обучения детально изучены на модели условного рефлекса у животных в экспериментах, где определенное поведение подкреплялось одинаково, немедленно и многократно. В то же время нейрофизиологические основы возможности обучения у человека в условиях нерегулярного или отсроченного подкрепления, несмотря на возросший в последние годы интерес к ним, остаются малоизвестными. Значительный вклад в разработку этой проблемы внесли исследования психических и психоневрологических расстройств. Так, специфические изменения отдельных аспектов в обучении с вероятностным подкреплением обнаружены у пациентов с болезнью Паркинсона, синдромом Туретта, шизофренией, депрессией, тревожными расстройствами. В частности, показано, что восприимчивость к положительному и к отрицательному подкреплению могут нарушаться независимо друг от друга. Исходя из патогенетических механизмов этих состояний, можно сделать вывод о том, что ключевой структурой для этого типа обучения является поясная и орбитофронтальная кора, вовлеченная в двустороннее взаимодействие с нижележащими структурами стрио-паллидарного комплекса, лимбической системы и ядер ретикулярной формации ствола мозга.

**Ключевые слова:** обучение с подкреплением, неопределенность, ошибка предсказания, фронтальная кора, дофамин, серотонин, норадреналин, психические расстройства.

#### Для цитаты:

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

#### For citation:

Kozunova G.L. Training in terms of probabilistic reinforcements and its role in adaptive and maladaptive human behavior [Elektronnyy resurs]. *Journal of Modern Foreign Psychology*, 2016. Vol. 5, no. 4, pp. 85–96. doi: 10.17759/jmfp. 2016050409 (In Russ., Abstr. in Engl.).

### Введение

Фундаментальным свойством психики, общим для человека и для животных, на котором базируется адаптация к меняющимся условиям, является способность устанавливать статистическую или причинную связь между событиями, которые совпадают по времени или следуют одно за другим с коротким временным интервалом [4]. Это позволяет организму сформировать оптимальный моторный или вегетативный ответ в соответствии не только с наблюдаемым, но и с ожидаемым событием — т. е., обучаться.

Зачаточные формы обучения с подкреплением обнаруживаются у животных с очень простой организацией нервной системы, например, у кольчатых червей [39]. Особенность обучения высших животных состоит в том, что они воспринимают избыточное количество сигналов, некоторые из которых стабильно сопряжены друг с другом во времени и пространстве, другие — не встречаются вместе никогда, а часть

из них сочетаются между собой с некоторой долей неопределенности. При этом, несмотря на свою избыточность, воспринимаемые сигналы могут быть противоречивыми и недостаточными для однозначного прогноза последующего события.

Типичными примерами адаптации к естественной неопределенности исхода является восприятие объектов в затрудненных условиях, а также — что особенно важно для человека — социальное взаимодействие. Действительно, внутренние состояния и намерения другого субъекта принципиально недоступны непосредственному наблюдению: о них можно только судить по совокупности противоречивых внешних признаков [31].

В настоящей статье под обучением подразумевается поведенческая адаптация к условиям частично неопределенного исхода. Его механизмы наиболее отчетливо можно продемонстрировать на моделях нарушенной поведенческой адаптации — т. е., психических и психоневрологических расстройств [18].

### Виды обучения с подкреплением

В общем значении процесс образования связи между нейтральными событиями — внешними сенсорными стимулами или собственными действиями субъекта с безусловно или субъективно значимыми событиями (непосредственно связанными с наградой или наказанием) называют обучением с подкреплением.

В зависимости от природы связываемых событий различают два основных взаимно влияющих друг на друга вида обучения с подкреплением: классическое павловское обусловливание и инструментальное (оперантное) обучение [12]. В случае классического павловского обусловливания связь образуется между нейтральным сенсорным стимулом и следующим за ним субъективно значимым событием, например, подачей пищи или электрическим ударом. Установление связи можно наблюдать по вегетативной/моторной реакции (слюноотделение, замирание, поворот головы), упреждающей появление значимого стимула. Такое обучение лежит в основе оценки объектов или событий как привлекательных или отталкивающих, в зависимости от того, были ли они связаны в предыдущем опыте с положительным или отрицательным подкреплением. Предполагают, что аномально усиленный механизм классического обусловливания лежит в основе посттравматического стрессового расстройства (ПТСР). ПТСР характеризуется неконтролируемыми приступами страха и агрессии при столкновении с событиями, напоминающими те обстоятельства (условные сигналы), которые сопутствовали острым травматическим переживаниям в прошлом опыте (безусловный биологически значимый раздражитель), но в настоящий момент больше не указывают на опасность [22]. Например, для ветерана, участвовавшего в боевых действиях за пределами страны, таким сигналом может быть иностранная речь.

Другой вид обучения с подкреплением — инструментальное или оперантное обучение, которое также называют обучением методом проб и ошибок. Основное отличие оперантного обусловливания от классического состоит в том, что первым из событий является не внешний сигнал, а собственное действие субъекта, а вторым — значимое событие. Подкрепление при таком типе обучения является результатом собственного действия и служит для субъекта показателем соответствия выполненного действия цели поведения, т. е. обратной связью. Классические эксперименты по формированию инструментальных навыков проведены Э. Торндайком. Он помещал кошек в клетку, предоставляя им свободно действовать до тех пор, пока они не совершали то действие, которое приводило к открытию клетки. Впоследствии, когда этих животных снова помещали в эту клетку, они немедленно воспроизводили те эмпирически обнаруженные ими формы поведения, которые приводили к желаемому освобождению [44]. Закрепление форм поведения, которые приводят к желаемому результату, Э. Торндайк назвал законом эффекта. Преимуществом этого типа обучения, по

сравнению с павловским обусловливанием, является то, что субъект имеет возможность не только заранее подготовиться к значимому событию, но и активно на него повлиять, повышая для себя вероятность желательных событий и минимизируя опасные и неприятные последствия.

Возможно, в силу этого преимущества в поведении животных отмечается стойкая тенденция формировать инструментальный ответ на условный сигнал, даже если он не влияет на подкрепление. Данный феномен описан в литературе как перенос условия на операцию (PIT — Pavlovian-to-instrumental transfer) [12]. Например, в одной ситуации крыса привыкла получать пищевое подкрепление после звукового сигнала (павловское обусловливание). В другой ситуации та же крыса научилась нажимать на рычаг для того, чтобы получать пищу (инструментальное обучение). Если же во втором случае крыса услышит звук, который в другом контексте предупреждал о подаче пищи, она начинает нажимать на рычаг значительно чаще (перенос условия на операцию).

Аналогичные формы поведения у человека можно наблюдать в повседневных ситуациях. Например, проходя мимо вывески буфета (условный стимул), в котором он перекусывал раньше (первичное пищевое подкрепление), человек может решить зайти туда и что-нибудь себе купить (инструментальное действие). Изначально действие не входило в его планы, и он не был голоден. Формирование подобных форм поведения смоделировано в экспериментальной компьютерной игре с денежным вознаграждением [5].

Аномальный перенос условия на операцию может рассматриваться как возможный патогенетический механизм формирования рецидивирующей алкогольной зависимости [35]. Так, если после лечения обстановка, в которой живет пациент, существенно не меняется, он склонен возвращаться к прежним формам поведения, в том числе к приему алкоголя.

