В.С. Левченков, Л.Г. Левченкова

СТАЦИОНАРНЫЕ СТРАТЕГИИ В МНОГОШАГОВОЙ ИГРЕ "ДИЛЕММА ЗАКЛЮЧЕННОГО"

Введение

Эволюция кооперации [1] представляет собой отдельную ветвь теории игр, возникшую при исследовании проблем кооперирования игроков, действия которых не ограничены рамками принципа рационального поведения, обычно используемого при теоретико-игровых построениях. Тем самым она изучает слабо формализованные области теории игр, в которых правила оптимального выбора стратегий игроками в процессе достижения ими согласованного поведения не являются определяющими, и требуется иной подход для описания соответствующего поведения. Экспериментальная база такого подхода сформирована компьютерными турнирами Р. Аксельрода [1], которые проводились на основе многошаговой игры "дилемма заключенного" (IPD).

Игра "дилемма заключенного" (PD) [2], являющаяся базисной для IPD, представляет собой классическую биматричную игру, где у каждого игрока имеется всего лишь по две стратегии C и D, исходы использования которых представлены в таблице 1.

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \left(\begin{array}{cc} (R,R) & (S,T) \\ (T,S) & (P,P) \end{array} \right) \end{array}$$

Таблица 1. Исходы в PD.

Вещественные числа R, S, T и P, определяющие конкретную величину выигрыша каждого их игроков в том или ином исходе, подчинены неравенствам

$$T > R > P > S, \tag{1}$$

а в остальном произвольны. В работах по эволюции кооперации добавляется еще одно условие на эти числа

$$R > \frac{T+S}{2}, \tag{2}$$

позволяющее считать ситуацию (C, C) примером кооперированного поведения игроков, а остальные ситуации – той или иной формой проявления ими несогласованного (эгоистического) поведения. Обычно в теории кооперации параметрам присваивают следующие стандартные значения $R = 3$, $S = 0$, $P = 1$, $T = 5$.

Правила соревнований, использованные Р. Аксельродом, отбирали программы определенного вида. Именно, это должны были быть программы, допускавшие реализацию их работы на компьютере и позволявшие без вмешательства человека осуществить l-шаговое (где l – длина траектории игры в IPD, заранее не заданная) взаимодействие любых двух представленных на турнир программ. Эти правила в основном выглядят так:

1) создателям программ известны значения параметров PD;

2) программы совершают ходы, не обладая информацией о текущем ходе противника;

3) в процессе взаимодействия каждая программа знает историю игры (ходы, уже совершенные обоими игроками) в любой ситуации на траектории игры;

4) победителем признается программа, получившая в ходе турнира наибольший суммарный выигрыш.

Анализ результатов турниров позволил Р. Аксельроду выработать определенные эвристические требования, выделяющие хорошо играющие программы. Для проверки этих требований на более широком круге программ были применены идеи генетического подхода к порождению новых алгоритмов [3]. Эти идеи были осуществлены в рамках исследования проблемы эволюции кооперации [4], когда вместо одного проводилась целая серия турниров. При переходе от одного турнира к следующему программы подвергались соответствующему отбору и видоизменению. В результате экспериментов искалась наиболее успешная популяция программ, возникавшая после некоторого числа последовательных турниров. Конечно, эти эксперименты не ставили своей целью нахождение абсолютно лучших программ (в такой игре их просто нет), а пытались достичь понимания поведения и эволюции заданной популяции программ, вид которых прост, но позволяет с определенным сходством воспроизводить ряд эффектов, свойственных эволюционному развитию в живой природе.

Оказалось, что успешно играющие программы принадлежат классу программ, учитывающих при выработке хода только конечное число

предыдущих ситуаций (т.е. историю конечной длины). Эти программы в контексте потенциально бесконечношаговой IPD соответствуют стационарным стратегиям в супериграх [5], обладающих рядом примечательных свойств. Например, любой исход, лежащий в выпуклой оболочке исходов PD может быть с какой угодно точностью реализован в игре парой таких стратегий, причем в виде равновесия Нэша (вывод аналогичный результату широко известной "народной теоремы" [6].

Использование стационарных программ дает возможность по иному взглянуть на проблему достижения кооперации, позволяя ввести понятие эффективной кооперации. При стандартных значениях параметров в PD (см. (1), (2)) примером эффективной ситуации является (C, C), которая лежит на границе Парето выпуклого замыкания исходов этой игры. Однако, если условие (2)) нарушено, то ситуация (C, C) уже не лежит на границе Парето и более эффективным поведением программ в IPD является поочередный выбор ситуаций (C, D) и (D, C) (см. раздел 1). Проблема достижения такого поведения требует более детального описания взаимодействия стационарных программ. В разделе 2 сформулирован соответствующий аппарат и показано, как на языке диаграмм взаимодействия описать режимы, возникающие в процессе игры двух программ, и найти соответствующие стационарные вероятностные распределения. Показывается, что описание программ на языке пространства состояний [7] легко укладывается в схему стационарных программ.

На достижение эффективного кооперативного исхода в IPD оказывает существенное влияние возможность неодновременных ходов противников. Модели этого вида возникают в рамках описания такого явления, как взаимный альтруизм [8] и получили документированное подтверждение в ряде наблюдений над объектами живой природы (см., например, [9] и [10]). Стремление преобразовать модель такого поведения к стандартной PD привело к фактическому переопределению исходной игры до новой игры "лидер-ведомый" с другим набором исходов [11], [12]. В отличие от такого подхода в разделе 3 мы исследуем поведение игроков в неодновременной стандартной PD, опираясь на асимметрию в информации, которая существует между игроками на каждом шаге игры. Строится обобщение программ Tit-for-tat и Pavlov на случай PD с поочередными ходами и исследуется возможность достижения этими программами режима чередования ситуаций (C, D) и (D, C).

В разделе 4 вводятся стационарные программы, позволяющие

находить эффективное кооперативное состояние как в случае стандартной PD, так и нестандартной. Такие программы характеризуются специальными параметрами (критическими уровнями игроков), определяющими момент перехода от режима кооперации в режим наказания противника, когда средний выигрыш игрока опускается ниже его критического уровня.

1 Понятие кооперативного поведения в игре PD

Возникновение кооперации между компьютерными программами в многошаговой игре "дилемма заключенного" (IPD) и дальнейшая эволюция наиболее успешных программ находится в центре внимания теории кооперации, получившей мощный стимул для развития после выхода в свет книги Р. Аксельрода [1].

Вывод о том, что в IPD ситуацией кооперации является только (C, C), в значительной степени обусловлен содержательной трактовкой игры, которой она обязана своим названием [2]. Однако, если использовать иное толкование исходов, представленных в табл. 1, то число ситуаций, которые могут быть оценены, как кооперативное поведение игроков, существенно расширится. К примеру, рассмотрим игру, имеющую следующее описание. Пусть два соседа по району проживания столкнулись с проблемой ремонта своих домов. Каждый из них умеет качественно сделать только часть необходимых работ. Обозначим через D стратегию владельца, отвечающую процессу самостоятельного ремонта дома, а C – с привлечением соседа для выполнения части работ. Ситуации (C, C) и (D, D) отвечают раздельной работе игроков: в первом случае, при выполнении работ в доме соседа, а во втором – в своем доме. Ситуации (C, D) и (D, C) отвечают совместной работе соседей по ремонту одного из домов. Очевидно, ситуацией кооперации в одношаговой игре будет не только (C, C), но также (C, D) и (D, C), поскольку в них соседи не только помогают друг другу, но и работают вместе, что позволяет в максимально возможной степени использовать их способности. Более того, для такой трактовки игры параметры R, T, S, P, удовлетворяя условиям (1), могут нарушать условие (2), поскольку при совместном ремонте обоими игроками качество ремонта домов может быть существенно выше, чем при выполнении тех же операций поотдельности. Таким образом, для такой

трактовки игры возможно выполнение противоположного (2) условия

$$R < \frac{T+S}{2}. \tag{3}$$

При многократном повторении игры к ситуациям кооперации будут относиться и любые последовательности ситуаций типа $\{(C,C),(C,C),(C,D),(D,C),(C,C),\ldots\}$ или $\{(C,D),(D,C),(C,D),(D,C),\ldots\}$ и т.д. Только ситуация (D,D) или последовательность $\{(D,D),(D,D),(D,D),\ldots\}$ с полной уверенностью не может считаться ситуацией кооперации. Даже последовательность $\{(D,D),(C,C),(D,D),(C,C),\ldots\}$ содержит в себе определенные черты кооперативного поведения игроков.

