УДК 336.76

РАСПОЗНАВАНИЕ ЭВРИСТИК И ОБУЧЕНИЕ В ИГРЕ «КАМЕНЬ-НОЖНИЦЫ-БУМАГА»: ЭКСПЕРИМЕНТАЛЬНЫЙ ПОДХОД

Иван С. Сусин, Григорий В. Чернов

Классическая теория обучения в повторяющихся играх рассматривает обучение как реакцию на успех или неуспех конкретного выбора в предыдущих периодах игры. Однако на практике возможны и другие правила обучения: например, люди могут подмечать характерные регулярности в поведении оппонента, и предсказывать на их основе его дальнейшее поведение. Мы изучаем успешность обучения в соответствии с правилами этого последнего типа на примере экспериментальной игры "Камень-ножницы-бумага". Участники нашего лабораторного эксперимента – 70 студентов и школьников из Москвы – играли в эту игру на протяжении 100 раундов против компьютерного алгоритма, который был запрограммирован играть оптимально против ограниченно рационального игрока. Мы показываем, что участники эксперимента распознают регулярность в поведении оппонента-компьтера, и способны обучаться действовать оптимально против такой программы. Успех распознавания прямо пропорционален степени предсказуемости алгоритма (доле неслучайных ходов комьпьютера). Кроме того, игроки лучше выучивают ту часть правил, которая помогает им выигрывать, и хуже учатся на поражениях. Результаты исследования говорят о том, что люди могут успешно использовать процедурно рациональные стратегии, основанные на обучении правилам (rule learning), а не только на реакции на предыдущие успехи или неудачи.

Ключевые слова: Обучение с шумом, экономические эксперименты, повторяющиеся игры, игры с нулевой суммой, "камень-ножницы-бумага", распознавание эвристик, нетранзитивность, k-уровни

Введение

Важной задачей теории игр в экономике является качественное описание и предсказание поведения участников. Чтобы достичь этой цели, недостаточно рассматривать только равновесия. Не менее важно как игроки приходят к пониманию того, какая именно стратегия будет для них оптимальной в игре против конкретного оппонента (причем, как показывают, в частности, [13], [14], эта стратегия не обязательно должна быть равновесной).

В литературе по обучению в играх (см.напр., [17]) широко используются моделей адаптации игрока к результатам предшествующей истории игры. Например, в моделях типа «обучения с подкреплением» (reinforcement learning) предполагается, что игроки чаще играет те стратегии, которые были в некотором смысле «успешны» в прошлом. Гораздо хуже исследованы процедурно рациональные стратегии, при которых игроки реагируют не просто

на платежи в предыдущие периоды, но делают предположения о стратегиях оппонентов, которые привели к этим платежам, и стремятся выбрать оптимальные ответы на эти стратегии. Такие модели должны описывать, как меняются действия игрока, когда он замечает изменения в поведении другого игрока, то есть адаптацию игрока к изменениям в стратегии оппонента. Для этого надо понимать, как формируется у игрока внутренняя модель поведения оппонента, как делается предсказание о его будущих действиях и находится наилучший отклик на эти действия.

Мы исследуем такую адаптацию в одном из простейших случаев – в конечной повторяющейся игре с единственным равновесием в смешанных стратегиях. В качестве таковой используется всем хорошо известная игра «Камень-ножницы-бумага», с единственным равновесием Нэша, где каждый игрок выбирает каждую из своих трех стратегий с равной вероятностью. Однако простота эта обманчива: в конечной повторяющей игре у каждого участника найдется множество стратегий, которые могут быть наилучшим ответом на те, или иные стратегии оппонента, будь эти последние реальными или мнимыми. Многие из этих стратегий каждый из нас, вероятно, может припомнить из собственной практики – и она же подсказывает, что проблема распознавания стратегии оппонента может быть очень сложной. Например, игроки могут принимать решения, исходя из собственной интерпретации поведения оппонента, своих ожиданий относительно его стратегического ответа на собственные решения и пр. Все эти особенности реальной игры представляют немалую сложность с точки зрения интерпретации поведения игроков. Чтобы изолировать эти поведенческие эффекты, в нашем эксперименте мы используем компьютерного оппонента (алгоритм), который выбирает или строго определенную и заданную заранее стратегию в повторяющейся игре, либо же ходит случайным образом, т.е. играет стратегию, равновесную по Нэшу. Такая постановка позволяет нам наблюдать в рамках эксперимента как человек распознает регулярности в поведении оппонента и адаптируется к ним. Кроме того, при такой постановке участник эксперимента может быть уверен, что если он выявит правило поведения оппонента, которое позволит его побеждать, оппонент не начнет адаптироваться вслед за ним. Это позволяет участнику эксперимента скорее укрепиться в понимании того, что делает его оппонент, и следовательно, дает ему больше оснований для следования той стратегии, которую он полагает оптимальной.

Мы показываем, что существует зависимость между регулярностью играемой стратегией и скоростью её распознавания человеком, а также зависимость переключения между стратегиями от динамики успешности игры, по аналогии с формированием условных рефлексов, работающих под влиянием механизмов подкрепления.

В нашем экспериментальном дизайне мы обращаемся к игре “Камень-Ножницы-Бумага”. Ее можно рассматривать как расширение игры “орлянка” на три возможные нетранзитивно друг друга превосходящие стратегии. Эта игра достаточно хорошо исследована в условиях ЧпЧ и МпМ (Человек против Человека и Машина против Машины), но не использовалась в предыдущих

исследованиях распознавания паттернов в стратегических играх (обычно как раз в режиме ЧпМ (Человек против Машины), вместо нее использовалась “орлянка” или “семейный спор”). Поведение людей в таких играх неполно описывается как случайным выбором [18], так и традиционными подходами по обучению, которые фокусируются вокруг концепций, основанных на равновесии и анализа последовательности игроками предыдущих исходов игры (например, belief learning [8], reinforcement learning [13]) или сочетания двух предыдущих - Experience Weighted Attraction (EWA) model [4]). Альтернативой служат модели распознавания правил в обучении (rule learning) и анализа паттернов [22]. В отличие от моделей, построенных на равновесии и неограниченных способностях агентов к вычислению, они предполагают в неявном виде наличие у агента представления о предпочитаемой стратегии другого игрока, а также условиях изменения им выбранной стратегии, что существенно обогащает игровой анализ. Еще одно достоинство именно игры ``камень-ножницы-бумага'' — из простых игр только в ней имитация действий оппонента не гарантирует от цикла проигрышей ([11]). Если игрок проигрывает, он может искать новую стратегию, которая будет более успешной. Наилучшее поведение алгоритма в ответ на его действия тогда не просто предсказание, основанное на серии ходов, а способность угадывать выбираемое правило, то есть то, как изменяет свое поведение проигрывающий человек и сразу подстроиться под эту измененную стратегию. Из [12] можно сделать вывод, что и стратегия tit-for-tat (делать то же, что оппонент, только если он успешен) тоже не гарантирует минимального успеха в этой игре.

