ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ВЫСШАЯ ШКОЛА ЭКОНОМИКИ

Ю.В. Автономов

МОДЕЛИРОВАНИЕ МОРАЛИ КАК ЭЛЕМЕНТА ВНУТРЕННЕЙ МОТИВАЦИИ ИНДИВИДОВ И МЕХАНИЗМА КОРРЕКЦИИ ПРОВАЛОВ РЫНКА

Препринт WP3/2006/06
Серия WP3
Проблемы рынка труда

Москва
ГУ ВШЭ
2006

УДК 33:303.7
ББК 65в6

Редактор серии WP3
“Проблемы рынка труда”
В.Е. Гимпельсон

Автономов Ю.В. Моделирование морали как элемента внутренней мотивации
А22 индивидов и механизма коррекции провалов рынка. Препринт WP3/2006/06. — М.: ГУ ВШЭ, 2006. — 76 с.

Работа посвящена описанию и анализу формальных моделей морали как элемента индивидуальной мотивации, построенных с помощью инструментов современного неоклассического микроэкономического анализа. Большинство рассматриваемых моделей были разработаны в 1990—2000 гг. Особое внимание уделяется моделированию неэгоистических предпочтений и чувства справедливости в экспериментальной экономике. В работе также затрагиваются проблемы интернализации моральных норм и их применения для компенсации провалов рынка, в частности, в качестве инструмента контроля за производством экстерналий и добровольного финансирования общественных благ.

УДК 33:303.7
ББК 65в6

Препринты ГУ ВШЭ размещаются на сайте:
http:/new.hse.ru/C3/C18/preprints ID/default.aspx

Введение[1]

Экспансия современной экономической науки в предметные области других общественных наук и интеграция их основных идей и результатов на основе единого исследовательского метода стимулируют интерес экономистов к человеческим мотивам, очевидно, играющим роль при принятии людьми экономических решений, но ранее не являвшимся предметом теоретических исследований и моделирования. Одним из таких мотивов являются нравственные установки и моральные нормы, которые во многих случаях регулируют поведение человека в экономической области и часто не могут быть сведены к иным мотивам. По мнению авторов наиболее основательного из существующих в настоящее время обзоров, посвященных проблемам и разработкам в области пересечения этики и экономической теории, Д. Хаусмана и М. Макферсона, “нравственные соображения экономических агентов влияют на их поведение и, следовательно, на конечные результаты их взаимодействия. Более того, нравственные соображения самих экономистов могут намеренно или ненамеренно влиять на мораль и поведение экономических агентов. Следовательно, поскольку экономистов интересуют конечные результаты взаимодействия экономических агентов, их должна интересовать и мораль”[2].

Поскольку “мораль” не является для большинства экономистов привычным термином, необходимо пояснить в каком смысловом аспекте мы будем использовать это понятие в рамках настоящей работы, не углубляясь при этом в сложные определения профессиональных философов. Наиболее продуктивным представляется подход к определению морали, представленный в работе известного исследователя экономического анализа права и автора одной из наиболее остроумных моделей функционирования моральных норм С. Шэйвелла. В разделе своей книги, посвященном описанию понятия морали[3], Шэйвелл отмечает, что мораль можно определить как инструмент оценки ситуаций, посредством которого правильное, верное, справедливое или моральное (в данном случае эти определения используются как синонимы) поведение или действие оценивается выше неправильного, неверного, несправедливого или аморального. В литературе часто подчеркивается, что моральные нормы определяются для ситуаций,

[1] Автор выражает благодарность Р.И. Капелюшникову и М.И. Левину за ценные замечания и комментарии.

[2] Hausman D., McPherson M. Taking Ethics Seriously: Economics and Contemporary Moral Philosophy // Journal of Economic Literature. 1993. Vol. 31. No. 2. P. 673.

[3] Shavell S. Foundations of Economic Analysis of Law. L.: Belknap Press, 2004.

когда поведение одного индивида влияет на благосостояние не только его самого, но и других людей.

Второй важный элемент определения морали или моральной нормы заключается в том, что с моральной нормой у человека должны ассоциироваться некоторые специфические эмоции ("нравственные чувства", moral sentiments). К таким эмоциям Шэйвелл относит чувство добродетельности или гордости собой (virtue), которое человек испытывает, соблюдая моральную норму, и противоположное чувство вины, которое он испытывает, нарушая ее. Такого рода эмоции могут переживаться и теми людьми, которые наблюдают за моральным или аморальным поступком. Отсюда следует, что индивидуальная полезность человека, обладающего набором моральных норм, может меняться от их соблюдения и несоблюдения им самим и другими людьми[4].

Предпосылки формального моделирования моральных норм в экономической теории были заложены в 1970-х гг., в период интенсивного расширения предметного поля экономической теории, и примерно к тому же времени относятся первые модели. В 1990-х гг. количество экономических исследований, учитывающих воздействие на поведение индивида моральных норм в формальных моделях, значительно возросло, и в настоящее время это направление в экономической литературе представлено достаточно широко.

Существует несколько обзорных работ, посвященных взаимоотношениям экономики и этики, среди которых можно назвать, например, книги А. Бьюкенена[5], А. Хэмлина[6], А. Сена[7] и уже упоминавшихся ранее Д. Хаусмана и М. Макферсона[8]. Однако эти работы имеют выраженный методологический уклон, и внимание в них уделяется в первую очередь фундаментальным проблемам, например, взаимосвязи морали и рациональности, роли морали в экономике благосостояния, концепциям права, свободы, справедливости и равенства в экономической теории, и т.п. Существующие разработки в области формального моделирования морали как фактора, влияющего на поведение индивидов, освещаются в них достаточно скудно. Данная работа призвана отчасти восполнить этот пробел. При этом она не претендует на формирование полной картины существующих направлений интеграции морали в экономическую теорию на формальном уровне, что могло бы стать предметом значительно более фундаментального по объему и теоретическому охвату исследования, примеры которого в современной экономической литературе нам не известны. Целью настоящей работы является, скорее, создание представления о тех попытках и инструментах формального моделирования морали, которые в целом находятся в границах современного неоклассического микроэкономического анализа и, следовательно, могут быть использованы для расширения его средств в анализе экономического поведения, имеющего очевидную моральную подоплеку.

Диапазон работы ограничивается подходами к формальному моделированию морали как *внутреннего* мотива, влияющего на поведение экономических агентов. В поле зрения, следовательно, не входит, во-первых, обширный корпус моделей, объясняющих существование потенциальных практических проявлений морали — кооперативного взаимодействия, честности, альтруистического поведения и т.п., — сугубо эгоистическими причинами. Во-вторых, под экономическими агентами понимаются прежде всего индивиды, а значит, влияние этических соображений на деятельность более сложных экономических агентов, таких как фирмы или общественные объединения, — несомненно существенное, — также лежит вне сферы рассмотрения.

В русле экономического анализа моральные нормы и предрасположенности могут быть инструментально оценены с различных сторон, как то: экономическая эффективность, распределительная справедливость, баланс сил и т.п. Пожалуй, наиболее естественным для экономической теории взглядом на мораль, соединяющим ее с традиционной логикой экономического анализа, является рассмотрение моральных норм с точки зрения экономической эффективности (Парето-оптимальности), и именно такая точка зрения наиболее широко представлена в рассматриваемых примерах моделирования морали.

Упомянутые работы достаточно разнообразны по степени абстрактности или конкретности анализируемых проблем. Основными из них являются влияние моральных норм на поведение экономических агентов при распределении благ, производство общественных благ и деятельность, связанная с производством экстерналий. К сожалению, в поле рассмотрения данной работы не вошла, вероятно, самая популярная область исследования влияния морали на экономическое поведение — влияние нравственных установок на мотивацию работников, форму контрактов и организацию фирмы. Эта тема, получившая во второй половине XX в. новый импульс к развитию в рамках теории контрактов, заслуживает отдельного представления и не может быть рассмотрена здесь вследствие обширности соответствующего материала.

---

[4] Представляется, однако, что желание человека следовать моральным нормам не является единственной причиной их существования. Как профессиональные философы, так и обычные люди часто обосновывают необходимость морали деонтологическими (основанными на понятии долга) соображениями.

[5] Buchanan A. Ethics, Efficiency and the Market. Totowa (NJ): Rowman and Allanfeld, 1985.

[6] Hamlin A. Ethics, Economics, and the State. N.Y.: St. Martin's Press, 1986.

[7] Sen A. On Ethics and Economics. Oxford: Basil Blackwell, 1987.

[8] Hausman D., McPherson M. Economic Analysis and Moral Philosophy. Cambridge: Cambridge University Press, 2002.

В большинстве работ мораль моделируется посредством изменения тех или иных элементов стандартной для основного течения современной экономической теории модели индивида, максимизирующего свои предпочтения при заданных ограничениях. (Под "стандартной моделью homo oeconomicus" мы будем понимать модель рационального максимизатора в том виде, в котором она описывается в учебниках по микроэкономике, без беккеровских или ланкастеровских доработок.) Наиболее частым объектом модификаций является именно система предпочтений.

Самый популярный вариант ее модификации — постулирование наличия у индивида так называемых "социальных предпочтений", т.е. определение предпочтений на множестве распределений ресурсов между индивидами, входящими в сообщество, рассматриваемое в конкретной задаче. При этом, помимо набора, достающегося ему самому, индивид судит о желательности того или иного распределения и по наборам, достающимся другим людям. Это позволяет моделировать роль индивидуального чувства справедливости в задачах, подразумевающих принятие экономическим агентом решений о распределении ресурсов. При этом элементами потребительского множества в задаче могут служить как сами распределения ресурсов, так и механизмы их распределения, в том числе стохастические. Таким образом, подобные подходы могут применяться при моделировании ситуаций, подразумевающих распределение неделимых благ. Область применения таких моделей потенциально достаточно обширна, но в настоящий момент они наиболее широко представлены в экспериментальной экономике.

Несколько менее популярным является подход, в соответствии с которым моральный аспект включается в экономическую задачу постулированием зависимости между предпочтениями и ограничениями индивида. Основная идея данного подхода состоит в том, что наличие у индивида некоторых желаний (в терминологии Маслоу — предпочтений высшего порядка) накладывает дополнительные ограничения на перечень действий, которые он может предпринять для их удовлетворения. Например, если вы цените собственный доход и вам нравится считать себя честным человеком, придется смириться с тем, что вы не сможете завышать объем ваших усилий перед работодателем. Конкретный механизм выбора в данных моделях четко не прописан, но можно предположить, что выбор становится двухступенчатым — например, индивид вначале выбирает оптимальный набор в рамках каждой из доступных ему систем "предпочтения — ограничения", а затем, сравнив оптимальные наборы, выбирает ту систему, которая принесет ему наибольшую "полезность"[9]. Таким образом, любое благо воспринимается в контексте того способа, которым оно получено.

[9] Термин "полезность" здесь достаточно условен и используется как традиционное обозначение единицы размерности целевой функции индивида в экономической теории.

Альтернативой концепции "социальных предпочтений" является использование реципрокности. Для агента соответствующих моделей в предпочтения некоторым образом включаются его субъективные ожидания относительно намерений партнера, т.е. одно и то же предложение партнера может иметь для агента разную ценность в зависимости от того, как он воспринимает множество его стратегий. Применимость концепции реципрокности для моделирования морали вызывает некоторые сомнения, поскольку в некоторых условиях реципрокность выливается в элементарное правило "ты мне, я тебе". В пользу реципрокности говорит то, что она (1) является настолько же базовым элементом человеческой психологии, как и собственный интерес, (2) ее можно толковать и как интернализованную норму платить добром за добро, (3) эти модели показывают хорошие результаты в объяснении неэгоистического поведения в игровых экспериментах.

Первые разделы данной работы посвящены анализу и описанию морали как компонента индивидуальной мотивации, далее мы обращаемся к моделям, акцентирующим внимание на различных аспектах функционирования моральных норм и имеющим в целом более теоретический характер. В число затрагиваемых в этих моделях проблем входят теоретическое описание механизма самоконтроля и самосовершенствования, формализация морально обусловленного выбора в условиях дискретного множества вариантов, оптимальное использование нравственных чувств, а также применение кантианских моральных правил как инструмента контроля за производством экстерналий и механизма, поддерживающего добровольное финансирование общественных благ.

Моделирование чувства справедливости на основе "социальных предпочтений"

Сама по себе идея социальных предпочтений в неоклассической микроэкономической теории имеет достаточно давнюю историю. Многие известные экономисты (например, Г. Беккер[10], К. Эрроу[11], П. Самуэльсон[12] и

[10] Becker G. A Theory of Social Interactions // Journal of Political Economy. 1974. No. 82. P. 1063—1093.

[11] Arrow K. Optimal and Voluntary Income Redistribution // Economic Welfare and the Economics of Soviet Socialism: Essays in Honor of Abram Bergson / Ed. by S. Rosefield. Cambridge: Cambridge University Press, 1981.

[12] Samuelson P. Altruism as a Problem Involving Group Versus Individual Selection in Economics and Biology // American Economic Review. 1993. No. 83. P. 143—148.

А. Сен[13]) высказывали мнение, что людей часто заботит чужое благосостояние, и эта озабоченность может иметь важные последствия для их собственного экономического поведения. Однако до недавнего времени эти соображения не были востребованы в мейнстримной экономической науке.

Одним из основных "проводников" этих идей в экономической науке являются исследования в области экспериментальной экономики. В настоящий момент исследователи располагают обширной коллекцией свидетельств расхождения человеческого поведения с гипотезой о строгом следовании собственному интересу. Во многих из этих экспериментов поведение участников позволяло предположить существенное влияние на их выбор нравственных соображений, таких как справедливость или ее более узкий вариант — реципрокность.

К основным играм, результаты которых свидетельствуют о влиянии на поведение участников честности и реципрокности, прежде всего следует отнести игры "Ультиматум", "Дарообмен", "Доверие", "Диктатор" и игры с добровольным финансированием общественных благ.

В игре "Ультиматум" игрок А предлагает игроку Б пропорцию, в которой между ними будет разделена некоторая сумма денег. Игрок Б может согласиться, и получить предложенную ему долю, или не согласиться, и тогда игроки А и Б не получат ничего. С точки зрения классической теории игр, в совершенном подыгровом равновесии игроку А выгодно предлагать игроку Б наименьшую положительную сумму, и игроку Б выгодно соглашаться. Однако результаты многочисленных экспериментов показывают, что игрок Б отвергает долю ниже 20% от общей суммы с вероятностью от 0,4 до 0,6, и чем больше предложенная доля, тем меньше вероятность того, что предложение будет отвергнуто. В одном из экспериментов было показано, что модальная доля, предлагаемая игроком А партнеру, примерно максимизирует его собственный ожидаемый доход[14].

В игре "Диктатор" игрок Б не имеет возможности отказаться от предложения игрока А. Таким образом, стандартная теория игр предсказывала бы, что игрок Б не получит ничего. Однако экспериментальные результаты показали, что игрок А предлагает игроку Б небольшую, но все же положительную сумму — значительно меньшую, чем в игре "Ультиматум".

Игра "Дарообмен" является экспериментальным переложением классических взаимоотношений принципала и агента. Принципал предлагает

[13] Sen A. Moral Codes and Economic Success // Market Capitalism and Moral Values / Ed. by C.S. Britten, A. Hamlin. Edward Eldar, Aldershot, 1995.

[14] Roth A., Prasnikar V., Okuno-Fujiwara M. et al. Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study // American Economic Review. 1991. No. 81. P. 1068—1095.

агенту некую сумму $w \in [\underline{w}, \overline{w}]$, которую тот может принять или не принять. Если агент отвергает платеж, оба не получают ничего. Если агент принимает платеж, ему предлагают выбрать "уровень усилий" $e \in [\underline{e}, \overline{e}]$, $\underline{e} > 0$. 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 предпосылках 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. В результате зависимость среднего усилия от w достаточно сильна, чтобы принципалам было выгодно предлагать высокие w.

В игре "Доверие" игрок А получает от экспериментатора сумму y, и может отдать часть ее, z, игроку Б, причем экспериментатор передает игроку Б $3z$, утроенную сумму, предложенную А. В ответ игрок Б может вернуть А некоторую часть полученных денег. В экспериментах многие игроки А передавали игроку Б ненулевое z, и возвращаемая им в ответ сумма коррелировала с величиной z.

Типичная игра с добровольным финансированием общественного блага разыгрывается между n участниками, одновременно решающими, какую долю своего первоначального запаса вложить в 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 продолжается 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. 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.

Перечисленные экспериментальные результаты вызвали ряд попыток теоретического объяснения поведения, отклоняющегося от гипотез рационального преследования узкозаданного собственного интереса[15].

Эти попытки делятся на несколько категорий. Некоторые сохраняют предпосылку о сугубо эгоистической мотивации и пытаются объяснить наблюдаемые результаты ограниченной рациональностью. Например, склонность пассивной стороны в игре "Ультиматум" наказывать активную объясняется тем, что потери первой от отказа относительно меньше, чем потери второй — а значит, стимул активной стороны адаптировать свое поведение, предлагая больше, более значителен. В свою очередь, если активная сторона начинает предлагать больше достаточно быстро, стимул пассивной стороны менять свое поведение еще больше сокращается. В таких условиях схождение игры к равновесию, совершенному в подыграх, может продолжаться очень долго. Возражением против такого объяснения может послужить то, что решение, принимаемое пассивной стороной в игре "Ультиматум", слишком простое для подобной череды ошибок. К тому же способность активных игроков подобрать стратегию деления так, чтобы в среднем максимизировать свой доход, заставляет предположить, что участники игры достаточно рациональны для поставленной им задачи.

В других вариантах описываемое поведение объясняют воздействием некой социальной нормы, выработавшейся на основе многочисленных случаев повторяющегося взаимодействия в реальной жизни, а затем переросшей в жесткое правило поведения. Ошибочное применение этого правила к одноперiодным лабораторным ситуациям, по логике объяснения, и порождает наблюдаемые результаты.

Как правило, этот аргумент в конечном счете сводится к тому, что участники игры ошибочно принимают экспериментальную одноперiодную игру за повторяющуюся. Критика этого подхода к объяснению, помимо систематических ошибок, якобы совершаемых индивидами, опирается также, во-первых, на существенные различия в поведении игроков в разных типах одноперiодных игр — например, в играх "Ультиматум" и "Диктатор", и во-вторых, на различия в поведении игроков в одноперiодных и повторяющихся экспериментальных ситуациях. Многие игроки после эксперимента жалуются, что одноперiодное взаимодействие не позволило им выработать более эффективную стратегию взаимодействия с партнером.

[15] Обзор этих теорий можно найти в работе Э. Фера и К. Шмидта (Fehr E., Schmidt K. Theories of Fairness and Reciprocity − Evidence and Economic Applications. Institute for Empirical Research in Economics. University of Zurich, 2001. Working paper No. 75). См. также: Falk A., Fehr E., Fischbacher U. (2000b). Testing Theories of Fairness – Intentions Matter. Institute for Empirical Research in Economics. University of Zurich, 2000. Working Paper No. 63.

Но чаще всего авторы теоретических объяснений рассматриваемого экспериментального поведения идут по пути модификации предпочтений, сохраняя предпосылку о полной рациональности. При этом наблюдается два подхода — агенты либо наделяются "социальными предпочтениями" и ситуация анализируется средствами традиционной теории игр, либо используется так называемая "реципрокность, основанная на намерениях" (intention-based reciprocity), и применяется психологическая теория игр. Данный раздел работы посвящен описанию первой, наиболее обширной группы моделей.