Возникает естественный вопрос, что означает фраза "игроки достигли в игре IPD кооперативного поведения"? Традиционный ответ, неявно содержащийся в работах по теории кооперации, сводится к тому, что таковым может быть признано только поведение, состоящее в выборе ситуации (C,C). Фактически для этого и было введено в описание PD условие (2). Однако, термин "кооперативное поведение игроков" заимствован из теории кооперативных игр, где под кооперацией понимается согласование игроками индивидуального выбора стратегий с целью достижения большего выигрыша, чем они способны гарантированно получить при произвольном поведении противника. Используя такое толкование кооперации, рассмотрим множество исходов PD и его выпуклое замыкание $V = con\{O_1, O_2, O_3, O_4\}$ (см. рис. 1, на котором рассмотрена PD с традиционными значениями параметров $R = 3$, $T = 5$, $P = 1$, $S = 0$).

Множество V содержит те точки, которые отвечают средним выигрышам игроков, приходящимся в IPD на один шаг игры. Ни одна из его точек, кроме O_4, не может быть гарантированно достигнута ни одним из игроков, т.е. все точки множества $V \setminus \{O_4\}$ следует признать исходами того или иного вида кооперативного поведения игроков. Конечно, не все эти точки обладают одинаковыми уровнями эффективности: для каждой из них, за исключением точек множества Парето $O_1O_2O_3$, существуют точки, приносящие больший выигрыш обоим игрокам.

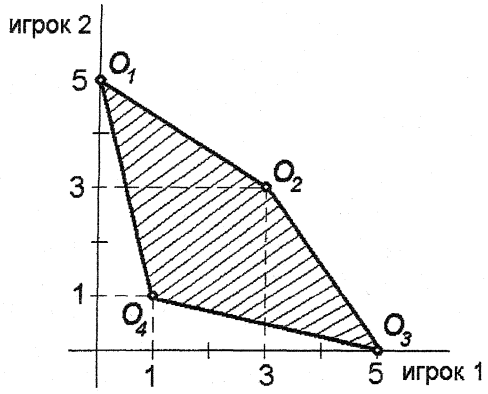

Рис. 1. Множество исходов в IPD.

Таким образом, даже если сузить понятие кооперативного исхода до точки множества Парето, то претендентов в PD будет 3, а в IPD – бесконечно много. Заметим, что для IPD с параметрами, удовлетворяющими (2), точка O_2, отвечающая ситуации (C, C), занимает особое положение в Парето-множестве. Во-первых, выигрыши игроков в ней равны, что является привлекательным свойством в свете теории эгалитаризма. А во-вторых, она достижима на основе выбора единственной чистой ситуации (C, C), что упрощает вид программ, приводящих к этой точке. Две другие точки в Парето-множестве, отвечающие чистым ситуациям (C, D) и (D, C), крайне несимметричны по выигрышам и могут рассматриваться игроками как промежуточный исход в процессе перехода игроков между ситуациями (C, D) и (D, C). Точки, лежащие внутри отрезков $[O_1, O_2]$ или $[O_2, O_3]$, достижимы только как средние исходы, возникающие в IPD при l-кратном повторении одной и той же базисной игры PD. Поскольку в реальных компьютерных экспериментах число l варьируется, поведение игроков в IPD должно задаваться так, чтобы результаты игры слабо зависели от выбора конкретного значения l (если, конечно, оно выбрано достаточно большим).

2 Методы описания программ игры в IPD

Прежде чем описывать программы, реализующие любую точку множества Парето, рассмотрим сначала некоторые частные виды компьютерных программ для игры IPD.

1. Равновесие Нэша (ALL-D).

Согласно этой программе, на каждом шаге игрок использует только одну стратегию, а именно, D. Эта стратегия входит в единственное равновесие Нэша (D, D) игры PD и является наиболее осторожным методом ее ведения. Эту программу можно описать также следующим образом: какая бы ситуация в игре не возникла в некоторый момент, на следующем шаге игрок всегда выбирает D, т.е. схема выбора хода имеет вид:

$$(C, C) \to D, \ \ (C, D) \to D, \ \ (D, C) \to D, \ \ (D, D) \to D. \qquad (4)$$

2. "Зуб-за зуб" (TFT).

В этой программе на первом ходе игрок играет C, а на всех последующих – ту стратегию, которую применил его противник на предыдущем ходе. Схематически это выглядит так

$$(C, C) \to C, \ \ (C, D) \to D, \ \ (D, C) \to C, \ \ (D, D) \to D. \qquad (5)$$

3. "Дают – бери, бьют – беги" (Pavlov, см. [14]).

Программа применяет принцип сохранения хода, использованного на предыдущем шаге игры в случае, когда выигрыш игрока достаточно велик (равен R или T), что достигается в ситуациях (C, C) или (D, C). Она рекомендует смену хода, если выигрыш мал (равен P или S). Схема зависимости выбора хода от ситуации в предыдущий момент игры имеет вид

$$(C, C) \to C, \ \ (C, D) \to D, \ \ (D, C) \to D, \ \ (D, D) \to C. \qquad (6)$$

3. "Толерантная" TFT (GTFT).

Жесткое поведение программы TFT, которая не оставляет без наказания выбор противником хода D, было смягчено в версии этой стратегии, называемой GTFT. Программа GTFT оставляет без наказания $1/10$ часть случаев выбора хода D противником, а в остальном действует

как TFT. Схема ее действия уже не детерминирована

$$(C,C)\to C;\ \ (C,D)\begin{array}{l}\nearrow D(\frac{9}{10})\\ \searrow C\,(\frac{1}{10})\end{array};\ \ (D,C)\to C;\ \ (D,D)\begin{array}{l}\nearrow D(\frac{9}{10})\\ \searrow C\,(\frac{1}{10})\end{array} \qquad (7)$$

и характеризуется вероятностным выбором хода в ситуациях (C,D) и (D,D). Соответствующая вероятность указана на схеме (7) после выбранного хода.

6. "Кающаяся" TFT (CTFT). Очень интересная программа была сформулирована на основе введения понятия "состояния игрока", заимствованного из теории автоматов. Суть идеи состоит в том, что выбор хода игрока зависит не только от истории игры, но и от "состояния", в котором рассматриваемый игрок находится ([7], [13] и [14]). Вводится три возможных состояния: 1 – "чувство вины", 2 – "удовлетворение", 3 – "раздражение".

Для того, чтобы описать эту программу схемой типа (4)-(7), обозначим через C_k, D_k возможные стратегии игрока в его k-том состоянии. Тогда программу CTFT можно реализовать на основе схемы

$$\begin{array}{l}(C_1,C)\to C_2,\ \ (C_1,D)\to C_2,\ \ (C_2,C)\to C_2,\ \ (C_2,D)\to D_3,\\ (C_3,C)\to D_2,\ \ (C_3,D)\to D_3,\ \ (D_1,C)\to C_1,\ \ (D_1,D)\to C_1,\\ (D_2,C)\to C_1,\ \ (D_2,D)\to C_2,\ \ (D_3,C)\to C_2,\ \ (D_3,D)\to D_3.\end{array} \qquad (8)$$

Эта схема отличается тем, что в ней базисная игра PD, изображенная в табл. 1 заменена другой базисной игрой. В ней игрок 1 имеет в своем распоряжении не две стратегиии C и D, а шесть: C_1, C_2, C_3, D_1, D_2, D_3. Его противник может быть наделен как двумя стратегиями C и D, так и большим их числом, в зависимости от числа его состояний, значения которых известны его противнику. Однако все используемые стратегии представляют собой одно из двух "действий", C или D, снабженных номером состояния.