Наша работа вносит вклад в три основные линии исследований - работы об обучении в играх в целом, обучении в режиме “Человек против робота” и обучении в игре “Камень-ножницы-бумага”. Только одна работа находится на пересечении всех трех - [30], но она концентрируется на моделировании памяти игроков, а не их модели принятия решений. Участники моделируются при помощи нейронной сети и авторы оценивают объем памяти, использованный для принятия решений против робота. Использовалось только две разновидности робота - с одноперiодной памятью (его участники уверенно обыгрывали) и двухпериодной памятью (выступавший почти на равных с людьми).

Также уже было оценено (в эксперименте [28]), насколько случайно люди играют в ``Камень-ножницы-бумага'' друг с другом. На примере этой работы можно отметить другую сторону значимости обучения в играх: если в поведении людей наблюдается регулярность, не обязательно эту регулярность можно эксплуатировать именно потому, что люди могут замечать регулярность в поведении оппонента и адаптироваться к ней. Авторы показали, что участники их эксперимента часто следовали так называемой ``павловской эвристике“, названной так в честь используемого в ней (и открытого И. Павловым) принципа формирования условного рефлекса - если действие получает позитивное подкрепление, оно повторяется чаще, если негативное - повторяется реже и заменяется другим действием. В ``Камень-ножницы-бумага'' под павловской эвристикой понимается повторение выбора после победы

(например, если против камня оппонент выбрал ножницы, то повторяется камень) и изменение выбора в случае ничьей или поражения (против камня оппонент выставил бумагу - в следующий раз выбирается бумага). Однако на вопрос, достаточно ли знания об этой эвристике для алгоритма, превосходящего человека в этой игре, наша работа (первая в литературе) дает отрицательный ответ, люди сами адаптируются к регулярностям в алгоритме и начинают действовать оптимальным против него способом.

Обзор литературы

Обучение в играх исследовалось подробно, но в основном в форме поиска простых правил, которые позволят агенту обучаться на основе успеха или неуспеха использованных в ближайшей истории стратегий.[1]

[22] рассматривает игру человека против нескольких простых алгоритмов, среди которых (помимо фиктивной игры (fictitious play) с одно- и двухпериодной памятью) эвристика обучения с подкреплением. В этой работе используется условное правило (если человек играет эвристику, похожую на павловскую, чаще половины раундов, то алгоритм играет наилучший на нее отклик, иначе алгоритм играет случайно), в то время как в данной статье используется безусловное правило, но с различными уровнями зашумленности. [24] использует асимметричное распределение вероятностей в игре «орлянка» и показывает распознавание этого искажения участниками, но их алгоритм не реагировал на действия участников.

[3] показали, что адаптация происходит быстрее, если алгоритм оппонента соответствует ожиданиям и когнитивным искажениям участников, сравнительно с ситуацией, когда он им противоречит. В нашей работе мы показываем, что часть алгоритма, позволяющую достигнуть цикла побед, люди находят лучше, чем ту часть, которая позволяет реагировать на поражения.

Литература по “Камень-ножницы-бумага” в соревнованиях между людьми стабильно находит циклические закономерности, в том числе павловскую эвристику ([12], [16], [28]). На основании нашего эксперимента можно сделать вывод, что эта регулярность — следствие более сложных процессов, чем просто невнимательность людей к регулярностям в поведении оппонента.

[9] используют ту же общую схему — игру человека против нескольких различных компьютерных алгоритмов. В этой работе используют игру с не полностью противоположными интересами (игра в олигополию) и алгоритмы обучаются по одному из классических правил (fictitious play, trial& error, best reply, imitation, reinforcement learning, EWA). Мы переносим эту схему на другую, более антагонистическую игру и иные наборы правил поведения алгоритмов.

---

[1] Таких, как фиктивная игра (fictitious play), обучение с подкреплением (reinforcement learning), взвешенное по опыту притяжение ('experience-weighted attraction' (EWA) [Camerer, Ho, 1999])[23] и [25], наилучший отклик (best reply), ``метод проб и ошибок'' (trial & error), имитация (imitation) ([9]), tit-for-tat ([2]).

Наконец, мы дополняем результаты исследований игры “RoShamBo” (принятое в литературе по искусственному интеллекту название игры “Камень-ножницы-бумага“), таких как [27] в чемпионатах между алгоритмами, поскольку оцениваем игру людей сравнительно с алгоритмами и пытаемся найти типично человеческие стратегии (в будущих исследованиях, возможно, мы обратимся к использованной в указанной работе стратегии “сопоставления историй“, использующей повторения в ближайшей истории ходов как наилучшую оценку текущего хода).

Другое важное для нашей работы направление исследований - распознавание изменений. [18] и [19] показывают, что в задаче статистического (не стратегического) распознавания люди склонны переоценивать полученные сигналы и недооценивать свойства системы, порождающей сигналы. Тут видна аналогия с использованием алгоритмов фиктивной игры (fictitious play), следующих истории нескольких последних раундов, однако возникает исследовательский вопрос, а переносим ли результат нестратегической ситуации на стратегическую. Возможно ли, что в стратегической игре участники будут больше ориентироваться не на конкретные недавние действия оппонентов (которые могут и быть направлены на то, чтобы сбить участника с толку), а на собственную модель поведения оппонента? Другая гипотеза, которую подсказывают эти исследования - они показали, что участники обучаются лучше в ситуациях, где было меньше ложноположительных сигналов и ниже информационная энтропия. Хотя наш экспериментальный дизайн пока не позволяет напрямую измерить именно ложноположительные сигналы, мы показываем, что шум влияет на динамику обучения предсказуемым образом.

Значительную долю исследований динамики в ``Камень-Ножницы-бумага" можно отнести к популяционным играм (как, например, в классической книге по эволюционным играм [29] и более новой работе [6]), но популяционные игры по построению не позволяют учитывать историю до предыдущего периода и будущее и поэтому не подходят для нашего анализа.