Итак, формально, если обозначить за $\{1, 2, ..., N\}$ множество индивидов, и обозначить вариант распределения натуральных ресурсов между N индивидами за $x = (x_1, x_2, ..., x_N)$ (где x_i соответствует количеству ресурсов у индивида i), изменение полезности индивида i при варьировании x_j, $j \neq i$, означает наличие у индивида i "социальных предпочтений".

В качестве самой простой модификации индивидуальных предпочтений агент наделяется альтруизмом. В приведенной выше формализации агент обладает альтруизмом, если частные производные его функции полезности ($U(x_1, x_2, \ldots, x_N)$) по доходу других людей $(\frac{\partial U}{\partial x_1}, \frac{\partial U}{\partial x_2} ... \frac{\partial U}{\partial x_N})$ строго положительны.

Чарнесс и Рэбин[16] 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":

$$U_i(x_1, x_2, \ldots, x_N) = (1 - \gamma)x_i + \gamma W(x_1, x_2, \ldots, x_N),$$

где упомянутая функция общественного благосостояния представляет собой выпуклую комбинацию утилитаристской функции благосостояния и роулсианского максиминимального критерия[17]:

[16] Charness G., Matthew R. Social Preferences: Some Simple Tests and a New Model. University of California at Berkeley, 2000 (Mimeo).

[17] Заметим, что Чарнесс и Рэбин не прибегают к нормализации утилитаристской компоненты W по величине группы N, а значит, при росте N и неизменных δ и γ вес максиминимального критерия в индивидуальной функции полезности падает.

$W(x_1, x_2, \ldots, x_N) = \delta min\{x_1, x_2, \ldots, x_N\} + (1 - \delta)(x_1 + x_2 + \ldots + x_N)$.

При такой функции полезности индивид меньше беспокоится о благосостоянии других, если их доход превышает его 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.

Альтернативная гипотеза, история которой в экономической науке восходит, по меньшей мере, к Веблену, состоит в том, 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". Болтон[18] формализовал эту гипотезу для игры с делением доллара между двумя игроками, наделив каждого функцией полезности вида $U_i(x_i, x_j) = u_i(x_i, x_i/x_j)$, с $\frac{\partial U_i}{\partial x_i} > 0$, $\frac{\partial U_i}{\partial \frac{x_i}{x_j}} \le 0$ при $x_j > x_i$, и $\frac{\partial U_i}{\partial \frac{x_i}{x_j}} = 0$ при $x_j \le x_i$. То есть индивид завидует благосостоянию партнера, если получает меньше его, но безразличен к положению партнера, если получает больше его. Такая формализация объясняла поведение участников в рассматриваемых Болтоном играх, но, разумеется, она противоположна по духу любым альтруистическим проявлениям. Несколько авторов, соединив воедино зависть и альтруизм, предложили модели, наделяющие индивида "неприязнью к неравенству". Это значит, что индивид альтруистичен по отношению к тем партнерам, чье благосостояние ниже "порога справедливости", но завидует тем, чье благосостояние выше этого порога. Для большинства экспериментальных игр естественно предполагать, что порог справедливости соответствует размеру платежа одному игроку, если все получают равную сумму.

Фер и Шмидт[19] формализуют это при помощи простой функции полезности вида

[18] Bolton G. A Comparative Model of Bargaining: Theory and Evidence // American Economic Review. 1991. No. 81. P. 1096—1136.

[19] Fehr E., Schmidt K. A Theory of Fairness, Competition and Co-operation // Quarterly Journal of Economics. 1999. No. 114. P. 817—868.

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

с $\beta_i \le \alpha_i$ и $\beta_i \le 1$. Заметим, что $\partial U_i / \partial x_j \ge 0$ тогда и только тогда, когда $x_i \ge x_j$, и неравенство сильнее беспокоит индивида в случае, когда благосостояние партнера выше собственного, чем если оно ниже собственного ($\alpha_i \ge \beta_i$). С такой функцией полезности возможно объяснить для всех рассматриваемых нами игр как альтруистические действия по отношению к партнерам, так и стремление наказать их. Но если бы все индивиды имели идентичные предпочтения, было бы невозможно объяснить, почему в некоторых случаях альтруизм, сотрудничество, стремление наказать несправедливость, не являясь лучшей эгоистической стратегией, наблюдаются в существенном объеме, а в других — едва ли имеют место. Фер и Шмидт показывают, что сочетание определенных распределений типов со стратегическим окружением может объяснить как неравномерные, 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[20]. Модель также объясняет ряд считающихся парадоксальными результатов, например, трансферты в игре "Диктатор" и обратные трансферты в игре "Дарообмен", а также отказ от предложения активной стороны в игре "Ультиматум". Они предлагают следующую функцию полезности (мотивационную функцию):

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

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

$$\sigma_i = \begin{cases} \dfrac{x_i}{\sum_{j=1}^{N} x_j} & if \quad \sum_{j=1}^{N} x_j \neq 0 \\ \dfrac{1}{N} & if \quad \sum_{j=1}^{N} x_j = 0 \end{cases}$$

[20] Bolton G., Ockenfels A. A Theory of Equity, Reciprocity and Competition // American Economic Review. 2000. No. 100. P. 166—193.

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. Болтон и Окенфельс не предлагают конкретной формы для своей функции полезности.

В случаях с двумя игроками две последние описанные 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 возникать неожиданная щедрость по отношению к игроку, намного более благополучному чем он, и зависть к игроку, который беднее его самого.

Левайн[21] объясняет трансферты в одних играх и желание наказать несправедливость иным путем. Рассмотрим функцию полезности вида

$$U_i = x_i + \sum_{j \neq i} \frac{x_j (a_i + \lambda a_j)}{1 + \lambda},$$ где $0 \leq \lambda \leq 1$ и $-1 < a_i < 1$ для всех $i \in \{1, \ldots, N\}$,

и предположим, что $\lambda = 0$. Тогда функция полезности превращается в $U_i = x_i + a_i \sum_{j \neq i} x_j$, и при $a_i > 0$ человек стремится увеличить 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. Однако в большинстве игр параметры a_j контрагентов ненаблюдаемы для игрока, и любая последовательная игра оказывается связанной с сигнализированием.

Откалибровав распределение a на данных игры "Ультиматум", Левайн оценил (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 рыночным играм. Однако, поскольку $a_i < 1$, модель не объясняет ненулевых трансфертов в игре "Диктатор".

[21] Levine D. Modeling Altruism and Spitefulness in Experiments // Review of Economic Dynamics. 1998. No. 1. P. 593—622.

Модели социальных предпочтений, рассмотренные выше, хорошо описывают человеческое поведение в достаточно широком перечне игр. Но основным их преимуществом перед другими подходами к моделированию честности, щедрости или неприязни к неравенству, вероятно, является потенциальная легкость, с которой они могут быть применены для анализа различных экономических проблем. По сравнению с конкурирующими моделями реципрокности, основанной на намерениях, они достаточно просты. Моделируя честность или неприязнь к неравенству через социальные предпочтения, мы имеем возможность пользоваться средствами стандартной теории игр, что упрощает анализ и выдвижение количественных и качественных прогнозов. Число параметров, которые необходимо оценивать, относительно невелико.

Моделирование чувства справедливости с помощью реципрокности, основанной на намерениях

Общей чертой расмотренных выше моделей "общественных предпочтений" является то, что любой исход оценивается изолированно, вне мотивов, обстоятельств и условий, породивших его. До некоторой степени такое упрощение является достоинством, делающим модели более операбельными. Однако нетрудно представить себе ситуации, когда объяснить или предсказать поведение индивидов, опираясь только на последствия их действий, окажется невозможным. В качестве примера рассмотрим две игры типа "Ультиматум", в которых множество стратегий активного игрока ограничено следующим образом. В первой игре активный игрок может предложить пассивному распределение 50:50 или 80:20. Во второй — 20:80 или 80:20. Из любых теорий, в которых предпочтения определяются только на множестве распределений, будет вытекать, что пассивный игрок, отвергший распределение 80:20 в первой игре, отвергнет его и во второй. Однако интуитивно можно было бы ожидать, что пассивный игрок скорее согласится принять распределение 80:20 во второй игре, чем в первой, где у его партнера действительно *была* возможность предложить более справедливый вариант. Фальк, Фер и Фишбахер экспериментально подтвердили, что во второй игре пассивные игроки существенно чаще соглашаются на распределение 80:20, чем в первой[22].

Данный раздел посвящен моделям, в которых проявления чувства справедливости у людей объясняются с помощью "реципрокности, основанной

[22] Falk A., Fehr E., Fischbacher U. Informal Sanctions. Institute for Empirical Research in Economics. University of Zurich, 2000. Working Paper No. 59.

на намерениях" (intention-based reciprocity). Авторы этих моделей 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. При этом для анализа взаимодействия игроков используется психологическая теория игр. Заметим, что одновременно существует постоянно растущий корпус литературы, объясняющей существование реципрокности эволюционными соображениями[23]. Подобные работы здесь не рассматриваются.

Одним из первых исследователей, предложивших рассматриваемую схему объяснения, был Рэбин[24], выдвинувший формальное определение реципрокности, основанной на намерениях. Он ограничивал свое внимание играми для двух игроков в нормальной форме, воспользовавшись психологической теорией игр, в которой полезность игроков зависит не только от платежей в конечных узлах игры, но и от их предположений о чужих намерениях. Простота модели позволяет нам почти полностью изложить здесь ее содержание.

Обозначим за A_1 и A_2 наборы смешанных стратегий игроков 1 и 2, и обозначим платежную функцию игрока i за x_i: $A_1 \times A_2 \to R$. Далее, определим убеждения на множестве стратегий. Пусть $a_i \in A_i$ обозначает стратегию игрока i, и выбирая ее, он должен иметь некую веру относительно стратегии, которую выберет j. Предположим, что $i \in \{1, 2\}$ и $j = 3 - i$. Обозначим как b_j веру индивида i относительно того, что сделает j. Чтобы вера b_j была рациональной, у индивида i должна быть и вера относительно того, что, по мнению j, предпримет он сам. Обозначим эту веру как c_i. Такая иерархия вер могла бы продолжаться, 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 выборе 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.

[23] См., например: Bowles S., Gintis H. The Evolution of Strong Reciprocity. University of Massachusetts at Amherst, 1999 (Mimeo); Sethi R., Somananthan E. Preference Evolution and Reciprocity // Journal of Economic Theory. 2000. No. 97. P. 273—297; Sethi R., Somananthan E. Understanding Reciprocity. Columbia University, 2000 (Mimeo).

[24] Rabin M. Incorporating Fairness into Game Theory and Economics // American Economic Review. 1993. No. 83 (5). P. 1281—1302.

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 веры игрока i относительно того, 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 $x_i^h(c_i) - x_j^l(c_i) = 0$. Обе функции доброты теперь 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 вер $b = (b_1, b_2)$ 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 полярных (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[25] расширили ее до игр с N игроками в расширенной форме, введя понятие последовательного реципрокного равновесия (Sequential Reciprocity Equilibrium, SRE). Основная идея заключалась в отслеживании вер по ходу игры и определении того, как должны формироваться веры на неиграемых 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 верами. Например, предложение активной стороной в игре "Ультиматум" такого распределения, которое будет заведомо отвергнуто, успешно объяснялось существованием у игроков вер относительно того, что оппонент хочет навредить им.

Авторы еще одного обобщения модели Рэбина, Фальк и Фишбахер[26], объединили заинтересованность индивида намерениями контрагента с наличием у него общественных 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, рассмотренных 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".

Предложенная Фальком и Фишбахером 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. Подобные предпочтения в сочетании с формой исходной игры формируют психологическую игру, подобную предложенным Геанакоплосом, Пирсом и Стакетти[27]. При таком подходе совершенное в подыграх равновесие по Нэшу, для некоторых диапазонов параметров, объясняет экспериментальное поведение во всех играх, перечисленных нами в начале предыдущего раздела.

Еще одна попытка комбинировать общественные предпочтения с реципрокностью принадлежит Чарнессу и Рэбину[28]. Расширяя предложенную ими ранее модель, которая рассматривалась выше, они дополняют предпочтения "профилем недостатков", $\rho \equiv (\rho_1, ..., \rho_N)$, 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

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

где $d, k, f \geq 0$ — три новых параметра. Если $d = k = f = 0$, предпочтения сводятся к квазимаксиминимальным, рассмотренным ранее. При больших d и k игрок i не захочет 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.

Взаимно 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, существенно 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 квазимаксиминимальную функцию общественного благосостояния. Наконец, сложность и обилие свободных параметров затрудняют эмпирическую проверку модели.

[25] Dufwenberg M., Kirchsteiger G. A Theory of Sequential Reciprocity. Discussion Paper. CentER, Tilburg University, 1998.

[26] Falk A., Fischbacher U. A Theory of Reciprocity. Institute for Empirical Research in Economics. University of Zurich, 1999. Working Paper No. 6.

[27] Geanakoplos J., Pearce D., Stacchetti E. Psychological Games and Sequential Rationality // Games and Economic Behavior. 1989. No. 1. P. 60—79.

[28] Charness G., Rabin M. Understanding Social Preferences with Simple Tests // Quarterly Journal of Economics. 2002. No. 117. P. 817—869.

В целом, к возможным областям применения моделей справедливости — как базирующихся на социальных предпочтениях, так и использующих реципрокность, основанную на намерениях, — следует отнести анализ человеческого поведения в семье, в рабочем коллективе, во взаимоотношениях с соседями и друзьями. 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.

Поскольку проблемы распределения обязанностей или выгод существуют не только в небольших сообществах или между знакомыми людьми, влияние соображений справедливости на поведение людей распространяется и на более широкие области (например, управление персоналом в организациях или государственная политика). Такие ситуации уже нашли отражение в экономической литературе. Например, было показано, что на размеры воровства среди работников и их общий моральный дух влияет то, что они думают о честности политики фирмы[29]. Влияние на мотивацию работников соображений честности и равенства может приводить к тому, что сокращать оплату труда напрямую становится невыгодно[30]. Общий уровень законопослушания, соблюдение контрактов и организационных правил существенно зависят от мнения людей о справедливости распределения материальных благ и процедур их распределения[31]. Мнение налогоплательщиков о справедливости системы налогообложения с определенной вероятностью влияет на объем уклонения от налогов[32]. Уровень поддержки населением решений о регулировании частных предприятий зависит от мнения о справедливости рыночной политики этих фирм[33].

[29] Bewley T. Why Wages Don't Fall During a Recession. Harvard: Harvard University Press, 1999; Greenberg J. Employee Theft as a Reaction to Underpayment Inequity: The Hidden Cost of Pay Cuts // Journal of Applied Psychology. 1990. No. 75. P. 56—568.

[30] Agell J., Lundborg P. Theories of Pay and Unemployment: Survey Evidence from Swedish Manufacturing Firms // Scandinavian Journal of Economics. 1995. No. 97. P. 295–308; Kahneman D., Knetsch J., Thaler R. Fairness as a Constraint on Profit Seeking: Entitlements in the Market // AER. 1986. LXXVI. P. 728—741.

[31] Fehr E., Gächter S., Kirchsteiger G. Reciprocity as a Contract Enforcement Device // Econometrica. 1997. No. 65. P. 833—860; Lind A., Tyler T. The Social Psychology of Procedural Justice. N.Y.; L.: Plenum Press, 1988.

[32] Andreoni J., Erard B., Feinstein J. Tax Compliance // Journal of Economic Literature. 1998. No. 36. P. 818–860; Alm J., Sanchez I., Juan de A. Economic and Noneconomic Factors in Tax Compliance // Kyklos. 1995. No. 48. P. 3—18; Frey B., Weck-Hannemann H. The Hidden Economy as an "Unobserved" Variable // European Economic Review. 1984. No. 26. P. 33—53.

[33] Zajac E. Political Economy of Fairness. Cambridge (Mass.): MIT Press, 1995.

Наконец, решение задач, требующих коллективного действия (например, правила доступа к общим ресурсам), существенно зависит от того, насколько справедливо эти правила распределяют издержки и выгоды[34].

Аксиоматические подходы к моделированию чувства справедливости

Все рассматриваемые в данной работе модели в большей или меньшей степени опираются на теоретический каркас неоклассической микроэкономической теории, и в том числе на фундаментальные аксиомы теории полезности. Используемые при этом специфические функции полезности, определенные на векторах материальных исходов и верах игроков относительно стратегий партнеров или их вер, правдоподобны с точки зрения человеческой психологии. Однако это вовсе не означает, что они сводятся к классическому набору аксиом теории полезности.

В этой связи появление попыток сформулировать аксиомы, достаточные для существования предпочтений, отражающих соображения справедливости или реципрокности, вполне естественно и логично. Однако, в отличие от неоклассического микроэкономического анализа, область моделирования морали и, в частности, справедливости и реципрокности, не имеет на данный момент единого, или базового, подхода. Поэтому различные наборы аксиом, которые могут поддерживать предпочтения, включающие соображения справедливости и реципрокности, часто привязаны к определенным функциональным формам задания таких предпочтений.

Общее количество таких работ достаточно невелико, и в данном разделе перечислена бóльшая их часть. Как правило, работы, посвященные аксиоматическому моделированию справедливости и реципрокности, достаточно сложны технически. Чтобы облегчить изложение данного раздела, технические детали были опущены. В случае необходимости читатели могут обратиться непосредственно к тексту соответствующих статей.

[34] Ostrom E. Collective Action and the Evolution of Social Norms // Journal of Economic Perspectives. 2000. No. 14. P. 137—158; Falk A., Fehr E., Fischbacher U. Appropriating the Commons. Institute for Empirical Research in Economics. University of Zurich, 2000. Working Paper No. 55.

Сигал и Собель[35] 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. 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 $\succ_{i,\sigma_j}$ на Σ_i, множестве собственных смешанных стратегий. Заметим, что оценка игроком собственных стратегий $\succ_{i,\sigma_j}$ зависит от стратегии, выбранной партнером — σ_j. Сигал и Собель показывают, что если отношения предпочтения $\succ_{i,\sigma_j}$ удовлетворяют аксиомам непрерывности и независимости, и если для заданного σ_j игрок, при фиксированном платеже партнера, имеет больший платеж для себя (предпосылка о собственном интересе), предпочтения $\succ_{i,\sigma_j}$ 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 $a_{i,\sigma j} = 0$. Если $a_{i,\sigma j}$ положителен, игрок является альтруистом, если отрицателен — игрок недоброжелателен. При этом коэффициент $a_{i,\sigma j}$ 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 противном случае. Технически эта аксиома означает, что при прочих равных коэффициент $a_{i,\sigma j}$ тем больше, чем лучший исход для индивида i обеспечивает стратегия его оппонента σ_j.

Фер и Шмидт замечают, что рассмотренные нами в предыдущем разделе формализации социальных предпочтений, основывающиеся на альтруизме, относительном доходе, неприязни к неравенству и квазимаксиминимальных предпочтениях, можно классифицировать как частные случаи предложенной Сигалом и Собелем функции полезности[36].

[35] 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.

[36] Fehr E., Schmidt K. Theories of Fairness and Reciprocity — Evidence and Economic Applications. Institute for Empirical Research in Economics. University of Zurich, 2001. Working paper No. 75. P. 24.

Более узкий частный случай представляет собой работа Нилсона[37], которая дает аксиоматическое обоснование модели неприязни к неравенству Фера и Шмидта[38]. Данная работа посвящена аксиоматическому моделированию предпочтений, зависимых от точки отсчета — как, например, предложенных Канеманом и Тверски[39] в теории ожиданий.