Таким образом, классическое павловское обусловливание и инструментальное обучение являются принципиально сходными процессами, в обоих случаях центральным компонентом ассоциации является эмоционально/биологически значимое событие (подкрепление) и формируется опережающая его моторная или вегетативная реакция.

### Условия формирования и распада ассоциативных связей

Образование и поддержание ассоциаций у человека и животных требует ряда условий, при невыполнении которых связь не формируется или уже сформированная связь распадается.

Следует отметить, что блокировка неактуальных ассоциаций и их распад являются такими же важными процессами для адаптивного поведения, как и их формирование. Так, например, установление связей между случайно или единично совпавшими событиями, не подтвержденное достаточным количеством повторе-

ний, может приводить к необоснованному и неадаптивному принятию решения (*jumping to conclusion*), что является характерной проблемой пациентов с шизофренией [19, 25].

С другой стороны, распад сформированных в прежнем опыте ассоциаций, которые со временем утратили свою актуальность (событие-сигнал перестало свидетельствовать о крайне значимом событии) нарушается при ПТСР, а также фобических, тревожных и обсессивно-компульсивных расстройствах, при которых у людей годами могут сохраняться оборонительные реакции на совершенно неопасные стимулы. Однако в типичных случаях оптимальный баланс между образованием и распадом условных связей достигается тремя основными условиями: временно-пространственной сопряженностью событий, их повторяемостью и фактором мотивации и внимания.

Прежде всего, ассоциация образуется между теми событиями, которые совпадают друг с другом по времени или следуют одно за другим с очень коротким временным интервалом. Эта простая идея, сформулированная еще в IV веке до нашей эры Аристотелем и получившая свое дальнейшее развитие в трудах британских философов-ассоцианистов XIX века, стала аксиомой, которая легла в основу современной экспериментальной психологии обучения [30].

Более того, как показывают экспериментальные исследования, временная сопряженность не только является непременным условием для связывания событий, но и детально кодируется как неотъемлемая часть психической репрезентации этой связи. В число кодируемых характеристик ассоциации входят длительность каждого события, порядок их следования, и временной интервал между ними [7]. Действительно, если события происходят с большими временными интервалами или нарушается их последовательность, то ассоциация не формируется [33]. Например, в большинстве экспериментов на крысах максимальным временным интервалом, достаточным для образования условной связи, было время не более 32–62 секунд, причем, чем короче этот временной интервал, тем эффективнее животные обучались [17]. Эти данные легли в основу гипотезы о том, что в процессе обучения связываются, по сути, не события А и В, а сенсорный след события А (*stimulus trace*) и события В [24].

Кроме временно-пространственной сопряженности событий, для формирования связи необходим достаточный уровень мотивации субъекта, поэтому в

классических экспериментах на животных используются пищевые или болевые подкрепления [11]. Например, одной из стандартных процедур подготовки к эксперименту с пищевым подкреплением является лишение животного пищи до снижения массы тела на 20 % от обычного веса [28], чем обеспечивается высокий уровень его восприимчивости к сигналам окружающей обстановки.

Следует отметить, что ключевым фактором обучения является не непосредственная биологическая значимость события, а его субъективная ценность для испытуемого с учетом его актуального состояния. Только в таком случае это событие можно назвать подкреплением или в психологической терминологии – мотивом [28].

Действительно, показано, что биологически необходимая пища выступает в качестве эффективного подкрепления поведения для голодной крысы только в том случае, если она ощущает ее вкус, но перестает выполнять эту функцию, когда пищу вводят непосредственно в желудок, минуя вкусовые рецепторы [15]. Наоборот, угрожающая здоровью инъекция наркотического вещества, приносящая при этом удовольствие, является крайне эффективным подкреплением поведения. Более того, одно и то же событие может выступать и подкреплением, и блокиратором поведения в зависимости от функционального состояния животного<sup>1</sup> [16].

Ввиду основной роли субъективной значимости подкрепления, эксперименты с участием людей позволяют использовать в качестве подкреплений денежные, социальные (надпись «Правильно!», изображение улыбающегося лица) или сенсорные знаки (красный цвет). При этом обычно воспроизводятся те же эффекты обучения, которые получены на животных.

Подкрепление включает в себя три относительно независимых компонента: эмоциональный (удовольствие), мотивационный (желание) и когнитивный (обучение), причем каждый из этих компонентов может нарушаться независимо от других [5]. Так, нарушения мотивационной системы при эндогенной депрессии характеризуются преимущественно утратой удовольствия при достижении результата, что снижает возможности обучения на основе положительной обратной связи. Феноменологически похожие трудности наблюдаются и у пациентов с болезнью Паркинсона, однако у них нарушается другой – мотивационный компонент подкрепления, т. е. само стремление выигрывать, в то время как эмоциональный компонент удовольствия остается сохранным [цит. по: 5].

<sup>1</sup> Ярким примером такой переоценки события может послужить один из недавних экспериментов на крысах. Крысы помещались в клетку, в разных частях которой находились три металлических рычага. Случайное нажатие животным одного из них приводило к насильственному впрыскиванию в рот высококонцентрированного раствора соли, контакт со вторым рычагом приводил к подаче раствора сахара, а третий (контрольный) рычаг не был связан с каким-либо последующим событием. В скором времени крысы стали сторониться первого рычага настолько, насколько им позволяло пространство клетки, и при этом многократно подходить ко второму рычагу, чтобы получить сладкий раствор, а третий – игнорировали. Однако разовая инъекция препаратов (дезоксикостерона и фуросемиды), имитирующих гормоны (ангиотестин II и альдостерон), сигнализирующие о недостатке в организме соли, необычным образом меняло поведение животных. Без какого-либо дополнительного обучения в той же самой клетке крысы начинали грызть и лизать тот рычаг, которого избегали ранее, несмотря на то, что рычаг был сделан из металла, и не обладал соленым вкусом, и на нем не оставалось никаких следов соли.

Третьим необходимым для формирования ассоциативной связи фактором является повторяемость сочетающихся событий, что проиллюстрировано многими экспериментами на животных и людях в виде характерных «кривых научения» [23]. Так, количество повторений, наряду с временным интервалом между событиями, позволяет рассматривать ассоциируемость событий как некую количественно измеримую величину, которая тем сильнее, чем короче интервал между событиями, чем больше количество сочетаний стимулов (и меньше их встречаемость отдельно друг от друга) и чем больше степень актуальной потребности субъекта и интенсивность стимулов.

В большинстве естественных ситуаций условие однозначного совпадения событий часто нарушается, поэтому происходит имплицитная статистическая оценка вероятности значимого события – фактически, Байесовская оценка вероятности. В 1970-х гг. появилось несколько многофакторных математических моделей обучения, рассматривающих его как постоянно обновляющуюся ассоциацию, которая по мере накопления типового опыта асимптотически стремится к максимально возможному предсказанию ожидаемых событий и контролю ситуации [43].

Наиболее важные выводы из этих моделей можно обобщить следующим образом: ассоциация формируется только с тем из множества окружающих стимулов, который наилучшим образом позволяет предсказать мотивационно значимое событие, обучение происходит лишь тогда, когда событие, наблюдаемое субъектом, не соответствует его ожиданиям [16]. В этом случае имеет место некоторая конкурентная борьба между актуальной информацией и информацией, накопленной в предыдущем индивидуальном опыте взаимодействия с объектами. Механизм байесовской оценки вероятности позволяет животным и человеку формировать прогноз последующего события, в том числе в таких условиях, когда совпадение событий носит вероятностный либо отсроченный характер, или в меняющихся условиях, требующих гибкой перестройки стратегии адаптивного поведения.