Это простейшая игра с нетранзитивностью, что представляет особый интерес также для исследования восприятия человеком нетранзитивности (см. [1], [15], [26]). Существующая литература, описывая ``павловскую эвристику", не может отличить ее от других, более сложных моделей поведения (так как предсказание в игре с только двумя стратегиями неизбежно бинарно), в то время как в нашей игре, наблюдая конкретную эвристику, мы можем оценить, насколько она оптимальна против конкретных закономерностей поведения оппонента.

Экспериментальный дизайн и методология

Участники эксперимента приглашались в аудиторию, оборудованную компьютерными терминалами. Эксперимент проводился с использованием программной экспериментальной среды oTree [7], участники из каждой группы находились в одной аудитории и одновременно работали с запущенными независимо друг от друга реализациями компьютерного алгоритма с

одинаковой степенью случайности (параметром β). При этом по условиям эксперимента всякое взаимодействие с другими участниками не допускалось.

Правила игры «Камень-Ножницы-Бумага» стандартные: В каждом раунде игрок выбирает одно из трех возможных действий: Камень, Ножницы или Бумага. Его оппонент (компьютерный алгоритм) делает то же самое независимо от действий игрока. Победитель в каждом раунде определяется по следующему правилу: Бумага побеждает камень («бумага камень завернет»); Камень побеждает ножницы («камень ножницы затупит»); Ножницы побеждают бумагу («ножницы бумагу разрежут»). Если оба игрока выбрали одно и то же действие, исходом раунда является ничья. Одна экспериментальная сессия состояла из 100 раундов для каждого участника.

По условиям эксперимента, в игре участвуют два игрока, один из которых человек, а второй - компьютерная программа. Экспериментальная выборка разделена на группы, в каждой из которых программа играет согласно некоторому алгоритму.

Функции выигрышей игроков в каждом раунде задаются матрицей выигрышей (приведена в таблице 1). В игре используется игровая валюта токены. (условные денежные единицы эксперимента) и итоговый результат изначально подсчитан в токенах. В конце экспериментальной сессии этот результат в токенах переводился в рубли по курсу токена к рублю 1 к 3 .

Таблица 1 **Матрица игры “Камень-ножницы-бумага” (в токенах)**

	Игрок выбрал Камень	Игрок выбрал Ножницы	Игрок выбрал Бумага
Программа выбрала Камень	1	0	2
Программа выбрала Ножницы	2	1	0
Программа выбрала Бумагу	0	2	1

В дополнение к выигрышу участников, вне зависимости от результатов эксперимента им выплачивалась плата за участие в размере 150 рублей.

Кроме того, для мотивации участников стараться на протяжении всей сессии из 100 игр за каждые следующие пять побед, начиная с 30, начислялся бонус в размере 15 токенов (см. таблицу 2)

Таблица 2 **Количество начисляемых бонусов в токенах по мере увеличения количества побед**

Победы	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100
Бонус	15	30	45	60	75	90	105	120	135	150	165	180	195	210	225

Экспериментальный дизайн построен на четырех возможных подходах к принятию решений и их сочетаниях.

Первый вариант - стандартное смешанное равновесие Нэша, то есть равновероятный выбор каждой из трех стратегий в каждом раунде. Другие базовые эвристики («павловская», «контр-павловская» и «оптимальная») задаются результатом предыдущего раунда. Из-за нетранзитивности удобно описывать их в терминах модулярной арифметики по основанию 3 — как смещение на плюс или минус единицу (будем считать, что если один выбор бьет другой, то этот выбор «больше» другого по модулю 3, например, если взять «камень» за 0, то «бумаге» соответствует 1, «ножницам» - 2). Правила модулярной арифметики означают, что смещение на 3 — то же самое, что повтор выбора, а смещение на +2 и на -1 тоже эквивалентны (в предыдущем примере получаем, что 0 больше 2 по модулю 3, так как 2+1(mod 3)=0, что верно - «камень» побеждает «ножницы»).

Павловской эвристикой мы назовем правило условного перехода, аналогичное выводам работы [12] – участник повторяет свой выбор в случае победы, в случае проигрыша «улучшает» свой выбор (то есть переходит к тому, который победит собственный предыдущий), а в случае ничьей «ухудшает». Записывая это модулярной арифметикой, получим правила — проигрыш =>+1, ничья =>-1, выигрыш =>0.

Опять же, следуя результатам работ [28] (которая тоже показывает склонность чаще менять неуспешный выбор и реже успешный) и [12], мы взяли именно эту эвристику в качестве нулевой гипотезы и оцениваем, насколько люди машинально следуют этой эвристике, а насколько адаптируются в зависимости от поведения оппонента. Для этого мы определяем еще две эвристики.

Первая задает поведение оппонента, оптимально играющего против павловской эвристики, поэтому мы назовем ее «контр-павловской» и именно такую регулярность в поведении проявляет используемый нами компьютерный алгоритм. Вторая, в свою очередь, является наилучшим ответом уже на «контр-павловскую» и как раз показывает адаптацию участника эксперимента к алгоритму (поэтому мы называем ее «оптимальной», это оптимальное с точки зрения выигрыша поведение участника в нашем эксперименте). Арифметически они задаются смещениями проигрыш=>-1 ничья=>0 выигрыш=>+1 для «контр-павловской» и проигрыш=>0 ничья=>+1 выигрыш=>-1 для «оптимальной».

Таблица 3 **Описание эвристических правил**

	Результат предыдущего раунда		
Смещение	Выигрыш	Ничья	Проигрыш
-1	**ОПТИМАЛЬНАЯ**	ПАВЛОВСКАЯ	**КОНТР**-ПАВЛОВСКАЯ
0	ПАВЛОВСКАЯ	**КОНТР**-ПАВЛОВСКАЯ	**ОПТИМАЛЬНАЯ**
+1	**КОНТР**-ПАВЛОВСКАЯ	**ОПТИМАЛЬНАЯ**	ПАВЛОВСКАЯ

Заметим, что все возможные исходы на каждом ходу (кроме первого в сессии, для которого не определен результат предыдущего раунда) однозначно и наблюдаемо раскладываются в одну из девяти ячеек таблицы 3, а значит, мы можем рассматривать ее как таблицу сопряженности.

Алгоритм игры компьютерного оппонента (смешанно-контр-павловская эвристика) задается как линейная комбинация (с коэффициентом β) смешанного равновесия Нэша и контр-павловской эвристики, то есть с внутренней вероятностью β алгоритм играет случайно, а с вероятностью (1- β) — согласно контр-павловской эвристике. Из-за того, что в игре только три возможных стратегии, случайный выбор совпадет с выбором, следующим эвристике, в трети случаев.