Рассмотрим игру с тремя участниками и предположим, что 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 $(x_1, 60, 60) \succ (x_1, 80, 40)$ 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 изменять 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 $u_0, \ldots, u_n$ на $v_0, \ldots, v_n$, полученные из $u_0, \ldots, u_n$ преобразованием $v_i = a + bu_i$, где a — скалярная величина, а b — строго положительная скалярная величина, предпочтения не изменятся.

Нетрудно заметить, что частным случаем функции $U(x) = u(x_0) + \sum_{i=1}^{n} u_i(x_i - x_0)$ является функция $U_i(x) = x_0 - \frac{\alpha}{n}\sum_{i=1}^{n} \max(x_i - x_0, 0) -$

[37] Neilson W. An Axiomatic Characterization of the Fehr-Schmidt Model of Inequity Aversion. Department of Economics, Texas A&M University, 2000 (Mimeo).

[38] Fehr E., Schmidt K. Theories of Fairness and Reciprocity — Evidence and Economic Applications. Institute for Empirical Research in Economics. University of Zurich, 2001. Working paper No. 75.

[39] Kahneman D., Tversky A. Prospect Theory: An Analysis of Decision Under Risk // Econometrica. 1979. No. 47. P. 263—291.

$-\frac{\beta}{n}\sum_{i=1}^{n}\max(x_i - x_0, 0)$, $0 \le \beta \le \alpha$, которую используют Фер и Шмидт в своей модели неприязни к неравенству[40].

Несколько условно среди аксиоматических подходов к моделированию чувства справедливости можно упомянуть модель, предложенную Оком и Кокесеном[41]. 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 моделирования щедрости в добровольном финансировании общественных благ или благотворительности.

Последняя работа, которая будет рассмотрена в данном разделе, выделяется из общего ряда специфическим объектом, по отношению к которому моделируется справедливость. Э. Карни и Ц. Сафра[42] предлагают аксиоматическую модель чувства справедливости, разработанную для механизмов случайного распределения дискретного блага между несколькими людьми, претендующими на него. То, что в поле рассмотрения авторов попадают лишь лотереи, безусловно, сужает возможный круг применения модели, однако лотереи достаточно часто используются при необходимости распределить неделимые блага. Авторы ссылаются на практику использования жребия при выборе года рождения мужчин, которые должны были отправляться на военную службу во Вьетнам, а также известный в США судебный прецедент "Соединенные Штаты Америки против Холмса", когда моряка по имени Холмс, в 1841 г. выбрасывавшего за борт пассажиров перегруженной спасательной шлюпки, осудили за то, что он сам выбирал, кого выбрасывать. В вердикте судья отметил, что поскольку права пассажиров на жизнь были абсолютно равны, выбирать из них того, кто должен был погибнуть, можно было лишь с помощью жребия.

[40] Тему аксиоматического моделирования неэгоистических предпочтений также развивали М. Сандбю (Sandbu M. Axiomatic Foundations for Reference Dependent Distributive Preferences. Harvard University, 2003 (Manuscript)), а также Нилсон и Стоу (Neilson W., Stowe J. Choquet Other-Regarding Preferences. Texas A&M University, 2004 (Manuscript)). В предложенных ими моделях при принятии индивидом решения имеет значение вся иерархия платежей, а не только личное место индивида в этой иерархии.

[41] Ok E., Kockesen L. Negatively Interdependent Preferences. Social Choice and Welfare. 2000. No. 17. P. 533—558.

[42] Karni E., Safra Z. Individual Sense of Justice: A Utility Representation // Econometrica. 2002. No. 70. P. 1; ABI/INFORM Global. P. 263.

Модель имеет следующий вид: в обществе, состоящем из N индивидов, необходимо распределить 1 единицу неделимого блага между $n \le N$ людьми, имеющими на него право ($n \ge 3$). Пусть e^i, единичный вектор, принадлежащий R^n, обозначает то распределение ex post, в котором благо получает индивид i. Пусть $X = \{ e^i | 1 \le i \le n\}$ — множество ex post распределений, а $P = \Delta(X) - n$-1-мерный симплекс, соответствующий множеству всех вероятностных распределений на X. С экономической точки зрения P соответствует множеству способов распределения неделимого блага. P наследует топологию R^{n-1}.

Каждый индивид обладает двумя полными, транзитивными и непрерывными наборами предпочтений на P: отношение $\succ$ соответствует традиционным предпочтениям, а отношение $\succ_F$ характеризует мнение индивида об относительной справедливости различных процедур. Практически отношение $\succ_F$ отражает мнение индивида о распределительных процедурах для групп, в которые не входит он сам. Таким образом, чувство справедливости моделируется как самостоятельная структура предпочтений, параллельная "полезностным" предпочтениям.

Для иллюстрации работы механизма сравнения лотерей по двум отношениям предпочтения рассмотрим пример, который приводят сами Карни и Сафра в начале статьи. Три индивида равно нуждаются в почке для трансплантации. Изобразим множество случайных механизмов распределения почки между ними в качестве треугольника, вершинам которого соответствуют лотереи, в которых один из трех индивидов получает почку с вероятностью "1" (рис. 1). Рассмотрим второго индивида. Если он сам не получит почку, ему безразлично, кто из остальных претендентов получит ее: для заданной вероятности p_2^0 получить почку самому процедуры $p = (\alpha p_2^0 (1 - \alpha - p_2^0))$ и $p' = ((1 - \alpha - p_2^0), p_2^0, \alpha)$ для него эквивалентны и (поскольку он не ценит благо ни одного из оставшихся больных выше, чем благо другого) эквивалентно справедливы. Однако процедуру p'', находящуюся ближе к равному распределению шансов между больными чем p, он признает более справедливой. Далее, точку, принадлежащую середине отрезка $p_2 = p_2^0$, он сочтет самой справедливой из всех процедур распределения, принадлежащих этому отрезку. Поскольку в модели постулируется непрерывность обоих предпочтений, индивид будет готов немного пожертвовать собственной вероятностью выигрыша в пользу более честной игры. Поэтому его кривая безразличия, проходящая через p и p' и отражающая как собственный интерес, так и соображения справедливости, будет симметрична относительно вертикальной линии, проходящей через $\hat{p}$, и выпукла вниз. Заметим также, что его традиционная, "сугубо эгоистическая" кривая безразличия — прямая линия, соответствующая личной вероятности выиграть p_2.

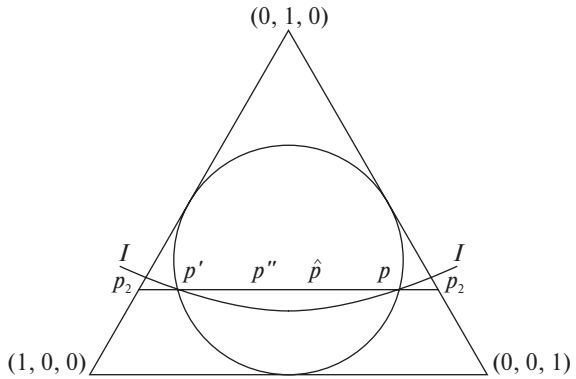

Рис. 1. Предпочтения на множестве случайных механизмов распределения неделимого блага между тремя претендентами

Для каждого из наборов предпочтений затем определяется собственный компонент в целевой функции индивида. Карни и Сафра разрабатывают аксиоматическую модель индивидуального выбора между процедурами распределения дискретного блага, где целевая функция, определяющая выбор, представима 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 оказывать влияние 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 разрабатывают формальную количественную меру интенсивности чувства справедливости и анализируют влияние интенсивности этого чувства на поведение[43]. Мера определяется для индивидов, имеющих

[43] Karni E., Safra Z. Intensity of the Sense of Fairness: Measurement and Behavioral Characterization // Journal of Economic Theory. 2002. No. 105. P. 318—338.

одинаковое понятие справедливости и ординально сравнимые[44] предпочтения в эквивалентных условиях выбора. Суть, коротко говоря, состоит в следующем — индивид, наделенный более сильным чувством справедливости, демонстрирует бóльшую готовность пожертвовать собственным интересом для реализации более справедливой процедуры распределения.

Моделирование нравственного самосовершенствования и принятия решений, влияющих на образ жизни

Экономические модели чувства справедливости, рассмотренные в первых двух разделах, оперировали только предпочтениями. Их субъекты, следовательно, принимали решения, сообразуясь с тем, насколько им “нравится” результирующее распределение ресурсов или механизм их распределения. Эти модели достаточно хорошо объясняют ряд экспериментальных результатов и могут быть применены для анализа широкого круга экономических проблем, перечисленных выше. Однако их применение для анализа моральных норм, с акцентом на последнем слове, по ряду причин является проблематичным.

Как и любая норма, моральная норма может быть нарушена, что влечет за собой возможные санкции, внешние (порицание, бойкот) и, в первую очередь, внутренние (угрызения совести). В рассмотренных нами моделях чувства справедливости мораль (точнее, представление о справедливости) играет роль одного из критериев оценки альтернатив, который можно “нарушить” только при недостатке информации или ограниченных когнитивных возможностях. Поскольку для удобства анализа критерии справедливости в целевых функциях непрерывны, возникает и такой вопрос: что трактовать как “нарушение”? И если включить в модели справедливости трактовку “нарушения” и внешние санкции не так трудно, то моделирование внутренних санкций обещает оказаться более проблематичным.

Модели, которые будут рассмотрены в данном разделе, описывают стратегические ситуации, в которых индивид принимает решения с широкими и потенциально долгосрочными последствиями, например, выбор образа жизни. С формальной точки зрения они выделяются тем, что вводят зависимость между предпочтениями и ограничениями. С точки зрения морали такая связь вполне реалистична — удовольствие от сознания своей правдивости одновременно ограничивает нас в возможностях, например, лгать жене или начальству. Выбор, который осуществляют рациональные субъ-

[44] В том смысле, что если для одного индивида процедура A предпочтительнее B по предпочтениям полезности и справедливости, то и для другого индивида это будет так.

екты обеих моделей, можно в общем охарактеризовать как принятие добровольного морального обязательства, интернализацию моральной нормы или отказ от нее.

В некотором смысле нравственные принципы являются разновидностью неформальных норм. Это дает повод думать, что функционирование нравственных правил связано с механизмом интернализации нормы, т.е. перехода ее из внешнего ограничения во внутреннее. Как замечает известный исследователь в области экономического анализа права Р. Кутер[45], в условиях постоянно усложняющейся организации общества, процессы взаимодействия в котором должны регулироваться законом, децентрализация закона выглядит все более привлекательно, по крайней мере, с точки зрения экономиста. Децентрализованный же закон лучше функционирует в ситуации, когда принуждение со стороны государства дополняется спонтанным соблюдением закона и частным правоприменением. И то, и другое работает более эффективно в условиях массовой интернализации моральных норм. Кутер задается вопросом, каким образом закон может стимулировать людей к интернализации норм.

По его мнению закон способен воздействовать на разум индивидов, поощряя их — в чисто эгоистических целях — развивать самоконтроль и прибегать к нравственному самосовершенствованию. Проиллюстрируем эту мысль с помощью его собственного удачного примера: зная, что родители не доверяют врунишкам и ограничивают их свободу, ребенок хочет выглядеть честным. Но врать, сохраняя правдоподобность, ребенку трудно, поэтому лучший способ выглядеть честным для него — это *стать* честным. В конечном счете в некоторых случаях происходит именно это, после чего ребенку предоставляют бóльшую свободу.

Итак, самоконтроль вводится как возможность рациональных индивидов менять свои предпочтения так, чтобы это создавало новые альтернативы, которые превосходили бы старые и по исходным, и по измененным предпочтениям. Для иллюстрации этой идеи Кутер пользуется следующей простой двухпериодной моделью.

В первом периоде индивид выбирает, будет ли он нарушать норму или нет. Нарушение дает выгоду b_1 в первом периоде и издержки c_2 во втором периоде, которые индивид дисконтирует по ставке r, учитывающей его межвременную норму замещения и оцениваемую им вероятность наказания. Следовательно, рациональный индивид нарушит норму, если $b_1 + \frac{c_2}{r} > 0$, и не нарушит ее в ином случае. Этим задается граничная

[45] Cooter R. Models of Morality in Law and Economics: Self-Control and Self-Improvement for the "Bad Man" of Holmes. Berkeley Program in Law & Economics, 1998. Working Paper No. 135 (http://repositories.cdlib.org/blewp/art135).

величина дисконт-фактора $r^* = \frac{c_2}{b_1}$, при которой индивиду безразлично, нарушать ли норму или соблюдать ее. Если $r > r^*$, индивид будет нарушать норму. Если $r < r^*$, он будет соблюдать норму. Предположим, что индивидуальная величина r для каждого человека изменяется во времени и представляет собой случайную величину с распределением $f(r)$. Чем выше ее дисперсия (которую Кутер напрямую связывает с переменчивостью настроения индивида и ставит в обратную зависимость от способности индивида к самоконтролю), тем больше вероятность, что r случайно превысит r^*, и индивид спонтанно нарушит норму[46].

Поскольку r меняется во времени, индивид может пожалеть о выборе, сделанном ранее (рис. 2).

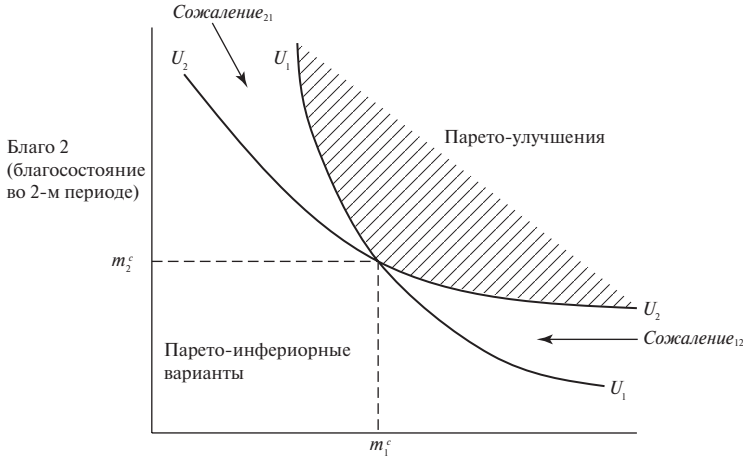

Рис. 2. Кривые безразличия для различных значений r

На рис. 2 изображены кривые безразличия одного и того же индивида в разные моменты времени, для двух разных значений r. Предположим, U_1 — кривая безразличия молодого индивида, который не слишком задумывается о будущем, а U_2 — его же кривая безразличия в более зрелом возрасте. В молодости набор, принадлежащий области "$Сожаление_{12}$", показался бы индивиду предпочтительным, но затем, с позиции кривой безразличия U_2, он сам бы сожалел о своем выборе. Если же предпочтения

[46] Дисперсию $f(r)$ можно считать и степенью, в которой ограничена рациональность индивида.

меняются постоянно, индивид гипотетически может попасть в "цикл сожалений": в пятницу, имея предпочтения U_1, я нарушаю норму, о чем сожалею в субботу с точки зрения предпочтений U_2, но в воскресенье мои предпочтения вновь меняются на U_1, и я сожалею о том, что в субботу никаких норм не нарушал.

Заштрихованная область, находящаяся над всеми возможными кривыми безразличия, проходящими через точку текущего выбора индивида (m_1^c, m_2^c) соответствует области Парето-улучшений, так как по отношению к точкам этой области индивид не может испытывать разочарований.

Очевидно, что с уменьшением дисперсии r сужается множество вариантов выбора, о которых индивид впоследствии может пожалеть, и расширяется множество Парето-улучшений по отношению к точке текущего выбора. Кутер отмечает, что настроение молодых людей в целом более переменчиво и поэтому молодежь в большей степени склонна к спонтанному нарушению норм. Семья, школа и другие институты социализации помогают молодым людям развивать способность к самоконтролю, снижая ожидаемую частоту поступков, о которых они могли бы пожалеть. Максимизируя свою ожидаемую полезность на некотором временном горизонте, индивид может сознательно вложить некоторое количество ресурсов в то, чтобы снизить волатильность собственных предпочтений r. Такова идея моделирования сознательного самоконтроля, предложенная в рассматриваемой здесь статье.

Заметим, что для простоты изложения в самой модели и во всех иллюстрациях используется межвременной выбор — индивид выбирает комбинацию дохода в настоящем и будущем периоде, оперируя параметрами распределения r — ставки дисконтирования, учитывающей норму межвременного замещения и субъективную вероятность наказания. Однако это далеко не единственный вариант применения модели. Случайный компонент предпочтений, параметрами распределения которого манипулирует индивид, может, например, характеризовать его трудовую или профессиональную этику, отношение к собственному потомству и близким, а также, возможно, интенсивность чувства справедливости, степень неприятия к неравенству, склонность к реципрокному поведению и другие параметры многих моделей морали, которые рассматриваются в настоящей работе.

В качестве следующего шага Кутер демонстрирует, каким образом нормы способны стимулировать людей к нравственному самосовершенствованию. Как и обсуждавшийся выше самоконтроль, самосовершенствование моделируется как сознательное изменение параметров $f(r)$ — с той разницей, что теперь изменения затрагивают и математическое ожидание этой величины. Стимулом для самосовершенствования на этот раз служит бюджетное ограничение, которое меняется в зависимости от параметров $f(r)$.

Так, если предположить, что другие люди могут наблюдать наш характер (не в точности, но с некоторой погрешностью), то распознавая в нас надежных, честных людей, они могут предложить нам сделки, которые были бы невозможны с партнером-оппортунистом. Причем иногда самым дешевым способом произвести впечатление "честного малого" оказывается стать им.

На рис. 3 изображены кривые безразличия и соответствующие им бюджетные ограничения для двух разных значений r. "Честный" оптимум (принадлежит области между кривыми U_1 и F_2) лежит в области Парето-улучшений по отношению к исходному оптимуму (точке касания кривых U_1 и F_1).

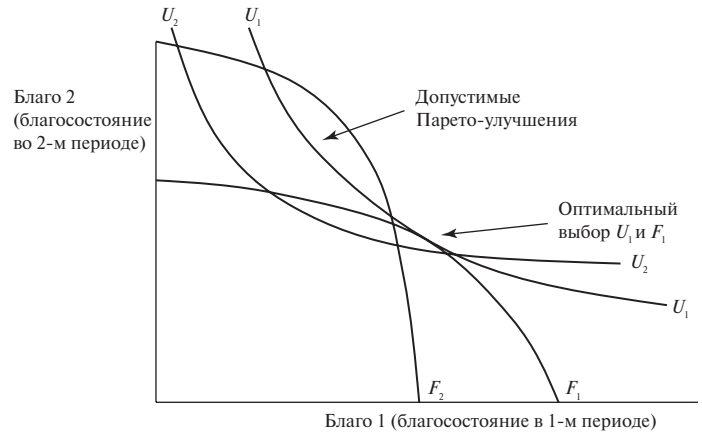

Рис. 3. Кривые безразличия и бюджетные ограничения для различных значений r

В качестве примера такого самосовершенствования Кутер приводит ситуацию с наемным работником, у которого есть выбор вступить в секту, считающую работу добродетелью, а праздность — грехом. Для работодателя такой работник представляет большую ценность, и он согласен платить ему больше. Таким образом, наемному работнику может быть выгодно искренне "полюбить" работу и деньги, даже если изначально это ему не свойственно.

Идеологически сходная модель была предложена Р. Дауэллом, Р. Гольдфарбом и У. Гриффитом[47] в аналитическом виде и более общей формули-

[47] Dowell R., Goldfarb R., Griffith W. Economic man as a moral individual // Economic Inquiry. 1998. No. 36. P. 4; ABI/INFORM Global. P. 645.