### Поведенческая адаптация к условиям вероятностного подкрепления

Принципиальная возможность животных и человека адаптировать поведение в соответствии с частотностью событий неоднократно показана в ряде экспериментов. Например, в одном из экспериментов испытуемым предъявляли световую вспышку, которая с разной вероятностью могла появиться или слева, или справа, и предлагали угадать, с какой стороны появится следующая вспышка. Оказалось, что прогнозы большинства испытуемых соответствовали частотности событий, т. е. если вспышка появлялась в 70% случаев справа, испытуемые примерно в 70% случаев «делали ставку» на правую сторону. Более того, когда экспериментатор без предупреждения начинал подавать вспышку в точном соответствии с прогнозом испытуемого, так чтобы он всегда «угадывал», испытуемые не замечали никакой

перемены и продолжали ожидать в 70% случаев вспышку справа, а в 30% случаев – слева [48]. Однако, как несложно посчитать, в реальности такая стратегия поведения не является оптимальной для успешного прогнозирования событий: если человек в 70% случаев выбирает стимул, при котором вероятность подкрепления составляет 70% (а в остальных 30% случаев выбирает альтернативу), тогда его прогноз окажется верным лишь в 58% случаев ( $(0.7 * 0.7) + (0.3 * 0.3)$ ). Более выгодной стратегией было бы каждый раз выбирать только более частотное событие – прогноз был бы верен в 70% случаев ( $0.7 * 1$ ). Именно такие выгодные стратегии вырабатывают животные (крысы) в подобных соотношениях вероятностей событий [26].

Объяснить природу этого различия, которое, как кажется, делает сравнение не в пользу человека, можно с позиции главного специфически человеческого фактора – адаптации к речевой среде.

Речевая среда, в отличие от предметной, характеризуется последовательным разворачиванием иерархической структуры. Например, глаголы используются вместе с существительными, наречия – с глаголами, предлоги сопутствуют существительным, обозначающим объект, а не субъект, и т. д. Благодаря этому по мере поступления естественного речевого материала и соответственно накопления лексических и грамматических признаков круг возможных вариантов последующего элемента сужается.

Эта особенность речи создает условия для непрерывного прогнозирования каждого последующего элемента речи, который предопределяется не только как наиболее часто встречающееся в этом контексте сочетание, но и как недостающее звено в имплицитно воспринимаемой иерархии.

По-видимому, прогнозирование на основе имплицитной иерархической структуры распространяется из сферы речевой коммуникации и на поведенческую адаптацию к любым чередующимся событиям, последовательность которых подчинена закономерности. Например, экспериментально показано, что у детей успешность в имплицитном освоении повторяющегося паттерна чередования зрительных стимулов является предиктором уровня вербальных способностей (особенно в сфере грамматической компетентности), а у детей с нарушениями речи такое обучение происходит значительно медленнее. Сходные трудности имплицитного обучения последовательности зрительных стимулов описаны при патологическом состоянии, для которого характерно нарушение грамматической организации речи – аграмматической афазии, связанной с поражением зоны Брока [13].

Интересно, что на ранних, дограмматических этапах освоения речи ребенок опирается не на иерархическую структуру языка, а преимущественно на частотные характеристики его элементов, позволяющие сегментировать поток речи на отдельные слова [40]. Иначе говоря, с точки зрения частотности слова могут быть представлены как наиболее часто встречающиеся в речи сочетания слогов.

В одном из обзоров приводится такой пример: сочетания, образующие слова «смешной» и «робот» встречаются в речи значительно чаще, чем сочетание «нойроб», образованное соединением последнего слога первого слова и первого слога второго слова [40]. Экспериментально показано, что способность сегментировать речь на основе частотности звукокомбинаций в зачаточном виде присуща и животным. Так, обезьяны обнаружили способность узнавать часто повторяющиеся псевдослова в 20-минутной аудиозаписи. Однако животным, по-видимому, недоступно восприятие иерархической структуры языка. Можно предполагать, что дограмматический уровень освоения речи обеспечивается преимущественно структурами правого полушария, которое может быть более чувствительно к частотности событий, а прогнозирование на основе иерархической структуры языка – левого, что может дать ключ к разгадке разных стратегий поведения человека и животных в условиях вероятностного подкрепления.

### Системные механизмы обучения в условиях вероятностного подкрепления

Механизм, по-видимому, специфически человеческой тенденции адаптировать поведенческий ответ в соответствии с ожидаемой частотностью событий может заключаться в особенностях функциональной специализации левого и правого полушарий. Разделение функций между полушариями мозга ярко проявляется лишь в патологических случаях, когда одно из полушарий повреждено, или когда межполушарные связи разрушаются оперативно (при лечении фармакорезистентной эпилепсии).

Возможности обучения в условиях вероятностного подкрепления у пациентов с повреждением межполушарных связей изучались группой исследователей под руководством Майкла Газзанига. Оказалось, что такие пациенты используют разные стратегии поведенческой адаптации к вероятностным событиям в зависимости от того, на какое из полушарий приходится функциональная нагрузка. Так, если обработка вероятностной информации протекает в правом полушарии, испытуемые постоянно ожидают события, которое они наблюдали чаще всего (оптимальная стратегия максимизации выигрыша, как у животных). Если аналогичная задача выполняется левым полушарием, происходит подстройка ожиданий к наблюдаемой частотности альтернативных событий [47].

Такие особенности поведенческой адаптации к условиям вероятностного подкрепления наблюдались не только при разрушении межполушарных связей. Сходные стратегии максимизации выигрыша присутствовали у пациентов с локализованными поражениями префронтальной коры правого или левого полушария. Эти наблюдения согласуются с многочисленными данными о ключевой роли орбитофронтальной и дорзальной префронтальной коры в вероятностном обучении [43].

Действительно, во время выполнения задач на вероятностное обучение регистрируется выраженная функ-

циональная активация латеральной префронтальной и передней поясной коры в ответ на получение обратной связи о последствиях собственного действия [34]. Электрофизиологическим коррелятом мониторинга эффективности собственного поведения считается негативный компонент вызванного потенциала, возникающий примерно через 250 миллисекунд после получения подкрепления или момента ожидаемого, но не поступившего подкрепления (FRN – feedback related negativity), который регистрируется на центральных лобных отведениях, предположительно соответствующих области передней поясной коры, преимущественно ее дорзальной части [38]. Чем более неожиданной, т. е. маловероятной с точки зрения индивидуального опыта субъекта, является полученная обратная связь, тем больше выражен по амплитуде этот компонент. Амплитуда FRN модулируется не только неожиданностью подкрепления, но и его величиной [43].

Также известно, что значительная часть нейронов передней поясной коры по-разному реагирует на позитивную и негативную обратную связь. Действительно, амплитуда ответа передней поясной коры зависит от знака ошибки предсказания: отсутствие ожидаемого подкрепления (негативная ошибка предсказания) обычно вызывает больший по амплитуде ответ, чем когда результат превосходит ожидания субъекта [45].

### Нейромодуляторные механизмы вероятностного обучения

Анатомически и функционально связи фронтальной коры с подкорковыми структурами, несомненно участвующими в кодировании подкрепления, имеют двусторонний характер.