Мы предполагаем, что участники будут изначально играть павловскую эвристику в игре против смешанно-контр-павловской эвристики компьютера с различными значениями коэффициента β (где $\beta=1$ - смешанное равновесие Нэша).

В каждой подгруппе экспериментальной сессии в качестве противников используются алгоритмы, играющие смешанно-контр-павловские эвристики с разными β (из таблицы 3) для разных участников.

Деление возможных эвристик на три базовые следует подходу k-levels models ([5], [20]). Этот класс моделей описывает многоуровневое стратегическое мышление, задавая базовый, "наивный" уровень и выстраивая каждый следующий уровень как наилучший отклик на предыдущий.

Выбирая павловскую эвристику в качестве базового уровня, мы получаем еще две эвристики, вводящие потенциально бесконечное число уровней стратегирования в нетранзитивный цикл из трех взаимно оптимальных эвристик. Наилучшим ответом на павловскую эвристику является анти-павловская, на нее - "оптимальная", а на "оптимальную" - снова павловская. Это не позволяет идентифицировать конкретный уровень, на котором думает участник, но позволяет сказать, адекватен ли этот уровень используемой компьютером эвристике, то есть адаптируется ли участник к конкретному оппоненту.

Процедура проведения эксперимента и результаты

В начале каждой экспериментальной секции участники читали инструкции, которые позже дублировались на экране компьютера, где игрок принимал решения. После чего экспериментаторы еще раз объясняли параметры игры, в которых игроку пояснялось, что в качестве оппонента с ним будет играть компьютерная программа. Игроку так же сообщалось, что действия программы подчиняются некоторому алгоритму, про который известно следующее:

- В момент принятия решения в каждом раунде алгоритму не известно действие игрока в текущем раунде
- Алгоритму доступна информация о действиях обоих участников (и ваших, и его) в предыдущих раундах.
- Алгоритм программы не меняется на протяжении всей игры
- Алгоритм может подчиняться некоторому правилу (закономерности), но может и не подчиняться. Знание закономерности позволит лучше предугадывать действия оппонента (компьютерной программы).

Одна экспериментальная сессия состояла из 100 раундов, по истечению которых участники проходили опрос, включающий в себя вопросы о возрасте, поле, образовании испытуемых, а так же вопросы, касающиеся угадывания правил алгоритма.

Эксперимент проводился в Лаборатории экспериментальной и поведенческой экономики НИУ ВШЭ в апреле 2018. Суммарно проводилось 4 сессии, часть из которых игрались со студентами часть со школьниками. Разные группы участников играли в разное время. Всего в эксперименте приняло участие 70 человек. Подробное описание сессий приведено в таблице 4.

Таблица 4 **Экспериментальные сессии**

Параметр бета	Кол-во игроков	Доля мужчин (%)	Тип сессии	Средний возраст	Миним. побед	Среднее кол-во побед	Макс. побед
β = 1	8	38	Студенты (с оплатой)	22,8	26	34	39
β = 0,6	9	33	Студенты (с оплатой)	23,8	30	37	50
β = 0,4	39	44	Школьники (без оплаты)	16,4	25	40	63
β = 0,4	14	21	Студенты (с оплатой)	21,2	31	43	59
Вся выборка	70	37	-	19	25	40	63

Поскольку использованная компьютерным оппонентом стратегия описывается результатом предыдущего хода и сдвигом следующего выбора, мы представляем все возможные исходы в виде матрицы сопряженности 3 на 3, где столбцы соответствуют результатам предыдущего хода (победа\ничья\поражение), а строки - сдвигу выбора хода (проигрывающий предыдущему\повтор предыдущего\выигрывающий у предыдущего).

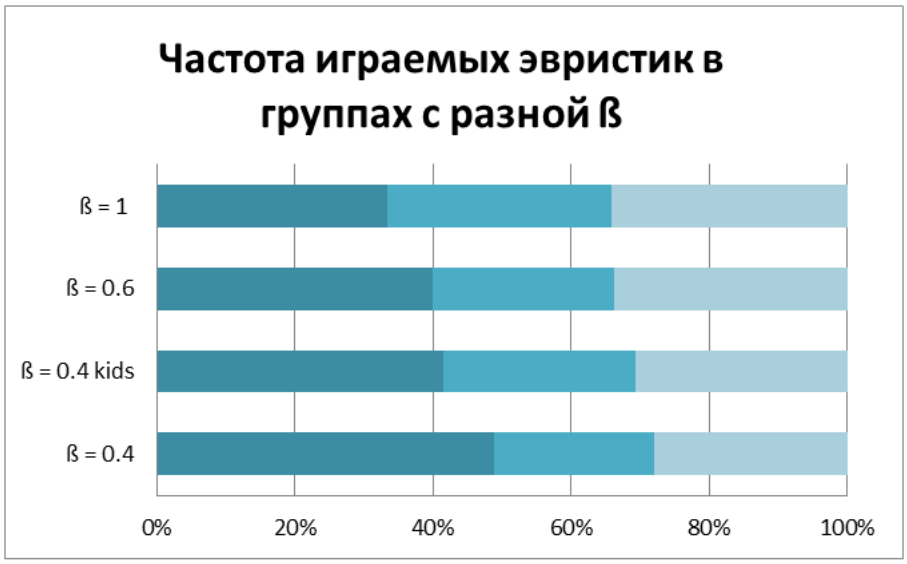

Рисунок 1

Как можно увидеть из графика, чем меньше случайности было в игре компьютерного оппонента от группы к группе, тем выше частота, с которой игралась оптимальная эвристика. При одном и том же значении коэффициента бета группа школьников играла в среднем менее оптимально, чем группа денежно мотивированных студентов. Поскольку группа школьников только одна, мы не можем сказать, какой фактор (денежный или возрастной) на это повлиял. При этом доля контр-павловской (дающей более частые ничьи) убывает медленнее, чем доля павловской.

