

А.Л.Блинов

Синергетика коллективной иррациональности

Синергетика — учение о сотрудничестве
*Р.Грэхем и Х.Хакен*¹

Синергетический пролог

В приложениях математической теории игр к экономике, социологии и политологии проблема сотрудничества проходит под именем «проблемы субоптимальности» или, по названию самой знаменитой субоптимальной ситуации, «проблемы Дилеммы заключенных». Никогда еще, насколько я знаю, эта последняя не рассматривалась в связи с междисциплинарным полем синергетики — как ее понимает Герман Хакен. Более того, высказывается мнение, что именно математическая теория игр остается наиболее антисинергетической дисциплиной в современной математике — ибо игнорирует нелинейные феномены².

Тем интереснее попробовать подготовить почву для включения проблематики субоптимальных ситуаций в поле зрения синергетики. Данная статья предназначена сделать шаг в этом направлении.

Рациональность как максимизация полезности

Что такое рациональное действие? Ответов наверняка имеется больше одного. Сосредоточимся на том, что стал центральным (если не единственным) в экономическом теоретизировании. Впрочем, и в других социальных науках он пользуется все большей популярностью. Его технический ярлык — «инструментальная рациональность».

Инструментальная рациональность — это соответствие средств цели. Цель принимается как данное, ее разумность не обсуждается. Обсуждать и выбирать можно средства — действия,

ведущие (или не ведущие) к достижению цели. Если действие приводит к заданной цели — оно инструментально рационально; если нет — нет. На русском языке инструментальную рациональность можно приблизительно охарактеризовать одним словом «целесообразность».

В чуть более общем и более удобном для многих приложений варианте вместо цели говорят об упорядочении возможных будущих положений дел (состояний мира) по их предпочтительности. Чем к более предпочтительному положению дел приводит то или иное действие, тем более оно инструментально рационально.

Можно ввести еще одно техническое удобство — репрезентировать порядок предпочтительности приписыванием каждому положению дел числа, отражающего степень полезности данного положения. Это удобно и допустимо, если не забывать, что (1) одному и тому же порядку предпочтительности соответствует бесконечное множество различных численных функций полезности; (2) сравнение численных величин полезности между разными людьми в принципе проблематично. Для наших целей оба проблематичных пункта не представляют опасности, и мы далее будем говорить о численной репрезентации полезностей. В этом случае смысл инструментальной рациональности можно передать так: инструментально рационально то действие, которое максимизирует полезность. Инструментальная рациональность есть максимизация полезности. Нужно отдавать себе отчет, что понятию инструментальной рациональности соответствует модель рационального действия, а всякая модель связана с идеализациями и упрощениями. Что же может дать нам модель, связанная с инструментальной рациональностью, и где проходят границы ее возможностей? Там, где имеет смысл рассматривать и оценивать действия одного изолированного деятеля («Робинзона»), — она дает очень много: смотри любой учебник по теории принятия решений (одного лица).

Робинзон встречается с Пятницей

Но как только мы переходим от одинокого деятеля к взаимодействию двух или нескольких лиц, — скажем, как только Робинзон встречается с Пятницей, — ситуация, как правило, выходит из-под контроля модели максимизации полезности (во всяком случае в том виде, как она формально разработана для случая одного деятеля). Оказывается, что при этом переходе вся

концепция рациональности как максимизации полезности сталкивается с принципиальными, драматическими и в общем-то предвидимыми трудностями.

Впрочем, легко предвидеть их задним числом, в конце 20-го века. Бентамов и Миллев утилитаризм, к примеру, никаких таких трудностей не предвидел. А между тем утилитаризм был, конечно, философским предтечей формальной модели рациональности как максимизации полезности. Не будет большим преувеличением сказать, что модель эта есть одна из возможных экспликаций ключевой идеи утилитаризма — действия должно оценивать по их последствиям.

Основатель утилитаризма Джереми Бентам (1748-1832) непостижимым (для нас, людей начала XXI века) образом смешал в одно (противоречивое) целое, во-первых, вышеупомянутый методологический принцип оценки действий по их последствиям — причем последствиям сугубо индивидуальным: боли и удовольствию, причиненным деятелю совершенным им действием, — и во-вторых, знаменитый этический принцип, гласящий, что то действие моральнее (и следовательно, предпочтительнее) всех прочих, последствием которого является «наибольшее счастье наибольшего числа людей». Бентам, таким образом, был уверен, что он успешно и без особых проблем совершил тот переход, который мы абзацем выше сочли в высшей степени проблематичным, — переход от максимизации полезности одиночного деятеля к максимизации полезности многих людей.