ровке. Они акцентируют свое внимание на дискретной природе некоторых принимаемых людьми решений, имеющих нравственную подоплеку, и предлагают метод моделирования подобных решений. Классическим примером такого дискретного решения служит выбор индивидом рода своих занятий. Представим себе двух молодых людей, сходных во всех отношениях, уроженцев одного и того же города, жители которого страдают от жестокой безработицы. Перед ними стоит выбор — искать малопрестижную и плохо оплачиваемую легальную работу, скажем, на автомойке, или заняться наркоторговлей. Очевидно, что эти варианты значительно различаются с точки зрения моральной оправданности, причем промежуточные решения между наркоторговлей и работой на мойке найти трудно: для честного человека торговля наркотиками по выходным или каждый день равно предосудительна. Помимо бюджетного ограничения, с выбором в пользу одного из вариантов меняются и жизненные установки индивида. Эту особенность и пытаются формализовать авторы модели.

В их работе перед индивидом ставится задача максимизации функции полезности, определенной на множестве нормальных благ X и, помимо этого, зависящей от бинарной переменной "честности".

Индивидуальная функция полезности имеет вид $U = U(X_1, X_2, ..., X_n, H)$, где $X_1, ..., X_n$ — нормальные блага, а H — бинарная (принимающая значения 1 или 0) переменная, отображающая "честность" индивида. Авторы предполагают, что при прочих равных обстоятельствах индивид скорее предпочтет поступать честно, поскольку чувство вины и угрызения совести мешают ему получать удовольствие от потребления обычных благ[48]. Здесь важно отметить, что собственно ординалистские предпочтения на множестве обычных благ в модели не меняются — индивид полностью сохраняет исходную структуру своих предпочтений. Его бюджетное ограничение выглядит следующим образом:

$W(H) = P_1X_1 + P_2X_2 + ... + P_nX_n$, где $W(1) \le W(0)$.

То есть, поступаясь нравственными соображениями, индивид получает бóльшую гибкость в выборе своего поведения, и расширяет собственные финансовые возможности.

Индивид захочет вести себя честно ($H = 1$), если:

$U(X_{1*}, X_{2*}, ..., X_{n*}, 1) > U(X_{1a}, X_{2a}, ..., X_{na}, 0)$

при: $P_1X_{1*} + P_2X_{2*} + ... + P_nX_{n*} = W(1) \le P_1X_{1a} + P_2X_{2a} + ... + P_nX_{na} = W(0)$,

где X_{i*} обозначает оптимальное потребление блага i при цене P_i и размере благосостояния $W(1)$, а X_{ia} соответственно — оптимальное потребление

[48] Доуэлл, Гольдфарб и Гриффит отмечают формальное сходство между собственной идеей эндогенного изменения предпочтений на множестве обычных благ и аналогичным изменением предпочтений в функциях полезности, зависимых от различных "состояний мира", применяемых в экономике неопределенности.

блага i при цене P_i и размере благосостояния $W(0)$. Очевидно, что поскольку в рамках модели сама структура предпочтений на множестве обычных благ не зависит от H, любые изменения в структуре потребления, которые мы могли бы наблюдать, будут связаны с эффектом дохода.

Описанная формализация выбора подчеркивает дискретный характер принимаемых нравственных решений: индивид либо занимается наркобизнесом, либо нет — нельзя быть "немножко наркодилером". Очевидно, такая постановка проблемы в большей степени подходит для единовременных решений, имеющих фундаментальные последствия либо для системы ценностей индивида, либо для его профессиональной или общественной репутации. Это представляет собой известное неудобство, сужая потенциальное поле практического применения рассматриваемой модели.

Авторы предлагают три возможных варианта ее расширения. Во-первых, можно предположить, что только серьезная нечестность отрицательно влияет на полезность от потребления обычных благ. В таком случае функция полезности принимает следующий вид:

$U = A(I)f(X_1, X_2, ..., X_n)$,

где i — индекс нечестности, возрастающий по мере того, как растет нечестность поведения ($I = 0$ соответствует безупречной честности), а A — убывающая функция, сдвигающая функцию полезности. Эта функция может иметь различный вид в зависимости от того, о каких прегрешениях идет речь и насколько серьезно сам индивид к ним относится.

Во-вторых, возможно, функция A меняется в зависимости от ситуации, в которой находится индивид, например, зависит от количества свидетелей его неблаговидного поступка. Так, многие люди предпочитают не показывать дурной пример собственным детям, хотя в их отсутствие вели бы себя иначе.

В-третьих, возможно, честность влияет не на уровень дохода или благосостояния, а на цены некоторых товаров или количество свободного времени. Например, решение индивида заботиться о пожилых родственниках (от которого, единожды приняв, как правило, не отказываются), часто требует от него прежде всего дополнительных временных, нежели финансовых или иных затрат.

Модели Кутера и Доуэлла — Голдфарба — Гриффита являются скорее иллюстрациями того, как можно подойти к анализу выбранных авторами проблем, нежели готовыми инструментами для экономического анализа или выработки эмпирически проверяемых прогнозов. В качестве идей для разработки собственных теорий и моделей они потенциально представляют интерес для экономистов, занимающихся исследованиями на стыке экономики и морали. С формальной точки зрения к их достоинствам можно отнести простоту, позволяющую самостоятельно доработать специфика-

цию моделей для анализа конкретной проблемы, а также относительную легкость, с которой они могут быть интегрированы в общий аналитический каркас неоклассической микроэкономической теории (в сравнении, например, с идеями множественных предпочтений, предложенными Сеном[49] и Этциони[50]).

Представляется, что модификация предпочтений индивида (следствием чего является изменение его ресурсных ограничений) может быть вызвана прежде всего решениями, серьезно влияющими на его образ жизни. Одним из наиболее очевидных примеров таких решений, к которому прибегают сами авторы обеих моделей, является выбор профессии. В этой связи полем их практического применения и развития мог бы стать экономический анализ таких человеческих решений, как выбор карьеры врача, решение заняться сбытом наркотиков или проституцией или — если отойти от темы профессионального развития, — усыновление ребенка, супружеская измена, пособничество террористам, вступление в религиозную общину и т.п.

Моделирование работы нравственных чувств как инструмента контроля за производством экстерналий

В данном разделе будет рассмотрена единственная модель, существенно отличающаяся от тех, что описывались выше. До этого момента общей целью рассматриваемых моделей был анализ и объяснение поведения отдельных людей, обусловленного различными нравственными соображениями. Работа Л. Каплоу и С. Шэйвелла[51] теоретически описывает построение и функционирование системы моральных норм как инструмента контроля за производством экстерналий. В этом смысле она дополняет рассмотренные нами выше модели самоконтроля Кутера и выбора образа жизни Доуэлла — Голдфарба — Гриффита.

Следует отметить, что роль морали как значимого фактора в решении проблемы контроля за производством экстерналий неоднократно освещалась и моделировалась в литературе, близкой к экономическому анализу

[49] Sen A. Choice, Orderings and Morality // Practical Reason / Ed. by S. Korner. New Haven: Yale University Press, 1974. P. 54–67; Sen A. Rational Fools: A Critique of the Behavioral Foundations of Economic Theory // Philosophy and Public Affairs. 1977. No. 6. P. 317—344.

[50] Etzioni A. The Case for a Multiple-Utility Conception // Economics and Philosophy. 1986. P. 159—183; Etzioni A. The Moral Dimension: Toward a New Economics. N.Y.: Macmillan, 1988.

[51] Kaplow L., Shavell S. Moral Rules and Moral Sentiments: Toward a Theory of an Optimal Moral System. 2001. NBER Working Paper 8688.

права. Например, в одной из самых ранних работ на подобную тему Лаффон[52] (о чем еще будет упомянуто в следующем разделе) предположил, что моральное правило, аналогичное кантианскому категорическому императиву, может использоваться для контроля за производством отрицательных экстерналий в больших сообществах. Ф. Карри и С. Монгрэйн[53] анализируют проблему оптимизации производства специфических экстерналий, порождаемых деятельностью, связанной с общественной стигмой.

Каплоу и Шэйвелл ставят акцент прежде всего на механизме принуждения к выполнению моральных норм — нравственных чувствах, — и предлагают, на наш взгляд, чрезвычайно интересную его формализацию. В их модели абстрактный регулятор (“общество”) сталкивается с достаточно сложной задачей: представляя общественные интересы, он должен при некоторых специфических ограничениях, характеризующих влияние нравственных чувств на мотивацию людей, сформулировать моральные нормы по отношению к ряду поступков, порождающих положительные или отрицательные внешние эффекты (практически, ассоциировать некоторые уровни вины и добродетельности с некоторыми классами этих поступков), таким образом, чтобы максимизировать благосостояние общества.

Предлагаемая авторами модель выглядит следующим образом. Пусть S — множество возможных ситуаций, в которых может оказаться индивид. В каждой ситуации он может совершить некий поступок или воздержаться от него (например, солгать или нет). Совершая поступок, индивид получает некоторую *непосредственную* полезность u, положительную или отрицательную, и причиняет внешний ущерб $h \geq 0$. Воздержавшись, он не получает полезности и не причиняет внешнего ущерба. Таким образом, ситуация характеризуется парой (u, h), а плотность распределения возможных ситуаций по значениям этих параметров задается функцией $f(u, h)$. Предполагается, что эта функция непрерывна, u принадлежит $(-\infty, \infty)$ а h принадлежит $[0, \infty)$. Наилучшим решением задачи максимизации общественного благосостояния было бы совершать любой поступок тогда и только тогда, когда $u > h$.

Общество может ассоциировать вину $g(u, h) \geq 0$ с совершением поступка в конкретной ситуации (u, h). Совершивший такой поступок будет получать отрицательную полезность $-g(u, h)$. Аналогичным образом общество способно ассоциировать добродетельность $v(u, h) \geq 0$ с воздержанием

[52] Laffont J.-J. Macroeconomic Constraints, Economic Efficiency and Ethics: An Introduction to Kantian Economics // Economica. New Series. 1975. Vol. 42. No. 168. P. 430—437.

[53] Curry Ph., Mongrain S. What you don’t see can’t hurt you: an economic analysis of morality laws. American Law & Economics Association Annual Meetings, 2004. Paper 48.

от поступка в ситуации (u, h). Воздержавшийся от поступка в ситуации (u, h) индивид получает полезность $v(u, h)$.

Индивид совершит поступок тогда и только тогда, когда $u - g(u, h) > v(u, h)$, или $u > g(u, h) + v(u, h)$, что означает превышение непосредственной полезностью общего веса нравственных санкций и вознаграждений.

Внушение вины и добродетельности связано с растущими предельными издержками $\alpha(g(u, h))$, где $\alpha(0) = 0$ и, для $g > 0$, $\alpha'(g) > 0$ и $\alpha''(g) \geq 0$ и, соответственно, $\beta(v(u, h))$, где $\beta(0) = 0$ и, для $v > 0$, $\beta'(v) > 0$ и $\beta''(v) \geq 0$. Эти издержки определены для каждого класса ситуаций, причем классы могут быть как широкими, так и очень узкими (каждая мыслимая ситуация — отдельный класс).

В модели вполне реалистично предполагается, что фактическая возможность ощущать вину ограничена: ожидаемые уровни ощущаемой вины и добродетельности не могут превышать $G \geq 0$ и $V \geq 0$ соответственно.

Авторы рассматривают четыре случая, отличающихся степенью специфицированности моральных правил (правила могут определяться с точностью до каждой мыслимой ситуации или с точностью до группы схожих ситуаций) и использованием нравственных чувств (исключительно вина или вина в сочетании с добродетельностью). Ниже будут рассмотрены только два наиболее интересных и показательных случая.

В случае с абсолютно специфицированными моральными правилами, подкрепляемыми как виной, так и добродетельностью, каждой ситуации (u, h) может быть независимо присвоен уровень вины $g(u, h)$ для совершения поступка, и уровень добродетельности $v(u, h)$ для воздержания от поступка, с соответствующими издержками внушения. Обозначим за A множество *совершаемых* поступков, т.е.

$$A = \{(u, h) \in S \mid u > g(u, h) + v(u, h)\},$$

и обозначим за N множество инцидентов *воздержания*, т.е.

$$N = \{(u, h) \in S \mid u - g(u, h) \leq v(u, h)\}.$$

Проблема максимизации общественного благосостояния состоит в выборе функций $g(u, h) \geq 0$ и $v(u, h) \geq 0$, максимизирующих общественное благосостояние с учетом нравственной составляющей

$$\iint_A \big(u - h - g(u,h)\big) f(u,h)\,du\,dh + \iint_N v(u,h) f(u,h)\,du\,dh -$$
$$- \iint_S \big(\alpha(g(u,h)) + \beta(v(u,h))\big) f(u,h)\,du\,dh,$$

при ограничениях

$$\iint_A g(u,h) f(u,h)\,du\,dh \leq G, \text{ и}$$

$$\iint_N v(u,h) f(u,h)\,du\,dh \leq V.$$

Первый член в максимизируемом выражении отражает последствия для общественного благосостояния от совершения поступков — непосредственную полезность и вину, испытываемые индивидами, и причиняемый внешний ущерб — взвешенные по частоте ситуаций. Второй член отражает влияние на общественное благосостояние добродетельности, которую ощущают индивиды, воздержавшиеся от поступков, взвешенное по частоте этих ситуаций. Третий член соответствует издержкам внушения.

Лагранжиан вышеописанной задачи выглядит следующим образом:

$$\iint_A \big(u - h - g(u,h)\big) f(u,h)\,du\,dh + \iint_N v(u,h) f(u,h)\,du\,dh -$$
$$- \iint_S \big(\alpha(g(u,h)) + \beta(v(u,h))\big) f(u,h)\,du\,dh -$$
$$- \lambda\left[\iint_A g(u,h) f(u,h)\,du\,dh - G\right] - \mu\left[\iint_N v(u,h) f(u,h)\,du\,dh - V\right],$$

где λ — множитель Лагранжа для ограничения на переживание вины, а μ — множитель Лагранжа для ограничения на переживание добродетельности.

Вина и добродетельность различаются по тому, как они переживаются людьми при их оптимальном использовании. Авторы показывают, что когда вина оптимально используется в качестве стимула, она не оказывает непосредственного влияния на полезность, так как все нежелательные поступки предотвращаются. Напротив, когда добродетельность успешно контролирует поведение людей, они фактически *ощущают* ее. Это одновременно преимущество и недостаток — можно внушать добродетельность исключительно для непосредственного создания полезности, но поскольку при успешном использовании добродетельность фактически ощущается людьми, ее использование сокращает запас добродетельности, доступный для контролирования прочих поступков, в то время как успешно используемая в качестве стимула вина людьми не переживается и, следовательно, доступного запаса вины не истощает.

Пусть $g^*(u, h)$ как и раньше обозначает оптимальную $g(u, h)$, а $v^*(u, h)$ обозначает оптимальную $v(u, h)$. Каплоу и Шэйвелл доказывают, что если в оптимуме для поступка (u, h) $\beta'(0) > (1 - \mu)f(u, h)$ (условие, делающее невозможным использование добродетельности кроме как для коррекции поведения), то для этого поступка:

a) положительный уровень вины или добродетельности внушается только если воздержание от поступка является первым наилучшим, и если вина или добродетельность внушаются, суммарный вес вины и добродетельности равен минимально необходимому для предотвращения поступка; т.е. если $g^*(u, h) > 0$ *или* $v^*(u, h) > 0$, *то* $u < h$ *и* $g^*(u, h) + v^*(u, h) = u$;

b) люди фактически никогда не ощущают вины и всегда ощущают добродетельность, если $v^*(u, h) > 0$ *и возникает ситуация* (u, h);

c) единственно возможным отклонением от первого наилучшего поведения является совершение нежелательных поступков;

d) в ситуациях с $u > 0$, *вину или добродетельность оптимально внушать тогда и только тогда, когда при* $g(u, h)$ *и* $v(u, h)$, *максимизирующих общественное благосостояние с учетом нравственной составляющей при условии* $g(u, h) + v(u, h) = u$,

$$\alpha(g(u,h)) + \beta(v(u,h)) < (h + (1-\mu)v(u,h) - u) f(u,h);$$

e) в общем случае (без ограничений на значения β *или значения* μ *в точке оптимума) ни один из вышеупомянутых пунктов не соблюдается с необходимостью, за исключением* (b) *— так как добродетельность может употребляться вне связи с экстерналиями, просто для повышения общественного благосостояния.*

В случае с общими моральными правилами (определенными с точностью до группы схожих поступков), вина и добродетельность могут быть независимо внушены только для n подмножеств поступков S_i, на которые делится множество ситуаций S. Обозначим как g_i и v_i единые уровни вины и добродетельности для поступков, принадлежащих S_i.

С учетом этих предпосылок задача общества — выбрать $g_i \geq 0$ и $v_i \geq 0$ для каждого подмножества S_i так, чтобы максимизировать общественное благосостояние с учетом ограничений на ощущение вины и добродетельности. Пусть $f_i(u, h)$ обозначает условную плотность ситуаций (u, h) на S_i, и пусть p_i обозначает вероятность того, что ситуация принадлежит S_i. Пусть $\alpha_i(g_i)$ и $\beta_i(v_i)$ обозначают издержки внушения вины g_i и добродетельности v_i для поступков, принадлежащих подмножеству S_i, и предположим, что производные $\alpha_i(g_i)$ и $\beta_i(v_i)$ имеют те же свойства, что и производные $\alpha(g)$ и $\beta(v)$. Тогда общественное благосостояние с учетом нравственной составляющей равно:

$$\sum_{i=1}^{n} W_i(g_i, v_i),$$

где $W_i(g_i, v_i) = p_i \left[\int_0^{\infty} \int_{g_i+v_i}^{\infty} (u - h - g_i) f_i(u,h)\,du\,dh + \int_0^{\infty} \int_0^{g_i+v_i} v_i f_i(u,h)\,du\,dh \right] -$

$$- \alpha_i(g_i) - \beta_i(g_i).$$

Ограничения на фактическое ощущение вины и добродетельности выглядят следующим образом:

$$\sum_{i=1}^{n} y_i(g_i, v_i) \leq G, \text{ и}$$

$$\sum_{i=1}^{n} z_i(g_i, v_i) \leq V,$$

где $y_i(g_i, v_i) = p_i \int_0^{\infty} \int_{g_i+v_i}^{\infty} g_i f_i(u,h)\,du\,dh = p_i g_i \left(1 - F_i(g_i + v_i)\right)$,

$$z_i(g_i, v_i) = p_i \int_0^{\infty} \int_0^{g_i+v_i} v_i f_i(u,h)\,du\,dh = p_i v_i F_i(g_i + v_i).$$

Лагранжиан задачи максимизации общественного благосостояния с учетом ограничений выглядит следующим образом:

$$\sum_{i=1}^{n} W_i(g_i, v_i) - \lambda \left[\sum_{i=1}^{n} y_i(g_i, v_i) - G \right] - \mu \left[\sum_{i=1}^{n} z_i(g_i, v_i) - V \right].$$

Условие первого порядка при $g_i^* > 0$ выглядит так:

$$p_i \left[\int_0^{\infty} \left(h + \lambda g_i - \mu v_i\right) f_i(g_i + v_i, h)\,dh - (1+\lambda)\left(1 - F_i(g_i + v_i)\right) \right] = \alpha_i'(g_i),$$

а условие первого порядка при $v_i^* > 0$ выглядит так:

$$p_i \left[\int_0^{\infty} \left(h + \lambda g_i - \mu v_i\right) f_i(g_i + v_i, h)\,dh + (1-\mu) F_i(g_i + v_i) \right] = \beta_i'(g_i).$$

Таким образом, вина и добродетельность порождают эффекты двух типов. Это маржинальные эффекты, состоящие из сокращения экстерналии и выгод или издержек, связанных с ограничениями (когда индивид воздерживается от совершения поступка, возникает выгода, связанная с тем, что теперь ощущается меньше вины, и издержки, связанные с тем, что ощущается больше добродетельности). Другая разновидность — инфрамаржинальные эффекты, относящиеся к тем людям, чье поведение не меняется; продолжающие совершать поступок ощущают бóльшую вину, а продолжающие воздерживаться — бóльшую добродетельность (для обоих условий первого порядка непосредственные влияния на полезность "маржинальных" индивидов равны нулю, поскольку для маржинальных индивидов $u = g_i + v_i$). Сумма этих двух эффектов, маржинального (или

эффекта сдерживания) и инфрамаржинального, приравниваются к непосредственным предельным издержкам внушения более высокого уровня вины или добродетельности.