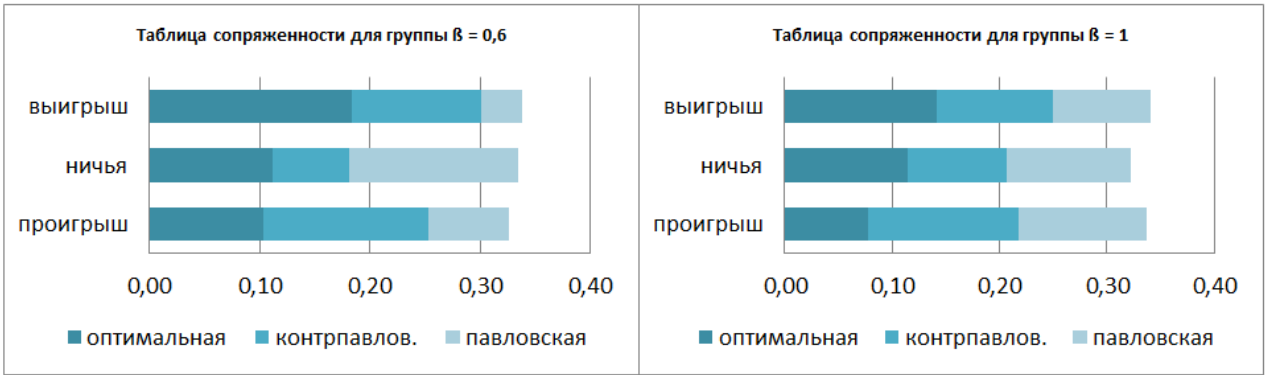

Рисунок 2

На следующих графиках мы можем посмотреть на действия участников подробнее. Снова заметим, что доля побед и поражений не является независимой переменной — она зависит от того, насколько участники могли

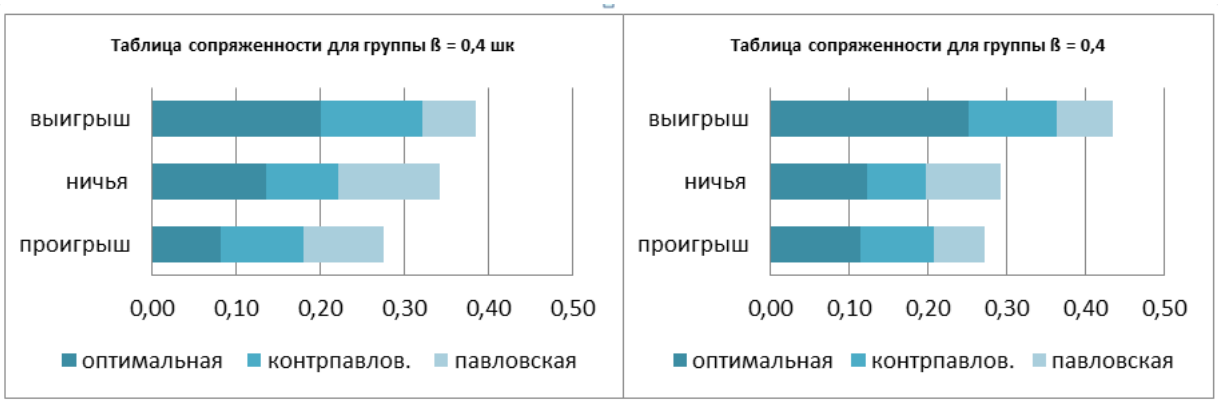

Рисунок 3

предсказать ходы компьютерного оппонента. Для групп со полностью случайным и умеренно случайным оппонентом можно увидеть примерно равные доли различных исходов и отсутствие явной тенденции предпочитать конкретное правило перехода. При этом и гипотезу о случайности можно отвергнуть (p-value по критерию хи-квадрат Пирсона об отличии от равномерности меньше 0.01) для обоих групп, доли различных эвристик в исходах сильно отличаются от тех, которые можно было бы ожидать при случайной игре.

Для оставшихся двух групп с более регулярным оппонентом четче просматривается закономерность — большой доле побед соответствует высокая частота применения оптимальной эвристики в случае победы. Если бы участники выучивали все три составляющие оптимальной эвристики, отношение долей частоты трех эвристик было бы одним и тем же для разных исходов, что, как мы можем увидеть на графиках, неверно. Мы можем уверенно сказать, что участники выучивают то правило, которое обеспечивает им серию побед (p-value критерия хи-квадрат Пирсона <0.01), но в случае случайно выпавшего поражения ведут себя гораздо более равномерно. Для ничьих пропорции более равномерны, чем для побед, но менее равномерны, чем для поражений.

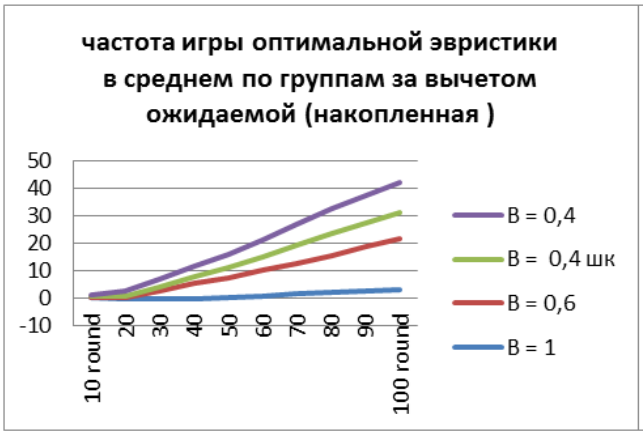

Рисунок 4

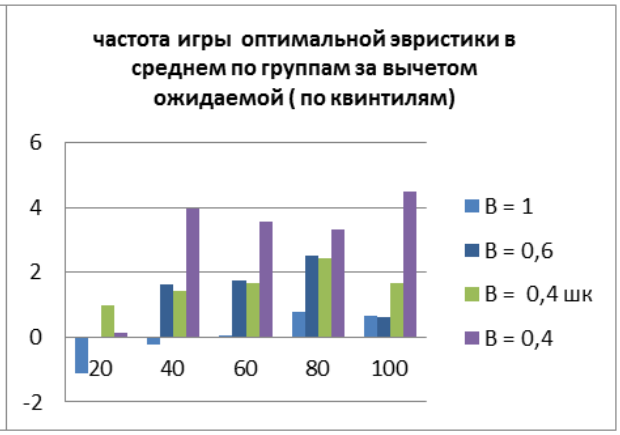

Рисунок 5

На этом графике мы можем увидеть процесс обучения уже в динамике, на промежутках в 10 раундов. Для каждой десятки в полной серии из ста раундов мы подсчитываем количество оптимально сыгранных раундов и вычитаем ту долю, которая может быть объяснена случайностью (треть). На графике приведена накопленная сумма таких величин для всех четырех групп. Видно, что частота оптимальной игры монотонно увеличивается с ростом регулярности и мотивации: для группы без регулярности накопленная частота растет в пределах ошибки округления, когда для остальных групп виден процесс обучения оптимальной игре.

