

Современные подходы к экономическому моделированию чувства справедливости

Автономов Ю.В.

В настоящий момент исследователи, работающие в области экспериментальной экономики, располагают обширными данными о том, что поведение людей в некоторых лабораторных экспериментах систематически расходится с опорной для микроэкономического анализа гипотезой рационального преследования собственного интереса, но может быть объяснено тем, что для людей имеют значение соображения справедливости или реципрокности. Эта статья посвящена обзору теорий и моделей, разработанных для объяснения и прогнозирования такого поведения, и содержит анализ преимуществ и недостатков различных существующих подходов.

1. Введение

Традиционный экономический анализ предполагает, что люди решают свои экономические задачи, рационально преследуя свой узко понимаемый интерес, и, в частности, поступают честно лишь постольку, поскольку это поведение таковому интересу соответствует. Однако в последние десятилетия XX в. в поле зрения исследователей стали попадать факты, позволяющие предположить, что в экономической деятельности, проходящей вне конкурентных рынков со стандартизованными товарами, – когда количество участников невелико, а информация несовершенна, в частности, внутри фирм и организаций, в условиях неполных контрактов – соображения честности и справедливости имеют самостоятельное влияние на экономическое поведение людей.

Большая часть таких свидетельств была получена в последние два десятилетия в рамках экспериментальной экономики. В ходе различных лабораторных экспериментов было обнаружено, что значительная доля участников, в ущерб собственному интересу, придает значение собственным и чужим обещаниям, проявляет заботу об интересах партнеров, поощряет реципрокное и стремится наказывать оппортунистическое поведение.

Для объяснения этих экспериментальных результатов был разработан ряд теоретических моделей, позволивших также по-новому взглянуть на такие традиционные экономические проблемы, как существование неполных контрактов и ставок заработной платы, устойчиво превышающих рыночные. Эти модели вызвали новую волну экспериментов, призванных проверить справедливость соответствующих гипотез.

Автономов Ю.В. – аспирант ГУ ВШЭ.

Работа выполнена при поддержке Научного фонда ГУ ВШЭ.

Статья поступила в Редакцию в июле 2006.

В данной работе предлагается краткий обзор последних теоретических разработок в области моделирования чувства справедливости, обсуждаются преимущества и недостатки различных подходов.

2. Основные результаты экспериментов

Перед тем, как перейти к изложению экспериментальных результатов, уточним, что в этой работе понятие «справедливость» (fairness) будет использоваться для характеристики решений индивида о распределении ресурсов или обязанностей между ним и другими людьми (в такой трактовке ее иногда называют «распределительной справедливостью» (distributive fairness), в противоположность «процедурной справедливости» (procedural fairness) – понятию, характеризующему процедуры и схемы принятия (коллективных) решений). Понятие «справедливость» также можно употреблять для характеристики поведения людей при выборе ответа на доброжелательные или недоброжелательные действия партнера, однако более корректно по отношению к таким ситуациям говорить о наличии или отсутствии реципрокности.

На текущий момент исследователи располагают обширным набором свидетельств того, что в некоторых ситуациях человеческое поведение расходится с гипотезой о строгом следовании собственному интересу, однако соответствует соображениям честности, справедливости или реципрокности. Тезис о влиянии этих соображений на поведение экономических агентов начал завоевывать позиции в экспериментальной экономике в 1980-х гг. благодаря результатам изучения двусторонних переговоров и взаимодействия людей в небольших группах [33, 47].

К основным играм, результаты которых свидетельствуют о том, что на поведение участников влияют соображения справедливости и честности, можно отнести игры «Ультиматум», «Дарообмен», «Доверие», «Диктатор», а также игры с добровольным финансированием общественных благ.

Вероятно, наиболее известной из перечисленных является игра «Ультиматум», в которой игрок А предлагает игроку Б пропорцию, в которой между ними будет разделена некоторая сумма денег, а Б либо соглашается, либо не соглашается (тогда оба не получают ничего). Результаты многочисленных экспериментов показывают, что с вероятностью от 0,4 до 0,6 игроки Б, в ущерб собственному материальному интересу, отвергают предложения меньше 20% общей суммы, и чем больше предложенная доля, тем меньше вероятность того, что предложение будет отвергнуто [9, 33]. Отказ принимать маленькие, но все равно выгодные предложения может быть истолкован как то, что игроки считают такие предложения «несправедливыми». В одном из экспериментов было показано, что модальная доля, предлагаемая А партнеру, примерно максимизирует его собственный ожидаемый доход – т.е. игроки А прогнозируют не соответствующую предпосылке рационального эгоизма реакцию партнеров [48].

Ненулевые предложения отмечаются и в игре «Диктатор», где Б не может отказаться от предложения А, хотя предлагаемые партнеру суммы в ней меньше, чем в игре «Ультиматум» [30].

Игра «Дарообмен» является экспериментальным переложением классических взаимоотношений принципала и агента. Принципал предлагает агенту некую сумму $w \in [\underline{w}, \bar{w}]$, которую тот может принять или не принять. Если агент отвергает платеж, оба не получают ничего. Если агент принимает платеж, ему предла-

гают выбрать «уровень усилий» $e \in [\underline{e}, \bar{e}]$, $\underline{e} > 0$. Выигрыши принципала и агента равны, соответственно $x^P = ve - w$ и $x^A = w - c(e)$, где v – предельная доходность усилия для принципала, а $c(e)$ – выпуклая возрастающая функция издержек агента. При стандартных предположениях агент будет соглашаться на любую w и выбирать самый низкий уровень усилий \underline{e} . Принципал, следовательно, будет выбирать $w = \underline{w}$.

Экспериментальные результаты свидетельствуют о поляризации группы агентов: для примерно 40–50% агентов средний размер усилий коррелирует с w , остальные преследуют свой собственный интерес – либо последовательно, либо со случайными отклонениями [13, 23, 27, 28, 31]. В результате зависимость среднего усилия от w достаточно сильна, чтобы принципалам было выгодно предлагать высокие w [27].

В игре «Доверие» игрок А получает от экспериментатора сумму y , и может отдать часть ее, z , игроку Б, причем экспериментатор утраивает этот трансферт. В ответ игрок Б может вернуть А некоторую часть полученных денег. В экспериментах многие игроки А передавали Б ненулевое z , и возвращаемая им в ответ сумма коррелировала с величиной z [15, 19].