Предельная выгода использования вины и добродетельности (первые члены в левой части условий первого порядка для v_i^* и g_i^*) одинаковы: в качестве сдерживающих стимулов вина и добродетельность полностью взаимозаменяемы. Предельные издержки внушения (правая часть условий первого порядка для v_i^* и g_i^*) симметричны, что склоняет к использованию той моральной санкции/вознаграждения, которая обладает меньшими предельными издержками внушения.

Однако если принять во внимание инфрамаржинальные эффекты (вторые члены в левой части условий первого порядка для v_i^* и g_i^*), качественная разница возникает. В том случае, который авторы берут за основу для сравнения (когда в оптимуме ограничение на использование добродетели носит жесткий характер и $\mu > 1$), оба вторых члена отрицательны, что свидетельствует о том, что большее фактическое использование и вины, и добродетели связано с издержками. Одно различие состоит в издержках на единицу используемой моральной санкции/вознаграждения, которые составляют $1 + \lambda$ для вины и $\mu - 1$ для добродетельности. Другое различие заключается в том, какой объем моральной санкции/вознаграждения реально используется: доля реально используемой вины составляет $1 - F_i(g_i + v_i)$, а добродетельности — $F_i(g_i + v_i)$. Таким образом, когда большинство индивидов воздержатся от совершения поступков, принадлежащих S_i, т.е. F_i будет велико, будет реально использоваться очень мало вины и значительное количество добродетельности (и то, и другое — в расчете на единицу внушаемой моральной санкции/вознаграждения). Соответственно, когда большинство поступков будет предотвращено, при прочих равных будет скорее оптимально использовать вину, а не добродетельность. Аналогично, когда лишь немногие индивиды воздержатся от совершения поступков, принадлежащих S_i и, следовательно, F_i будет мало́, — использовать добродетельность, а не вину[54]. И поскольку эффект от увеличения g_i или v_i для инфрамаржинальных издержек может быть большим, даже если первоначально $g_i = 0$ или $v_i = 0$, вполне может оказаться оптимальным опираться исключительно на вину в первом случае, и исключительно на добродетельность — во втором.

Это наиболее значительный из всех относительно неочевидных выводов, извлекаемых из модели, причем он согласуется с наблюдаемым использованием вины и добродетельности. С одной стороны, индивиды, совершающие ряд нежелательных поступков (от попытки пройти без очереди до физического насилия в отношении тех, с кем у них возникли разногласия), обычно чувствуют себя виноватыми, и действительно, от этих поступков большинство индивидов удается, как правило, удержать. Но индивиды, похоже, не ощущают себя особенно добродетельными, воздерживаясь от таких поступков, поскольку ожидается, что так поступит каждый. С другой стороны, индивиды, подвергающие свою жизнь опасности во имя спасения других людей, и те, кто посвящает свою жизнь, скажем, помощи бедным в менее развитых странах, ощущают себя добродетельными, в то время как большинство из нас, не посвящающих бóльшую часть времени или ресурсов помощи чужим людям (и нас нелегко было бы к этому склонить), в целом не испытывают при этом чувства вины. Обычно не наблюдается сколько-нибудь существенного использования вины и добродетельности в отношении одного и того же решения одновременно. Таким образом, модель помогает понять, почему с некоторыми поступками может ассоциироваться вина, а с другими — добродетельность, различие, которое не так легко объяснить иными причинами.

[54] Авторы сравнивают этот вывод с мнением Д. Уиттмана, предложившего выбирать между наградами и санкциями на основе того, какой из инструментов экономит большую величину административных издержек, определяющихся частотой применения (Wittman D. Liability for Harm or Restitution for Benefit? // Journal of Legal Studies. 1984. No. 13. P. 57—80).

Применение кантианских моральных правил при контроле за производством экстерналий и добровольном финансировании общественных благ

В предыдущем разделе было показано, каким образом система моральных норм и подкрепляющих ее нравственных чувств могла бы использоваться как инструмент контроля за производством положительных и отрицательных внешних эффектов в интересах всего общества в целом. Мораль в соответствующей модели выступала в роли инструмента координации. Однако построение системы моральных норм и ее оптимизация проводились от лица и с точки зрения некоторого гипотетического агента, представлявшего, по словам авторов модели, интересы всего общества. В результате моделировалось то, "как средства морального наказания и поощрения — чувство вины или ощущение добродетельности — ассоциировались бы с конкретными поступками или естественными группами поступков, если бы целью была максимизация общественного благосостояния"[55]. При том, что аналогию упомянутому гипотетическому агенту возможно найти в реальном мире (на уровне небольших сообществ, где формируется бóльшая

[55] Kaplow L., Shavell S. Moral Rules and Moral Sentiments: Toward a Theory of an Optimal Moral System. 2001. NBER Working Paper 8688. P. 46.

часть моральных норм, коллективный интерес как действующая сила более реален), она все же выглядит несколько искусственно.

Продолжая тему моделирования морали как инструмента координации человеческого поведения, в данном разделе мы рассмотрим примеры альтернативных подходов, формально не требующих участия гипотетического анонимного максимизатора общественного благосостояния. Во всех рассматриваемых моделях индивид координирует собственное поведение с поведением других людей, судя о нем с помощью правил, основанных на самой известной из нескольких сформулированных Кантом форм категорического императива — "действуй только в соответствии с такой максимой, которую ты желал бы возвести в ранг универсального закона".

Как правило, экономисты склонны считать, что значимость неэгоистической (и в том числе моральной) мотивации падает с расширением круга контрагентов и деперсонализацией трансакций. В ситуации, когда действующими лицами в модели оказывается население целой страны, практически всегда предполагается, что индивиды ведут себя сугубо эгоистически. Ж.-Ж. Лаффон[56] еще в 1975 г. усомнился в реализме такой предпосылки, рассматривая поведение очень большой группы в условиях макроэкономических ограничений, когда каждый член группы действует, казалось бы, в противоречии с собственным эгоистическим интересом.

В подобной ситуации объяснения, основанные на теориях кооперации, разработанных для небольшого числа агентов, оказываются либо слишком сложны, либо нереалистичны в силу жестких предпосылок о доступности информации и возможности общаться. Более простым объяснением, которое и предложил Лаффон, была гипотеза о том, что агенты, во-первых, осведомлены о наличии макроэкономического ограничения в той среде, где они действуют (предел загрязнения окружающей среды, когда ущерб становится необратимым, предел перегруженности автодорог, по достижении которого передвижение становится почти невозможным, и т.п.), и во-вторых, пользуются аналогом кантианского категорического морального императива для прогнозирования поведения себе подобных.

В качестве практического примера Лаффон использует большой пляж с расставленными на расстоянии 100 метров друг от друга мусорными баками, который ежедневно посещает большое количество отдыхающих. В некоторых странах реальные пляжи, подобные рассматриваемому, остаются относительно чистыми. Однако если ограничить круг предпосылок только тем, что (1) индивид страдает от общей загрязненности пляжа в целом,
(2) ленится дойти до ближайшего бака и (3) предельное влияние несколь-

[56] Laffont J.-J. Macroeconomic Constraints, Economic Efficiency and Ethics: An Introduction to Kantian Economics // Economica. New Series. 1975. Vol. 42. No. 168. P. 430—437.

ких выброшенных им самим банок на его полезность невелико, объяснить этот факт нелегко. Эффект "хорошего примера" вряд ли имеет место, поскольку большинство посетителей не знакомы друг с другом.

Для анализа этого примера автор предлагает следующую простую модель. Рассматривается замкнутая экономика с измеримым пространством агентов $A = [0, 1]$, наделенным лебеговой мерой μ. В экономике только два блага — X и Y, первоначальные запасы каждого агента составляют (1, 1). Цена X нормализуется к 1. Благо Y можно производить из блага X по технологии $y \le \alpha x$. Цена Y в таких условиях будет равна $p = 1/\alpha$, причем функции полезности у всех одинаковы — они дифференцируемые, возрастающие и выпуклые. Далее, функции полезности зависят от совокупного потребления y: $\int_A y d\mu$ — таким образом в модель вводится внешний эффект.

Задача потребителя выглядит так: $\max U(x, y, \int_A y d\mu)$ при ограничении $x + (1/\alpha)y = 1 + (1/\alpha)1$, что дает условие первого порядка $U_2/U_1=1/\alpha$, если решение внутреннее.

Макроэкономическими ограничениями этой модели служит ограниченность ресурсов в экономике: $\int_A x d\mu = 1$ и $\int_A y d\mu = 1$. Некооперативный исход в этих условиях будет неэффективен[57]. Условиями Парето-оптимума при равном распределении благ было бы $U_2/U_1=1/\alpha - U_3/U_1$.

Далее вводится предпосылка о том, что каждый человек предполагает, что его сограждане рассуждают и поступают подобно ему самому. Таким образом, выбирая блага Y, он знает что $\int_A y d\mu = y$, и его задача максимизации превращается в следующую:

$\max U(1 + 1/\alpha - (1/\alpha)y, y, y)$,

для этой задачи условия первого порядка имеют вид:

$U_2/U_1=1/\alpha - U_3/U_1$.

Так, поведение в соответствии с кантианским принципом позволяет реализовать оптимум, который при эгоистическом поведении можно было бы реализовать только с помощью налога $t = - U_3/U_1$ на потребление блага Y.

Для вышеописанного примера с пляжем подобная формализация вполне приемлема. Считая налогообложение нереальным для этого примера, автор противопоставляет два возможных решения: основанную на правилах и санкциях разрешительную систему и модификацию предпочтений

[57] Laffont J.-J., Laroque G. Effets externs et théorie de l'équilibre général. Cahiers du Séminaire d'Econométrie. Paris: CNRS, 1972.

посетителей путем информационной кампании с целью склонить их к кантианскому поведению — подобно тому, как в Средние века этический кодекс торговли с помощью некоторого формально-институционального принуждения был привит большинству купцов.

Лаффон приводит несколько примеров, в которых навязывание людям кантианского поведения давало бы Парето-оптимальные равновесные исходы. В частности, предлагается модель перекрывающихся поколений, в которой осознание людьми связи между получаемыми ими от государства трансфертами и инфляцией приводит к нейтральности денег в долгосрочном периоде, и ситуация сбора налогов, когда уклонение от налогов заставляло бы правительство повышать ставки налогообложения до неэффективного уровня. Следует отметить, что определение кантианских правил для случая неоднородных групп может сопровождаться существенными затруднениями, что ограничивает их потенциальную применимость случаями, когда положение участников можно считать относительно равным.

Эти затруднения попытались решить М. Билодо и Н. Гравель[58], расширив анализ кантианского поведения для случая неоднородных групп. В работе, посвященной существованию и Парето-эффективности кантианских моральных норм в играх с добровольным предоставлением общественных благ, они обобщают более ранние попытки анализа кантианского поведения, в частности, работы Лаффона (см. выше) и Бординьона[59], вводя дифференциацию агентов по предпочтениям и доступным множествам стратегий.

Билодо и Гравель, опираясь на часто приписываемое самому Канту замечание о том, что нравственное поведение приводит к наилучшему результату, если оно подобающим образом *универсализировано*, предлагают трактовать кантианский категорический императив следующим образом. Во-первых, универсальный закон, в ранг которого необходимо мысленно возводить свое поведение, должен предписывать индивидам не одно и то же, а *эквивалентное* поведение. Во-вторых, при всеобщем соблюдении такой универсальный закон должен обеспечивать каждому наиболее предпочитаемый им исход.

Эти условия Билодо и Гравель используют для формализации понятия кантианской максимы, вводимой как совместное ограничение на наборы стратегий всех участников игры. Для того чтобы иметь возможность судить о моральной эквивалентности поступков различных людей, авторы

[58] Bilodeau M., Gravel N. Voluntary provision of a public good and individual morality // Journal of Public Economics. 2004. No. 88. P. 645—666.

[59] Bordignon M. Was Kant right? Voluntary provision of public goods under the principle of unconditional commitment // Economic Notes. 1990. No. 3. P. 342—372.

формализуют *систему универсализации* (экзогенно заданную схему морального сравнения поступков).

Билодо и Гравель показывают, что существование кантианской максимы при неоднородности агентов не обязательно, однако для игр с добровольным финансированием общественного блага, при довольно специфических (хотя и не невероятных) условиях на учет предпочтений и богатства людей найдется как минимум одна система универсализации, применительно к которой можно определить кантианскую максиму, которая будет Парето-эффективной.

Формальная модель, используемая авторами, в сжатом виде выглядит следующим образом. Рассматриваются игры в стратегической форме $G = \{N, \times_{i=1}^{n} S_i, \langle V_i(.)\rangle_{i=1}^{n}\}$, где $N = \{1,\ldots, n\}$ — конечное множество игроков, S_i — множество стратегий игрока i, а $V_i : \times_{j=1}^{n} S_j \to \mathbb{R}$ — платежная функция игрока i. Относительно игрока $i \in N$, обозначим за $S_{-i} = \times_{j\neq i} S_j$ множество всех комбинаций стратегий, которые могут играть остальные игроки. Также обозначим за $(s_i; s_{-i}) \in S_i \times S_{-i}$ комбинацию стратегий в которой игрок i играет s_i, а все остальные играют комбинацию $s_{-i} \in S_{-i}$.

Определение. Зададим *жесткую систему универсализации* как отношение эквивалентности M на множестве $U_{i\in N}\{i\} \times S_i$, удовлетворяющее свойству: для любых $i, j \in N$ и любого $s \in S_i$ $\#\, E_M(i, s) \cap \{j\} \times S_j = 1$.

Множество $E_M(i, s)$ обозначает множество всех индивидуальных поступков, морально эквивалентных поступку s индивида i. Чтобы сделать систему универсализации полной, авторы требуют, чтобы пересечение этого множества с множеством стратегий любого другого индивида было непустым. Выражение $(i, s)\ M\ (j, s')$ будет означать "стратегия s для игрока i морально эквивалентна стратегии s' для игрока j". В жесткой системе универсализации у любой стратегии индивида i есть лишь одна морально эквивалентная стратегия индивида j. Отношения моральной эквивалентности по определению рефлексивны (любое действие морально эквивалентно себе), симметричны (если $(i, s)\ M\ (j, s')$, то и $(j, s')\ M\ (i, s)$) и транзитивны (если s игрока i эквивалентно s' игрока j, которое в свою очередь эквивалентно s'' игрока h, то s игрока i эквивалентно s'' игрока h).

Лемма. *Пусть M — жесткая система универсализации на $U_{i\in N}\{i\} \times S_i$. Тогда для любого индивида h существует n функций $\Psi_i^h : S_h \to S_i$ (для $i \in N$) таких, что для любого $i,\ j \in N$, любого $s \in S_i$, $s' \in S_j$, $(i,s)M(j,s') \Leftrightarrow s = \Psi_i^h(\Psi_h^j(s'))$ (где $\Psi_h^j : S_j \to S_h$ обратно Ψ_j^h, т.е. $s_h = \Psi_h^j(s_j) \Leftrightarrow s_j = \Psi_j^h(s_h)$).*

Эта лемма позволяет представить систему универсализации в виде n функций, сопоставляющих стратегиям одного индивида морально эквивалентные стратегии всех остальных. Далее вводится понятие максимы.

Определение. Обозначим за $P(A)$ множество всех непустых подмножеств множества A. Максимой $\mu \in \times_{i=1}^{n} P(S_i)$ является упорядоченный список подмножеств множеств индивидуальных стратегий.

"Строгой" предлагается называть максиму, предписывающую каждому игроку только одну возможную стратегию. Принадлежность максимы к классу "кантианских" определяется на основе двух свойств.

Свойство 1. (Моральная эквивалентность) $\forall s \in \mu$, $(i, s_i)\, M\, (j, s_j)$ $\forall i, j \in N$.

Второе свойство заключается в том, что индивиды должны хотеть сделать максиму универсальным законом, а для этого максима должна обеспечивать игроку наивысший платеж при морально эквивалентном поведении остальных игроков.

Свойство 2. (Универсализированная рациональность) $\forall i \in N$, $\forall s \in \mu$,

$$\int_{E_M(i,\, s_i)|\{i\}\times S_i} V_i(s_i; s_{-i})\pi_i^{s_i} ds_{-i} \geq \int_{E_M(i,\, \overline{s}_i)|\{i\}\times S_i} V_i(\overline{s}_i; \overline{s}_{-i})\pi_i^{\overline{s}_i} d\overline{s}_{-i}$$

для $\forall \overline{s}_i \in S_i$. Для каждого $s \in S_i, \pi_i^s$ является борелевской мерой вер игрока i на $E_M(i, s)|\{i\} \times S_i$.

Понятие Парето-эффективности применительно к рассматриваемому примеру формулируется следующим образом.

Определение. (Парето-эффективность) Профиль стратегий $\hat{s}_1, \ldots \hat{s}_n \in \times_{i=1}^{n} S^i$ эффективен по Парето тогда и только тогда, когда для каждого $(s_1, \ldots, s_n) \in \times_{i=1}^{n} S^i, V_i(s_1, \ldots, s_n) > V_i\ (\hat{s}_1, \ldots, \hat{s}_n)$ для некоторых $i \in N$ соблюдается $V_j(\hat{s}_1, \ldots, \hat{s}_n) > V_j(s_1, \ldots, s_n)$ для некоторых $j \in N$.

Влияние морали в описанной выше формализации на добровольное финансирование общественных благ анализируется на основе следующей модели. В экономике существует n индивидов и два блага: частное, количество которого обозначается как x, и общественное, объем которого обозначается как Z. Каждый индивид $i \in N = \{1, \ldots, n\}$ обладает ω_i единицами частного блага, которые он может использовать для собственного частного потребления x_i или вложения в производство общественного блага z_i. Общественное благо производится только за счет таких добровольных пожертвований, $Z = \sum_{i \in N} z_i$. Суммарное благосостояние общества обозначается как $\omega = \sum_{i \in N} \omega_i$. Предпочтения любого индивида описываются дифференцируемой, строго квазивогнутой и строго монотонной функцией полезности $U_i : R_+^2 \to R$. Обозначим за $MRS^i(\overline{Z}, \overline{x})$ предельную норму

замещения индивида i в точке $(\overline{Z}, \overline{x})$, $MRS^i(\overline{Z}, \overline{x}) = \dfrac{\dfrac{\partial U_i(\overline{Z}, \overline{x})}{\partial Z}}{\dfrac{\partial U_i(\overline{Z}, \overline{x})}{\partial x}}$.

Множеством стратегий индивида i являются все возможные размеры пожертвований на добровольное финансирование общественного блага, которые ему доступны $S_i = [0, \omega_i]$. Платежной функцией индивида

$V_i : \times_{j \in N}[0, \omega_j] \to R$ является его функция полезности

$U_i(\sum_{j \in N} z_j, \omega_i - z_i)$. Обозначим за $(N, \times_{i \in N}[0, \omega_i], \langle V_i(.)\rangle_{i \in N})$ типичную игру с добровольным финансированием общественных благ. Далее, наложим некоторые ограничения на класс рассматриваемых игр.