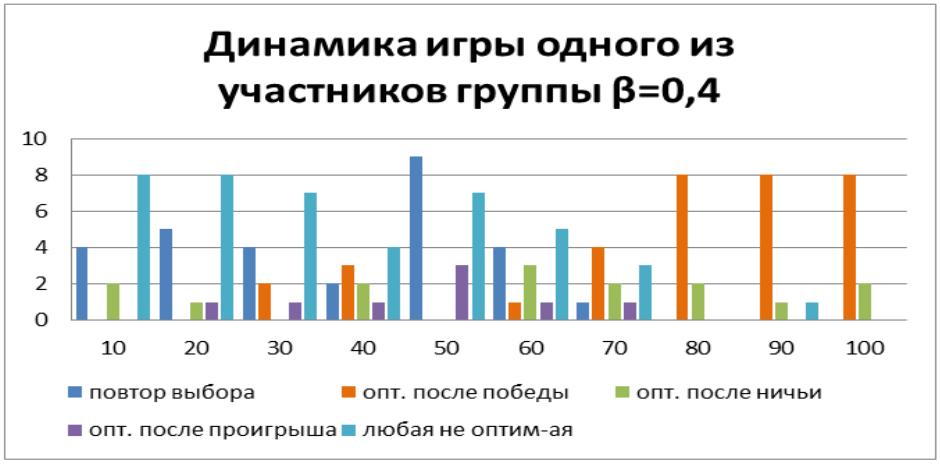

Рисунок 6

Динамика на уровне отдельных участников тоже позволяет исключить случайность. Например, на рисунке 6 можно заметить интересную динамику: сначала участник (из группы с β=0,4) играл относительно случайно, но между раундами 40 и 50 он 9 раз подряд ходил то же, что в предыдущий раз (серия из 9 «камней» подряд). После этого можно заметить, что частота игры им оптимальной эвристики резко возрастает (из последних 30 раундов оптимально сыграно 29) и это обеспечивается серией побед. Когда случайный ход компьютера принес участнику поражение, он неоптимально на него реагировал. Такая последовательность действий не объясняется существующими теоретико-игровыми моделями обучения и явно является следствием целенаправленного «экспериментирования» со стороны участника.

После экспериментальной сессии участникам было предложено заполнить анкету, где спрашивалось, заметили ли они в игре компьютера закономерность, алгоритм или нет. Распределение ответов по группам - 5 (из 8) заметивших закономерность в группе бета=1, 7 из 9 в группе бета =0.6, 34 из 39 в группе школьников (бета=0.4) и 12 из 14 в группе студентов с бета=0.4. Примеры ответов (орфография оригинала сохранена):

"да, использовал. Ничья и победа - не меняет тактику; поражение - меняет. После серии из 4-5 поражений меняет тактику, но затем возвращается к ней." (участник, который "экспериментировал")

"Если много раз подряд выбирать один и тот же элемент, то компьютер будет выбирать то же самое" (неполное выучивание правила, участник распознал только оптимальный ход при ничьей. группа школьников)

"Компьютер проиграл - выкидывает то, то проигрывает его предыдущему выбору, при ничьей - повторял свой ход, при выигрыше ходил с того, что проигрывает моему предыдущему ходу"(полностью распознал все три составляющих алгоритма. группа школьников)

"не знаю. Иногда казалось, что помогает повторять за алгоритмом. Иногда что он сам повторяет предыдущее действие. Но мне кажется, что скорее что-то другое." (группа с бета=1)

"да, можно заметить закономерность в какую сторону играет компьютер

после того как выигрывает или проиграет" (распознан принцип построения алгоритма. группа бета=0.6)

"Да. Чтобы выиграть необходимо нажимать то что выбирал до этого компьютер" (распознано только правило для серии побед. группа школьников)

"Нет, компьютер исходил из моего решения: если у меня был выигрыш, то компьютер думал, что я пойду тем же ходом и ходил наперед. Когда он проигрывал 3 и более раз подряд, то менял алгоритм" (два примечательных факта: участник описывает стратегию в терминах "оппонент думал, что", что и было целью экспериментального дизайна, второй факт - участник описывает правило, которое не было заложено в алгоритм, а могло появиться случайно в конкретной серии игр. группа бета=0.6)

Обсуждение дальнейших исследований и выводы

Возможности использованного экспериментального дизайна не исчерпываются проведенным исследованием, возможно несколько направлений исследований в рамках того же подхода.

Первое и самое простое возможное продолжение — использовать другие правила поведения алгоритма и изучать игру людей против них. Это позволит ответить на вопрос, насколько сложные детерминированные алгоритмы люди могут отличать и замечать, какие алгоритмы более или менее успешны в игре против людей. Сюда же можно отнести вариант экспериментального дизайна, когда уже распознанный алгоритм меняется или модифицируется прямо в процессе эксперимента (встраивается переключатель, меняющий алгоритм при выполнении простого условия, например, количества ходов).

Второе направление — учет эмоциональной составляющей. Хотя можно ожидать, что человек помнит свой последний ход, маловероятно, чтобы вся история игры сохранялась в памяти настолько же ярко, в предыдущих работах ([30]) уже проводилась грубая оценка этой рабочей памяти. Однако вполне возможно, что в принятии решений человеком в качестве скрытой переменной присутствует «счет», усредненный (по неизвестным нам пока правилам) показатель успеха в прошлом. Возможно, именно этот показатель объясняет, почему человек не проигрывает алгоритму — когда алгоритм эксплуатирует неслучайность в игре человека, значение скрытой переменной снижается и человек чаще меняет стратегию. Для этого будет уже недостаточно фиксированных алгоритмов и их переключений, потребуется разработать код, позволяющий алгоритму адаптивно меняться в процессе эксперимента.

Третье направление, результаты которого могут помочь и для первого и второго — сравнение людей с обучающимися алгоритмами. Мы можем сказать, что человек лучше адаптируется к нашей экспериментальной среде, чем алгоритм фиктивной игры, однако чтобы сравнить игру человека, например, с оптимальным обучением байесовского агента (который наиболее полно и точно эксплуатирует всю доступную ему информацию и не подвержен эмоциям или искажениям), потребуются дальнейшие исследования.

Наконец, четвертое, наиболее далекое, амбициозное и объединяющее — построение общей модели обучения в играх, учитывающей как поведенческие особенности и искажения, так и адаптивность человека.

Перечисленные ниже примеры продолжения исследований проранжированы по простоте постановки и проведения дальнейших экспериментов.

В данной работе мы экспериментально показали, что участники способны адаптироваться к алгоритму оппонента даже при значительном уровне шума (хотя распознавание и снижается с увеличением шума). Эта адаптация не равномерна, а преимущественно направлена на распознавание условий, позволяющих раз за разом выигрывать. Кроме простой адаптации, мы нашли некоторые свидетельства активного экспериментирования со стороны участников.