Финансирование общественных благ является очень популярным предметом исследований в экспериментальной экономике, и анализу этих игр посвящено множество статей (за обзором можно обратиться, например, к работам Ледьярда [40] или Доуса и Талера [16]). Типичная игра с добровольным финансированием общественного блага разыгрывается между n участниками, одновременно решающими, какую долю своего первоначального запаса пустить на финансирование общественного блага. Выигрыш игрока i равен $x_i = y_i - g_i + m \sum g_j$, где y – первоначальный запас, g – вклад в финансирование общественного блага, а m – денежный эквивалент единицы предоставляемого общественного блага, $m < 1 < nm$. Доминирующей стратегией было бы не финансировать общественное благо вообще, хотя общий выигрыш максимизируется, если каждый участник полностью отдает свой первоначальный запас на финансирование общественного блага. В большинстве экспериментов игра ведется 10 периодов, в каждом из которых состав группы случайным образом меняется. Если ограничить внимание поведением в последнем периоде игры (абстрагируясь от повторяющихся игр и обучения в процессе игры), оказывается, что 75% игроков не финансируют общественное благо, а оставшиеся дают очень мало. Однако, если в игру вводится возможность наказания, ситуация меняется драматически [26]. При возможности наказания, после каждого раунда игры игрок, зная размер вклада каждого из партнеров в общественное благо, может наказать любого партнера на сумму до 10 очков. Каждое очко наказания снижает доход наказываемого на 3 единицы, но связано для наказывающего с издержками, описываемыми возрастающей выпуклой функцией. Таким образом, игрок-эгоист воздержался бы от наказания, и все игроки, зная это, не вкладывали бы ничего в финансирование общественного блага. Однако экспериментальные результаты показывают, что к последнему периоду все участники группы отдают примерно 75% своих первоначальных запасов на финансирование общественного блага.

При том, что перечисленные закономерности поведения участников в играх «Ультиматум», «Дарообмен», «Доверие» и играх с добровольным финансированием общественных благ были подтверждены многочисленными экспериментами, их также можно объяснить и не прибегая к соображениям честности или справедливости у игроков. Можно, например, предположить, что ставки в игре недостаточно высоки. Однако при попытках проверить эту гипотезу, увеличив ставки вплоть до размеров среднего дохода игроков за несколько месяцев, поведение игроков менялось незначительно [10, 29, 34, 53,].

Еще одним альтернативным объяснением описанных выше закономерностей могла бы быть ограниченная рациональность агентов. В частности, Рот и Эрв [46], а также Бинмор, Гейл и Самуэльсон [3] предполагают, что предложение ненулевых долей и отказ принимать низкие доли в игре «Ультиматум» есть следствие постепенной адаптации игроков: игроки А, теряя существенные деньги в случае отказа Б, быстро учатся делать «справедливые» предложения, а Б, отвергая «несправедливо маленькие» доли, теряют гораздо меньше, и потому медленнее приходят к пониманию, что даже такие предложения выгодны. В то время как такая гипотеза представляется достаточно правдоподобной, анализ экспериментальных результатов показывает, что игроки А предвидят действия Б гораздо чаще и точнее, чем это предсказывалось бы чисто адаптивными ожиданиями [48].

Отклонение от рационального преследования собственного интереса также может объясняться социальными нормами. Поскольку в повседневной жизни большинство ситуаций выбора для людей являются повторяющимися, а при повторяющихся взаимодействиях вполне возможно реализовать в качестве равновесия и щедрость игроков А, и несговорчивость игроков Б – это фиксируется в социальных нормах, которые участники игр ошибочно (с точки зрения экономической рациональности) применяют к неповторяющимся лабораторным ситуациям [2]. У этого объяснения также есть недостатки. Во-первых, значительные различия в поведении людей в достаточно похожих играх заставляют предположить, что социальные нормы слишком уж многообразны. Во-вторых, участники экспериментов демонстрируют четкое различие в поведении в однопериодных и многопериодных играх, т.е. определенно различают эти типы взаимодействия.

Наиболее перспективной гипотезой для объяснения описанных выше экспериментальных закономерностей в итоге представляется полная рациональность в сочетании с элементами неэгоистической мотивации.

Интересные косвенные подтверждения рационального неэгоистического поведения и веры в неэгоистическое поведение партнеров были получены на стыке нейробиологии и экспериментальной экономики в ходе многопериодной игры «Социальная дилемма» [39]. В этой разновидности игры «Доверие» игроки располагают равными начальными запасами. Начинаящий игру А может оставить начальный запас себе или отдать партнеру, в случае чего передаваемая сумма удваивается экспериментатором. Игрок Б, наблюдая действия А, также может оставить начальный запас себе или отдать А. Таким образом, в игре моделируется ситуация двустороннего обмена в отсутствие механизмов принуждения к исполнению контракта.

Авторы показали, что введение участникам гормона окситоцина, подавляющего инстинкт избегания себе подобных и ингибирующего защитные реакции у животных [35], увеличивает долю максимально доверяющих партнеру игроков А до 45%, по сравнению с 21% в контрольной группе – однако не сказывается на

поведении игроков Б. Влияние окситоцина сокращается до 20% после эконометрического устранения эффектов его воздействия на спокойствие, бдительность и настроение испытуемых.

3. Подходы к моделированию индивидуального чувства справедливости

Среди существующих подходов к моделированию индивидуального чувства справедливости выделяются два основных направления. Представители первого предполагают наличие у индивида так называемых «социальных предпочтений» (social preferences), т.е. предпочтений, определенных на множестве распределений ресурсов между индивидами, входящими в сообщество. Вторым направлением является использование так называемой «реципрокности, основанной на намерениях» (intention-based reciprocity). Для агента соответствующих моделей в предпочтения некоторым образом включаются его субъективные ожидания относительно намерений партнера, т.е. практически одно и то же предложение партнера имеет для агента разную ценность в зависимости от того, как он воспринимает множество стратегий партнера.

3.1. «Социальные предпочтения»

«Социальные предпочтения» имеют место, если полезность индивида меняется при изменении распределения ресурсов между другими людьми. При этом, если функция полезности индивида возрастает по доходу других, «социальные предпочтения» альтруистичны.

Чарнесс и Рэбин [14] рассматривают специфическую форму альтруизма, которую они называют «квазимаксиминимальными предпочтениями». Индивидуальные предпочтения представляют собой выпуклую комбинацию его собственного дохода и так называемой «беспристрастной функции общественного благосостояния» – выпуклой комбинации утилитаристской функции благосостояния и роулианского максиминимального критерия:

$$U_i(x_1, x_2, \dots, x_N) = (1 - \gamma)x_i + \gamma(\delta \min\{x_1, x_2, \dots, x_N\} + (1 - \delta)(x_1 + x_2 + \dots + x_N)).$$

Наделенный такой функцией полезности индивид заботится о благосостоянии сообщества в целом, а также благосостоянии своего наиболее бедного собрата. Однако в остальном он совершенно невосприимчив к неравенству в распределении доходов среди членов группы.

Для случаев, когда игроков только два, эта функция полезности приобретает вид

$$U_i(x_1, x_2) = \begin{cases} x_i + \gamma(1 - \delta)x_j & \text{если } x_i < x_j \\ (1 - \gamma\delta)x_i + \gamma x_j & \text{если } x_i \geq x_j. \end{cases}$$

В этой форме важной ее особенностью становится то, что индивид меньше беспокоится о благосостоянии партнера, если его доход превышает собственный.