Покажем, что аналогичным образом можно изобразить поведение программ для IPD, описываемых так называемыми стратегиями пространства состояний [7]. Для этого вводится пространство состояний X, пространство действий $A=\{C,D\}$ и две функции: функция действия r и функция переходов f

$$r:\ X\to A,\quad f:\ X\times A^2\to X.$$

Функция r показывает, какое действие $r(x)$ совершает первый игрок, находясь в состоянии $x \in X$, а функция f предписывает, в какое состояние $f(x,(a,a'))$ он переходит из состояния x, если первый игрок и его партнер совершают действия a и a', соответственно. Для такой программы базисная игра PD заменяется игрой, в которой первый игрок имеет множество стратегий $S_1 = \{[a]_x\}_{\substack{x\in X\\ a\in A}}$ и следующую схему игры: $\forall x \in X$, $\forall a, b \in A$

$$([a]_x, b) \to [r(x')]_{x'}, \quad \text{где } x' = f(x,(a,b)). \tag{9}$$

Кроме (4)-(8) возможны также схемы, учитывающие при выборе следующего хода не только информацию о предыдущей ситуации игры, но и более глубокую историю. Например, при формировании генетического алгоритма, применяемого для описания эволюции программ игры в IPD, Р. Аксельрод [4] конструирует программы, использующие при выборе хода знание трех предыдущих ситуаций.

Анализ схем переходов (4)-(8) позволяет описать взаимодействие игроков на основе следующих диаграмм. Для их построения используем тот факт, что схемы (4)-(8) построены однотипным образом: зная ситуацию игры на предыдущем шаге, мы можем определить ситуацию на следующем шаге, если используем схемы, применяемые обоими игроками. Например, для схем (4)-(6) существуют всего четыре ситуации (C,C), (C,D), $D,C)$ и (D,D). Все траектории игры можно описать, рассмотрев пути в графе с вершинами ab $(a,b \in \{C,D\})$ и дугами, проведенными из вершины ab в вершину $a'b'$, если в схеме игрока 1 есть переход $(a,b) \to a'$, а в схеме игрока 2 – переход $(a,b) \to b'$. Для иллюстрации, диаграмма взаимодействия программ TFT и ALL-D представляет собой граф, изображенный на рис. 2.

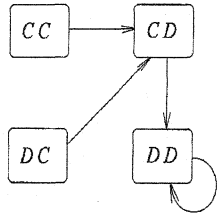

Рис. 2. Диаграмма взаимодействия TFT и ALL-D

На диаграмме отчетливо видно, что с каких-бы стратегий игроки не начинали эту игру, не далее, чем на третьем ходе они попадают в ситуацию (D, D) и продолжают в ней оставаться неограниченно долго. Иные диаграммы возникают при взаимодействии программы TFT с TFT или с программой Pavlov (рис. 3,4).

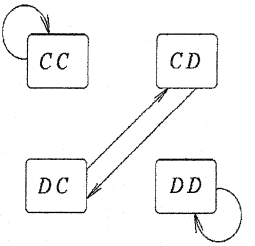

Рис. 3. TFT против TFT

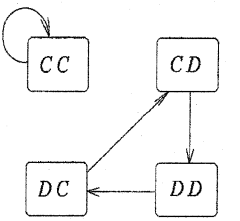

Рис. 4. TFT против Pavlov

Из рис. 3 следует, что в зависимости от начального хода возможны три режима в игре TFT с TFT: 1) выбор только ситуации (C, C); 2) "перепрыгивание" из ситуации (C, D) в (D, C) и обратно; 3) повторение равновесия Нэша (D, D). Обычно последние два режима исключаются условием выбора начального хода. Однако, при наличии ошибок при выборе ходов игроками возможны и другие режимы. Это обстоятельство оказывает определяющее влияние на результат игры, достигаемый в условиях "шума" [13]. Аналогичный эффект имеет место и при игре TFT с программой Pavlov (см. рис. 4). Эти недостатки в игре TFT (или Pavlov) элиминируются при использовании "толерантной" TFT (см. рис. 5, 6).

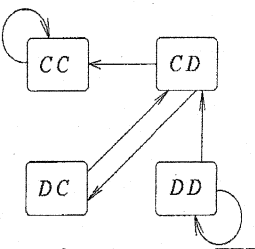

Рис. 5. GTFT против TFT

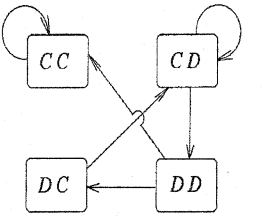

Рис. 6. GTFT против Pavlov

С какой бы ситуации теперь ни начинали игроки, они неизбежно переходят в ситуацию (C, C) и в ней остаются.

Аналогичное рассмотрение применимо и к схеме (8), только число вершин у графа будет больше. Однако если считать, что CTFT будет работать без ошибок, то схему (8) легко упростить. Для этого заметим, что в ней нет переходов к стратегиям C_3, D_1, D_2 ни из каких ситуаций. Значит, эти стратегии не встретятся на траектории игры и тем самым (8) преобразуется в схему

$$C_2C \to C_2, \;\; C_2D \to D_3, \;\; D_3C \to C_2, \;\; D_3D \to D_3,$$

которая после отождествления $C_2 \equiv C$, $D_3 \equiv D$ сводится к схеме (5).

Таким образом, CTFT без ошибок работает точно также, как и TFT. Различие возникает, только если в ее работе возникают ошибки (например, совершается не то действие, которое предписывается состоянием согласно схеме (8)), либо имеется вероятностный "шум", приводящий к возможности выбора другой стратегии вместо предписываемой детерминированной схемой. При этом наличие различных состояний у программы (и тем самым возникновение эффекта расширения множества стратегий у базисной игры) приводит к более широким модификациям работы программы при учете влияния различных ошибок, поскольку при таких ошибках могут меняться не только действия (т.е. C или D), но и состояния. С этой точки зрения вводимые состояния должны быть наблюдаемыми, т.е. имеющими физическую интерпретацию, а не появляться искусственным образом.

Схемы (4)-(8) реализуют частные примеры так называемых стационарных стратегий, введенных для описания супериг в [5]. В общем случае, стационарные стратегии представляют собой условные вероятностные распределения $\nu_i(x|s^{(1)}...s^{(h)})$, заданные на множествах стратегий игроков S_i $(i = 1, 2)$ и показывающие, с какой вероятностью игроки выбирают ту или иную возможную стратегию, если им известны h предыдущих ситуаций игры $[s^{(1)}...s^{(h)}]$ (здесь h – целое неотрицательное число, называемое глубиной истории игры). Для схем (4)-(8) $h = 1$, а для схем из [4] $h = 3$. Схемы (4)-(6), (8) детерминированы, и для них ν_i принимают только значения 0 или 1, а в схеме (7) для историй (C, D) и (D, D) возникают вероятностные распределения $\nu_i(D|(C, D)) = \nu_i(D|(D, D)) = 9/10$; $\nu_i(C|(C, D)) = \nu_i(C|(D, D)) = 1/10$.