Условие 1. Для каждого $i \in N$, $U_i(\omega_i, 0) > U_i(0, \omega)$.

Данное условие означает, что индивид скорее согласится покупать только частное благо и совсем не получать общественного, чем тратить все свои деньги на финансирование общественного блага, даже если все другие игроки поступят также. Это возможно, если частное благо включает некий элемент первой необходимости, например, еду. Кроме того, введенное условие означает, что вектор пожертвований не может быть Парето-эффективным.

Обозначим за P подкласс игр, удовлетворяющих этому условию. Билодо и Гравель доказывают, что кантианская максима в жесткой системе универсализации в игре класса P с необходимостью является строгой.

Лемма. Пусть $(N, \times_{i \in N}[0, \omega_i], \langle V_i(.)\rangle_{i \in N}) \in P$. Тогда максима μ удовлетворяет свойствам моральной эквивалентности и универсализированной рациональности в жесткой системе универсализации только если $\#\mu = 1$.

Билодо и Гравель также пользуются условием Самуэльсона для Парето-оптимальности в рассматриваемом классе игр.

Лемма. Пусть $(z_1, \ldots, z_n) \in R_+^n$ — вектор пожертвований, причем $z_i \in [0, \omega_i]$ для любого i, и $\sum_{i \in N} MRS(\sum_{j \in N} z_j, \omega_i - z_i) = 1$ для игр с добровольным финансированием общественных благ $(N, \times_{i \in N}[0, \omega_i], \langle V_i(.)\rangle_{i \in N}) \in P$.

Тогда $(z_1, \ldots, z_n)$ эффективен по Парето.

Авторы задаются вопросом: в общем случае игры с добровольным финансированием общественных благ, когда предпочтения и уровни дохода игроков различны, будет ли кантианская максима предписывать Парето-оптимальный уровень пожертвований. Для того чтобы ответить на этот вопрос, системы универсализации предполагаются дифференцируемыми,

а составляющие их функции — монотонно возрастающими[60]. В экономической интерпретации это означает, что в глазах любого игрока увеличение размера пожертвования одним человеком морально эквивалентно увеличению размера пожертвования кем угодно другим. Таким образом, никто не должен считать общественное благо бесполезным или вредным.

Предположение 1. Пусть $(N, \times_{i\in N}[0,\omega_i], \langle V_i(.)\rangle_{i\in N})$ — игра с добровольным финансированием общественных благ, принадлежащая классу P, и пусть M — жесткая, дифференцируемая система универсализации на $\cup_{i\in N}\{i\}\times[0,\omega_i]$. Тогда, если максима μ удовлетворяет свойствам моральной эквивалентности и универсализированной рациональности по отношению к M, она является Парето-эффективной.

Для экономики с двумя индивидами это иллюстрируется следующим образом.

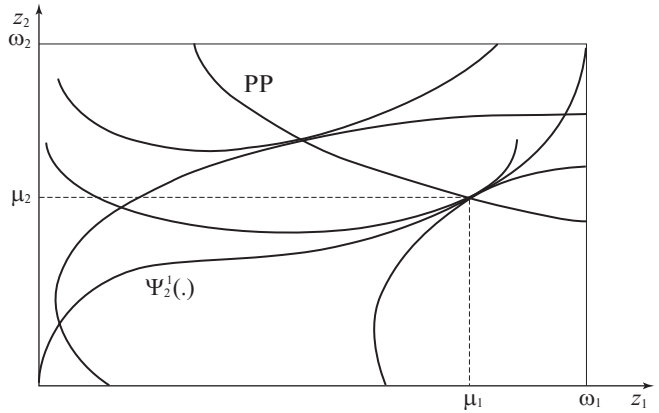

Рис. 4. Множество Парето-оптимальных и множество морально эквивалентных распределений расходов на финансирование общественного блага

В плоскости (z_1, z_2) изображены U-образные кривые безразличия игрока 1, и C-образные кривые безразличия игрока 2. Линия PP соединяет все точки касания кривых безразличия, являясь множеством Парето-оптимальных комбинаций пожертвований. Система универсализации представлена

[60] Дифференцируемость жесткой системы универсализации определяется как дифференцируемость каждой из n составляющих ее функций $\Psi_i^h : S_h \to S_i$. Авторы доказывают, что в играх класса P при дифференцируемой жесткой системе универсализации кантианская максима μ может существовать только если все n функций $\Psi_i^h : [0,\omega_h] \to [0,\omega_i]$, составляющие эту систему универсализации, монотонно возрастают.

(дифференцируемой и монотонно возрастающей) функцией $\Psi_2^1(.)$, соединяющей точки (0, 0) и (ω_1, ω_2).

Чтобы максима удовлетворяла универсализированной рациональности, она должна предписывать каждому индивиду выбирать такой размер пожертвования, при котором его кривая безразличия касательна к этой функции[61]. Чтобы удовлетворять моральной эквивалентности, пожертвования обоих игроков должны находиться в одной и той же точке кривой $\Psi_2^1(.)$. Это может соблюдаться только если они касательны в этой точке, а значит, кантианская максима должна предписывать размер пожертвований, соответствующий точке пересечения $\Psi_2^1(z_1)$ и PP[62].

Затем Билодо и Гравель задаются вопросом о существовании дифференцируемой жесткой системы универсализации, которая бы поддерживала кантианские максимы. На несколько ограниченном графическом примере с двумя игроками (см. рис. 4), кантианская максима поддерживается, если углы наклона кривых безразличия и функции $\Psi_2^1(.)$ совпадают в точке $(\hat{z}_1, \hat{z}_2)$. Таким образом, геометрически проблема поиска жесткой дифференцируемой системы универсализации, которая поддерживала бы кантианскую максиму, сводится к поиску непрерывной функции, соединяющей точки (0, 0) и (ω_1, ω_2) и разделяющей кривые безразличия, касающиеся в $(\hat{z}_1, \hat{z}_2)$. Эта проблема аналогична поиску цен Линдаля в экономике с общественным благом, которые на рис. 4 выглядели бы как луч, выходящий из начала координат и разделяющий кривые безразличия в некоторой точке на PP. Из стандартных теорем о существовании равновесия Линдаля в экономиках с общественными благами известно, что по крайней мере один такой луч будет существовать. Однако случай с системой универсализации несколько сложнее, поскольку соответствующая функция должна соединять (0, 0) и (ω_1, ω_2).

Авторы доказывают следующее утверждение:

Утверждение. Пусть $(N, \times_{i\in N}[0,\omega_i], \langle V_i(.)\rangle_{i\in N})$ — игра с добровольным финансированием общественных благ, принадлежащая классу P. Тогда существует жесткая непрерывная система универсализации M, поддерживающая кантианскую максиму “жертвуйте вашу равновесную по Линдалю сумму”.

[61] Угловые решения рассматриваются авторами отдельно.

[62] Геометрическая иллюстрация этого примера не может быть расширена до трех и более игроков. Для рассматриваемого класса игр эффективное внутреннее распределение ресурсов характеризуется условием Самуэльсона, согласно которому сумма углов наклона кривых безразличия в 2-мерной плоскости “частное благо – общественное благо” равняется единице. Для двух игроков это эквивалентно касанию их кривых безразличия на плоскости (z_1, z_2), но и только.

Это утверждение дает надежду на возможность использовать индивидуальную мораль в качестве средства предотвратить “проблему безбилетника” — до тех пор пока общепринятая система моральной эквивалентности учитывает различие в предпочтениях и богатстве индивидов. Однако условия, в которых кантианские правила могут быть успешно задействованы, достаточно узки — так, у людей должно быть достаточно информации для того, чтобы вычислить свое “равновесное по Линдалю пожертвование”.

Вторым проблематичным моментом являются стимулы людей интернализировать кантианскую максиму. Очевидно, что в одиночку никто не будет заинтересован в этом. Стимулы могут появиться только если все индивиды смогут выиграть, коллективно интернализировав максиму. Тогда с помощью образования или пропаганды их можно было бы убедить сделать это. В этой связи необходимо как минимум выяснить, будет ли существовать жесткая система универсализации, которая поддерживала бы кантианскую максиму, обеспечивающую такой результат, который доминировал бы некооперативный исход игры по Парето. Очевидно, такая система будет существовать не всегда.

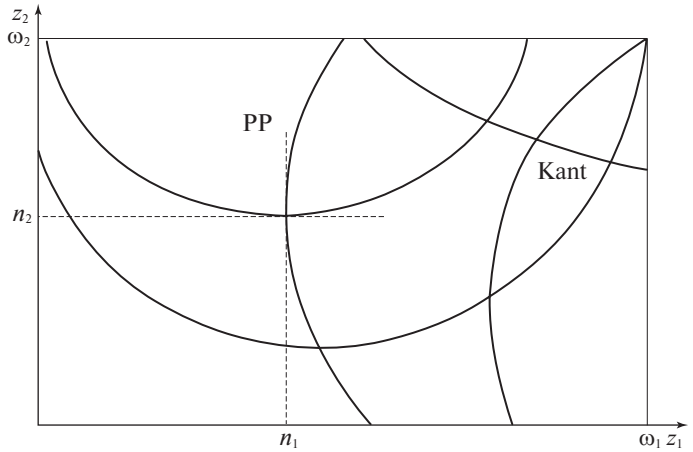

Рис. 5. Пример экономики, в которой кантианскими максимами невозможно улучшить положение игроков по сравнению с некооперативным равновесием

На рис. 5 для индивида 1 любое распределение, поддерживаемое кантианскими максимами (участок контрактной кривой, обозначенный на рисунке как “Kant”), хуже того, что он получает в равновесии по Нэшу — (n_1, n_2).

Добровольное финансирование общественных благ, обусловленное эндогенной моральной нормой

Моделирование влияния моральных норм при анализе добровольного финансирования общественных благ является достаточно популярным направлением исследований. В круг проблем, стоящих перед экономистами в этой области, входят собственно объяснение периодически наблюдаемого на практике добровольного финансирования (или участия в производстве) общественных благ, а также объяснение жизнеспособности подобной практики в условиях анонимности пожертвований и при большом числе жертвователей, взаимосвязь между моральной мотивацией, экономическими стимулами и государственной политикой, объяснение механизма вытеснения добровольного финансирования государственным, влияние окружения на поведение индивидуального жертвователя, и др.

Мораль может включаться в задачу экономических агентов в качестве элемента системы предпочтений[63], ограничения на множество доступных стратегий[64], иногда — самой формы игры[65] или эвристического правила для прогнозирования поведения контрагентов[66]. Следует отметить, что исследователи, стремящиеся проанализировать влияние морали на экономическую эффективность (Парето-оптимальность), трактуют мораль именно как правило в противовес морали как компоненту системы предпочтений, что, вероятно, связано с трудностью определения Парето-оптимума в моделях с существенно модифицированными предпочтениями; те же авторы, которые пытаются доказать, что мораль может послужить инструментом реализации Парето-оптимальных состояний, на том или ином этапе рассуждений приходят к вариации равновесия по Линдалю[67]. Самой популярной формой используемого морального правила следует назвать кантиан-

[63] Andreoni J. Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving // The Economic Journal. 1990. Vol. 100. No. 401. P. 464; Nyborg K., Howarth R., Brekke K. Green Consumers and Public Policy: on Socially Contingent Moral Motivation. Working Paper No. 31. University of Oslo, Dept. of Economics, 2003.

[64] Sugden R. Reciprocity: The Supply of Public Goods Through Voluntary Contributions // The Economic Journal. 1984. Vol. 94. No. 376. P. 772–787; Bilodeau M., Gravel N. Voluntary Provision of a Public Good and Individual Morality // Journal of Public Economics. 2004. No. 88. P. 645—666.

[65] Young D.J. A “Fair share” Model of Public Good Provision // Journal of Economic Behavior and Organization. 1989. No. 11. P. 137—147.

[66] Brekke K., Kverndokk S., Nyborg K. An Economic Model of Moral Motivation // Journal of Public Economics. 2003. No. 87.

[67] Young D.J. “Fair share” Model of Public Good Provision; Bilodeau M., Gravel N. Voluntary Provision of a Public Good and Individual Morality; Sugden R. Reciprocity: The Supply of Public Goods Through Voluntary Contributions.

ский категорический императив, который используется в различных моделях как инструмент прогнозирования чужого поведения, моральной оценки исходов в играх или координационный механизм. Его недостатком является безусловность, порождающая моральный конфликт: “Мой долг требует финансировать общественное благо независимо от поведения остальных, однако то, что остальные этого не делают, несправедливо”.

Достаточно часто проблема добровольного финансирования общественных благ рассматривается с точки зрения экономической политики государства: различные авторы пытаются объяснить реакцию агентов на внешние стимулы через те мотивы, которые побуждают людей добровольно финансировать общественное благо. При этом исследователей закономерно волнует проблема вытеснения внешними стимулами внутренних, поскольку за счет этого явления некоторые меры государственной политики могут приводить к совершенно неожиданным результатам.

В этом разделе мы рассмотрим модель, в которой добровольное финансирование общественных благ объясняется через наличие у индивидов желания считать себя социально ответственными членами общества. Индивиды судят о собственном уровне “социальной ответственности”, сравнивая свое фактическое поведение с эндогенным “нравственно безупречным” образцом. Для определения этого образца они используют комбинацию упрощенного кантианского категорического императива и бентамианской функции общественного благосостояния.

Отправной точкой для разработки модели послужил подход к объяснению альтруистического поведения, предложенный Андреони[68] и известный под названием “warm glow giving”, или “смешанный альтруизм” (“impure altruism”). Теория “смешанного альтруизма”, сформулированная применительно к проблеме добровольного финансирования общественных благ, или благотворительности, предполагала, что помимо самого объема общественного блага, которое финансирует жертвователь, источником полезности для него служит и само пожертвование (человек испытывает удовлетворение от поступка, который он считает достойным). Смешанный альтруист, таким образом, будет продолжать финансировать общественное благо даже если его индивидуальное влияние на благосостояние других людей ничтожно мало.

Однако в описанном выше виде у теории “смешанного альтруизма” существуют некоторые проблемы: она не позволяет объяснить поведение, наблюдающееся у людей в ситуациях так называемого “вытеснения внутренней мотивации” (intrinsic motivation crowding-out)[69]. Например, Б. Фрей и Ф. Оберхольцер-Ги[70] показали, что готовность людей терпеть строительство в их районе завода по переработке опасных отходов при выплате им компенсации была гораздо меньше, чем в случае, когда компенсации не предлагалось. У. Гнизи и А. Рустичини[71] привели экспериментальные результаты, свидетельствующие о немонотонном влиянии объема вознаграждения на усилия агента: при увеличении вознаграждения с исходной нулевой величины производительность снижалась, а при увеличении с исходной положительной величины — росла.

Адаптировав концепцию “warm glow giving” для случаев вытеснения внутренней мотивации, К.А. Брекке, С. Квердокк и К. Нюборг[72] предложили модель, объясняющую связь между финансовыми ограничениями и частным добровольным финансированием общественных благ, когда это финансирование обусловлено тем, что потребителям нравится считать себя социально ответственными членами общества. Присутствующее в модели государство может влиять на размер добровольных пожертвований двояко: во-первых, за счет изменения относительных цен и бюджетных (или вре-

[68] Andreoni J. Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. P. 464—477.

[69] Заметим, что эта тема достаточно популярна в литературе, посвященной экономическому анализу морально обусловленного поведения. Одной из первых попыток анализа в этой области стала работа Р. Титмусса, посвященная сравнению эффективности сложившихся систем донорства крови в Великобритании и США. Сопоставив частоту случаев заражения гепатитом при переливании крови в США, где донорство оплачивалось, и Великобритании, где оно было добровольным, он обнаружил практически четырехкратную разницу не в пользу США. Отнеся этот факт к отсутствию у добровольных доноров мотивации скрывать свою болезнь, Титмусс подчеркнул возможную опасность вытеснения внутренней мотивации внешней: предлагаемый вывод состоял в том, что введение денежной платы за донорство крови фактически не увеличило, а сократило предложение качественной (неинфицированной) крови. Последующая критика обнаружила, что анализ Титмусса был некорректен, но собственно проблема поведения наделенных значимой внутренней мотивацией агентов в ситуациях moral hazard и, на более общем уровне, взаимодействия внутренней и внешней мотивации вызвала значительный интерес. Впоследствии поведение агентов, обладающих значимой внутренней мотивацией и, в частности, проблему вытеснения внутренней мотивации внешней исследовал Б. Фрей. Он предложил теорию, потенциально способную объяснить закономерности вытеснения внутренней мотивации внешней, и составил обзор эмпирических свидетельств за и против нее (см.: Frey B.S., Jegen R. Motivation Crowding Theory: A Survey of Empirical Evidence. Institute for Empirical Research in Economics. University of Zurich, 1999. Working Paper). Анализ представленных нами моделей с точки зрения проблемы вытеснения внутренней мотивации заслуживает самостоятельного исследования, поэтому здесь она лишь упоминается.

[70] Frey B.S., Oberholzer-Gee F. The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-out // American Economic Review. 1997. No. 87 (4). P. 746—755.

[71] Gneezy U., Rustichini A. Pay Enough or don’t Pay at All // Quarterly Journal of Economics. 2000. CXV (3). P. 791–810.

[72] Brekke K., Kverndokk S., Nyborg K. An Economic Model of Moral Motivation // Journal of Public Economics. 2003. No. 87.

менных) ограничений и, во-вторых, косвенно, за счет изменения размера пожертвования, которое можно считать "нравственно безупречным".

В модели существует N одинаковых индивидов с возрастающими, строго квазивогнутыми функциями полезности вида

$$U_i = u(x_i, l_i, G, I_i), \quad (1)$$

где x_i — объем потребления частных благ индивидом i, l_i — объем досуга, G — количество общественного блага, а I_i — степень, в которой индивид считает себя социально (или морально) ответственным (заметим, что такое удовлетворение от сознания собственной моральной ответственности не возникает само по себе, а также культивируется специфическими средствами морального воздействия). Индивидуальное предложение труда и доход принимаются экзогенно заданными, и задача индивида будет заключаться в распределении времени между досугом и вкладом в общественное благо. Индивидуальное временное ограничение имеет вид

$$l_i + e_i = T, \quad (2)$$

где e_i соответствует вложениям индивида в общественное благо (в виде времени, затраченного, например, на сортировку своего мусора), а T — одинаковое для всех индивидов свободное от работы время. Количество производимого в экономике общественного блага определяется как сумма государственного и частного финансирования:

$$G = G_p + \Sigma_i g_i, \quad (3)$$

где

$$g_i = \gamma(e_i, \theta) \quad (4)$$

соответствует возможностям индивида i производить общественное благо. θ — экзогенный параметр, отвечающий за технологические или институциональные условия производства. Предположим, $\gamma(0, \theta) = 0$, $\gamma_e > 0$, $\gamma_{ee} < 0$, $\gamma_\theta > 0$, $\gamma_{e\theta} > 0$.

Обозначим размер морально безупречного пожертвования индивида i за e^*_i. Тогда образ индивида в собственных глазах будет определяться следующим выражением:

$$I_i = f(e_i, e^*_i) = -a(e_i - e^*_i)^2, \; a > 0 \quad (5)$$

глобальный максимум этого выражения равен нулю и достигается при $e_i = e^*_i$.