Альтруизм вообще и в квазимаксиминимальных предпочтениях в частности объясняет отличные от минимальных трансферты в игре «Диктатор», ненулевые

ответные трансферты в игре «Дарообмен», а также добровольное финансирование общественных благ, но не объясняет желание людей отомстить за проявленную другими «несправедливость» (игры «Ультиматум» и игры с добровольным финансированием общественных благ с возможностью санкций). Поэтому в последующих работах Чарнесс и Рэбин дополняют квазимаксиминимальные предпочтения реципрокностью.

Болтон [6], объясняя результаты игры с делением доллара между двумя игроками, предположил, что для индивида имеет значение не только абсолютное количество денег, которое получает он лично, но и то, как это количество соотносится с тем, что получают другие, т.е. его место в «иерархии благосостояния»¹⁾. Он наделил каждого функцией полезности вида

$$U_i(x_i, x_j) = u_i(x_i, x_i/x_j),$$

производная которой по соотношению доходов была отрицательной при $x_j > x_i$ и нулевой при $x_j \leq x_i$. То есть индивид завидует благосостоянию партнера, если получает меньше его, но безразличен к положению партнера, если получает больше. Такая формализация объясняла поведение участников в рассматриваемых Болтоном играх, но, разумеется, она противоположна по духу любым альтруистическим проявлениям, а следовательно, не способна объяснить многие из экспериментальных результатов, описанных в разделе 2.

Несколько авторов объединили мотивы зависти и альтруизма, предложив модели, наделяющие индивида «неприятностью к неравенству», т.е. альтруизмом по отношению к тем партнерам, чье благосостояние ниже «порога справедливости», и завистью к тем, чье благосостояние выше этого порога. Для большинства экспериментальных игр естественно предполагать, что этот «порог справедливости» соответствует размеру платежа одному игроку, если все получают равную сумму.

Фер и Шмидт [24] формализуют это при помощи простой функции полезности вида

$$U_i(x_1, \dots, x_N) = x_i - \frac{\alpha_i}{N-1} \sum_{j \neq i} \max\{x_j - x_i, 0\} - \frac{\beta_i}{N-1} \sum_{j \neq i} \max\{x_i - x_j, 0\}$$

с $\beta_i \leq \alpha_i$ и $\beta_i \leq 1$. Заметим, что индивид альтруистичен только по отношению к более бедным партнерам: $U_i/\partial x_j \geq 0$ тогда и только тогда, когда $x_i \geq x_j$, и его интерес к благосостоянию партнеров асимметричен: неравенство сильнее беспокоит индивида в случае когда благосостояние партнера выше собственного, чем если оно ниже собственного ($\alpha_i \geq \beta_i$).

С такой функцией полезности возможно объяснить как альтруистические действия по отношению к партнерам, так и стремление наказывать их, для всех рассматриваемых нами игр. Но если бы все индивиды имели идентичные предпочтения, было бы невозможно объяснить, почему в некоторых случаях альтруизм, сотрудничество, стремление наказывать несправедливость, не являясь лучшей эгоистической стратегией, наблюдаются в существенном объеме, а в других едва ли имеют место. Фер и Шмидт показывают, что сочетание определенных распределенных типов со стратегическим окружением может объяснить как неравномерные,

¹⁾ История этого подхода в экономической науке восходит по меньшей мере к теории демонстративного потребления Веблена.

так и очень эгалитаристские исходы. Например, в игре «Ультиматум» с конкуренцией активных игроков даже популяция, состоящая из очень честных индивидов (с высокими значениями α и β) не сможет предотвратить очень неравномерных исходов, поскольку ни один из игроков не может обеспечить более справедливый исход за счет собственных действий. Наоборот, в игре с финансированием общественного блага при возможности наказаний небольшой доли честных игроков достаточно, чтобы при правдоподобной угрозе наказания игроки-эгоисты участвовали в финансировании общественного блага.

Похожую модель «неприязни к неравенству» разработали независимо Болтон и Окенфельс [5]. Они предлагают следующую функцию полезности:

$$U_i = U_i(x_i, \sigma_i),$$

где σ_i отражает относительный платеж, или долю индивида в общем объеме платежей, и определяется как

$$\sigma_i = \begin{cases} \frac{x_i}{\sum_{j=1}^N x_j} & \text{если } \sum_{j=1}^N x_j \neq 0 \\ \frac{1}{N} & \text{если } \sum_{j=1}^N x_j = 0. \end{cases}$$

При заданном σ_i функция полезности возрастает и вогнута по доходу игрока x_i . При заданном x_i функция полезности строго вогнута по доле игрока в общем доходе σ_i и достигает максимума при $\sigma_i = 1/N$. Таким образом, при любом размере собственного платежа игрок предпочел бы, чтобы этот платеж совпал со средним значением по совокупности участников игры. Этот подход также объясняет ряд считающихся парадоксальными результатов, как-то трансферты в игре «Диктатор» и обратные трансферты в игре «Дарообмен», а также отказ от предложения активной стороны в игре «Ультиматум». Болтон и Окенфельс не предлагают конкретной формы для своей функции полезности.

В случаях с двумя игроками две последних описанных модели дают качественно похожие результаты, однако в случаях с многими игроками возникают интересные различия. Так, Фер и Шмидт предполагают, что в этой ситуации индивид сравнивает себя с каждым партнером по отдельности, а для модели Болтона – Окенфельса имеет значение лишь средний уровень дохода. Поэтому у моделируемого ими индивида в некоторых условиях может возникнуть неожиданная щедрость по отношению к игроку, намного более благополучному, чем он, и зависть к игроку, который беднее его самого.

Левайн [41] объясняет трансферты в одних играх и желание наказать несправедливость иным путем, замечая, что индивиду в общем случае безразлично, с каким партнером он имеет дело. Левайн рассматривает функцию полезности вида

$$U_i = x_i + \sum_{j \neq i} \frac{x_j (a_i + \lambda a_j)}{1 + \lambda},$$

где $0 \leq \lambda \leq 1$, $-1 < a_i < 1$ для всех $i \in \{1, \dots, N\}$. Если предположить, что $\lambda = 0$, эта функция превращается в $U_i = x_i + a_i \sum_{j \neq i} x_j$, и при $a_i > 0$, человек стремится увели-

чить благосостояние других, а при $a_i < 0$ – недоброжелателен к окружающим. Но как объяснить альтруизм в одних обстоятельствах и недоброжелательность в других? Предположим, что $\lambda > 0$. В этом случае чем более альтруистичен контрагент, тем больше индивид заинтересован в его благосостоянии. Если $-\lambda a_j > a_i$, индивид будет недоброжелателен к контрагенту. Поскольку в большинстве игр параметры a_j контрагентов ненаблюдаемы для игрока, любая последовательная игра оказывается связанной с сигнализированием.