К фронтальной коре подходят обширные восходящие проекции нейромодуляторных систем, которые оказывают влияние на функции исполнительного контроля: дофаминовой (от прилежащего ядра и вентральной тегментальной области), серотониновой (от ядра шва и миндалины), норадреналиновой (от голубого пятна) и ацетилхолиновой системы.

За последние десятилетия накоплен ряд доказательств главенствующей, хотя не исключительной, роли дофаминовой нейромодуляторной системы стриатума и прилежащего ядра в обеспечении обучения с подкреплением [46].

Так, активность дофамин-чувствительных нейронов прилежащего ядра и стриатума отражает все возможные виды ошибки предсказания [41]. Если вероятность подкрепления высока и предсказуема и субъект контролирует ситуацию, эти нейроны поддерживают неизменно высокий тонический (фоновый) уровень активности [2]. Если же неожиданно результат превосходит прогноз с точки зрения вероятности подкрепления, его величины или времени появления, они отвечают резким фазическим (функциональным) повышением активности [8]. Если произошла негативная ошибка предсказания, т. е. наблюдаемый результат оказался «хуже», чем ожидал субъект, активность таких нейронов временно пода-

вляется, в их тонической активности увеличиваются временные интервалы между разрядами [9].

По-видимому, такая разноплановая система кодирования для положительных и отрицательных ошибок предсказания является наиболее метаболически экономной [32].

Небольшая часть дофамин-чувствительных нейронов дает фазический ответ не только на положительное, но и на отрицательное подкрепление, а также на любые новые, неожиданные, интенсивные, привлекающие внимание сенсорные стимулы безотносительно к наличию подкрепления. Это дает основания полагать, что основная часть этих нейронов кодирует положительный или отрицательный знак события (valence), а другая часть — его интенсивность и/или неожиданность (salience) [43].

Интересно, что искусственная стимуляция обеих групп нейронов у животных может способствовать формированию у них условного рефлекса даже в тех условиях, в которых ассоциативные связи, как правило, не формируются, как бы имитируя неожиданность полностью предсказуемого подкрепления [43].

Так, обычно животные не устанавливают ассоциативной связи между подкреплением и стимулом В, если оно полностью предсказуемо другим условным сигналом А, даже если В всегда следует за сигналом А. При такой последовательности событий сигнал В не имеет никакого прогностического значения, чем оправдан эффект блокировки обусловливания [3]. Однако при электрической стимуляции дофамин-чувствительных нейронов в момент появления стимула В условный рефлекс все же формируется.

Стимуляции нейронов стриатума за счет неожиданности или новизны события обеспечивается взаимодействием между дофаминовой и серотониновой нейромодуляторными системами: возбуждающие проекции от серотонинэргических нейронов ядра шва в вентральную тегментальную область опосредованно могут вызывать активацию дофамин-чувствительных нейронов прилежащего ядра [3].

Однако функции серотониновой системы в регуляции механизмов подкрепления этим не ограничиваются. Ядро шва содержит около 65% серотонинэргических нейронов, которые также проявляют фазическую активность в ответ на обратную связь и посылают восходящие проекции в регуляторные области коры (префронтальную и переднюю поясную), а также миндалину.

Отличительной особенностью серотонинэргических нейронов является то, что уровень их тонической активности повышается в ответ на сигнал о последующем подкреплении пропорционально величине подкрепления и не снижается весь период его ожидания [20]. По достижении подкрепления серотонинэргические нейроны дают фазический ответ вне зависимости от фактора предсказуемости.

Влияние серотониновой нейромодуляторной системы на поведение определяется свойствами тонической и фазической активности этих нейронов. Тонический уровень серотонина, повышение которого

сопровождает ожидание подкрепления, напрямую связан с положительным эмоциональным настроением, и его выраженное снижение наблюдается при эндогенной депрессии [32]. Фазическая модуляция этой системы функций фронтальной коры может лежать в основе механизма поддержания мотивации субъекта в условиях отсроченного подкрепления.

Действительно, блокировка этого нейромодуляторного пути приводит к импульсивному поведению у людей и животных: отсроченное подкрепление утрачивает для них привлекательность, они предпочитают только те действия, которые дают немедленный результат, даже если величина отсроченного подкрепления несопоставимо выше [29].

Вторая важнейшая функция серотониновой нейромодуляторной системы — участие в психических процессах, требующих когнитивной гибкости [32].

Основной парадигмой для исследования гибкости поведения является задача на реверсивное обучение [42]. Человек или животное учится выбирать из двух стимулов тот, который чаще другого приносит подкрепление (обычно в 80% случаев > 20%). Затем этот стимул перестает так часто приносить подкрепление (вероятность подкрепления снижается до 20%), в то время как альтернативный стимул начинает подкрепляться чаще.

Показано, что уровень серотонина в орбитофронтальной коре, как у крыс, так и у людей, предопределяет индивидуальные различия в выполнении задачи на реверсивное обучение. Субъекты с низким уровнем серотонина демонстрировали множественные персеверации. После смены условий подкрепления они инертно продолжали выбирать прежний стимул, который уже перестал давать желаемый результат [27].

Изложенные выше свойства серотонина привели к появлению гипотезы о том, что функциональный уровень этого нейромодулятора определяет взаимные переходы между двумя основными формами инструментального поведения. Переход от привычных действий к целенаправленному поведению требует повышения фазической активности серотониновой системы, тогда как закрепление нового паттерна поведения и превращение его в привычку, наоборот, сопровождается ее снижением [29].

Третьей нейромодуляторной системой, от которой зависит процесс обучения, являются норадренэргические нейроны голубого пятна (locus coeruleus), регулирующие общий уровень возбуждения нервной системы (arousal) и внимания [6].

Общая восприимчивость организма к сигналам окружающей среды (готовность к обучению) зависит от уровня тонической активности НА-нейронов, но не линейно, а по типу колоколообразной кривой в соответствии классическим законом Йеркса—Додсона. Так, крайне низкий уровень норадреналина сопровождается невнимательностью и сонливостью, а очень высокий, наоборот, приводит к хаотичному возбуждению и отвлекаемости [14]. Фазический ответ норадренэргических нейронов наблюдается при целенаправленном поведении в ответ на появление условных

сигналов о подкреплении и имеет два разнесенных во времени пика активности [28]. Первый пик может отражать переход от непроизвольного «автоматического» внимания, привлекаемого новизной и интенсивностью внешних стимулов, к произвольному вниманию, направляемому внутренними мотивами субъекта [14].

Поскольку в норадренэргических нейронах голубого пятна первый пик активности возникает позже, чем в орбитофронтальной коре, можно предполагать, что активность этой нейромодуляторной системы усиливается напрямую через нисходящие проекции орбитофронтальной коры. Особенностью второго пика активности является то, что его амплитуда прямо зависит от ожидаемой величины подкрепления, и ее увеличение сопровождается повышением скорости и точности реагирования субъекта [10]. Модуляция второго пика активности этих нейронов может обеспечиваться через обширные структурно-функциональные связи голубого пятна с вентральной тегментальной областью, которая играет ключевую роль в «кодировании» величины подкрепления [41].

По-видимому, второй пик фазической активности норадренэргических нейронов голубого пятна может отражать соотношение степени мобилизации сенсорных, моторных и регуляторных ресурсов мозга с силой мотивации субъекта [27].