Индивид судит о e^*_i следующим образом. Он максимизирует общественную функцию полезности, выраженную бентамовской суммой $W = u_1 + ... + u_N$ при ограничениях (1)—(5) и дополнительном ограничении $e_i = e_j$, $j \neq i$, $j = 1, ..., N$ (упрощенная версия кантианского категорического императива: "как я поступлю в этой ситуации, если все остальные поступят так же"). Таким образом, утилитаризм определяет то, как должно выглядеть хорошее общество, а кантианский категорический императив — то, как следует думать о поведении других, определяя свое поведение. Условие первого порядка, одинаковое для всех индивидов, имеет вид

$$u_l = N \times u_G \times \gamma_e, \quad (7)$$

т.е. если объем пожертвований соответствует морально безупречной величине, предельная полезность одного часа досуга равна общественной оценке дополнительного количества общественного блага, которое индивид может произвести за тот же час. Следующим шагом индивид определяет свой реальный вклад в производство общественного блага, максимизируя функцию (1) с ограничениями (2)—(5) и принимая $e_{j\neq i}$ и e^* как заданные. Это дает равновесие по Нэшу, описываемое следующим условием первого порядка:

$$u_l = u_G \times \gamma_e + u_I(-2a\,(e_i - e^*_i)), \quad (8)$$

сформулированным для равновесного по Нэшу количества общественного блага $G = G_p + Ng'$, где g' соответствует равновесному пожертвованию репрезентативного индивида. Из сравнения (7) и (8) видно, что максимизация индивидуальной полезности не может привести к общественно оптимальному результату — для этого (7) и (8) должны быть эквивалентны. Поскольку второй член в правой части (8) равен нулю, (7) и (8) были бы эквивалентны при $(N-1)u_G \times \gamma_e = 0$, т.е. или $N = 1$, или $u_G = 0$, или $\gamma_e = 0$, что противоречит предпосылкам модели. Таким образом, даже при потребности людей в соответствии образу морально ответственного гражданина общественное благо недопроизводится.

Предположим, что θ возрастает (например, государство устанавливает мусорные контейнеры, упрощающие сортировку мусора), и за одно и то же потраченное на сортировку мусора время e индивид может достичь большего сокращения количества отходов g. Предположим, что функция полезности аддитивно сепарабельна по x, l, G и I, и предельная полезность I равна 1. Функция полезности приобретает вид

$$U = u(x, l) + v(G) + I, \quad (9)$$

где u и v возрастающие и вогнутые. Влияние θ на идеальный *объем усилий* находится дифференцированием условия первого порядка (7) по θ:

$$e^*_\theta = N\frac{v_G\gamma_{e\theta} + Nv_{GG}\gamma_e\gamma_\theta}{-u_{ll} - N^2v_{GG}\gamma_e^2 - Nv_G\gamma_{ee}}. \quad (10)^{73}$$

Числитель выражения (10) положителен, но в знаменателе наблюдаются два противоположных эффекта: во-первых, с ростом θ предельная продуктивность пожертвований растет (общественное благо стало дешевле в производстве), что влечет рост размера идеального усилия, а во-вторых, для заданного e рост θ увеличивает g и общий объем общественного блага G. При этом, естественно, падает его предельная полезность и вместе с ней — размер идеального усилия. Таким образом, невозможно сказать, как

[73] Чтобы получить это условие, e представляется функцией θ: $u_l(x, T - e(\theta)) = Nv_G(N\gamma(e(\theta),\theta))\,\gamma_e(e(\theta),\theta)$. Дифференцируя по θ и преобразуя, получаем (10).

рост θ повлияет на идеальное усилие, однако идеальный индивидуальный вклад $g^* = \gamma(e^*, \theta)$ определенно возрастет, поскольку второй из вышеописанных эффектов к нему не относится, а эффективность усилий возрастает ($\gamma_\theta > 0$).

Чтобы определить последствия изменившейся технологии для реального вклада, дифференцируем (8) по θ. Предположим, что люди неспособны увидеть, как их собственные усилия меняют состояние окружающей среды: $u_G \times \gamma_e = 0$. Тогда условие первого порядка равновесия по Нэшу будет выглядеть как $u_l = -2a(e - e^*)$, т.е. индивид тратит время на производство общественного блага до тех пор, пока получаемая предельная отдача (в виде улучшения собственного образа как морально ответственного человека) не сравняется с предельной полезностью досуга. Тогда предельный эффект технологического сдвига на фактический объем усилий равен

$$e_\theta = \frac{2ae_\theta^*}{-u_{ll} + 2a}. \qquad (12)$$

При стандартных предпосылках о вогнутости, знаменатель положителен. А поскольку e_θ^* может быть положительным или отрицательным, реальное усилие будет реагировать на изменение θ так же, как и идеальное.

Новая информация — например, сведения о том, что переработка вторичного сырья действует гораздо эффективнее, чем принято считать — работает похожим образом. Пусть, как и раньше, e_i соответствует индивидуальным усилиям, а g_i — их последствиям для окружающей среды. Предположим, усилия наблюдаемы, а их последствия — даже при наблюдаемом G — нет. Тогда рост θ можно интерпретировать как рост *наблюдаемых* последствий частных усилий. Последствия роста θ для индивидуальной полезности в равновесии по Нэшу, описываемом выражением (8), заданы

$$U_\theta = \upsilon_G\gamma_\theta + (N - 1)\upsilon_G(\gamma_e e_\theta + \gamma\theta) - 2a(e - e^*)e_\theta^*. \qquad (13)$$

Предположим, что $e_\theta > 0$ и $e_\theta^* > 0$. Тогда первые два слагаемых в правой части положительны, т.е. рост эффективности увеличит полезность, повысив G. Но последний член отрицателен и общий знак выражения неизвестен.

Итак, если в аналогичных моделях с экзогенно заданным идеальным поведением изменение эффективности никогда не уменьшило бы индивидуальную полезность, в данной модели индивидуальная полезность может упасть в силу возросших моральных требований к индивидуальным усилиям. Рост θ, следовательно, способен фактически снизить уровень общественного благосостояния, поскольку с ростом моральных требований к идеальному усилию растет разрыв между ним и фактическим усилием.

Развивая модель, авторы также исследовали влияние штрафов и поощрений на добровольное финансирование общественных благ. Их основные

выводы заключались в том, что эффект внешних стимулов зависит от восприятия их агентами: если агент считает, что размер штрафа позволяет нанять рабочего, чтобы выполнить его работу, введение штрафа приведет к отказу от добровольного производства общественных благ.

Используя данные специальных опросов о переработке бытовых отходов и добровольном участии норвежцев, являющихся членами различных клубов и организаций, в работах для нужд этих клубов, авторы не нашли противоречия ни одному из своих выводов. Предложенная ими модель также объясняла распределение ответов существенно лучше, чем подход с точки зрения смешанного альтруизма в том виде, в котором он был предложен Андреони. Однако бóльшая часть вопросов, предложенных респондентам в ходе этих исследований, касалась гипотетических ситуаций. Фрей и Гётте[74], используя данные о фактической работе добровольцев в Швейцарии, обнаруживают, что сам факт предложенного вознаграждения сокращает активность добровольцев, но размер вознаграждения увеличивает ее.

Можно также отметить некоторую нереалистичность модели с точки зрения ее требований к когнитивным способностям агентов. Последовательность принятия решения делится на два этапа: вначале индивид определяет некое "утопическое состояние", в котором благосостояние группы было бы максимальным, и на основе этого состояния рассчитывает стандарт для оценки своих усилий. Затем он решает собственную задачу, в которой учитывает расхождение своего поведения с вышеупомянутым идеалом. При попытке верификации модели предпосылки первого этапа могут представлять проблему, так как в реальной жизни ни предпочтения, ни ограничения не однородны, и необходимые расчеты обещают оказаться чрезмерно сложными. Это указывает на возможное направление доработки модели, когда схема принятия индивидом решения дополнялась бы механизмом прогнозирования предпочтений и ограничений других членов сообщества.

Заключение

При экономическом анализе индивидуального поведения стандартно предполагается, что каждый агент руководствуется исключительно тем, какой выигрыш (полезность) непосредственно ему может принести тот или иной вариант действий. Однако в некоторых случаях реальное экономическое поведение и результаты лабораторных экспериментов регулярно расходят-

[74] Frey B.S., Goette L. Does Pay Motivate Volunteers? Institute for Empirical Research in Economics. University of Zurich, 1999 (November). Working Paper Series No. 7.

ся с этой гипотезой. В игровых экспериментах некоторая доля участников стабильно демонстрирует поведение, которое выглядит парадоксально с точки зрения преследования собственного интереса, но естественно объясняется присущим им чувством справедливости — игроки заботятся о выигрышах партнеров, не имеющих возможности повлиять на распределение ресурсов в играх “Диктатор”, в ущерб собственному выигрышу отвергают “нечестные” предложения в играх “Ультиматум” и играх с последовательными переговорами, проявляют щедрость и отвечают добром на добро в играх “Дарообмен” и “Доверие”, демонстрируют заботу об интересах группы, участвуя в финансировании общественных благ и наказывая “безбилетников” в играх с добровольным финансированием общественных благ. В реальной жизни люди добросовестно платят налоги на доход, который могли бы скрыть, ходят на выборы, зная, что влияние их голоса ничтожно мало, протестуют против ущемления чужих прав, оставляют чаевые в ресторанах, делают пожертвования, чтобы помочь людям, которых никогда не увидят, сортируют свой мусор и сообща поддерживают в надлежащем состоянии территорию вокруг своих домов, отвечают на повышение заработной платы выше рыночного уровня честной работой при минимальном внешнем контроле, мирятся с ее сокращением, если знают, что фирме грозит банкротство, читают бесплатные лекции и являются донорами крови. Присущие всем людям способности испытывать сострадание, благодарность, угрызения совести или ощущение гордости от выполненного долга, судить о справедливости непосредственно влияют на форму существующих в нашем обществе институтов.

С расширением предметного поля современной экономической теории в экономической литературе отмечается растущий интерес к анализу подобного поведения и стоящих за ним мотивов, среди которых важнейшее место занимают человеческие нравственные ценности и представления о морали. В данной работе были описаны подходы к интеграции таких мотивов в стандартную экономическую модель принятия индивидуальных решений. Предпочтение отдавалось современным теоретическим разработкам (большинство из представленных моделей были разработаны в последние 5 лет), близким к традиционному аналитическому аппарату неоклассической микроэкономической теории. При этом в цели данной работы не входило создание единой картины существующих направлений интеграции морали в формальный инструментарий экономической теории. Учитывая, что на сегодняшний момент попытки формального анализа морально обусловленного поведения в различных областях экономической науки относительно разрозненны, это могло бы стать предметом более глубокого исследования, основанного не только на экономическом, но и философском анализе.

Среди различных областей экономической науки, в которых существует интерес к морали как одному из факторов, влияющих на человеческое поведение, и, в частности, к формальному моделированию этого влияния, прежде всего выделяется экспериментальная экономика. Обилие экспериментальных данных о поведении, не согласующемся с гипотезой узкого преследования 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. При формализации справедливости и заботы об интересах других агента наделяют “социальными предпочтениями”, на основе которых он судит о различных вариантах по их последствиям для благосостояния партнеров или схемой оценки стратегий партнеров, в соответствии с которой он судит об их намерениях. В связи с тем, что поведение участников экспериментов зачастую обнаруживает их заинтересованность как последствиями тех или иных поступков партнеров, так и их намерениями, авторы последних теоретических разработок склонны комбинировать оба метода в своих моделях.

Ни одна из имеющихся на сегодняшний день моделей справедливости не позволяет реалистично описать весь массив экспериментальных результатов. Однако в совокупности они объясняют и позволяют предсказывать ряд принципиальных 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 рационирования и перераспределения доходов и др. Следует, однако, отдавать себе отчет, что при исследовании реального человеческого поведения, предположительно обусловленного моральными мотивами, социальные нормы, часто определяемые на том же множестве ситуаций, могут оказаться конкурирующим, если не более значимым, объясняющим фактором, и изолировать влияние тех и других не всегда возможно.

Помимо экспериментальной экономики, влияние морали на индивидуальное поведение очевидным образом представляет интерес применительно к таким областям исследований, как экономический анализ права и экономика благосостояния. В описанных нами моделях моральные нормы рассматриваются прежде всего в качестве координационного механизма, позволяющего корректировать провалы рынка, в частности, контролировать производство экстерналий и обеспечивать финансирование общественных

благ. При этом, в отличие от рассмотренных выше моделей индивидуального чувства справедливости, предпочтения индивида могут оставаться совершенно эгоистическими. Это относится в первую очередь к моделям децентрализованного контроля за производством экстерналий и добровольного финансирования общественных благ, в которых агенты используют аналог кантианского категорического императива только как когнитивную рутину, чтобы прогнозировать поведение себе подобных. Теоретически это позволяет достичь Парето-эффективных исходов в ситуациях, где при стандартных предпосылках это было бы невозможно. Интересно, что такая трактовка моральных норм косвенно отражает еще одну из их возможных функций, а именно экономию внимания и мыслительных усилий при принятии решений, требующих оценки поведения других людей.

Приложение

Краткие характеристики рассматриваемых моделей

Работа	Рассматриваемая проблема	Базовая модель/игра	Предлагаемое решение/гипотеза	Отличие агента модели от стандартной модели homo oeconomicus
Charness and Rabin, 2000	Противоречие поведения участников лабораторных экспериментов гипотезе максимизации собственного выигрыша в случаях, когда их выбор влияет на выигрыши других участников	Игры “Диктатор”, “Дар-обмен”, игры с добровольным финансированием общественных благ	Агенты заботятся об общем благосостоянии группы и отчасти — благосостоянии отдельных игроков	Полезность индивида зависит от общего дохода группы и дохода игрока, находящегося в наихудшем положении
Bolton, 1991	Противоречие поведения участников лабораторных игр с последовательными перегоорами гипотезе максимизации собственного платежа	Двухпериодное “деление доллара” с последовательными переговорами и дисконтированием (оба игрока дисконтируют платежи второго периода по индивидуальной ставке, резервная полезность равна нулю)	Помимо абсолютной величины собственного платежа, агент принимает во внимание соотношение между своей долей и долей контрагента (проявляет зависть)	Полезность индивида отрицательно зависит от соотношения собственного платежа и платежа партнера в случае если платеж партнера больше

Работа	Рассматриваемая проблема	Базовая модель/игра	Предлагаемое решение/гипотеза	Отличие агента модели от стандартной модели homo oeconomicus
Fehr and Schmidt, 1999	Поведение участников лабораторных экспериментов в различных играх подтверждает несколько конкурирующих гипотез. В частности, интерес к справедливости в двусторонних переговорах совмещается с совершенно конкурентным поведением в рыночных играх. Аналогично в играх с добровольным финансированием общественных благ введение возможности санкций повышает вероятность кооперативного исхода, несмотря на то что использование санкций невыгодно для наказывающих	Игры с двусторонними переговорами, рыночные игры, кооперативные игры с санкциями и без санкций, игры “Диктатор” и “Дарообмен”	Некоторые агенты предпочитают более равномерные (“справедливые”) распределения выигрышей и издержек. Присутствие некоторой доли таких агентов в группе при определенных условиях может обеспечить кооперативное поведение всех игроков	Полезность индивида зависит от разницы между доходом каждого из игроков и его собственным. Зависимость асимметрична — индивид более подвержен зависти, нежели сочувствию
Bolton and Ockenfels, 2000	Поведение участников лабораторных экспериментов в различных играх подтверждает несколько конкурирующих гипотез. В зависимости от игры, наблюдается озабоченность соображениями справедливости, реципрокное или совершенно конкурентное поведение	Игры “Ультиматум”, “Диктатор”, “Дилемма заключенного”, “Дарообмен”, дуополии Бертрана	(Расширение Bolton, 1991) Помимо абсолютной величины собственного выигрыша, агенты предпочитают, чтобы их доля в доходе группы была близка к средней	Полезность индивида отрицательно зависит от разницы между его собственным и средним по группе выигрышем

Работа	Рассматриваемая проблема	Базовая модель/игра	Предлагаемое решение/гипотеза	Отличие агента модели от стандартной модели homo oeconomicus
Levine, 1998	Поведение участников лабораторных экспериментов в различных играх подтверждает несколько конкурирующих гипотез. В то время как в рыночных играх поведение успешно предсказывается гипотезой максимизации собственного выигрыша, в играх “Ультиматум” и играх с финансированием общественных благ наблюдается альтруизм или недоброжелательность	Игры “Ультиматум”, “Аукцион”, “Сороконожка”, игры с добровольным финансированием общественных благ	Агенты обладают индивидуальной предрасположенностью к альтруизму или недоброжелательности. Определяя свое отношение к выигрышу каждого из контрагентов, индивид также учитывает предполагаемое значение этого признака у них	Полезность индивида зависит от собственной предрасположенности к альтруизму или недоброжелательности, предрасположенности к альтруизму или недоброжелательности контрагентов, их выигрышей и восприимчивости индивида к чужому альтруизму или недоброжелательности
Rabin, 1993	Поведение участников лабораторных экспериментов в ряде игр отражает желание жертвовать собственным доходом для того, чтобы вознаградить или наказать контрагента за доброжелательное или недоброжелательное поведение по отношению к себе. Эмпирические исследования показывают, что такая реакция ослабевает в случаях, когда поведение контрагента было вынужденным	Игры с двумя участниками и полной информацией в нормальной форме	Агент пытается интерпретировать свои намерения по отношению к партнеру и намерения партнера по отношению к нему. При этом он отвечает доброжелательностью на то, что воспринимают как доброжелательность, и наоборот	Полезность индивида зависит от вер первого и второго порядков о намерениях партнера. Равновесия в игре с такими участниками определяются как пара стратегий, являющихся наилучшими ответами друг на друга, и пара наборов рациональных вер, поддерживающих эти стратегии

Работа	Рассматриваемая проблема	Базовая модель/игра	Предлагаемое решение/гипотеза	Отличие агента модели от стандартной модели homo oeconomicus
Dufwenberg and Kirchsteiger, 1998	В динамической игре модель, предложенная Рэбином (Rabin, 1993), предсказывает неоптимальное поведение на неиграемых ветвях. Помимо этого, в динамической игре веры игроков (и обусловленное ими реципрокное поведение) могут меняться на протяжении игры	Динамические игры с N участниками в развернутой форме	См. Rabin, 1993	Основное отличие от модели Рэбина (Rabin, 1993) состоит в том, что по достижении любой подыгры веры игрока меняются (в зависимости от того, как повел себя контрагент), и в любой подыгре он руководствуется теми верами, которые сформировались у него именно в начальном узле подыгры
Falk and Fischbacher, 1999	Поведение участников лабораторных экспериментов свидетельствует о значимости для людей как исходов игры, так и намерений партнера. Модели, учитывающие только один из этих двух факторов (например, Rabin, 1993 или Fehr and Schmidt, 1999), не в состоянии одновременно объяснить эти факты	Динамические игры с N участниками в развернутой форме	Для агента имеют значение как последствия выбранной партнером стратегии, так и его предполагаемые намерения	Полезность зависит от общей склонности агента к реципрокному поведению, склонности к реципрокному поведению в отношении каждого из партнеров, последствий их поведения (оцениваемых по разнице между собственным выигрышем и их выигрышем) и оценки доброжелательности их намерений