Стационарные стратегии несмотря на то, что они носят

вероятностный характер, связаны с единственной траекторией игры, реализуемой в соответствии с ходами, осуществляемыми игроками. Таким образом, результат взаимодействия игроков определяется не вероятностным процессом, а всего лишь одной его реализацией. Это обстоятельство накладывает определенные ограничения на процедуру оценки выигрыша, получаемого игроками. В этом случае нельзя говорить о среднем выигрыше, получаемом игроками от осуществляемого ими вероятностного процесса взаимодействия, а всего лишь об арифметическом среднем выигрыша, вычисляемого по реализованной траектории. Вообще говоря, этот выигрыш может зависеть как от конкретной реализации траектории, так и от числа шагов игры, использованном игроками до ее остановки (на практике, IPD всегда имеет конечное число ходов). Оказывается, что если использовать стационарные стратегии, определенные выше, то, во-первых, арифметические средние будут всегда иметь предел при числе шагов, стремящемся к бесконечности, и во-вторых, этот предел будет одним и тем же (почти) для всех возможных траекторий игры. Таким образом, стационарные стратегии $\nu_i(x|s^{(1)}...s^{(h)})$ порождают на ситуациях игры совместное распределение вероятностей, показывающее асимптотическую (при числе шагов игры $n \to \infty$) частоту появления различных ситуаций почти в каждой траектории игры. Эти распределения могут быть эффективно найдены из соответствующей системы линейных уравнений [5]. Например, для $h = 3$ совместное распределение $\rho(s)$ на $S_1 \times S_2$ при определенных условиях на вид ν_1 и ν_2, находится из системы уравнений

$$\rho(s) = \sum_{s',s'' \in S_1 \times S_2} \beta(s', s'', s),$$

где при $s = (x, y)$

$$\beta(s', s'', s) = \sum_{z \in S_1 \times S_2} \nu_1(x|z, s', s'')\nu_2(y|z, s', s'')\beta(z, s', s'').$$

Если $h = 1$, то эти уравнения переходят в более простые соотношения

$$\rho(s) = \sum_{s'} \rho(s')\nu(s', s), \tag{10}$$

в которых

$$\nu(s', (x, y)) = \nu_1(x|s')\nu_2(y|s') \tag{11}$$

является неразложимой матрицей. В силу того, что $\nu_i(x|s)$ – вероятностное распределение, эта матрица стохастична

$$\sum_{s \in S_1 \times S_2} \nu(s', s) = 1.$$

Вместе с условиями нормировки

$$\sum_{s \in S_1 \times S_2} \rho(s) = 1$$

соотношения (10) показывают, что при $h = 1$ выбор ситуаций производится игроками согласно асимптотическому распределению конечной эргодической марковской цепи с переходными вероятностями (11). Это обстоятельство широко использовалось для описания результата взаимодействия программ в условиях шума (см. [12], [15], [16]).

Средние значения выигрышей игроков $\bar{v}_i$ находятся по $\rho(s)$ из очевидных соотношений

$$\bar{v}_i = \sum_{s \in S_1 \times S_2} \rho(s) v_i(s), \tag{12}$$

где $(v_1(s), v_2(s))_{s \in S_1 \times S_2}$ – биматрица исходов игры (см. табл. 1).

В [5] показано, что они также могут быть получены, например, по значениям $\beta(s, s', s'')$ из соотношений

$$\bar{v}_i = \frac{1}{3} \sum_{s, s', s'' \in S_1 \times S_2} \beta(s, s', s'')(v_i(s) + v_i(s') + v_i(s'')). \tag{13}$$

Ограничение вида программ схемами с историей конечной глубины учитывает то обстоятельство, что реальное разыгрывание IPD не может проводиться бесконечно, а ограничивается некоторым конечным числом шагов l [1]. Для того, чтобы эксперименты, проводимые в рамках эволюции кооперации, слабо зависили от значения l, необходимо, чтобы поведение игроков было стационарным, т.е. они довольно быстро достигали некоторого устойчивого образа действий, приводящего к фиксированному значению выигрыша на один шаг игры. Тем самым, при выборе достаточно большого l ($l \geq l_0$) экспериментатор мог надежно выявлять соответствующее поведение игроков и находить значение выигрыша, которое приобретало вид объективной характеристики, слабо зависящей от выбора конкретной величины l (но не l_0). Стационарные стратегии как раз и обеспечивают реализацию такого подхода, поскольку какая бы ни возникала конкретная реализация траектории игры, почти для всех таких реализаций (т.е. для "типичных" траекторий) арифметические средние выигрышей игроков на данном куске траектории сходятся к величинам (12), вычисляемым по стационарному распределению для некоторой марковской цепи. Однако, в задачах, где существует ненулевая (хотя и малая) вероятность прерывания игры, уравнения

более сложны. При рассмотрении игр с фиксированной вероятностью окончания обычно предполагают, что на любом шаге игры возможно ее прерывание с некоторой вероятностью ϵ ($\epsilon \ll 1$) (см., например, [1] или [15]). Это предположение адекватно идее дисконтирования выигрышей, используемой при анализе суперигр, но в контексте теории кооперации является чрезвычайно упрощенным. Именно, с какой стати игра равновероятно останавливается на любом шаге, если, например, игроки достигли состояния кооперации и продолжение игры является обоюдовыгодным. Другое дело, если игра развивается так, что один из игроков недоволен ее течением и не видит смысла продолжения игры при возникшем типе действий партнера. Например, при игре TFT с ALL-D игроки быстро переходят в тупиковую ситуацию (D, D), чего можно избежать, введя малую вероятность ϵ прерывания игры после попадания в это состояние. При такой модификации нас интересует результат игры двух программ на тех траекториях, где прерывание не происходит. С точки зрения теории марковских цепей мера указанных траекторий равна нулю, поэтому их свойства нельзя изучать на основе этой теории. Однако, используя подход [5], можно оценить арифметическое среднее выигрышей игроков по таким траекториям. Оказывается, что на них существует стационарное распределение $\rho(s)$, но оно описывается иной, чем (10) системой уравнений. Именно

$$\begin{gathered}\rho(s) = \alpha(s)\beta(s), \\ \sum_{s \in S_1 \times S_2} \nu(s', s)\alpha(s) = \lambda_0 \alpha(s'), \\ \sum_{s' \in S_1 \times S_2} \beta(s')\nu(s', s) = \lambda_0 \beta(s),\end{gathered} \tag{14}$$

где λ_0 – максимальное собственное значение матрицы $\nu(s, s')$.

Например, при взаимодействии двух программ TFT (с шумом p и вероятностью ϵ прерывания игры в ситуации (D, D)) эта матрица с точностью до членов порядка p ($p \ll \epsilon \ll 1$) имеет вид (см. [5])

$$\begin{array}{c} \\ CC \\ CD \\ DC \\ DD \end{array} \begin{array}{c} \begin{array}{cccc} CC & CD & DC & DD \end{array} \\ \begin{pmatrix} 1-2p & p & p & 0 \\ p & 0 & 1-2p & 0 \\ p & 1-2p & 0 & p \\ 0 & p & p & 1-2p-2\epsilon \end{pmatrix} \end{array} \tag{15}$$

Согласно (14),

$$\rho(s) = \left(\frac{1}{2} - \frac{\sqrt{2}p}{8\ \epsilon},\ \frac{1}{4} + \frac{\sqrt{2}p}{16\ \epsilon},\ \frac{1}{4} + \frac{\sqrt{2}p}{16\ \epsilon},\ 0\right) \approx \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0\right),$$

что резко контрастирует с ситуацией игры двух TFT с шумом уровня p без прерывания, где распределение $\rho(s)$ оказывается равным $\rho(s) = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4},\right)$ (см. [16]).

3 Проблема кооперации в нестандартной PD

Особенно ярко проблема достижения состояния кооперации возникает для IPD, в которой условие (2) не выполнено. Это имеет место, например, когда $T = 8$, а остальные параметры имеют стандартные значения. Расположение соответствующих исходов представлено на рис. 7.

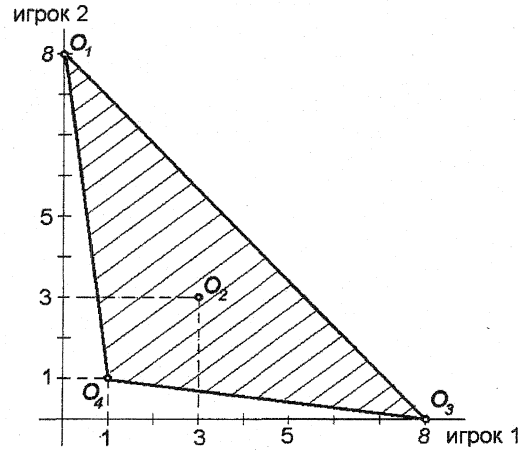

Рис. 7. Исходы для нестандартной IPD.