Таким образом, нейрофизиологические механизмы обучения с вероятностным подкреплением имеют сложную многокомпонентную организацию, в основе которой лежит петля реципрокного взаимодействия поясной и орбитофронтальной коры и распределенных нейромодуляторных систем. Различные пути этого взаимодействия вносят собственный вклад в оценку текущей и долгосрочной значимости событий, регуляцию уровня внимания и мотивации. Эти же системы мозга играют ключевую роль в патогенезе большинства психических и психоневрологических расстройств [32]. Неудивительно, что для всех этих патологических состояний характерны нарушения обучения, специфика которых в каждом конкретном случае будет подробнее рассмотрена ниже.

### **Нарушения обучения при психических и психоневрологических расстройствах**

Наиболее показательной моделью для исследования роли дофаминэргической системы в кодировании положительного подкрепления может служить болезнь Паркинсона, так как характерная для этого состояния обширная потеря дофамин-чувствительных нейронов является его несомненным патогенетическим механизмом и основным метаболическим нарушением. Эксперименты, направленные на сравнительное изучение обучения на основе положительного и отрицательного подкрепления с участием пациентов с болезнью Паркинсона, показали, что эти две системы подкрепления могут нарушаться независимо друг от друга. Так, у пациентов с болезнью Паркинсона страдает способность обучаться на основе положительного подкрепления, в то время как чувствительность к негативному подкреплению у них остается такой же, как у здоровых людей того же возраста.

Более того, лечение болезни Паркинсона препаратами, приводящими к подъему уровня дофамина в стриопаллидарной системе мозга, могут менять эти особенности обучения на противоположные. То есть агонисты дофамина способны настолько повысить чувствительность к положительному подкреплению, что под влиянием лечения пациенты с болезнью Паркинсона превосходят здоровых людей того же возраста по способности учиться на положительном опыте. Одновременно с этим повышенный уровень дофамина снижает чувствительность человека к отрицательной обратной связи [21].

Противоположный болезни Паркинсона патогенетический механизм лежит в основе синдрома Туретта, при котором причиной характерных тиков, навязчивых движений и вокализаций является аномальное повышение уровня дофамина в стриатуме.

Показательно, что в обучении с вероятностным подкреплением у этой группы пациентов и у пациентов с болезнью Паркинсона до терапии наблюдаются противоположные психологические особенности. Первые эффективнее обучаются на основе положительного подкрепления (также как больные болезнью Паркинсона после лечения). Лечение пациентов с синдромом Туретта блокаторами рецепторов дофамина (типа D2) делает их более чувствительными к отрицательному подкреплению, чем к положительному (как у пациентов с болезнью Паркинсона до лечения) [36].

По-видимому, болезнь Паркинсона и синдром Туретта являются крайними вариантами отклонения уровня дофамина, в то время как сбалансированный уровень дофамина у здоровых людей обычно отражается в одинаковой успешности обучения на основе положительного и отрицательного подкрепления [21].

Однако уровень дофамина в стриопаллидарной системе не всегда однозначно определяет его в префронтальной коре. Например, у пациентов с шизофренией на фоне повышенного количества рецепторов к дофамину в стриатуме понижен уровень дофамина на уровне префронтальной коры [21]. Кроме того, у них отмечаются спонтанные повышения активности дофамин-чувствительных нейронов безотносительно к подкреплению, что как бы имитирует эффект «неожиданности» стимула или позитивную ошибку предсказания, способствуя формированию избыточных связей [1].

Примером неадаптивного установления ассоциаций при шизофрении может послужить экспериментально обнаруженный у них феномен аберрантного обучения. При выполнении задачи на вероятностное обучение пациенты ошибочно связывали подкрепление с второстепенным признаком стимулов, который в действительности не имел отношения к подкреплению [19].

В силу этих особенностей пациенты с шизофренией в условиях вероятностного подкрепления часто меняли стратегию принятия решения, и как следствие у них были снижены показатели обучения на основе как положительного, так и отрицательного подкрепления [37].

Еще одним патологическим состоянием, для которого также характерны аномалии дофаминовой и норадре-

налиновой нейромодуляторных систем, является синдромом гиперактивности с дефицитом внимания (СДВГ). У этой категории пациентов снижен уровень как тонической, так и фазической активности дофаминэргических нейронов как в стриатуме, так и в префронтальной коре, однако повышен уровень норадреналина [1]. Характерной поведенческой особенностью этих детей и молодых людей является отвлекаемость и импульсивность, понимаемая как нечувствительность к отсроченным подкреплениям. Фактически пациенты с СДВГ могут эффективно обучаться только на основе немедленного подкрепления, что может быть обусловлено сниженным уровнем дофамина в орбитофронтальной коре. Сходные тенденции в поведении отмечаются и у людей с наркотической и игровой зависимостью [32].

Приведенный обзор особенностей обучения с вероятностным подкреплением при различных психических и психоневрологических расстройствах демонстрирует роль распределенных нейромодуляторных систем мозга в обеспечении оптимального функционирования высших регуляторных областей мозга. При этом следует отметить, что наиболее подробно изучена петля двустороннего взаимодействия лобной коры с системой дофамин-чувствительных нейронов стриатума. Однако несомненный вклад остальных нейромодуляторных систем в процессы обучения, а также их взаимодействие между собой до настоящего времени остаются недостаточно изученными и спорными вопросами, которые требуют дальнейшего экспериментального исследования.

### Благодарности

Работа выполнена при базовом финансировании МЭГ-центра Министерством образования и науки РФ. Выражаю признательность Татьяне Александровне Строгановой за помощь в подготовке статьи.