Откалибровав распределение a на данных игры «Ультиматум», Левайн оценил предположительно единое для всех игроков λ и показал, что с найденными параметрами модель хорошо ложится на экспериментальные данные по играм «сороконожка», играм с финансированием общественных благ и поведению участников экспериментальных рынков. Однако, поскольку $a_i < 1$, модель не объясняет ненулевых трансфертов в игре «Диктатор».

3.2. Реципрокность, основанная на намерениях или типе

Авторы моделей, объясняющих проявления чувства справедливости у людей с помощью «реципрокности, основанной на намерениях», предполагают, что агент не просто оценивает действия партнера, но и пытается их интерпретировать – если он сочтет, что с ним поступили хорошо, он захочет ответить партнеру тем же, и наоборот. Для анализа взаимодействия игроков при этом используется психологическая теория игр²⁾. Заметим, что одновременно существует постоянно растущее число публикаций, объясняющих существование реципрокности эволюционными соображениями [8, 51, 52]. Подобные работы здесь не рассматриваются.

Одним из первых исследователей, предложивших описываемую схему объяснения, был Рэбин, выдвинувший формальное определение реципрокности, основанной на намерениях [45]. Он ограничивал свое внимание играми для двух игроков в нормальной форме.

Обозначим через A_1 и A_2 наборы смешанных стратегий игроков 1 и 2, и обозначим платежную функцию игрока i через x_i : $A_1 \times A_2 \rightarrow \mathbb{R}$. Далее определим убеждения на множестве стратегий. Пусть $a_i \in A_i$ обозначает стратегию игрока i , и, выбирая ее, он должен иметь некое убеждение о стратегии, которую выберет j . Будем предполагать, что $i \in \{1, 2\}$ и $j = 3 - i$. Обозначим через b_j убеждение индивида i относительно того, что сделает j . Чтобы убеждение b_j было рациональным, у индивида i должно быть и убеждение относительно того, что, по мнению j , примет он сам. Обозначим это убеждение о намерениях контрагента как c_i . Такая иерархия убеждений могла бы продолжаться далеко, но нам достаточно двух уровней.

Рэбин вводит «функцию доброты», $f_i(a_i, b_j)$, определяющую, насколько хорошо игрок i относится к игроку j . Если i верит, что j выберет стратегию b_j , он выбирает платеж j из множества $[x_j^l(b_j), x_j^h(b_j)]$, где $x_j^h(b_j)$ и $x_j^l(b_j)$ обозначают наибольший (наименьший) платеж игрока j , которого может добиться i при выбо-

²⁾ Популярный термин, введенный Геанакопелосом, Пирсом и Стакетти [32] для обозначения игр, в которых платежи игроков меняются в зависимости от их убеждений, в том числе убеждений о намерениях друг друга.

ре оппонентом стратегии b_j . «Справедливым» Рэбин называет платеж $x_j^f(b_j)$, являющийся средним арифметическим низшего и высшего возможных (впрочем, за исключением Парето-доминируемых платежей). Заметим, что $x_j^f(b_j)$ не зависит от того, что получает сам i . Доброта, испытываемая i к j , измеряется отклонением фактического платежа, который i дает j , от «справедливого», деленным на диапазон возможных платежей:

$$f_i(a_i, b_j) \equiv \frac{x_j(b_j, a_i) - x_j^f(b_j)}{x_j^h(b_j) - x_j^l(b_j)}.$$

При этом, если $x_j^h(b_j) - x_j^l(b_j) = 0$, то $f_i(a_i, b_j) = 0$. Заметим, что $f_i(a_i, b_j) > 0$ тогда и только тогда, когда i дает j больше, чем «справедливый» платеж.

Наконец, Рэбин определяет убеждения игрока i о том, насколько честно к нему относится игрок j . Если i убежден, что j выбирает b_j и думает, что i выберет c_i , доброта игрока j в глазах i определяется функцией:

$$f_j'(b_j, c_i) \equiv \frac{x_i(c_i, b_j) - x_i^f(c_i)}{x_i^h(c_i) - x_i^l(c_i)},$$

где $j = 3 - i$ и $f_j(b_j, c_i) = 0$, если $x_i^h(c_i) - x_i^l(c_i) = 0$. Обе функции доброты используются в функции полезности

$$U_i(a, b_j, c_i) = x_i(a, b_j) + f_j'(b_j, c_i)[1 + f_i(a_i, b_j)],$$

где $a = (a_1, a_2)$.

Если i считает, что j недоброжелателен к нему, i сам постарается быть недружелюбнее, и наоборот. Заметим также, что значения функций доброты должны принадлежать $[-1, 0,5]$, и функции полезности оказываются чувствительны к положительным аффинным преобразованиям. Значимость доброты также убывает с ростом платежей. «Честным равновесием» в такой игре называется пара стратегий (a_1, a_2) , являющихся наилучшими ответами друг на друга, и набор рациональных убеждений $b = (b_1, b_2)$ и $c = (c_1, c_2)$.

Предложенная Рэбином модель стала первым формальным определением и исследованием последствий реципрокности, основанной на намерениях. К сожалению, она была слабо приспособлена для прогнозирования, имея много равновесий, часто совершенно противоположных (например, равновесные исходы «взаимный альтруизм» и «взаимный эгоизм» в одной и той же игре) и каждое подкреплялось самосбывающимися пророчествами, а также ряд других странностей. Теория Рэбина была определена для игр с двумя игроками в нормальной форме, и при столкновении с нормальными формами последовательных игр давала парадоксальные результаты. Например, в последовательной дилемме заключенного безусловная кооперация второго игрока является частью равновесия, поскольку предложенное Рэбином определение равновесия не обязывает второго игрока вести себя оптимально на неиграемых ветвях игры.

Эти недостатки способствовали появлению ряда обобщений модели Рэбина. Дюфвенберг и Кирхштайгер [17] расширили ее до игр с N игроками в расширен-

ной форме, введя понятие последовательного реципрокного равновесия (Sequential Reciprocity Equilibrium, SRE). Основной идеей было отслеживать убеждения по ходу игры и определить, как должны формироваться убеждения на неиспользуемых ветвях игры. При этом стратегии игроков должны были составлять честное равновесие в каждой подыгре. Предложенная модель сняла часть парадоксов, но осталась очень громоздкой для анализа и по-прежнему давала множественные равновесия, подкрепленные самосбывающимися убеждениями. Например, предложение активной стороной в игре «Ультиматум» такого распределения, которое будет заведомо отвергнуто, успешно объяснялось существованием у игроков убеждения, что оппонент хочет навредить им.