Для IPD с такими параметрами множество эффективных ситуаций кооперации совпадает с отрезком O_1O_3, а традиционная точка (C, C) не является эффективной и уступает по выигрышам эгалитарному исходу $(4, 4)$. Любая точка отрезка O_1O_3 может быть реализована как средний исход IPD на траектории игры, содержащей только ситуации O_1 и O_3, появляющиеся в этой траектории с заданными частотами. Содержательно, такая траектория может быть также проинтерпретирована

как последовательность чередующихся (с некоторой частотой) актов благотворительности одного игрока по отношению к другому, когда выгода от кооперативного взаимодействия равна нулю для первого участника и максимальна для второго. Именно такие взаимодействия рассматривались в контексте ансамбля игр в [17] в рамках рассмотрения проблем кооперации без взаимности. Важная проблема, которая стоит при изучении таких игр – это исследование программ, достигающих эффективной кооперации в таких условиях. Заметим, что две TFT согласно схеме их взаимодействия содержат цикл $(C,D)\overset{\rightarrow}{\leftarrow}(D,C)$, позволяющей реализовать эгалитарную точку $(4,4)$, лежащую на отрезке O_1O_3. Однако, проблема состоит в том, что без специального "усилия" эти программы не могут выйти на соответствующий режим: если их первый ход C, то они реализуют цикл $(C,C)\overset{\rightarrow}{\leftarrow}(C,C)$; а если первый ход D, то возникает цикл $(D,D)\overset{\rightarrow}{\leftarrow}(D,D)$. Только ошибка в выборе хода одной из программ может привести к желаемому циклу. Можно слегка модифицировать TFT, убрав ее стремление к выбору хода C в ситуации (C,C), применив более жесткую схему

$$(C,C)\to D,\ \ (D,C)\to C,\ \ (C,D)\to D,\ \ (D,D)\to D. \qquad (16)$$

Тогда диаграмма взаимодействия такой программы с TFT имеет вид (рис. 8), показывающий, что при безошибочной игре программы приходят к циклу $(C,D)\overset{\leftarrow}{\rightarrow}(D,C)$.

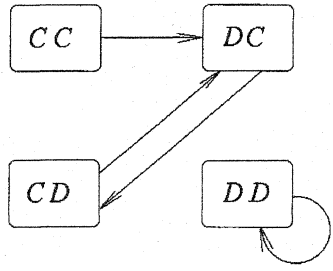

Рис. 8.

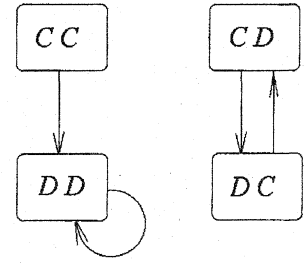

Рис. 9.

Однако, две программы со схемами (16) приводят к диаграмме, изображенной на рис. 9. С какого бы (одного и того же) хода ни начинали эти программы, они неизбежно попадают в ситуацию (D,D).

Важным элементом, позволяющим преодолеть трудности подобного типа, является использование модификации игры, допускающей неодновременность выбора хода партнерами. Одновременность ходов в PD связана со специальной исходной трактовкой этой игры [2] и совсем не характерна для многих типов взаимодействий, где целью игроков является достижение кооперативного эффекта. Обычно, в таких случаях намерение (или ход) партнера становится известным игроку до выбора своего действия. В частности, в теории сделок используется процедура поочередного выдвижения предложений партнерами. Соответствующий вариант IPD обычно называют дилеммой заключенного с переменной последовательностью ходов (APD). В стратегическом смысле наиболее принципиальное отличие APD и одновременной IPD состоит в различной информированности игроков в момент совершения ими очередного хода: игрок, делающий ход первым, знает только историю игры, а его партнер кроме истории знает и текущий ход противника. Это обстоятельство существенно меняет возможности игроков при достижении того или иного типа кооперативного поведения.

Рассмотрим сначала APD с поочередными ходами, например, пусть игрок 1 в нечетные моменты игры ходит первым, а игрок 2 делает это в четные моменты. Тогда стратегия игрока 1 для истории $\varphi^h = [s^{(1)} \dots s^{(h)}]$ $s^{(i)} \in S_1 \times S_2$ содержит две функции, $\nu_1^-(x|\varphi^h)$ и $\nu_1^+(x|\varphi^h y)$, определяющие его ходы в нечетные и четные моменты игры, соответственно. Поскольку в нечетный момент он ходит первым, то вероятность $\nu_1^-(x|\varphi^h)$ выбора его хода $x \in S_1$ зависит только от истории игры φ^h глубины h. В четные моменты он ходит вторым и вероятность $\nu_2^+(x|\varphi^h y)$ выбора его хода x зависит не только от истории игры, но и от хода противника $y \in S_2$. Аналогично, поведение второго игрока определяется двумя функциями $\nu_2^-(y|\varphi^h)$ и $\nu_2^+(y|\varphi^h x)$, показывающими вероятность выбора хода $y \in S_2$ в четные и нечетные моменты игры, соответственно. Использование двух функций вместо одной при выборе ходов в IPD расширяет возможности программ и позволяет осуществить их выход на цикл $CD \overrightarrow{\leftarrow} DC$.

Для иллюстрации, рассмотрим обобщение TFT на случай APD, положив

$$\nu^-(C|(C,C)) = \nu^-(D|(C,D)) = \\ = \nu^-(C|(D,C)) = \nu^-(D|(D,D)) = 1, \tag{17}$$

с начальным условием: в первый момент игры игрок выбирает стратегию C. Очевидно, сотношения (17) совпадают с традиционной схемой (5) для

TFT.

Функцию ν^+ выберем в виде

$$1 = \nu^+(D|(C,C)C) = \nu^+(C|(C,C)D) = \nu^+(C|(D,C)C) =$$

$$= \nu^+(C|(D,C)D) = \nu^+(D|(C,D)C) = \tag{18}$$

$$= \nu^+(D|(C,D)D) = \nu^+(C|(D,D)C) = \nu^+(D|(D,D)D)$$

с начальным условием: если в первый момент противник играет C, то применяется стратегия D; в противном – C.

Назовем для краткости программу с выбором ходов вида (17) и (18) "упорядоченной TFT" (STFT).

Из (18) видно, что при наличии информации о ходе противника STFT при историях (C,C), (C,D) и (D,C) поддерживает предложение противника разыгрывать цикл на ситуациях (C,D) и (D,C) и наказывает его ходом D, если тот пытается извлечь одностороннюю выгоду.

Диаграмма взаимодействия двух STFT имеет вид, показанный на рис. 10.

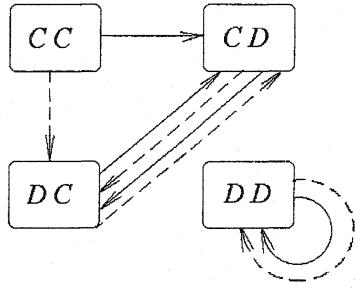

Рис. 10.

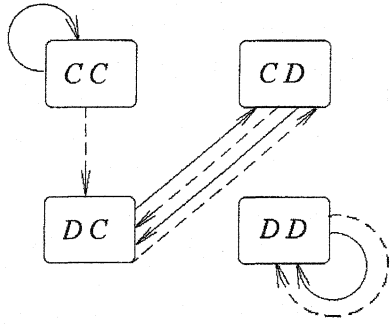

Рис. 11.

Сплошная стрелка указывает направление перехода, если первый ход при указанной в вершине графа истории делает игрок 1, а пунктирная линия – если первый ход делает второй игрок.

Аналогичная диаграмма для STFT и TFT представлена на рис. 11: она отличается только одним из переходов для истории (C,C).

Траектории игры восстанавливаются по диаграммам на рис. 10, 11 следующим образом. Каждая траектория представляет собой

последовательность ситуаций, достигаемых на основе чередующихся смежных сплошных и пунктирных стрелок. Конкретный вид траектории зависит от очередности ходов игроков. Из диаграмм видно, что STFT достигает цикла $DC \rightleftarrows CD$ при безошибочной игре с программами TFT и STFT. Однако, при игре с программой Pavlov этого не происходит (см. диаграмму на рис. 12).

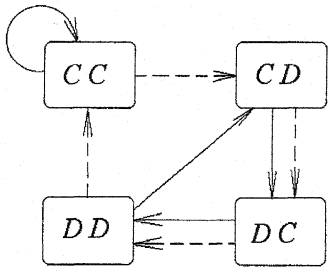

Рис. 12.