Насколько я могу судить, первым увидел здесь теоретическую проблему Джон фон Нейман, основатель (в том числе) теории игр, который во Введении к своему основополагающему труду «Теория игр и экономического поведения» (1944, в соавторстве с Оскаром Morgenштерном) писал:

«Проблема Робинзона Крузо (т.е. проблема максимизации полезности, стоящая перед одиночным деятелем — А.Б.) — это обычная проблема максимизации, и трудности ее носят чисто технический, а не концептуальный характер. (С другой стороны, проблема, с которой сталкивается индивид в ситуации взаимодействия с другими индивидами — А.Б.) — это уж точно не проблема максимизации, а какая-то странная и приводящая в замешательство смесь нескольких проблем максимизации, находящихся в конфликте друг с другом. Каждый участник действует, исходя из своего собственного, не контролирующего весь набор переменных, затрагивающих его интересы. Классическая математика никогда не сталкивалась с проблемой такого рода»³.

Обобщение Принципа максимизации на ситуацию нескольких деятелей: Парето-оптимальность

Итак, как решить проблему перехода от одинокого Робинзона к взаимодействию Робинзона с Пятницей? Первое, что приходит в голову, — это идея естественного обобщения Принципа максимизации индивидуальной полезности на ситуацию нескольких взаимодействующих агентов.

Применительно к одинокому Робинзону Принцип максимизации индивидуальной (МИ) полезности гласит:

МИ-Робинзон! Если действие R_1 принесет тебе большую полезность, чем действие R_2 , предпочти действие R_1 действию R_2 .

Обобщение этого принципа на ситуацию взаимодействия Робинзона и Пятницы (МК) могло бы звучать примерно так:

МК-Робинзон и Пятница! Если пара ваших действий R_1 и P_1 принесет каждому из вас большую полезность, чем пара ваших действий R_2 и P_2 , предпочтите первую пару второй.

Есть ситуации, создающие впечатление, что Принцип МК может служить руководством к действию. Рассмотрим одну из них:

		Пятница	
		P_1	P_2
Робинзон	R_1	4; 4	3; 3
	R_2	2; 2	1; 11

Рис. 1

В ситуации, схематически представленной на рис. 1, у каждого из двух деятелей есть выбор между двумя действиями: у Робинзона — между R_1 и R_2 ; у Пятницы — между P_1 и P_2 . В каждой клетке матрицы слева от точки с запятой указана численная величина полезности, которую получит Робинзон, если будут совершены два соответствующих действия; справа — величина полезности, получаемой Пятницей. Например, если Робинзон совершит действие R_2 , а Пятница — P_1 , то Робинзон получит две единицы полезности, и Пятница получит также две единицы. Однако, конечно, это был бы не самый рациональный выбор для обоих. Если бы они следовали Принципу МК, Робинзон должен был бы выбрать действие R_1 , а Пятница — P_1 . Это дало бы каждому из них по 4 единицы полезности, а 4 больше 2.

Таким образом, результат действий R_1 и P_1 для каждого из них предпочтительней, чем результат действий R_2 и P_1 . На жаргоне теории принятия решений говорят, что результат действий R_1 и P_1 есть Парето-улучшение результата действий R_2 и P_1 . Исход, который уже не может быть Парето, — улучшен, называется Парето-оптимальным.

Принцип МК, стало быть, можно назвать Принципом коллективного предпочтения Парето-оптимальных ситуаций. «Коллективного» — потому что ни Робинзону, ни Пятнице не под силу реализовать этот принцип в одиночку; он обращен сразу к ним обоим: «Робинзон и Пятница, предпочтите такие-то свои действия таким-то!»