### ЛИТЕРАТУРА

1. A dynamic developmental theory of attention-deficit/hyperactivity disorder (ADHD) predominantly hyperactive/impulsive and combined subtypes / T. Sagvolden [et al.] // Behavioral and Brain Sciences. 2005. Vol. 28. № 3. P. 397–418. doi: 10.1017/S0140525X05000075
2. A causal link between prediction errors, dopamine neurons and learning / E.E. Steinberg [et al.] // Nature neuroscience. 2013. Vol. 16. № 3. P. 966–973. doi: 10.1038/nn.3413
3. A glutamatergic reward input from the dorsal raphe to ventral tegmental area dopamine neurons / J. Qi [et al.] // Nature communications. 2014. Vol. 5. Art. 5390. doi: 10.1038/ncomms6390
4. Alloy L.B., Tabachnik N. Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information // Psychological review. 1984. Vol. 91. №. 1. P. 112–149. doi: 10.1037/0033-295X.91.1.112
5. Assessment of reward responsiveness in the response bias probabilistic reward task in rats: Implications for cross-species translational research / A. Der-Avakian [et al.] // Translational psychiatry. 2013. Vol. 3. № 8. doi: 10.1038/tp.2013.74
6. Aston-Jones G., Cohen J.D. An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance // Annual Review of Neuroscience. 2005. Vol. 28. P. 403–450. doi: 10.1146/annurev.neuro.28.061604.135709
7. Balsam P.D., Drew M.R., Yang C. Timing at the start of associative learning // Learning and Motivation. 2002. Vol. 33. № 1. P. 141–155. doi: 10.1006/lmot.2001.1104
8. Bayer H.M., Glimcher P.W. Midbrain dopamine neurons encode a quantitative reward prediction error signal // Neuron. 2005. Vol. 47. № 1. P. 129–141. doi: 10.1016/j.neuron.2005.05.020
9. Bayer H.M., Lau B., Glimcher P.W. Statistics of midbrain dopamine neuron spike trains in the awake primate // Journal of Neurophysiology. 2007. Vol. 98. № 3. P. 1428–1439. doi: 10.1152/jn.01140.2006
10. Bouret S., Richmond B.J. Sensitivity of locus ceruleus neurons to reward value for goal-directed actions // The Journal of Neuroscience. 2015. Vol. 35. № 9. P. 4005–4014. doi: 10.1523/JNEUROSCI.4553-14.2015
11. Bourgeois A., Chelazzi L., Vuilleumier P. How motivation and reward learning modulate selective attention // Progress in Brain Research. 2016. Vol. 229. P. 325–342. doi: 10.1016/bs.pbr.2016.06.004
12. Cartoni E., Puglisi-Allegra S., Baldassarre G. The three principles of action: A Pavlovian-instrumental transfer hypothesis // Frontiers in behavioral neuroscience. 2013. Vol. 7. P. 1–11. doi: 10.3389/fnbeh.2013.00153
13. Conway C.M., Christiansen M.H. Sequential learning in non-human primates // Trends in cognitive sciences. 2001. Vol. 5. № 12. P. 539–546. doi: 10.1016/S1364-6613(00)01800-3
14. Corbetta M., Patel G., Shulman G.L. The reorienting system of the human brain: From environment to theory of mind // Neuron. 2008. Vol. 58. № 3. P. 306–324. doi: 10.1016/j.neuron.2008.04.017
15. Cytawa J., Trojnar W. The state of pleasure and its role in instrumental conditioning // *Activitas nervosa superior*. 1976. Vol. 18. № 1–2. P. 92–96.
16. Dayan P., Berridge K.C. Model-based and model-free Pavlovian reward learning: Revaluation, revision, and revelation // Cognitive, Affective, & Behavioral Neuroscience. 2014. Vol. 14. № 2. P. 473–492. doi: 10.3758/s13415-014-0277-8
17. Dickinson A., Watt A., Griffiths W.J.H. Free-operant acquisition with delayed reinforcement // Comparative and Physiological Psychology. 1992. Vol. 45. № 3. P. 241–258.
18. Dimensional psychiatry: Mental disorders as dysfunctions of basic learning mechanisms / A. Heinz [et al.] // Journal of Neural Transmission. 2016. Vol. 123. № 8. P. 809–821. doi: 10.1007/s00702-016-1561-2

19. Do patients with schizophrenia exhibit aberrant salience? / J.P. Roiser [et al.] // *Psychological medicine*. 2009. Vol. 39. № 2. P. 199–209. doi: 10.1017/s0033291708003863
20. Dorsal raphe neurons signal reward through 5-HT and glutamate / Z. Liu [et al.] // *Neuron*. 2014. T. 81. № 6. P. 1360–1374. doi: 10.1016/j.neuron.2014.02.010
21. Frank M.J., Seeberger L.C., O'reilly R.C. By carrot or by stick: Cognitive reinforcement learning in parkinsonism // *Science*. 2004. Vol. 306. № 5703. P. 1940–1943. doi: 10.1126/science.1102941
22. From Pavlov to PTSD: The extinction of conditioned fear in rodents, humans, and anxiety disorders / M.B. VanElzakker [et al.] // *Neurobiology of learning and memory*. 2014. Vol. 113. P. 3–18. doi: 10.1016/j.nlm.2013.11.014
23. Gallistel C.R., Fairhurst S., Balsam P. The learning curve: Implications of a quantitative analysis // *Proceedings of the national academy of Sciences of the united States of America*. 2004. Vol. 101. № 36. P. 13124–13131. doi: 10.1073/pnas.0404965101
24. Gershman S.J. A Unifying Probabilistic View of Associative Learning // *PLoS Computational Biology*. 2015. Vol. 11. № 11. P. 1–20. doi: 10.1371/journal.pcbi.1004567
25. Guillain O., Abi-Dargham A., Laruelle M. Neurobiology of dopamine in schizophrenia // *International review of neurobiology*. 2007. Vol. 78. P. 1–39. doi: 10.1016/S0074-7742(06)78001-1
26. Hinson J.M., Staddon J.E.R. Matching, maximizing, and hill climbing // *Journal of the experimental analysis of behavior*. 1983. Vol. 40. № 3. P. 321–331. doi: 10.1901/jeab.1983.40-321
27. Hofmeister J., Sterpenich V. A role for the locus ceruleus in reward processing: Encoding behavioral energy required for goal-directed actions // *The Journal of Neuroscience*. 2015. Vol. 35. № 29. P. 10387–10389. doi: 10.1523/JNEUROSCI.1734-15.2015
28. Holroyd C.B., Coles M.G.H. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity // *Psychological review*. 2002. Vol. 109. № 4. P. 679–709. doi: 10.1037/0033-295X.109.4.679
29. Homberg J.R. Serotonin and decision making processes // *Neuroscience & Biobehavioral Reviews*. 2012. Vol. 36. № 1. P. 218–236. doi: 10.1016/j.neubiorev.2011.06.001
30. Kirkpatrick K., Balsam P.D. Associative learning and timing // *Current opinion in behavioral sciences*. 2016. Vol. 8. P. 181–185. doi: 10.1016/j.cobeha.2016.02.023
31. Ma W.J., Jazayeri M. Neural coding of uncertainty and probability // *Annual review of neuroscience*. 2014. Vol. 37. P. 205–220. doi: 10.1146/annurev-neuro-071013-014017
32. Maia T.V., Frank M.J. From reinforcement learning models to psychiatric and neurological disorders // *Nature neuroscience*. 2011. Vol. 14. № 2. P. 154–162. doi: 10.1038/nrn.2723
33. Molet M., Miller R.R. Timing: An attribute of associative learning // *Behavioural processes*. 2014. Vol. 101. P. 4–14. doi: 10.1016/j.beproc.2013.05.015
34. Neural mechanisms supporting flexible performance adjustment during development / E.A. Crone [et al.] // *Cognitive, Affective, & Behavioral Neuroscience*. 2008. Vol. 8. № 2. P. 165–177. doi: 10.3758/CABN.8.2.165
35. Pavlovian-to-instrumental transfer in alcohol dependence: A pilot study / M. Garbusow [et al.] // *Neuropsychobiology*. 2014. Vol. 70. № 2. P. 111–121. doi: 10.1159/000363507
36. Pharmacological modulation of subliminal learning in Parkinson's and Tourette's syndromes / S. Palminteri [et al.] // *Proceedings of the National Academy of Sciences*. 2009. Vol. 106. № 45. P. 19179–19184. doi: 10.1073/pnas.0904035106
37. Probabilistic reversal learning in schizophrenia: Stability of deficits and potential causal mechanisms / L.F. Reddy [et al.] // *Schizophrenia bulletin*. 2016. Vol. 42. № 4. P. 942–951. doi: 10.1093/schbul/sbv226
38. Reinforcement-related brain potentials from medial frontal cortex: Origins and functional significance / S. Nieuwenhuis [et al.] // *Neuroscience & Biobehavioral Reviews*. 2004. Vol. 28. № 4. P. 441–448. doi: 10.1016/j.neubiorev.2004.05.003
39. Robinson J.S. Stimulus substitution and response learning in the earthworm // *Journal of comparative and physiological psychology*. 1953. Vol. 46. № 4. P. 262–266. doi: 10.1037/h0056151
40. Saffran J.R., Aslin R.N., Newport E.L. Statistical learning by 8-month-old infants. 1996. Vol. 274. № 5294, P. 1926–1928.
41. Schultz W. Predictive reward signal of dopamine neurons // *Journal of neurophysiology*. 1998. Vol. 80. № 1. P. 1–27.
42. The neural basis of reversal learning: An updated perspective / A. Izquierdo [et al.] // *Neuroscience*. 2016. doi: 10.1016/j.neuroscience.2016.03.021
43. The processing of unexpected positive response outcomes in the mediofrontal cortex / N.K. Ferdinand [et al.] // *The Journal of Neuroscience*. 2012. Vol. 32. № 35. P. 12087–12092. doi: 10.1523/JNEUROSCI.1410-12.2012
44. Thorndike E.L. *Animal intelligence: Experimental studies* // Transaction Publishers. 1965.
45. Walsh M.M., Anderson J.R. Learning from delayed feedback: Neural responses in temporal credit assignment // *Cognitive, Affective, & Behavioral Neuroscience*. 2011. Vol. 11. № 2. P. 131–143. doi: 10.3758/s13415-011-0027-0
46. Weismüller B., Bellebaum C. Expectancy affects the feedback-related negativity (FRN) for delayed feedback in probabilistic learning // *Psychophysiology*. 2016. Vol. 53. № 11. P. 1739–1750. doi: 10.1111/psyp.12738
47. Wolford G., Miller M.B., Gazzaniga M. The left hemisphere's role in hypothesis formation [Электронный ресурс] // *Journal of Neuroscience*. 2000. Vol. 20. № 6. P. 1–4. URL: <http://www.jneurosci.org/content/jneuro/20/6/RC64.full.pdf> (дата обращения: 27.12.2016).
48. Yellott J.I. Probability learning with noncontingent success // *Journal of mathematical psychology*. 1969. Vol. 6. № 3. P. 541–575. doi: 10.1016/0022-2496(69)90023-6