Для преодоления этой проблемы, программу Pavlov можно модифицировать аналогично STFT. Именно, определим программу SPavlov со схемой переходов

$$1 = \nu^-(C|(C,C)) = \nu^-(D|(C,D)) = \nu^-(D|(D,C)) = \nu^-(C|(D,D)) \tag{19}$$

(первый ход: C),

$$\begin{aligned} 1 &= \nu^+(D|(C,C)C) = \nu^+(D|(C,C)D) = \nu^+(D|(C,D)C) = \\ &= \nu^+(C|(D,C)D) = \nu^+(C|(C,D)D) = \nu^+(C|(D,C)C) = \\ &= \nu^+(C|(D,D)C) = \nu^+(D|(D,D)D) \end{aligned} \tag{20}$$

(первый ход: если C, то D; если D, то C).

Диаграммы взаимодействия SPavlov с программами TFT, SPavlov или STFT представлены на рис. 13, 14 и 15, соответственно.

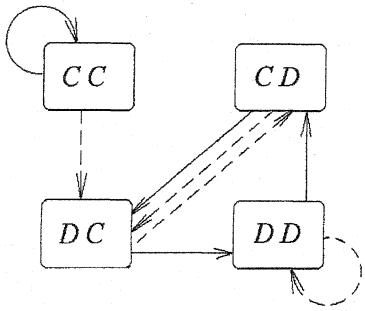

Рис. 13.

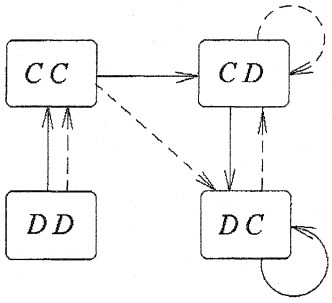

Рис. 14.

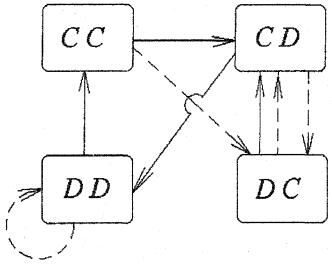

Рис. 15.

Из диаграмм видно, что в последних двух случаях программы переходят в цикл $(C,D) \rightleftarrows (D,C)$ из любых ситуаций, а в первом случае – только если они исходно оказываются в ситуации (C,D) и первым ходит игрок 1, что, впрочем, выполняется согласно начальному условию.

Если рассматривать диаграммы на рис. 10-15 в условиях "шума" [13], то все вершины графов будут связаны друг с другом соответствующими переходами. Возникает естественный вопрос об оценке средних выигрышей по реализациям возникающего вероятностного процесса. В отличие от прежнего рассмотрения, стратегии (17)-(20) не являются стационарными, так как вид стратегии, применяемой игроком, зависит от четности номера момента игры. Однако, даже в этом случае стационарное распределение будет существовать, если изменить понятие ситуации игры, как это было предложено в [11] в контексте специально определенной игры "лидер-ведомый" с переменой ролей в ходе разыгрывания. Именно,

для программ с глубиной истории h ($h \geq 1$) и строгим чередованием обязательства делать ход первым, выберем в качестве элемента игры $z = \{(x_1, y_1), (x_2, y_2)\}$ упорядоченную пару последовательных ситуаций (x_1, y_1) и (x_2, y_2) на траектории игры. В этих терминах траектория игры $\theta = (z_i)_{i=0}^{\infty}$ оказывается бесконечной последовательностью элементов

$$z_i = \{(x_1^{(i)}, y_1^{(i)}), (x_2^{(i)}, y_2^{(i)})\}, \quad \text{где} \quad x_1^{(i)}, x_2^{(i)}, y_1^{(i)}, y_2^{(i)} \in \{C, D\}.$$

Пусть $\nu_1^-(x|\varphi^h)$, $\nu_1^+(x|\varphi^h y)$, $\nu_2^-(x|\varphi^h)$, $\nu_1^+(x|\varphi^h y)$ – стратегии игроков для истории φ^h при условии, что в нечетный момент первым делает ход игрок 1, а в четный момент – игрок 2. Поскольку в каждом элементе в любой момент игры первый ход делает игрок 1, то при любом h почти всякая траектория игры θ будет характеризоваться определенной частотой появления в ней заданного элемента $z \in (S_1 \times S_2)^2$. Для частного случая $h = 1$ эта частота может быть найдена из системы (10) с матрицей ν, имеющей следующий вид. Пусть $\varphi^2 = [\varphi_0 \varphi_1]$, где $\varphi_i \in S_1 \times S_2$, а $\psi^2 = \{(x_1, y_1), (x_2, y_2)\}$. Тогда

$$\nu\left(\varphi^2, \psi^2\right) = \nu_1^-\left(x_1, \varphi_1\right) \nu_2^+\left(y_1|\varphi_1 x\right) \nu_1^+\left(x_2|(x_1, y_1) y_2\right) \nu_2^-\left(y_2|(x_1, y_1)\right).$$

Для неразложимой матрицы ν и почти всякой $\theta = (z_i)_{i=0}^{\infty}$ существует единственное стационарное распределение $\rho(z)$ на элементах игры, являющееся левым собственным вектором этой матрицы. Средние выигрыши игроков ($i = 1, 2$) находятся по $\rho(z)$ согласно сооотношениям

$$\bar{v}_i = \frac{1}{2} \sum_{\varphi_0, \varphi_1} [(v_i(\varphi_0) + v_i(\varphi_1)] \rho(\{\varphi_0, \varphi_1\}). \tag{21}$$

Ситуация будет иной, если обязательство делать ход первым формируется согласно некоторому вероятностному распределению. В этом случае выигрыш игрока на различных траекториях игры будет, вообще говоря, различным и кроме того, зависящим от числа шагов игры. Поскольку для достижения кооперативного поведения важным обстоятельством является предсказуемость поведения партнеров, то естественным режимом в неодновременной IPD является режим поочередного выбора первого хода игроками. Такой режим особенно важен, как мы видим из рис. 10-15, для IPD с нестандартными исходами (см. рис. 7), где эффективная кооперация достигается только на отрезке $[O_1 O_3]$ путем поочередного выбора ситуаций O_1 и O_3. Заметим, что для этого достаточно, чтобы средние выигрыши игроков лежали в какой-то определенной точке этого отрезка, а не только в $(4, 4)$. Выход на такое

решение достигается специальными стратегиями, вид которых мы опишем в следующем разделе.

4 Эффективная кооперация в IPD

Согласно "народной" теореме (см., например, [6]) любая точка множества исходов, изображенного на рис. 7, может быть реализована как равновесие Нэша в IPD. Однако, возможность наказания противника, если он не следует благоприятному для вас способу действия, разрушает устойчивость равновесия Нэша: уклоняясь от такого равновесия при использовании стратегии наказания, вы, возможно, уменьшаете свой выигрыш, но одновременно, с гарантированной возможностью уменьшаете выигрыш противника. По этой причине многошаговые игры требуют введения дополнительных параметров, характеризующих игроков и позволяющих выделить некоторое равновесие Нэша, как основу для соглашения при выборе решения игры [16]. В качестве таких параметров естественно использовать величины, определяющие момент перехода игроков к стратегии наказания в ходе продвижения по траектории игры. Этого можно достигнуть, полагая, например, что каждый игрок i устанавливает некоторый предельный (критический) уровень c_i своего среднего дохода, ниже которого он не согласен опуститься при использовании стационарных стратегий, обеспечивающих согласованные действия игроков. Наличие такого уровня автоматически приводит к существованию соответствующего числа f_i, ограничивающего сверху возможный уровень среднего дохода противника. Число f_i выбирается так, что точка с координатами c_i, f_i лежит на границе Парето $[O_1O_3]$ множества $V = [O_1O_3O_4]$ – выпуклой оболочки множества всех исходов игры (см. рис. 16).

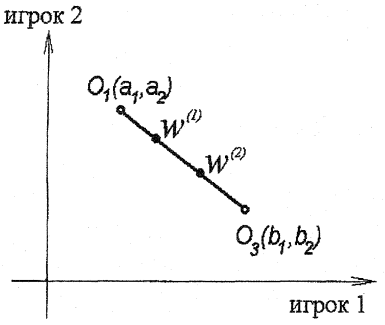

Рис. 16. Критические точки игроков.