Авторы еще одного обобщения модели Рэбина Фальк и Фишбахер [20] объединили заинтересованность индивида намерениями контрагента с наличием у него социальных предпочтений, рассматривая игры с N участниками и неполной информацией в экстенсивной форме. «Доброта» в их модели измеряется в тех же терминах, что и неприязнь к неравенству: j считает стратегию i доброжелательной, если в результате ее реализации платеж j превышает платеж i . Это в корне отличает модель Фалька и Фишбахера от других моделей реципрокности, описывавшихся выше. Более того, Фальк и Фишбахер оценивают, способен ли контрагент повлиять на неравное распределение и изменить его или нет. В последнем случае вес компоненты «доброжелательности» при принятии решения падает. Но даже если контрагент не способен изменить исход игры, вес компоненты «доброжелательности» положителен. Таким образом, реципрокность, основанная на намерениях, превращается в «неприязнь к неравенству». Модель Фалька и Фишбахера, для некоторых диапазонов параметров, объясняет экспериментальное поведение во всех играх, перечисленных в разделе 2. Платой за эту гибкость является то, что модель довольно сложна. В каждом узле игры, чтобы принять решение, индивид i должен оценить доброжелательность контрагента, зависящую от ожидаемой разницы в выигрыше между ними и от того, что контрагент мог или не мог бы сделать с этой разницей. Компонент доброжелательности домножается на компонент взаимности, положительный – если игрок i доброжелателен к этому контрагенту, и отрицательный – в ином случае. Это произведение далее домножается на параметр, определяющий значимость для индивида i соображений взаимности по отношению к стремлению увеличить собственный доход.

Еще одна попытка комбинировать социальные предпочтения с реципрокностью принадлежит Чарнессу и Рэбину [11]. Расширяя предложенную ими ранее модель, которая описывалась выше, они дополняют предпочтения «профилем недостатков», $\rho \equiv (\rho_1, \dots, \rho_N)$, где $\rho_i \in [0, 1]$ измеряет то, чего заслуживает игрок i по мнению всех остальных. Чем меньше ρ_i , тем меньший вес игрок i имеет в функциях полезности других игроков. При заданном профиле недостатков функция полезности игрока имеет вид

$$U_i(x_1, x_2, \dots, x_N / \rho) = (1 - \gamma)x_i + \gamma \left[\delta \min \{x_i, \min_{j \neq i} \{x_j + d\rho_j\}\} + (1 - \delta) \left(x_i + \sum_{j \neq i} \max \{1 - k\rho_j, 0\} x_j \right) - f \sum_{j \neq i} \rho_j x_j \right],$$

где $d, k, f \geq 0$ – три новых параметра. Если $d = k = f = 0$, предпочтения сводятся к квази-максиминимальным, описанным ранее. При больших d и k игрок i не захо-

чет способствовать благосостоянию игрока j , а при большом f – даже захочет навредить ему. Ключевой шаг – эндогенизация профиля недостатков ρ , в модели осуществляется сравнением стратегии игрока j с экзогенно заданным объективным стандартом. Чем сильнее расхождение, тем выше параметр недостойности ρ_j .

Реципрокно справедливым равновесием (reciprocal fairness equilibrium, RFE) называется профиль стратегий и профиль недостатков, такие, что каждая стратегия является наилучшим ответом на другие при заданном профиле недостатков, и профиль недостатков совместим с равновесными стратегиями.

Концепция RFE обладает рядом недостатков, сильно ограничивающих ее практическое применение. Так, предпочтения заданы только в равновесии (т.е. для равновесного профиля ρ), и как оценивать множественные равновесия или неравновесные состояния, неясно. Далее, чтобы определить профиль ρ , игроки должны иметь идентичные функции полезности и признавать одну и ту же квазимаксимимальную функцию общественного благосостояния. Наконец, сложность и обилие свободных параметров затрудняет эмпирическую проверку модели.

3.3. Аксиоматические подходы

Описанные в предыдущих разделах функции полезности могут быть правдоподобны с точки зрения человеческой психологии, но не обязаны сводиться к классическому набору аксиом теории полезности. В этом свете естественно появление попыток сформулировать аксиомы, достаточные для существования предпочтений, отражающих соображения справедливости или реципрокности.

Сигал и Собель [50] рассматривают взаимодействие двух игроков, которые играют в некоторую игру с известным пространством исходов X_i , множеством стратегий s_i и множеством смешанных стратегий Σ_i , где i – номер игрока X_i . Платежная функция O переводит комбинации смешанных стратегий $\Sigma_1 \times \Sigma_2$ в пространство лотерей на множестве исходов, $\Delta(X_1) \times \Delta(X_2)$. Каждый игрок наделяется двумя наборами предпочтений. Это, во-первых, набор «эгоистических» предпочтений \succ_i^{sel} на $\Delta(X_i)$, множестве лотерей на X_i , который удовлетворяет аксиомам фон Неймана – Моргенштерна. И во-вторых, это полный и транзитивный набор предпочтений \succ_{i,σ_j} на Σ_i , множестве собственных смешанных стратегий. Заметим, что оценка игроком собственных стратегий \succ_{i,σ_j} зависит от стратегии, выбранной партнером, σ_j .

Сигал и Собель показывают, что если отношения предпочтения \succ_{i,σ_j} удовлетворяют аксиомам непрерывности и независимости и если для заданного σ_j игрок, при фиксированном платеже партнера, может иметь больший платеж для себя (предпосылка о собственном интересе), предпочтения \succ_{i,σ_j} на Σ_i могут быть представлены функцией полезности вида

$$u_i(\sigma_i, \sigma_j) = v_i(\sigma_i, \sigma_j) + a_{i,\sigma_j} v_j(\sigma_i, \sigma_j).$$

Для стандартного *homo oeconomicus*³⁾ $a_{i,ij} = 0$. Если $a_{i,ij}$ положителен, игрок является альтруистом, если отрицателен – игрок недоброжелателен. При этом коэффициент $a_{i,ij}$ зависит от σ_j . То есть различные стратегии оппонента могут вызвать у игрока альтруизм или недоброжелательность. Таким образом, в предпочтения может быть интегрирована реципрокность. Чтобы зафиксировать это в своем наборе аксиом, Сигал и Собель вводят дополнительную аксиому «реципрокного альтруизма», которая заставляет игроков при прочих равных, поощрять оппонента, если он выбирает «доброжелательные» стратегии, и наказывать его в обратном случае. Технически эта аксиома означает, что при прочих равных коэффициент $a_{i,ij}$ тем больше, чем лучший исход для индивида i обеспечивает стратегия его оппонента σ_j .

Более узкий частный случай представляет собой работа Нилсона [43], которая дает аксиоматическое обоснование модели неприязни к неравенству Фера и Шмидта [24]. Сама статья посвящена аксиоматическому моделированию предпочтений, зависящих от точки отсчета, как, например, предложенных Канеманом и Тверски [36] в теории ожиданий.