Полученные результаты проливают свет на ряд вопросов, важных для задач создания стимулов, поскольку четкое понимание механизмов возникновения паттернов позволит не только предсказывать поведение оппонента, но и в какой-то мере управлять им, что может быть использовано в дизайне механизмов, создания схем подталкивании и т. п. Результат данного исследования может помочь прояснить известный парадокс поведенческой экономики - одновременное существование ошибки игрока gamblers fallacy и hot-hand fallacy - ошибки горячей руки [21]. Одна из них говорит, что люди продолжают эксплуатировать успешную в прошлом стратегию, даже если вероятность говорит об обратном, а вторая - о склонности игроков ставить на тот результат, который реже выпадал в прошлом.

Асимметрия в выучивании правил при проигрыше и выигрыше в нашем эксперименте говорит о том, что обучение с подкреплением может объяснять эти искажения - «ошибка горячей руки» возникает при серии успехов и игрок интерпретирует ее не как случайность, а как то, что он «выучил правило», нашел регулярность в поведении оппонента. Напротив, проигрывая, игрок воспринимает случайность как повод перестроиться, повысить непредсказуемость своего поведения, чтобы прервать серию побед оппонента.

Список литературы

[1] Поддьяков А. Н. Нетранзитивность превосходства и ее использование для обмана и тренировки мышления // Психолого-экономические исследования. 2016. Т. 3 (9). № 4. С. 43-50.

[2] Axelrod, R. and Hamilton, W.D., 1981. The evolution of cooperation. science, 211(4489), pp.1390-1396.

[3] Abrahamyan, A., Silva, L.L., Dakin, S.C., Carandini, M. and Gardner, J.L., 2016. Adaptable history biases in human perceptual decisions. Proceedings of the National Academy of Sciences, 113(25), pp.E3548-E3557.

[4] Camerer, C. and Hua Ho, T., 1999. Experience-weighted attraction learning in normal form games. Econometrica, 67(4), pp.827-874.

[5] Camerer, C.F., Ho, T.H. and Chong, J.K., 2004. A cognitive hierarchy

model of games. The Quarterly Journal of Economics, 119(3), pp.861-898.

[6] Cason, T.N., Friedman, D. and Hopkins, E., 2013. Cycles and instability in a rock–paper–scissors population game: a continuous time experiment. Review of Economic Studies, 81(1), pp.112-136.

[7] Daniel L. Chen, Martin Schonger, Chris Wickens, oTree—An open-source platform for laboratory, online, and field experiments, Journal of Behavioral and Experimental Finance, Volume 9, 2016, Pages 88-97, ISSN 2214-6350, https://doi.org/10.1016/j.jbef.2015.12.001.

[8] Cheung, Y.W. and Friedman, D., 1997. Individual learning in normal form games: Some laboratory results. Games and Economic Behavior, 19(1), pp.46-76.

[9] Duersch, P., Kolb, A., Oechssler, J. and Schipper, B.C., 2010. Rage against the machines: how subjects play against learning algorithms. Economic Theory, 43(3), pp.407-430.

[10] Duersch, P., Oechssler, J. and Schipper, B.C., 2014. When is tit-for-tat unbeatable?. International Journal of Game Theory, 43(1), pp.25-36.

[11] Duersch, P., Oechssler, J. and Schipper, B.C., 2012. Unbeatable imitation. Games and Economic Behavior, 76(1), pp.88-96.

[12] Dyson, B.J., Wilbiks, J.M.P., Sandhu, R., Papanicolaou, G. and Lintag, J., 2016. Negative outcomes evoke cyclic irrational decisions in Rock, Paper, Scissors. Scientific reports, 6, p.20479.

[13] Roth, A.E. and Erev, I., 1995. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. Games and economic behavior, 8(1), pp.164-212.

[14] Erev, I. and Roth, A.E., 1998. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. American economic review, pp.848-881.

[15] Fishburn, P.C., 1991. Nontransitive preferences in decision theory. Journal of risk and uncertainty, 4(2), pp.113-134.

[16] Frey, S. and Goldstone, R.L., 2013. Cyclic game dynamics driven by iterated reasoning. PloS one, 8(2), p.e56416.

[17] Fudenberg, D. and Levine, D.K., 1998. The theory of learning in games. MIT press.

[18] Li, Y., Massey, C. and Wu, G., 2014. Learning to Detect Change.

[19] Massey, C. and Wu, G., 2005. Detecting regime shifts: The causes of under-and overreaction. Management Science, 51(6), pp.932-947.

[20] Nagel, R., 1995. Unraveling in guessing games: An experimental study. The American Economic Review, 85(5), pp.1313-1326.

[21] Rabin, M. and Vayanos, D., 2010. The gambler's and hot-hand fallacies: Theory and applications. The Review of Economic Studies, 77(2), pp.730-778.

[22] Spiliopoulos, L., 2013. Strategic adaptation of humans playing computer algorithms in a repeated constant-sum game. Autonomous agents and multi-agent systems, 27(1), pp.131-160.

[23] Spiliopoulos, L., 2012. Pattern recognition and subjective belief learning in a repeated constant-sum game. Games and economic behavior, 75(2), pp.921-935.

[24] Shachat, J. and Swarthout, J.T., 2004. Do we detect and exploit mixed

strategy play by opponents?. Mathematical Methods of Operations Research, 59(3), pp.359-373.

[25] Shachat, J. and Swarthout, J.T., 2012. Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control, 36(3), pp.383-402.

[26] Tversky, A., 1969. Intransitivity of preferences. Psychological review, 76(1), p.31.

[27] Sony E. Valdez, V. J. D. Barayuga and P. L. Fernandez (2014). The Effectiveness of using a Historical Sequence-based Predictor Algorithm in the First International RoShambo Tournament. International Journal of Innovative Research in Information Security (IJIRIS), Volume 1, Issue 5, November 2014, pp. 59-65.

[28] Wang, Z., Xu, B. and Zhou, H.J., 2014. Social cycling and conditional responses in the Rock-Paper-Scissors game. Scientific reports, 4, p.5830.

[29] Weibull, J.W., 1997. Evolutionary game theory. MIT press.

[30] West, R.L. and Lebiere, C., 2001. Simple games as dynamic, coupled systems: Randomness and other emergent properties. Cognitive Systems Research, 1(4), pp.221-239.