Попытка игрока превысить уровень f_i обязательно приводит со стороны противника к использованию стратегии наказания. Таким образом, эффективная кооперация игроков может возникнуть в такой точке на границе V, в которой их выигрыши лежат в интервалах (c_1, f_2) и (c_2, f_1), соответственно. Точки $w^{(1)} = (c_1, f_1)$ и $w^{(2)} = (f_2, c_2)$ назовем *критическими* точками игроков (см. рис. 16).

Точки $w^{(1)}$ и $w^{(2)}$, лежащие на отрезке $[O_1O_3]$, могут быть аналитически описаны линейными комбинациями соответствующих концевых точек

$$w_i^{(1)} = \alpha a_i + (1-\alpha)b_i,$$
$$w_i^{(2)} = \beta a_i + (1-\beta)b_i, \tag{22}$$

здесь α и β – некоторые вещественные числа из отрезка $[0, 1]$.

Предположим, что игроки знают положение критической точки противника и хотят реализовать некоторое равновесие Нэша в стационарных стратегиях. То, каким оно будет, зависит от конкретного расположения точек $w^{(1)}$ и $w^{(2)}$ на границе Парето. Будем считать, что при выборе стационарных стратегий игроки исходят из глубины истории h, а их критические точки заданы выражениями (22), в которых

$$\alpha = \frac{s}{h} \quad \text{и} \quad \beta = \frac{r}{h} \tag{23}$$

являются некоторыми рациональными числами.

В этом случае, например, разыгрывание точки $w^{(1)}$ состоит в совместном выборе s раз исхода (C, D) и $(h - s)$ раз исхода (D, C). Будет ли игрок 2 согласен с таким выбором стационарных стратегий зависит от положения его критической точки $w^{(2)}$. Согласно смыслу

критических уровней для существования кооперативного поведения необходимо выполнение следующих условий

$$w_1^{(1)} \le w_1^{(2)}, \qquad w_2^{(1)} \ge w_2^{(2)}. \tag{24}$$

При нарушении одного из неравенств (24) игроки неизбежно переходят в режим взаимного наказания друг друга, поскольку в V не оказывается ни одной точки, в которой средний выигрыш каждого участника был не менее его критического уровня.

Поскольку игроки знают расположение критической точки противника, то каждый из них хотел бы реализовать равновесие, отвечающее именно этой точке: для игрока 1 наилучший исход – это $w^{(2)}$, а для игрока 2 – $w^{(1)}$. Тогда игрок 2, например, стремится $h-s$ раз сыграть O_3 и s раз – O_1. Однако, игроку 1 выгодно сыграть O_1 только r раз и перейти в точку O_3. Поскольку расположение критических точек на рис. 16 таково, что $h > s > r > 0$, то игроки будут вынуждены r раз играть O_1 и $h-s$ раз играть O_3, в противном случае их действия окажутся не согласованными и приведут к использованию стратегий наказания. В результате будет реализован исход w^*, такой что

$$w_i^* = \gamma a_i + (1-\gamma) b_i, \tag{25}$$

где $\gamma = r/(r+h-s)$.

Заметим, что при $s > r$

$$\frac{s}{h} - \frac{r}{r+h-s} = \frac{(s-r)(h-s)}{h(h+r-s)} > 0$$

и

$$\frac{h-r}{h} - \frac{(h-s)}{h+r-s} = \frac{r(s-r)}{h(h+r-s)} > 0.$$

Таким образом, точка w^* находится на отрезке $[O_1 O_3]$ между точками $w^{(1)}$ и $w^{(2)}$.

Выбор (25) в качестве решения игры можно обосновать также следующим образом. Поскольку игроки стремятся сыграть Парето-оптимальным образом, их стратегиями становятся числа ходов, в течении которых они выбирают точки O_1 и O_3. При этом смена выбора точки находится в руках одного из игроков. Действительно, точка O_1 дает игроку 2 больший выигрыш, чем точка O_3. Поэтому он готов находиться в ней произвольное число раз, и ее смена на O_3 определяется только игроком 1, которому O_3 выгоднее, чем O_1. Пусть игрок 1 выбирает точку

O_3 r' раз. Тогда $r' \geq r$, т.к. в противном случае на истории длины h будет реализована точка v', находящаяся ближе к O_3, чем $w^{(2)}$, что заставит игрока 2 применить стратегию наказания (т.к. $w^{(2)}$ его критическая точка). Аналогично, игрок 2 не может выбирать точку O_3 менее, чем $(h-s)$ раз. Таким образом, если он выбирает точку O_3 s' раз, то $s' \geq h-s$. В результате, стратегиями игроков (назовем эти стратегии *согласованными с критическими уровнями*) становится выбор пары чисел: r' для первого игрока и s' для второго, при этом $r' \geq r$, $s' \geq h-s$.

Нетрудно показать, что если игроки обладают критическими точками (22) и (23), лежащими на отрезке $[O_1O_3]$ и выполнено условие $s \geq r$, то в стратегиях, согласованных с критическими уровнями, точка w^* оказывается равновесием Нэша.

Выбор решения игры в виде (25) использует предположение о том, что каждый игрок знает критический уровень противника. Что можно сказать об их поведении, если такое знание отсутствует? Пусть критические точки игроков $w^{(1)}$ и $w^{(2)}$ расположены на отрезке $[O_1O_3]$ (см. рис. 16) и представимы на основе точек O_1 и O_3 в виде

$$w_i^{(1)} = \frac{s}{h}a_i + \frac{h-s}{h}b_i \quad \text{и} \quad w_i^{(2)} = \frac{r}{h}a_i + \frac{h-r}{h}b_i,$$

где $s > r$.

Игрок 1 знает значение s, но не знает величину r. В свою очередь, игроку 2 известна величина r, но он не имеет точного представления об s. Оба игрока знают, что точки $w^{(1)}$ и $w^{(2)}$ лежат на отрезке $[O_1O_3]$. Если игроки заинтересованы в скорейшем достижении согласованного поведения, то единственная возможность для этого состоит в использовании величин s и r. Именно, если траектория игры находится в ситуации O_1 (которая более выгодна игроку 2 по сравнению с ситуацией O_3), то решение о переходе из O_1 в ситуацию O_3 будет принимать игрок 1. В силу того, что он не знает положения точки $w^{(2)}$, единственная информация, которой он может руководствоваться в этом случае – это положение точки $w^{(1)}$. Чтобы не опустить свой выигрыш ниже критического уровня, он должен находиться в ситуации O_1 не более s шагов игры. Заметим, что игрок 1 не может находиться в O_1 меньше, чем s шагов, поскольку в этом случае велика вероятность, что выигрыш игрока 2 опустится ниже его критического уровня (это будет, например, тогда, когда $r = s$). Таким образом, игрок 1, стремясь к согласованному поведению, вынужден находиться в ситуации O_1 s шагов игры. По аналогичным соображениям игрок 2 будет будет находиться в ситуации O_3 в течение

$(h-r)$ шагов игры. В результате, без специального соглашения оба игрока придут к совместной стратегии, дающей решение игры с исходом

$$\tilde{w}^* = \frac{s}{s+h-r}a + \frac{h-r}{s+h-r}b.$$

Сопоставляя $\tilde{w}^*$ с w^* из (25), мы видим. что при $v > s$ знак разности

$$\frac{r}{r+h-s} - \frac{s}{s-h-r} = \frac{(r-s)(h-r-s)}{(h-r+s)(h+r-s)}$$

зависит только от знака $(h-r-s)$.

Если $h > r+s$, то точка w^* на отрезке $[O_1O_3]$ находится ближе к точке O_1, чем точка $\tilde{w}^*$. Значит, решение игры при отсутствии информации о критической точке противника будет выгоднее игроку 2, чем решение (25). Заметим, что условие $h > r+s$ означает, что расстояние критической точки $w^{(1)}$ игрока 1 от точки O_1 больше, чем расстояние критической точки $w^{(2)}$ игрока 2 до точки O_3. Если же $h < r+s$, то отсутствие информации о критической точке будет более выгодно игроку 1.