Нилсон отталкивается от примера игры с тремя участниками, в которой первый игрок, из любви к равенству, предпочитает распределение (60, 60, 60) распределению (60, 80, 40). Стандартная аксиома аддитивной сепарабельности гласит, что при сравнении двух наборов имеют значение только те компоненты, количество которых неодинаково. В нашем примере это означает, что $(x_1, 60, 60) > (x_1, 80, 40)$ для любых x_1 . Но если $x_1 = 50$, индивид может проявить нежелание перераспределять ресурсы до распределения (50, 60, 60) и тем самым гарантировать себе самый низкий доход из всех участников. Он может не захотеть изменить распределение ресурсов между другими участниками игры, если это изменит его собственный ранг в распределении ресурсов. Предложенная Нилсоном аксиома сепарабельности по внутренней точке отсчета (self-reference separability, SRS), в отличие от стандартной аддитивной сепарабельности, учитывает озабоченность индивида своим рангом в распределении.

Аксиома SRS в сочетании с обычными аксиомами полноты, транзитивности и непрерывности обеспечивает существование функции полезности вида
$$U(x) = u(x_0) + \sum_{i=1}^n u_i(x_i - x_0),$$
 аддитивно сепарабельной по самой референтной переменной (здесь – x_0) и разнице между ней и остальными переменными ($x_i - x_0$). Функции u_i при этом уникальны до совместного положительного линейного преобразования, т.е. при замене u_0, \dots, u_n на v_0, \dots, v_n полученные из u_0, \dots, u_n преобразованием $v_i = a + bu_i$, где a – скалярная величина, а b – строго положительная скалярная величина, предпочтения не изменятся.

³⁾ Под стандартной моделью *homo oeconomicus* здесь понимается модель рационального максимизатора в том виде, в котором она описывается в учебниках по микроэкономике, без беккеровских или ланкастеровских доработок.

Нетрудно заметить, что частным случаем функции $U(x) = u(x_0) + \sum_{i=1}^n u_i(x_i - x_0)$, является функция $U_i(x) = x_0 - \frac{\alpha}{n} \sum_{i=1}^n \max(x_i - x_0, 0) - \frac{\beta}{n} \sum_{i=1}^n \max(x_i - x_0, 0)$, $0 \leq \beta \leq \alpha$, которую используют Фер и Шмидт в своей модели неприязни к неравенству.

Тему аксиоматического моделирования неэгоистических предпочтений также развивали Сандбю [49], Нилсон и Стоу [42]. В отличие от описанного выше варианта Нилсона, в предложенных ими моделях при принятии индивидом решения имеет значение вся иерархия платежей, а не только личное место индивида в этой иерархии.

Несколько условно среди аксиоматических подходов к моделированию чувства справедливости можно упомянуть модель, предложенную Оком и Кокесеном [18]. Их работа посвящена аксиоматическому моделированию предпочтений, которые учитывали бы желание людей занимать субъективно лучшее положение по сравнению с другими. Поскольку в качестве критерия для сравнения Ок и Кокесен выбирают уровень дохода, их анализ выливается в моделирование зависти. Однако при выборе другого критерия их аксиомы потенциально могут представлять интерес, например, для объяснения проявлений щедрости в добровольном финансировании общественных благ или благотворительности.

Последняя работа, которая будет упомянута в этом разделе, выделяется из общего ряда тем, что предметом моделирования выступает скорее процедурная справедливость, нежели распределительная. Э. Карни и Ц. Сафра [37] предлагают аксиоматическую модель чувства справедливости, разработанную для механизмов случайного распределения дискретного блага между несколькими людьми, претендующими на него.

Карни и Сафра наделяют индивида двумя полными, транзитивными и непрерывными наборами предпочтений на множестве процедур распределения: один соответствует традиционным предпочтениям, а второй отражает мнение индивида об относительной справедливости распределительных процедур для групп, в которые не входит он сам. Более справедливыми признаются процедуры, обеспечивающие претендентам более равные шансы на получение блага.

Целевая функция, определяющая выбор, представляется функцией от двух компонент – линейной функции полезности и квазивогнутой функции ощущаемой справедливости. При некоторых условиях целевая функция оказывается также аддитивно сепарабельной по этим компонентам.

Поскольку авторы задавали предпочтения справедливости в достаточно общем виде, модель оставляет возможность учесть в предлагаемой целевой функции индивида различные частные аспекты понятия справедливости, например беспристрастность. В заключение Карни и Сафра отмечают, что если предпочтения справедливости у людей достаточно похожи, форма этих предпочтений должна влиять на социальную политику и институты сильнее, чем эгоистический компонент человеческой мотивации, поскольку при усреднении противоречащих друг другу эгоистических интересов они могут нейтрализовывать друг друга.

Развивая описанную выше модель, в 2002 г. авторы опубликовали работу, в которой разработали формальную количественную меру интенсивности чувства справедливости и проанализировали влияние этого показателя на поведение [38].

Мера определяется для индивидов, имеющих одинаковое понятие справедливости и ординально сравнимые⁴⁾ предпочтения. Вкратце, индивид, наделенный более сильным чувством справедливости, в большей степени готов пожертвовать собственным интересом для реализации более справедливой процедуры распределения.

4. Преимущества и недостатки различных подходов

Поскольку большинство описанных в предыдущем разделе моделей было разработано в последние пять лет, мы располагаем лишь ограниченными данными для их оценки и сравнения. О некоторых сильных и слабых их сторонах можно судить и на основе известных закономерностей в поведении участников лабораторных игр, а также общего строения моделей. Элементы такого сопоставления делаются, в частности, в работах Чарнесса и Рэбина [11, 14], Фалька, Фера и Фишбахера [21, 22] и Фера и Шмидта [25].

Все модели чувства справедливости предполагают, что индивид сравнивает себя с другими игроками. Для игр с большим числом участников имеет значение вопрос о том, с кем именно себя сравнивает индивид. Наименее требовательны модели Чарнесса – Рэбина и Болтона – Окенфельса: субъект первой сравнивает свой доход с доходом беднейшего игрока в группе и коллективным доходом группы, а субъект второй сравнивает свое положение с положением «среднестатистического участника игры». Эта простота значительно снижает когнитивные требования к субъектам, делая модели более реалистичными в ситуациях с большим числом участников, но противоречит некоторым экспериментальным данным. В частности, в играх с добровольным финансированием общественных благ и возможностью наказания участники склонны наказывать своих партнеров по игре за некооперативное поведение, в то время как субъекту Болтона – Окенфельса безразлично, кого наказывать – важна только разница между средним доходом и собственным.