Информация об авторах

Иван С. Сусин – аспирант, аспирантская школа по экономике НИУ ВШЭ (Москва, isusin@nes.ru)
Григорий В. Чернов – аспирант, аспирантская школа по экономике НИУ ВШЭ, стажер-исследователь международная лаборатория экспериментальной и поведенческой экономики НИУ ВШЭ (Москва, gr.chernov54@ya.ru)

Ivan S. Susin, Grigory V. Chernov

Heuristics Recognition And Learning In Rock-Paper-Scissors Game: Experimental Study

Classic theory of learning in repeated games considers

learning as a reaction to success or failure of specific choice in previous rounds. However in practice there can be other rules of learning: for example, people can spot specific regularities in opponents' behavior and thus predict his future behavior. We study to what degree this type of learning is successful on the example of laboratory game of "rock-paper-scissors". Our participants - 70 students and schoolchildren from Moscow - played this game for 100 rounds against a computer algorithm, which was programmed to play optimally against boundedly rational player. We show that participants succesfully recognize regularities in computer opponent and are able to learn to optimally reply to such program. Success in recognition is directly proportional to program's predictability (share of nonrandom moves by computer). Moreover, players are better at learning the pattern that allows them to win and are worse at studying from defeats. Results suggest that people can succesfully use procedurally-rational strategies that are based on rule learning, not only on reactions to past successes and failures.

Keywords: Learning with noise, experiments, repeated games, pattern detection, zero sum games, 'rock-paper-scissors', heuristic recognition ,k-levels, rule learning, nontransitivity

Приложение 1.

Инструкции

Добро пожаловать на экспериментальную сессию. Вам предстоит принять ряд решений, и Вы получите возможность заработать деньги. То, сколько Вы заработаете, будет зависеть как от Ваших решений, так и от решений вашего оппонента. Поэтому очень важно, чтобы Вы внимательно прочитали данные инструкции.

Деньги, которые будут вам положены по итогам эксперимента, будут выплачены вам наличными в конце экспериментальной сессии.

Ваши решения, так же как и ваши результаты, являются анонимными. Мы гарантируем конфиденциальность ваших решений и ответов, и будем анализировать их только в обезличенном виде.

Эти инструкции предназначены только для Вашего личного использования. На протяжении всей экспериментальной сессии Вам запрещено общаться с другими участниками. В случае нарушения этого правила Вы можете быть удалены из эксперимента и лишитесь всех денег, полагающихся за игру.

В случае если у Вас возникнут какие-либо вопросы, поднимите, пожалуйста, руку. Мы подойдем к Вашему рабочему месту и ответим на Ваши вопросы в индивидуальном порядке.

Во время эксперимента мы не будем пользоваться рублями, а будем использовать токены (условные денежные единицы эксперимента). Это значит, что Ваш итоговый результат изначально будет подсчитан в токенах. В конце экспериментальной сессии этот результат в токенах будет переведен в рубли по курсу, который будет отображаться на экране в течении всего эксперимента и составляет:

1 токен = 3 руб

В доп
результатов экспериментальной сессии. В конце сессии каждый участник получит свои деньги в индивидуальном порядке.

Есть ли у вас вопросы?

Эта экспериментальная сессия состоит из 100 раундов. Вам предстоит принять 1 решение в каждом раунде, в игре Камень-Ножницы-Бумага. На каждый ход вам будет отведено 45 секунд.

Правила

Правила игры стандартные:

В каждом раунде Вы выбираете одно из трех возможных действий: Камень, Ножницы или Бумага. Ваш оппонент делает то же самое независимо от вас. Победитель в каждом раунде определяется по следующему правилу

Бумага побеждает камень («бумага камень завернет»);

Камень побеждает ножницы («камень ножницы затупит»);

Ножницы побеждают бумагу («ножницы бумагу разрежут»).

Если оба игрока выбрали одно и то же действие, исходом раунда является ничья.

Оппонент

В качестве оппонента с вами будет играть компьютерная программа.

Действия программы подчиняются некоторому алгоритму, про который известно следующее:

- В момент принятия решения в каждом раунде ему не известно Ваше действие в текущем раунде
- Ему доступна информация о действиях обоих участников (и ваших, и его) в предыдущих раундах.
- Алгоритм программы не меняется на протяжении всей игры
- Алгоритм *может* подчиняться некоторому правилу (закономерности), но может и не подчиняться. Знание закономерности позволит лучше предугадывать действия оппонента (компьютерной программы).

Больше про ее алгоритм Вам ничего не известно.

Вознаграждение

Ваше вознаграждение определяется следующим образом

За каждый выигрыш у программы в каждом раунде Вам начисляется 2 токена, за ничью — 1 токен, за проигрыш Вам - 0 токенов. Количество токенов, начисляемое за исход раунда в зависимости от результата, приведено в таблице ниже.

	Вы выбрали Камень	Вы выбрали Ножницы	Вы выбрали Бумагу
Программа выбрала Камень	1	0	2
Программа выбрала Ножницы	2	1	0
Программа выбрала Бумагу	0	2	1

Кроме того, за каждые пять побед начиная с 30 начисляется бонус в размере 15 токенов. Например, бонус за 4 победы составит 0 токенов, за 30 побед – 15 токенов, за 41 побед 30 токенов. Максимум бонусных очков можно получить за серию из 100 побед, тогда бонус составит 225 токенов.

То есть ваш суммарный выигрыш в токенах будет вычисляться по формуле:

Выигрыш в токенах = (Количество Побед)*2 + (Количество ничьих)*1 + (Бонус)*15

Где Бонус – неполное частное при делении Количества Побед, начиная с 30 на 5. Количество начисляемых бонусов в токенах по мере увеличения количества побед приведены в таблице:

Победы	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100
Бонус	15	30	45	60	75	90	105	120	135	150	165	180	195	210	225

Примеры:

При результате игры 33 победы 33 ничьи 33 поражения игрока, его выигрыш составит: (33)*2+(33)*1+([33/5]) *15=66+33+15=114 токен, тогда выигрыш в рублях без учета оплаты за участие составит – 342 рубля, вместе с оплатой за участие 492 рубля.

При результате игры 50 победы 25 ничьи 25 поражения игрока, его выигрыш составит: (50)*2+(25)*1+([50/5]) *15=100+25+75=200 токен, тогда выигрыш в рублях без учета оплаты за участие составит – 600 рублей, вместе с оплатой за участие 750 рублей.

При результате игры 100 победы 0 ничьи 0 поражения игрока, его выигрыш составит: (100)*2+(0)*1+([100/5]) *15=200+0+225=425 токен, тогда выигрыш в рублях без учета оплаты за участие составит – 1275 рублей, вместе с оплатой за участие 1425 рублей.

Есть ли у Вас вопросы?