## Reinforcement learning in probabilistic environment and its role in human adaptive and maladaptive behavior

**Kozunova G.L.,**

*candidate of psychological sciences, Senior Research Fellow, Centre for neuro-cognitive studies (MEG-center),  
Moscow State University of Psychology and Education,  
chukhutova@gmail.com*

The article discusses human training in conditions of partly uncertain outcomes of his/her actions that models one of the mechanisms of adaptive behavior in natural environment. Basic learning mechanisms are studied in details through modelling conditional reflexes of animals in experiments, where a certain behavior is reinforced similarly, immediately and repeatedly. At the same time, neurophysiological foundations of learning opportunities in humans under conditions of irregular or delayed reinforcements, despite increased interest to them in recent years, remain poorly studied. Research of mental and neuropsychiatric disorders has made a significant contribution to the development of this problem. Thus, the specific changes in some aspects of learning with probabilistic reinforcement were found in patients with Parkinson's disease, Tourette's syndrome, schizophrenia, depression, and anxiety disorders. In particular, it is shown that susceptibility to positive and negative reinforcement can be violated independently. Taking into consideration the pathogenetic mechanisms of these conditions, it can be concluded that the key structure for this type of training is the cingulate cortex and orbito-frontal cortex involved in bilateral interaction with underlying structures of striatal system, the limbic system and cores of reticular formations of the brain stem.

**Keywords:** reinforcement learning, uncertainty, prediction error, frontal cortex, dopamine, serotonin, norepinephrine, mental disorders.

### Acknowledgements

This work was supported by core funding to MEG-Center from the Ministry of Education and Science of the Russian Federation. I am grateful to Tatiana Stroganova for her help in preparation this article.

### REFERENCES

1. Sagvolden T. et al. A dynamic developmental theory of attention-deficit/hyperactivity disorder (ADHD) predominantly hyperactive/impulsive and combined subtypes. *Behavioral and Brain Sciences*, 2005. Vol. 28, no. 3, pp. 397–418. doi: 10.1017/S0140525X05000075
2. Steinberg E.E. et al. A causal link between prediction errors, dopamine neurons and learning. *Nature neuroscience*, 2013. Vol. 16, no. 3, pp. 966–973. doi: 10.1038/nn.3413
3. Qi J. et al. A glutamatergic reward input from the dorsal raphe to ventral tegmental area dopamine neurons. *Nature communications*, 2014. Vol. 5, art. 5390. doi: 10.1038/ncomms6390
4. Alloy L.B., Tabachnik N. Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological review*, 1984. Vol. 91, no. 1, pp. 112–149. doi: 10.1037/0033-295X.91.1.112
5. Der-Avakian A. et al. Assessment of reward responsiveness in the response bias probabilistic reward task in rats: implications for cross-species translational research. *Translational psychiatry*, 2013. Vol. 3, no. 8. doi: 10.1038/tp.2013.74
6. Aston-Jones G., Cohen J.D. An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, 2005. Vol. 28, pp. 403–450. doi: 10.1146/annurev.neuro.28.061604.135709
7. Balsam P.D., Drew M.R., Yang C. Timing at the start of associative learning. *Learning and Motivation*, 2002. Vol. 33, no. 1, pp. 141–155. doi: 10.1006/lmot.2001.1104
8. Bayer H.M., Glimcher P.W. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, 2005. Vol. 47, no. 1, pp. 129–141. doi: 10.1016/j.neuron.2005.05.020
9. Bayer H.M., Lau B., Glimcher P.W. Statistics of midbrain dopamine neuron spike trains in the awake primate. *Journal of Neurophysiology*, 2007. Vol. 98, no. 3, pp. 1428–1439. doi: 10.1152/jn.01140.2006
10. Bouret S., Richmond B.J. Sensitivity of locus coeruleus neurons to reward value for goal-directed actions. *The Journal of Neuroscience*, 2015. Vol. 35, no. 9, pp. 4005–4014. doi: 10.1523/JNEUROSCI.4553-14.2015
11. Bourgeois A., Chelazzi L., Vuilleumier P. How motivation and reward learning modulate selective attention. *Progress in Brain Research*, 2016. Vol. 229, pp. 325–342. doi: 10.1016/bs.pbr.2016.06.004
12. Cartoni E., Puglisi-Allegra S., Baldassarre G. The three principles of action: A Pavlovian-instrumental transfer hypothesis. *Frontiers in behavioral neuroscience*, 2013. Vol. 7, pp. 1–11. doi: 10.3389/fnbeh.2013.00153
13. Conway C.M., Christiansen M.H. Sequential learning in non-human primates. *Trends in cognitive sciences*, 2001. Vol. 5, no. 12, pp. 539–546. doi: 10.1016/S1364-6613(00)01800-3