Эти рассуждения показывают, что обладание дополнительной информацией о критической точке противника будет выгодно одному из игроков. С этой точки зрения для "рациональных игроков" решение с исходом $\tilde{w}^*$ будет неустойчивым, поскольку оно дает возможность игрокам узнать положение обеих критических точек, а значит, один из игроков обязательно перейдет к разыгрыванию решения (25). Отсюда следует, что на начальных шагах игры участники будут стремиться выяснить положение критической точки противника и скрыть положение своей. Этот процесс получения информации порождает у каждого игрока j свою собственную оценку $\tilde{w}^{(i)}$ $(i \neq j)$ критической точки противника i. Если эти оценки удовлетворяют условиям типа (24), а именно, справедливо

$$w_2^{(1)} \geq \tilde{w}_2^{(1)} \geq \tilde{w}_2^{(2)} \geq w_2^{(2)}$$

и

$$w_1^{(2)} \geq \tilde{w}_1^{(2)} \geq \tilde{w}_1^{(1)} \geq w_1^{(1)},$$

то вместо стационарного решения (25) может возникнуть "квазистационарное" решение вида

$$\tilde{w}_i^* = \tilde{\gamma}a_i + (1-\tilde{\gamma})b_i,$$

где $\tilde{\gamma} = \tilde{r}/(\tilde{r}+h-\tilde{s})$. Параметры $\tilde{r}$ и $\tilde{s}$ определяются расположением точек $\tilde{w}^{(1)}$ и $\tilde{w}^{(2)}$ на отрезке $[O_1O_3]$

$$\tilde{w}_i^{(1)} = \frac{\tilde{r}}{h}a_i + \frac{h-\tilde{r}}{h}b_i; \qquad \tilde{w}_i^{(2)} = \frac{\tilde{s}}{h}a_i + \frac{h-\tilde{s}}{h}b_i.$$

В процессе игры при получении дополнительной информации об истинных значениях критических точек игроков "квазистационарное" решение будет стремиться к решению (25). Конкретное протекание этого процесса зависит от ряда индивидуальных особенностей игроков, например, их умения скрывать свои критические точки (способности к блефу), а также от их стремления быстрее или медленнее придти к согласованному поведению.

Заключение

Рассмотрение проблемы возникновения кооперации в IPD характеризуется двумя особенностями. Первое состоит в том, что в отличие от PD, при многократном повторении игры размывается понятие кооперативной ситуации. Если в PD это фактически только ситуация (C, C), то в IPD кооперативное взаимодействие образует в пространстве исходов целую область (см. рис. 1,7). Даже в случае сужения понятия кооперации до эффективного по Парето множество состояний кооперации оказывается весьма обширным. Это обстоятельство несомненно влияет на разрешение вопроса о том, при каких именно условиях и какими программами ведения игры достигается некоторое состояние кооперации.

Вторая особенность обусловлена спецификой многошагового разыгрывания IPD. С точки зрения теории эволюции кооперации IPD не может иметь бесконечное число шагов (такая игра просто не осуществима в реальной жизни), а является существенно конечношаговой. Возникает естественный вопрос о том, как выбирать длину игры, чтобы ее выбор оказывал слабое влияние на процесс достижения кооперации, например, слабо влиял на средний одношаговый выигрыш игрока. Более того, поскольку кроме детерминированных программ в теории рассматриваются также их вероятностные аналоги (например, изучается поведение программ в условиях "шума"), то этот средний выигрыш должен слабо зависеть также от выбора конкретной реализации траектории игры. Оказывается, что программы, использующие для выработки следующего хода информацию об истории игры конечной фиксированной длины преодолевают указанные трудности. В силу того, что выбор хода в таких программах не зависит от момента принятия решения, их взаимодействие друг с другом приводит в общем случае к стационарным распределениям, заданным на пространстве ситуаций

игры. Это означает, что частоты появления всевозможных ситуаций в заданной траектории игры фиксированы и не зависят (почти наверное) от выбора этой траектории.

Используя стационарные стратегии, игроки достаточно быстро попадают в окрестность некоторого фиксированного среднего (на один шаг игры) выигрыша, который и служит естественной оценкой результата игры. Однако, только программы специального вида могут обеспечить эффективность кооперации, когда средние выигрыши партнеров оказываются на границе Парето множества исходов. Удобство стандартной PD (с параметрами, удовлетворяющими условиям (1) и (2)) состоит в том, что таким свойством обладает чистая ситуация (C, C). Для нестандартной PD при многошаговом разыгрывании достижение эффективной кооперации требует появления на траектории игры последовательности, содержащей уже две чистые ситуации: (C, D) и (D, C). Программы, реализующие такие траектории, опираются на новые возможности, предоставляемые неодновременным разыгрыванием PD.

Другой класс программ, позволяющий достичь эффективной кооперации, составляют описанные в разделе 4 программы с заданными критическими уровнями игроков. В результате взаимодействия они находят единственную точку (если она существует) на множестве Парето, средние выигрыши в которой оказываются не менее значений их критических уровней. Обобщения этого подхода на случай "рациональных" игроков приводит к тому, что достигаемый ими средний выигрыш зависит не только от значений их собственных критических уровней, но и от их представлений о значении критического уровня партнера. Эта возможность приводит к усложнению процесса достижения состояния кооперации, вводя в игру ряд предварительных игровых этапов, связанных с изучением положения критического уровня партнера или использования ходов, направленных на создание ложного представления о реальном расположении своего критического уровня. Эти действия позволяют подготовленному игроку достигнуть более выгодной точки на множестве Парето и тем самым повысить свой средний выигрыш.

Литература

1. Axelrod R. The Evolution of Cooperation. Basic Books,New York, 1984.
2. Льюис Р.В., Х. Райфа. Игры и решения. М: ИЛ, 1961.
3. Holland, J. (1975). Adaptation in Natural and Artificial Systems. Ann Arbor: Univ. of Michgan Press.
4. Axelrod R. The Evolution of Strategies in the Iterated Prisoner's Dilemma, in Lawrence Davis (ed.) "Genetic Algorithms and Simulated Annealing", (London: Pitman, and Los Alton, CA: Morgan Kaufman, 1987), pp. 32-41.
5. Левченков В.С. (2004) Стационарные стратегии в супериграх. ДАН, т. 397, N2, с.181-185.
6. Gibbons, R. A Primer in Game Theory. Harvester Wheatsheaf, 1992.
7. Leimar, O. (1997). Repeated Games: A State Space Approach. J. theor. Biol. 184, p.471-498.
8. Trivers, R. (1985) Social Evolution. Menlo Park, CA: Benjamin Cummings).
9. Willinson, G.S. (1984) Reciprocal Food-Sharing in the Vampire Bat. Nature, Lond. 308, 181-184.
10. Packer, C. (1977) Reciprocal Alttruism in Papio Annubis. Nature, Lond. 265, 441-443.
11. Nowak, M.A. and K. Sigmund (1994) The Alternating Prisoner's Dilemma. J. Theor. Biol. 168, p. 219-226.
12. Hanert C.H. and H.G. Schuster (1998). Extending the Iterated Prisoner's Dilemma without Synchrony. J. Theor. Biol. 193, P. 155-166.
13. Wu J., and Axelrod R. How to Cope with Noise in the Iterarated Prisoner's Dilemma. Journal of Conflict Resolution, V. 39, N1, 1995, p. 183-189.
14. Sugden, R. (1986). The Economics of Rights, Co-operation and Welfare, Oxford: Basil Blackwell.

15. Boerlijst, M.C., M.A. Nowak and K. Sigmund. (1996). The Logic of Contrition. J. theor. Biol. 185, p.281-293.

16. Molander, P. (1985) The Optimal Level of Generosity in a Selfish, Uncertain Environment. Journal of Conflict Resolution, V. 29, pp. 611-618.

17. Riolo, R.L., M.D. Cohen&R. Axelrod. Evolution of Cooperation without Reciprocity. Nature,v.414, p.441-443 (2001).

18. Левченков В.С. (2005) Выбор решения в биматричных супериграх. ДАН, т. 403, N1.