Провал Парето-оптимальности: Соблазн предательства

Но всегда ли дела обстоят так гладко? Рассмотрим следующую ситуацию:

		Пятница	
		P_1	P_2
Робинзон	R_1	80; 80	2; 100
	R_2	100; 2	4; 4

Рис. 2

Исход $\langle 4; 4 \rangle$ допускает Парето-улучшение $\langle 80; 80 \rangle$; это — единственный не Парето-оптимальный исход ситуации рис. 2. Стало быть, первая проблема, с которой сталкиваются попытки применить МК к ситуации рис. 2, состоит в том, что здесь не один, а целых три Парето-оптимальных исхода. К которому из них должны стремиться Робинзон и Пятница? Сам принцип МК об этом умалчивает.

Но предположим, что у наших двух деятелей есть (1) здравый смысл; (2) возможность переговорить друг с другом перед тем, как сделать свой выбор. Переговоры проходили бы, пожалуй, примерно так.

Робинзон: Послушай, Пятница, тут у нас с тобой оказалось целых три Парето-оптимальных исхода. Какой будем выбирать?

Пятница: Только не $\langle 100; 2 \rangle$.

Робинзон: И не $\langle 2; 100 \rangle$.

Пятница: Стало быть, $\langle 80; 80 \rangle$?

Робинзон: Само собой!

Пятница: Стало быть, я выбираю действие P_1 .

Робинзон: А я — R_1 .

Теперь настало время заметить, что концепция инструментальной рациональности включает в себя два фундаментальных пункта: (1) положение о том, что деятель всегда стремится максимизировать величину своей индивидуальной полезности; (2) допущение о том, что он обладает идеальными (=неограниченными и непогрешимыми) интеллектуальными способностями: (а) он знает (или, в слегка ослабленном варианте: способен узнать) о своей ситуации все, что нужно знать для вычисления действия, максимизирующего полезность; (б) он рассуждает логически непогрешимым образом; (в) он умеет безошибочно осуществить любое вычисление за пренебрежимо малое время.

Конечно, условия (а)-(в) — идеализации, недостижимые в реальной жизни. Но это, по-видимому, те естественные и эффективные идеализации, без которых не может обойтись ни одна наука. Они эффективны в том смысле, что допускают постепенное аппроксимирование к реальным условиям, сопровождающееся усложнением технического аппарата теории. Но начинать теорию естественно с наиболее простых случаев.

Можно сказать, что условия (а)-(в) призваны гарантировать идеальную когнитивную рациональность деятеля. Когнитивную рациональность не следует смешивать с инструментальной рациональностью. Инструментальная рациональность деятеля состоит в его стремлении (=решимости) максимизировать величину получаемой им полезности. Его когнитивная рациональность состоит в том, что для надобностей максимизации полезности он способен узнать, рассудить и вычислить все, что ему нужно.

Когда речь идет о взаимодействии нескольких людей, а не об одиночном деятеле, то обычно вводят еще одно допущение (3) — допущение о взаимном знании: все участники ситуации взаимодействия знают, что все они инструментально и когнитивно рациональны; и все они знают, что все они это знают; и все они знают, что все они знают, что все они это знают... И т.д. *ad infinitum*.

Теперь, дорогой читатель, с этими тремя условиями в руках взглянем на положение, в котором оказались Робинзон и Пятница после проведения переговоров.

Какое действие должен выбрать инструментально и когнитивно рациональный Робинзон?

Вы скажете: « R_1 , ибо так они договорились с Пятницей».

А я скажу: « R_2 , ибо Робинзон, по допущению, стремится максимизировать величину своей (а вовсе не Пятницевой) полезности (=инструментальная рациональность!) и умеет логически рассуждать (когнитивная рациональность!)».

Умея рассуждать, Робинзон рассудит так (следите за логикой!):

Посылка 1: Пятница выберет либо P_1 , либо P_2 . Других возможностей выбора у него нет.

Посылка 2: Если Пятница выберет P_1 , то мое действие R_2 принесет мне больше полезности, чем P_2 (ибо $100 > 80$).

Посылка 3: Если Пятница выберет P_2 , то мое действие R_2 принесет мне больше полезности, чем P_2 (ибо $4 > 2$).

Заключение 1: Стало быть, в любом случае мое действие R_2 максимизирует полезность для меня.

Посылка 4: Я стремлюсь максимизировать полезность.

Заключение 2: Следовательно, я должен выбрать R_2 . Это — мой рациональный выбор.