14. Corbetta M., Patel G., Shulman G.L. The reorienting system of the human brain: From environment to theory of mind. *Neuron*, 2008. Vol. 58, no. 3, pp. 306–324. doi: 10.1016/j.neuron.2008.04.017
15. Cytawa J., Trojnar W. The state of pleasure and its role in instrumental conditioning. *Activitas nervosa superior*, 1976. Vol. 18, no. 1–2, pp. 92–96.
16. Dayan P., Berridge K.C. Model-based and model-free Pavlovian reward learning: Revaluation, revision, and revelation. *Cognitive, Affective, & Behavioral Neuroscience*, 2014. Vol. 14, no. 2, pp. 473–492. doi: 10.3758/s13415-014-0277-8
17. Dickinson A., Watt A., Griffiths W.J.H. Free-operant acquisition with delayed reinforcement. *Comparative and Physiological Psychology*, 1992. Vol. 45, no. 3, pp. 241–258.
18. Heinz A. et al. Dimensional psychiatry: Mental disorders as dysfunctions of basic learning mechanisms. *Journal of Neural Transmission*, 2016. Vol. 123, no. 8, pp. 809–821. doi: 10.1007/s00702-016-1561-2
19. Roiser J.P. et al. Do patients with schizophrenia exhibit aberrant salience? *Psychological medicine*, 2009. Vol. 39, no. 2, pp. 199–209. doi: 10.1017/s0033291708003863
20. Liu Z. et al. Dorsal raphe neurons signal reward through 5-HT and glutamate. *Neuron*, 2014. Vol. 81, no. 6, pp. 1360–1374. doi: 10.1016/j.neuron.2014.02.010
21. 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, pp. 1940–1943. doi: 10.1126/science.1102941
22. VanElzakker M.B. et al. From Pavlov to PTSD: The extinction of conditioned fear in rodents, humans, and anxiety disorders. *Neurobiology of learning and memory*, 2014. Vol. 113, pp. 3–18. doi: 10.1016/j.nlm.2013.11.014
23. Gallistel C.R., Fairhurst S., Balsam P. The learning curve: Implications of a quantitative analysis. *Proceedings of the national academy of Sciences of the united States of america*, 2004. Vol. 101, no. 36, pp. 13124–13131. doi: 10.1073/pnas.0404965101
24. Gershman S.J. A Unifying Probabilistic View of Associative Learning. *PLoS Computational Biology*, 2015. Vol. 11, no. 11, pp. 1–20. doi: 10.1371/journal.pcbi.1004567
25. Guillin O., Abi-Dargham A., Laruelle M. Neurobiology of dopamine in schizophrenia. *International review of neurobiology*, 2007. Vol. 78, pp. 1–39. doi: 10.1016/S0074-7742(06)78001-1
26. Hinson J.M., Staddon J.E.R. Matching, maximizing, and hill-climbing. *Journal of the experimental analysis of behavior*, 1983. Vol. 40, no. 3, pp. 321–331. doi: 10.1901/jeab.1983.40-321
27. Hofmeister J., Sterpenich V. A role for the locus ceruleus in reward processing: encoding behavioral energy required for goal-directed actions. *The Journal of Neuroscience*, 2015. Vol. 35, no. 29, pp. 10387–10389. doi: 10.1523/JNEUROSCI.1734-15.2015
28. Holroyd C.B., Coles M.G.H. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological review*, 2002. Vol. 109, no. 4, pp. 679–709. doi: 10.1037/0033-295X.109.4.679
29. Homberg J.R. Serotonin and decision making processes. *Neuroscience & Biobehavioral Reviews*, 2012. Vol. 36, no. 1, pp. 218–236. doi: 10.1016/j.neubiorev.2011.06.001
30. Kirkpatrick K., Balsam P.D. Associative learning and timing. *Current opinion in behavioral sciences*, 2016. Vol. 8, pp. 181–185. doi: 10.1016/j.cobeha.2016.02.023
31. Ma W.J., Jazayeri M. Neural coding of uncertainty and probability. *Annual review of neuroscience*, 2014. Vol. 37, pp. 205–220. doi: 10.1146/annurev-neuro-071013-014017
32. Maia T.V., Frank M.J. From reinforcement learning models to psychiatric and neurological disorders. *Nature neuroscience*, 2011. Vol. 14, no. 2, pp. 154–162. doi: 10.1038/nn.2723
33. Molet M., Miller R.R. Timing: An attribute of associative learning. *Behavioural processes*, 2014. Vol. 101, pp. 4–14. doi: 10.1016/j.beproc.2013.05.015
34. Crone E.A. et al. Neural mechanisms supporting flexible performance adjustment during development. *Cognitive, Affective, & Behavioral Neuroscience*, 2008. Vol. 8, no. 2, pp. 165–177. doi: 10.3758/CABN.8.2.165
35. Garbusow M. et al. Pavlovian-to-instrumental transfer in alcohol dependence: A pilot study. *Neuropsychobiology*, 2014. Vol. 70, no. 2, pp. 111–121. doi: 10.1159/000363507
36. Palminteri S. et al. Pharmacological modulation of subliminal learning in Parkinson's and Tourette's syndromes. *Proceedings of the National Academy of Sciences*, 2009. Vol. 106, no. 45, pp. 19179–19184. doi: 10.1073/pnas.0904035106
37. Reddy L.F. et al. Probabilistic reversal learning in schizophrenia: Stability of deficits and potential causal mechanisms. *Schizophrenia bulletin*, 2016. Vol. 42, no. 4, pp. 942–951. doi: 10.1093/schbul/sbv226
38. Nieuwenhuis S. et al. Reinforcement-related brain potentials from medial frontal cortex: Origins and functional significance. *Neuroscience & Biobehavioral Reviews*, 2004. Vol. 28, no. 4, pp. 441–448. doi: 10.1016/j.neubiorev.2004.05.003
39. Robinson J.S. Stimulus substitution and response learning in the earthworm. *Journal of comparative and physiological psychology*, 1953. Vol. 46, no. 4, pp. 262–266. doi: 10.1037/h0056151
40. Saffran J.R., Aslin R.N., Newport E.L. Statistical learning by 8-month-old infants. *Science*. 1996. Vol. 274, no. 5294, pp. 1926–1928.
41. Schultz W. Predictive reward signal of dopamine neurons. *Journal of neurophysiology*, 1998. Vol. 80, no. 1, pp. 1–27.
42. Izquierdo A. et al. The neural basis of reversal learning: An updated perspective. *Neuroscience*, 2016. doi: 10.1016/j.neuroscience.2016.03.021

43. Ferdinand N.K. et al. The processing of unexpected positive response outcomes in the mediofrontal cortex. *The Journal of Neuroscience*, 2012. Vol. 32, no. 35, pp. 12087–12092. doi: 10.1523/JNEUROSCI.1410-12.2012
44. Thorndike E.L. Animal intelligence: Experimental studies. *Transaction Publishers*, 1965.
45. Walsh M.M., Anderson J.R. Learning from delayed feedback: Neural responses in temporal credit assignment. *Cognitive, Affective, & Behavioral Neuroscience*, 2011. Vol. 11, no. 2, pp. 131–143. doi: 10.3758/s13415-011-0027-0
46. Weismüller B., Bellebaum C. Expectancy affects the feedback related negativity (FRN) for delayed feedback in probabilistic learning. *Psychophysiology*, 2016. Vol. 53, no. 11, pp. 1739–1750. doi: 10.1111/psyp.12738
47. Wolford G., Miller M.B., Gazzaniga M. The left hemisphere's role in hypothesis formation [Electronic resource]. *Journal of Neuroscience*, 2000. Vol. 20, no. 6, pp. 1–4. URL: <http://www.jneurosci.org/content/jneuro/20/6/RC64.full.pdf> (Accessed 27.12.2016).
48. Yellott J.I. Probability learning with noncontingent success. *Journal of mathematical psychology*, 1969. Vol. 6, no. 3, pp. 541–575. doi: 10.1016/0022-2496(69)90023-6