Работа	Рассматриваемая проблема	Базовая модель/игра	Предлагаемое решение/гипотеза	Отличие агента модели от стандартной модели homo oeconomicus
Charness and Rabin, 2002	Экспериментальные результаты проверки гипотезы о неприязни к неравенству на широком множестве игр свидетельствуют о том, что поведение игроков скорее обусловлено реципрокностью и озабоченностью благосостоянием группы, в особенности самых бедных ее членов. В этой связи для более надежного объяснения результатов широкого класса игр влияние на поведение агентов максиминимального критерия благосостояния группы и реципрокности целесообразно рассматривать вместе	Авторы не указывают конкретных игр, для анализа которых предназначена модель	Агент заботится об общем благосостоянии группы и отчасти — благосостоянии отдельных игроков, но при этом учитывает только выигрыши тех партнеров, которые ведут себя похожим образом	Полезность индивида зависит от суммарного выигрыша участников игры и выигрыша игрока, находящегося в наихудшем положении, однако при этом выигрышу каждого игрока присваивается вес, соответствующий степени, в которой сам этот игрок проявляет озабоченность интересами группы. Этот вес рассчитывается сравнением стратегии игрока с экзогенно заданным стандартом
Segal and Sobel, 1999	Стандартный микроэкономический набор аксиом, использующийся для характеристики индивидуальных предпочтений, не объясняет существование предпочтений, порождающих альтруистическое, недоброжелательное или реципрокное поведение при стратегическом взаимодействии	Игры с двумя участниками и полной информацией в нормальной форме	Агент обладает двумя наборами предпочтений — стандартными эгоистическими предпочтениями на множестве исходов и предпочтениями на множестве собственных смешанных стратегий, зависящих от стратегии контрагента	

Работа	Рассматриваемая проблема	Базовая модель/игра	Предлагаемое решение/гипотеза	Отличие агента модели от стандартной модели homo oeconomicus
Neilson, 2000	При выборе вариантов распределения ресурсов между членами группы, индивид зачастую стремится сохранить свой ранг в распределении. Использование стандартной аксиомы сепарабельности при моделировании предпочтений индивида не позволяет объяснить такое поведение	Широкий набор игр, где отдельные агенты способны влиять на распределение выигрышей	Предпочтения, по которым индивид оценивает разные варианты распределения, зависят от точки отсчета (личного выигрыша самого индивида)	Вместо стандартной аксиомы сепарабельности используется предложенная автором аксиома “сепарабельности по внутренней точке отсчета” (выигрышу индивида). В результате предпочтения индивида представляются функцией полезности, аддитивно сепарабельной по собственному выигрышу и разнице между собственным выигрышем и выигрышами других игроков
Karni and Safra, 2002a; Karni and Safra, 2002b	Стандартный микроэкономический набор аксиом, использующийся для характеристики индивидуальных предпочтений, не позволяет объяснить желание индивида жертвовать собственным благосостоянием для более справедливого распределения ресурсов между другими или его способность судить о “справедливости” механизмов распределения ресурсов	Распределение неделимого блага между несколькими индивидами с помощью лотереи	Агент обладает двумя наборами предпочтений — стандартными эгоистическими предпочтениями на множестве лотерей и так называемыми “предпочтениями справедливости” на том же множестве. При принятии решения традиционные эгоистические предпочтения и предпочтения справедливости взвешиваются	

Работа	Рассматриваемая проблема	Базовая модель/игра	Предлагаемое решение/гипотеза	Отличие агента модели от стандартной модели homo oeconomicus
Cooter, 1998	Теоретическое объяснение того, каким образом формальные и неформальные нормы способны стимулировать людей к рациональному развитию самоконтроля и совершенствованию своих предпочтений	Двухпериодная модель межвременного выбора. Нарушение нормы приносит выгоду в первом периоде и влечет санкции во втором	Если некий параметр в предпочтениях индивида колеблется, индивид способен спонтанно нарушать нормы, а потом об этом жалеть. Чтобы не испытывать внешних или внутренних (совесть) санкций, он может “инвестировать” в сокращение дисперсии этого параметра. Если же значение этого параметра наблюдается контрагентами, и эти сигналы влияют на сделки, которые предлагаются индивиду, последний может быть заинтересован инвестировать в изменение математического ожидания параметра	Предпочтения содержат случайную величину (ею может быть, например, норма межвременного замещения), на параметры распределения которой индивид может повлиять
Dowell, Goldfarb and Griffith, 1998	Предельный анализ во многих ситуациях выбора, имеющего нравственную подоплеку, бывает затруднен, так как с точки зрения морали множество вариантов выбора не является непрерывным	Модели выбора профессии	Нравственно обусловленные решения индивида дискретно меняют его бюджетное ограничение и полезность, получаемую от различных наборов благ. Например, изначально честный человек может заняться наркодилерством, но независимо от масштабов этой деятельности будет испытывать угрызения совести	Аргументы функции полезности включают дискретную переменную “честности”. С разными значениями этой переменной ассоциируются различные бюджетные ограничения

Работа	Рассматриваемая проблема	Базовая модель/игра	Предлагаемое решение/гипотеза	Отличие агента модели от стандартной модели homo oeconomicus
Kaplow and Shavell, 2001	С эволюционной точки зрения устойчивые нормы (в том числе моральные) должны способствовать выживанию сообщества (повышению общественного благосостояния). В этой связи каким образом можно объяснить использование нравственных чувств вины и добродетельности для обеспечения соблюдения различных моральных норм?	Модель производства экстерналий	Индивид может воздержаться от нежелательного поступка, потому что предчувствует угрызения совести (чувство вины) или потому что предвидит, как будет гордиться собой потом (добродетельность). Вина вычитается из общественного благосостояния, а добродетельность прибавляется к нему. Но психические способности людей реально переживать вину или ощущать добродетельность ограничены, и это необходимо учитывать при формулировке моральных норм	Агент способен ощущать внушенную обществом вину или добродетельность, совершая поступки, подпадающие под соответствующее правило
Laffont, 1975	В некоторых случаях объем производства отрицательных экстерналий в большой группе незнакомых друг с другом агентов ощутимо ниже того, что предсказывает гипотеза максимизации собственного выигрыша. В подобных ситуациях применение теорий кооперативного поведения, разработанных для небольшого числа агентов, затруднено ввиду их сложности или чрезмерно жестких предпосылок о доступности информации и возможности общаться	Модель совместного использования общего ограниченного ресурса (разновидность трагедии общин) с большим числом одинаковых агентов	Агенты знают об ограниченности общего ресурса и пользуются интернализированным кантианским категорическим императивом для оценки последствий собственного выбора	Индивид судит о действиях других с помощью кантианского категорического императива

Работа	Рассматриваемая проблема	Базовая модель/игра	Предлагаемое решение/гипотеза	Отличие агента модели от стандартной модели homo oeconomicus
Bilodeau and Gravel, 2004	Теоретически, с помощью интернализированного кантианского категорического императива можно обеспечивать исходы, которые Парето-доминируют некооперативные равновесия, например, при добровольном финансировании общественных благ. Однако определение “кантианских правил” затруднительно, если сообщество неоднородно	Модель экономики с общественным благом, производство которого финансируется исключительно добровольно	Кантианские правила должны предписывать каждому морально эквивалентное (а не строго одинаковое) поведение. Чтобы индивид был заинтересован возвести кантианское правило “в ранг универсального закона”, оно должно гарантировать ему наивысший платеж при морально эквивалентном поведении остальных	Нет
Kjell, Kverndokk and Nyborg, 2003	Для объяснения добровольного финансирования общественных благ был предложен ряд моделей, однако они не объясняют, почему введение дополнительных внешних стимулов (например, вознаграждение за участие в производстве общественного блага) иногда отрицательно влияет на объем усилий	Модель экономики с общественным благом и государством. Общественное благо финансируется государством и однородными потребителями (за счет своего свободного времени), причем государство может влиять на технологию производства общественного блага	Участие индивидов в производстве общественного блага объясняется их желанием считать себя социально ответственными членами общества. Уровень “социальной ответственности” определяется на основе эндогенного эталона, на который могут влиять внешние факторы, в том числе стимулы	Полезность индивида зависит от “индекса социальной ответственности”. Индекс рассчитывается на основе разницы между вложением индивида в финансирование общественного блага, и “морально безупречным” вложением. Размер последнего индивид рассчитывает, максимизируя общественное благосостояние в предположении, что остальные вложат в производство общественного блага столько же, сколько и он

Литература

1. Agell J., Lundborg P. Theories of Pay and Unemployment: Survey Evidence from Swedish Manufacturing Firms // Scandinavian Journal of Economics. 1995. No. 97. P. 295−308.

2. Alm J., Sanchez I., Juan de A. Economic and Noneconomic Factors in Tax Compliance // Kyklos. 1995. No. 48. P. 3−18.

3. Andreoni J. Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving // The Economic Journal. 1990. No. 100. P. 464—477.

4. Andreoni J., Erard B., Feinstein J. Tax Compliance // Journal of Economic Literature. 1998. No. 36. P. 818−860.

5. Arrow K. Optimal and Voluntary Income Redistribution // Economic Welfare and the Economics of Soviet Socialism: Essays in Honor of Abram Bergson / Ed. by S. Rosenfield. Cambridge: Cambridge University Press, 1981.

6. Becker G. A Theory of Social Interactions // Journal of Political Economy. 1974. No. 82. P. 1063−1093.

7. Bewley T. Why Wages Don't Fall During a Recession. Harvard: Harvard University Press, 1999.

8. Bilodeau M., Gravel N. Voluntary Provision of a Public Good and Individual Morality // Journal of Public Economics. 2004. No. 88. P. 645—666.

9. Bolton G., Ockenfels A. A Theory of Equity, Reciprocity and Competition // American Economic Review. 2000. No. 100. P. 166−193.

10. Bolton G. A Comparative Model of Bargaining: Theory and Evidence // American Economic Review. 1991. No. 81. P. 1096−1136.

11. Bordignon M. Was Kant right? Voluntary Provision of Public Goods Under the Principle of Unconditional Commitment // Economic Notes. 1990. No. 3. P. 342—372.

12. Bowles S., Gintis H. The Evolution of Strong Reciprocity. University of Massachusetts at Amherst, 1999 (Mimeo).

13. Brekke K., Kverndokk S., Nyborg K. An Economic Model of Moral Motivation // Journal of Public Economics. 2003. No. 87. P. 1967—1983.

14. Buchanan A. Ethics, Efficiency and the Market. Totowa (NJ): Rowman and Allanfeld, 1985.

15. Charness G., Matthew R. Social Preferences: Some Simple Tests and a New Model. University of California at Berkeley, 2000 (Mimeo).

16. Charness G., Rabin M. Understanding Social Preferences with Simple Tests // Quarterly Journal of Economics. 2000. No. 117. P. 817—869.

17. Cooter R. Models of Morality in Law and Economics: Self-Control and Self-Improvement for the "Bad Man" of Holmes. Berkeley Program in Law & Economics, 1998. Working Paper No. 135.

18. Curry Ph., Mongrain S. What you don't see can't hurt you: an Economic Analysis of Morality Laws. American Law & Economics Association Annual Meetings, 2004. Paper 48.

19. Dowell R., Goldfarb R., Griffith W. Economic Man as a Moral Individual // Economic Inquiry. 1998. No. 36. P. 4; ABI/INFORM Global. P. 645.

20. Dufwenberg M., Kirchsteiger G. A Theory of Sequential Reciprocity. Discussion Paper. CentER, Tilburg University, 1998.

21. Etzioni A. The Moral Dimension: Toward a New Economics. N.Y.: Macmillan, 1988.

22. Etzioni A. The Case for a Multiple-Utility Conception // Economics and Philosophy. 1986 (October). P. 159—183.

23. Falk A., Fischbacher U. A Theory of Reciprocity. Institute for Empirical Research in Economics. University of Zurich, 1999. Working Paper No. 6.

24. Falk A., Fehr E., Fischbacher U. Appropriating the Commons. Institute for Empirical Research in Economics. University of Zurich, 2000. Working Paper No. 55.

25. Falk A., Fehr E., Fischbacher U. Informal Sanctions. Institute for Empirical Research in Economics. University of Zurich, 2000. Working Paper No. 59.

26. Falk A., Fehr E., Fischbacher U. Testing Theories of Fairness — Intentions Matter. Institute for Empirical Research in Economics. University of Zurich, 2000. Working Paper No. 63.

27. Fehr E., Schmidt K. A Theory of Fairness, Competition and Cooperation // Quarterly Journal of Economics. 1999. No. 114. P. 817—868.

28. Fehr E., Schmidt K. Theories of Fairness and Reciprocity − Evidence and Economic Applications. Institute for Empirical Research in Economics. University of Zurich, 2001. Working paper No. 75.

29. Fehr E., Gächter S., Kirchsteiger G. Reciprocity as a Contract Enforcement Device // Econometrica. 1997. No. 65. P. 833−860.

30. Frey B.S., Goette L. Does Pay Motivate Volunteers? Institute for Empirical Research in Economics. University of Zurich, 1999 (November). Working Paper Series No. 7.

31. Frey B.S., Oberholzer-Gee F. The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out // American Economic Review. 1997. No. 87 (4). P. 746—755.

32. Frey B.S., Weck-Hannemann H. The Hidden Economy as an "Unobserved" Variable // European Economic Review. 1984. No. 26. P. 33—53.

33. Frey B.S., Jegen R. Motivation Crowding Theory: A Survey of Empirical Evidence. Institute for Empirical Research in Economics. University of Zurich, 1999. Working Paper.

34. Geanakoplos J., Pearce D., Stacchetti E. Psychological Games and Sequential Rationality // Games and Economic Behavior. 1989. No. 1. P. 60—79.

35. Gneezy U., Rustichini A. Pay Enough or don't Pay at all // Quarterly Journal of Economics. 2000. CXV (3). P. 791—810.

36. Greenberg J. Employee Theft as a Reaction to Underpayment Inequity: The Hidden Cost of Pay Cuts // Journal of Applied Psychology. 1990. No. 75. P. 56—568.

37. Hamlin A. Ethics, Economics, and the State. N.Y.: St. Martin's Press, 1986.

38. Hausman D., McPherson M. Taking Ethics Seriously: Economics and Contemporary Moral Philosophy // Journal of Economic Literature. 1993. Vol. 31. No. 2. P. 671—731.

39. Hausman D., McPherson M. Economic Analysis and Moral Philosophy. Cambridge: Cambridge University Press, 2002.

40. Kahneman D., Tversky A. Prospect Theory: An Analysis of Decision Under Risk // Econometrica. 1979. No. 47. P. 263—291.

41. Kahneman D., Knetsch J., Thaler R. Fairness as a Constraint on Profit Seeking: Entitlements in the Market // American Economic Review. 1986. LXXVI. P. 728−741.

42. Kaplow L., Shavell S. Moral Rules and Moral Sentiments: Toward a Theory of an Optimal Moral System. 2001. NBER Working Paper 8688.

43. Karni E., Safra Z. Individual Sense of Justice: A Utility Representation // Econometrica. 2002. No. 70. P. 1; ABI/INFORM Global. P. 263.

44. Laffont J.-J., Laroque G. Effets Externs et Théorie de l'équilibre General. Cahiers du Séminaire d'Econométrie. Paris: CNRS, 1972.

45. Laffont J.-J. Macroeconomic Constraints, Economic Efficiency and Ethics: An Introduction to Kantian Economics // Economica. New Series. 1975. Vol. 42. No. 168. P. 430—437.

46. Levine D. Modeling Altruism and Spitefulness in Experiments // Review of Economic Dynamics. 1998. No. 1. P. 593−622.

47. Lind A., Tyler T. The Social Psychology of Procedural Justice. N.Y.; L.: Plenum Press, 1988.

48. Neilson W., Stowe J. Choquet Other-Regarding Preferences. Texas A&M University, 2004 (Manuscript).

49. Neilson W. An Axiomatic Characterization of the Fehr-Schmidt Model of Inequity Aversion. Department of Economics. Texas A&M University, 2000.

50. Nyborg K., Howarth R., Brekke K. Green Consumers and Public Policy: on Socially Contingent Moral Motivation. Working Paper No. 31. University of Oslo, Dept. of Economics, 2003.

51. Ok E., Kockesen L. Negatively Interdependent Preferences // Social Choice and Welfare. 2000. No. 17. P. 533—558.

52. Ostrom E. Collective Action and the Evolution of Social Norms // Journal of Economic Perspectives. 2000. No. 14. P. 137—158.

53. Rabin M. Incorporating Fairness into Game Theory and Economics // American Economic Review. 1993. No. 83 (5). P. 1281—1302.

54. Roth A., Prasnikar V., Okuno-Fujiwara M. et al. Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study // American Economic Review. 1991. No. 81. P. 1068—1095.

55. Samuelson P. Altruism as a Problem Involving Group Versus Individual Selection in Economics and Biology // American Economic Review. 1993. No. 83. P. 143−148.

56. Sandbu M. Axiomatic Foundations for Reference Dependent Distributive Preferences. Harvard University, 2003 (Manuscript).

57. 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.

58. Sen A. Choice, Orderings and Morality // Practical Reason / Ed. by S. Korner. New Haven: Yale University Press, 1974. P. 54—67.

59. Sen A. Rational Fools: A Critique of the Behavioral Foundations of Economic Theory // Philosophy and Public Affairs. 1977. No. 6. P. 317—344.

60. Sen A. On Ethics and Economics. Oxford: Basil Blackwell, 1987.

61. Sen A. Moral Codes and Economic Success // Market Capitalism and Moral Values / Ed. by C.S. Britten, A. Hamlin. Edward Eldar, Aldershot, 1995.

62. Sethi R., Somananthan E. Understanding Reciprocity. Columbia University, 2000 (Mimeo).

63. Sethi R., Somananthan E. Preference Evolution and Reciprocity // Journal of Economic Theory. No. 97. P. 273—297.

64. Shavell S. Foundations of Economic Analysis of Law. L.: Belknap Press, 2004.

65. Sugden R. Reciprocity: The Supply of Public Goods Through Voluntary Contributions // The Economic Journal. 1984 (December). Vol. 94. No. 376. P. 772—787.

66. Wittman D. Liability for Harm or Restitution for Benefit? // Journal of Legal Studies. 1984. No. 13. P. 57—80.

67. Young D.J. A "Fair share" Model of Public Good Provision // Journal of Economic Behavior and Organization. 1989. No. 11. P. 137—147.

68. Zajac E. Political Economy of Fairness. Cambridge (Mass.): MIT Press, 1995.

Оглавление

Введение 3

Моделирование чувства справедливости на основе “социальных предпочтений” 7

Моделирование чувства справедливости с помощью реципрокности, основанной на намерениях 15

Аксиоматические подходы к моделированию чувства справедливости 21

Моделирование нравственного самосовершенствования и принятия решений, влияющих на образ жизни 27

Моделирование работы нравственных чувств как инструмента контроля за производством экстерналий 34

Применение кантианских моральных правил при контроле за производством экстерналий и добровольном финансировании общественных благ 41

Добровольное финансирование общественных благ, обусловленное эндогенной моральной нормой 51

Заключение 57

Приложение 61

Литература 70

Препринт WP3/2006/06
Серия WP
Проблемы рынка труда

Ю.В. Автономов

Моделирование морали как элемента внутренней мотивации индивидов и механизма коррекции провалов рынка

Публикуется в авторской редакции

Выпускающий редактор *А.В. Заиченко*
Технический редактор *Ю.Н. Петрина*

ЛР № 020832 от 15 октября 1993 г.
Отпечатано в типографии ГУ ВШЭ с представленного оригинал-макета.
Формат 60×84$^{1}/_{16}$. Бумага офсетная. Тираж 150 экз. Уч.-изд. л. 5.
Усл. печ. л. 4,4. Заказ № . Изд. № 623.

ГУ ВШЭ. 125319, Москва, Кочновский проезд, 3
Типография ГУ ВШЭ. 125319, Москва, Кочновский проезд, 3
Тел.: (495) 772-95-90; 772-95-73

Для заметок