Многие модели чувства справедливости опираются на сравнение возможных распределений с неким эталонно справедливым (в первом приближении часто предполагают, что таковым является равномерное распределение). Из этого ряда выделяется модель квазикасиминимальных предпочтений Чарнесса и Рэбина, где в расчет принимается только суммарное благосостояние группы. Ряд экспериментов, посвященных сравнению этой модели с моделями неприязни к неравенству (Болтона – Окенфельса и Фера – Шмидта), показал, что не доминирующая, но существенная часть игроков готова отдать часть своего дохода и усилить неравенство не в свою пользу, если это увеличит общий доход в играх «Диктатор» [1, 4, 14, 15]. Однако нужно заметить, что эта игра существенно отличается от большинства ситуаций реального экономического взаимодействия (при обмене один игрок редко бывает полностью во власти другого).

Модели справедливости, опирающиеся на альтруизм (или квазикасиминимальный критерий), вступают в конфликт с результатами лабораторных экспериментов, в которых участники проявляли желание наказывать отказавшихся сотрудничать партнеров. В остальных моделях такое поведение допустимо и объясняется желанием снизить неравенство доходов или наказать партнера за оппортунистическое поведение. При этом модели, использующие реципрокность, осно-

⁴⁾ В том смысле, что если для одного индивида процедура А лучше процедуры В по предпочтениям полезности и справедливости, то и для другого индивида это будет так.

ванную на намерениях, предсказывают, что субъект захочет наказать «жадного» партнера вне зависимости от того, как это скажется на распределении дохода, но наказания не будет в случае, если распределение определяется некой третьей стороной или безличной силой. Экспериментальные результаты в целом подтверждают обе закономерности [21, 22, 44]. Однако в ряде экспериментов некоторая доля игроков наказывала партнера и в том случае, когда распределение генерировалось компьютером. Таким образом, желание сократить платежи других игроков может быть спровоцировано несправедливостью самого распределения. Это, равно как и желание части игроков наказывать партнера независимо от того, изменит ли это распределение дохода [21], нельзя трактовать в пользу моделей справедливости, использующих реципрокность основанную на намерениях или типа.

Модели, опирающиеся на намерения или тип партнера, хуже согласуются с экспериментальными данными о поощряющем поведении. Большинство авторов [7, 12, 14, 15, 44] не обнаруживают влияния намерений или типа партнера на желание игрока поощрить его, за единственным исключением [22]. Эти модели по построению не могут объяснять поведение игроков, поощряющих партнера первыми, так как предполагают, что игрок должен вначале получить сигнал о доброжелательных намерениях партнера. Исключение – модель Левайна, где такое поведение может объясняться стремлением субъекта увеличить общее благосостояние.

5. Заключение

Обилие экспериментальных данных о поведении, не согласующемся с гипотезой преследования узко понятого собственного интереса, но позволяющем предположить наличие у игрока чувства справедливости, стимулирует непрерывный процесс разработки теоретических моделей, объясняющих такое поведение. Относительная простота, с которой большинство из этих моделей поддаются проверке, способствует их динамичному развитию и постоянному появлению новых данных. Моделируя чувство справедливости и заботу об интересах других, экономисты наделяют агента «социальными предпочтениями», на основе которых он судит о различных вариантах по их последствиям для благосостояния партнеров, или придают ему механизм оценки стратегий партнеров, который позволяет ему судить о намерениях партнеров. В связи с тем, что поведение участников экспериментов зачастую обнаруживает их интерес как к последствиям тех или иных поступков партнеров, так и к их намерениям, авторы последних теоретических разработок склонны комбинировать оба вышеперечисленных метода в своих моделях.

К достоинствам моделей, опирающихся на «социальные предпочтения», следует отнести относительную простоту, нетребовательность к когнитивным способностям субъектов и возможность пользоваться средствами стандартной теории игр. При их калибровке достоинством является небольшое число параметров, которые необходимо оценивать. Они также лучше предсказывают доброжелательное поведение в ситуациях, когда намерения партнера или его личные качества по каким-либо причинам не имеют значения. Некоторые из них также могут быть адаптированы для оценки справедливости институтов и процедур. Модели, использующие реципрокность, основанную на намерениях или типе партнера, в свою очередь, значительно лучше предсказывают стремление людей наказывать оппортунистическое поведение.

Ни одна из имеющихся на сегодняшний день моделей индивидуального чувства справедливости не позволяет реалистично описать весь массив экспериментальных результатов. Однако в совокупности они объясняют и позволяют предсказывать ряд важных закономерностей человеческого поведения, потенциально важных для более глубокого понимания таких экономических проблем, как существование и функционирование неполных контрактов, взаимоотношения работодателей и работников, построение оптимальных мотивационных схем в коллективах, стратегическое взаимодействие при несовершенной конкуренции, общественный выбор, строение механизмов рационализации и перераспределения доходов и др. Справедливость очевидно имеет значение при распределении обязанностей и выигрышей между членами одного коллектива или участниками одного проекта, особенно если их взаимоотношения регулируются неполными или неявными контрактами.

* *
*

СПИСОК ЛИТЕРАТУРЫ

1. *Andreoni J., Miller J.* Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism // *Econometrica*. 2002. Vol. 70. № 2. P. 737–753.
2. *Binmore K.* Game Theory and the Social Contract: Just Playing, Cambridge, Massachusetts: MIT Press, 1998.
3. *Binmore K., Gale J., Samuelson L.* Learning to be Imperfect: The Ultimatum Game // *Games and Economic Behavior*. 1995. № 8. P. 56–90.
4. *Bolle F., Kritikos A.* Self-Centered Inequality Aversion versus Reciprocity and Altruism: Discussion Paper. 1998. № 143. Europa-Universität Viadrina.
5. *Bolton G., Ockenfels A.* A Theory of Equity, Reciprocity and Competition // *American Economic Review*. 2000. № 100. P. 166–193.
6. *Bolton G.* A Comparative Model of Bargaining: Theory and Evidence // *American Economic Review*. 1991. № 81. P. 1096–1136.
7. *Bolton G., Brandts J., Ockenfels A.* Measuring Motivations for the Reciprocal Responses Observed in a Simple Dilemma Game // *Experimental Economics*. 2000. № 3. P. 207–221.
8. *Bowles S., Gintis H.* The Evolution of Strong Reciprocity // *Theoretical Population Biology*. 2004. № 65. P. 17–28.
9. *Camerer C., Thaler R.* Ultimatums, Dictators and Manners // *Journal of Economic Perspectives*. 1995. № 9. P. 209–219.
10. *Cameron L.* Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia // *Economic Inquiry*. 1999. Vol. 37. № 1. P. 47–59.
11. *Charness G., Rabin M.* Understanding Social Preferences with Simple Tests // *Quarterly Journal of Economics*. 2002. № 117. P. 817–869.
12. *Charness G.* Attribution and Reciprocity in a Labor Market: An Experimental Investigation // *Journal of Labor Economics*. 2004. № 22. P. 665–688.
13. *Charness G.* Responsibility and Effort in an Experimental Labor Market // *Journal of Economic Behavior and Organization*. 2000. № 42. P. 375–384.

14. *Charness G., Rabin M.* Social Preferences: Some Simple Tests and a New Model: Working Paper E00-283. University of California at Berkeley, 2000.
15. *Cox J.* Trust and Reciprocity: Implications of Game Triads and Social Contexts: Discussion Paper. University of Arizona, 2000.
16. *Dawes R., Thaler R.* Cooperation // *Journal of Economic Perspectives*. 1988. II. P. 187–197.
17. *Dufwenberg M., Kirchsteiger G.* A Theory of Sequential Reciprocity // *Games and Economic Behavior*. 2004. № 47. P. 268–298.
18. *Ok E., Kockesen L.* Negatively Interdependent Preferences // *Social Choice and Welfare*. 2000. № 17. P. 533–558.
19. *Fahr R., Irlenbusch B.* Fairness as a Constraint on Trust in Reciprocity: Earned Property Rights in a Reciprocal Exchange Experiment // *Economics Letters*. 2000. № 66. P. 275–282.
20. *Falk A., Fischbacher U.* A Theory of Reciprocity // *Games and Economic Behavior*. 2006. № 54. P. 293–315.
21. *Falk A., Fehr E., Fischbacher U.* Informal Sanctions: Working Paper № 59. Institute for Empirical Research in Economics. University of Zürich, 2000.
22. *Falk A., Fehr E., Fischbacher U.* Testing Theories of Fairness – Intentions Matter: Working Paper № 63. Institute for Empirical Research in Economics. University of Zürich, 2000.
23. *Fehr E., Falk A.* Wage Rigidity in a Competitive Incomplete Contract Market // *Journal of Political Economy*. 1999. № 107. P. 106–134.
24. *Fehr E., Schmidt K.* A Theory of Fairness, Competition and Co-operation // *Quarterly Journal of Economics*. 1999. № 114. P. 817–868.
25. *Fehr E., Schmidt K.* Theories of Fairness and Reciprocity - Evidence and Economic Applications: Working paper № 75. Institute for Empirical Research in Economics, University of Zürich, 2001.
26. *Fehr E., Gächter S.* Cooperation and Punishment in Public Goods Experiments // *American Economic Review*. 2000. № 90. P. 980–994.
27. *Fehr E., Kirchsteiger G., Riedl A.* Does Fairness Prevent Market Clearing? An Experimental Investigation. *Quarterly Journal of Economics*. 1993. CVIII. P. 437–460.
28. *Fehr E., Kirchsteiger G., Riedl A.* Gift Exchange and Reciprocity in Competitive Experimental Markets // *European Economic Review*. 1998. № 42. P. 1–34.
29. *Fehr E., Fischbacher U., Tougareva E.* Do High Monetary Stakes Remove Reciprocal Fairness? Experimental Evidence from Russia: Working Paper № 120. Institute for Empirical Economic Research, University of Zürich, 2002.
30. *Forsythe R., Horowitz J., Savin N., Sefton M.* Fairness in Simple Bargaining Games // *Games and Economic Behavior*. 1994. № 6. P. 347–369.
31. *Gächter S., Falk A.* Reputation or Reciprocity?: Working Paper № 19. Institute for Empirical Research in Economics, University of Zürich, 1999.
32. *Geanakoplos J., Pearce D., Stachetti E.* Psychological Games and Sequential Rationality // *Games and Economic Behavior*. 1989. № 1. P. 60–79.
33. *Güth W., Schmittberger R., Schwarze B.* An Experimental Analysis of Ultimatum Bargaining // *Journal of Economic Behavior and Organization*. 1982. III. P. 367–88.
34. *Hoffman E., McCabe K., Smith V.* On Expectations and Monetary Stakes in Ultimatum Games // *International Journal of Game Theory*. 1996. № 25. № 289–301.
35. *Insel T., Young L.* The Neurobiology of Attachment // *Nature Reviews Neuroscience*. 2001. Vol. 2. № 2. P. 129–36.
36. *Kahneman D., Tversky A.* Prospect Theory: An Analysis of Decision Under Risk // *Econometrica*. 1979. № 47. P. 263–291.

37. *Karni E., Safra Z.* Individual Sense of Justice: A Utility Representation // *Econometrica*. 2002. Vol. 70. № 1; ABI/INFORM Global. P. 263.
38. *Karni E., Safra Z.* Intensity of the Sense of Fairness: Measurement and Behavioral Characterization // *Journal of Economic Theory*. 2002. № 105. P. 318–337.
39. *Kosfeld M., Heinrichs M., Zak P., Fischbacher U., Fehr E.* Oxytocin Increases Trust in Humans // *Nature*. 2005. June 2. P. 673–676.
40. *Ledyard J.* Public Goods: A Survey of Experimental Research // *Handbook of Experimental Economics* / A. Roth, J. Kagel (eds.), Princeton: Princeton University Press, 1995.
41. *Levine D.* Modeling Altruism and Spitefulness in Experiments // *Review of Economic Dynamics*. 1998. № 1. P. 593–622.
42. *Neilson W., Stowe J.* Choquet Other-Regarding Preferences: Manuscript. Texas A&M University, 2004.
43. *Neilson W.* An Axiomatic Characterization of the Fehr-Schmidt Model of Inequity Aversion: Working Paper. Department of Economics, Texas A&M University, 2000.
44. *Offerman T.* Hurting Hurts More Than Helping Helps // *European Economic Review*. 2002. № 46. P. 1423–1437.
45. *Rabin M.* Incorporating Fairness into Game Theory and Economics // *American Economic Review*. 1993. Vol. 83. № 5. P. 1281–1302.
46. *Roth A., Erev I.* Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term // *Games and Economic Behavior*. 1995. № 8. P. 164–212.
47. *Roth A., Malouf M., Murningham J.* Sociological Versus Strategic Factors in Bargaining // *Journal of Economic Behavior and Organization*. 1981. № 2. P. 153–177.
48. *Roth A., Prasnikar V., Okuno-Fujiwara M., Zamir S.* Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: Tokyo: An Experimental Study // *American Economic Review*. 1991. № 81. № 1068–1095.
49. *Sandbu M.* Axiomatic Foundations for Reference Dependent Distributive Preferences. Manuscript. Harvard University, 2003.
50. *Segal U., Sobel J.* Tit for Tat: Foundations of Preferences for Reciprocity in Strategic Settings: Discussion Paper 99-10. University of California at San Diego, 1999.
51. *Sethi R., Somanathan E.* Understanding Reciprocity // *Journal of Economic Behavior and Organization*. 2003. Vol. 50. № 1. P. 1–27.
52. *Sethi R., Somanathan E.* Preference Evolution and Reciprocity // *Journal of Economic Theory*. 2000. № 97. P. 273–297.
53. *Slonim R., Roth A.* Financial Incentives and Learning in Ultimatum and Market Games: An Experiment in the Slovak Republic // *Econometrica*. 1997. № 65. P. 569–596.