Но Пятница, по допущению, инструментально и когнитивно рационален не менее Робинзона, а поскольку ситуация симметрична, он будет рассуждать в точности как Робинзон. Стало быть, он выберет P_2 . Результатом пары действий P_2 и R_2 будет исход $\langle 4; 4 \rangle$ — единственный не Парето-оптимальный исход из четырех возможных.

Заметьте, что к этому Парето-субоптимальному исходу (и к нарушению данных друг другу обещаний) наших героев привела их (1) инструментальная рациональность и (2) идеальная когнитивная рациональность.

Ситуация, изображенная на рис. 2, представляет собой, таким образом, трагический парадокс инструментальной и когнитивной рациональности: парадокс — ибо героям этой ситуации не удалось максимизировать свои полезности именно потому, что они стремились их максимизировать, используя всю необходимую информацию и свои неограниченные интеллектуальные возможности; трагический — потому, что они не просто остались у разбитого корыта, но еще и совершили взаимное предательство, нарушив данное друг другу слово.

Устранение соблазна: Выбор Нэш-равновесия

Если исход p интерактивной ситуации с таков, что коллективный договор сторон о выборе действий, ведущих к p , способен породить хотя бы у одной из сторон соблазн предательства, то на техническом жаргоне p называется Нэш-неравновесным исходом*. Если все участники ситуации инструментально и когнитивно рациональны, то им договариваться о совершении действия, ведущих к Нэш-неравновесному исходу, — только попусту время терять, ибо они сами знают, что те из них, в ком Нэш-неравновесный исход порождает соблазн предательства, — предадут; и Нэш-неравновесный исход не будет достигнут.

Договариваться им, стало быть, имеет смысл только о Нэш-равновесных исходах; но если в ситуации с таковых имеется ровно один, то предварительные переговоры вообще теряют смысл, ибо их результат предрешен, и все участвующие стороны знают это. Зная результат возможных переговоров, стороны и безо всяких переговоров выберут действия, ведущие к единственному Нэш-равновесному исходу — ведь Нэш-равновесный договор не порождает соблазна предательства; в этом и состоит его равновесие.

Если в ситуации с имеется единственный Нэш-равновесный исход p , то p называется решением (solution) ситуации s . Решение ситуации есть тот ее исход, к которому придут инструментально и когнитивно рациональные деятели, т.е. реалистически мыслящие максимизаторы своей полезности.

Существование субоптимальных ситуаций

И вот данные рис. 2 свидетельствуют, что существуют такие интерактивные ситуации, решение которых Парето-субоптимально. Будем впредь называть сами такие ситуации Парето-субоптимальными.

Вдумаемся еще раз в суть Парето-субоптимальной ситуации — например, той, что представлена на рис. 2. Стороны хотели бы прийти к чему-то лучшему, чем $\langle 4; 4 \rangle$ — и, абстрактно говоря, они вполне могут получить результат $\langle 80; 80 \rangle$ (в 20 раз лучший для каждого чем $\langle 4; 4 \rangle$): для этого достаточно, чтобы каждый из них выбрал соответствующие действия: P_1 и R_1 . Но они выберут P_2 и R_2 — и они знают это. Почему? Потому что они идеально разумны и умеют считать и за себя, и за партнера.

* Джон Нэш ввел это понятие, которое, впоследствии стало центральным в математической теории игр.

Были бы они поглупее и поневежественнее в арифметике и логике — вполне могли бы договориться о $\langle 80; 80 \rangle$ и реализовать договор на деле! Горе от ума!

— Да, да, дорогой читатель, в этом месте я слышу твой негодующий крик. Я даже слышу два твоих крика:

Первый: — Да это все потому, что они оба — отъявленные эгоисты. Не ума у них избыток, а совести и уважения к интересам партнера недостаток.

Второй: — Бросьте вы! Это все — абстрактные модели, пустая арифметика. В жизни так не бывает, и никакого интереса для реальной жизни ваши субоптимальные ситуации не представляют.

Отвечу на оба крика по очереди:

Ответ на крик 1: Субоптимальность для альтруистов

Модель максимизации полезности разделяет со всеми «просвещенными» вариантами гедонизма и утилитаризма, — например, с гедонизмом Эпикура и утилитаризмом Джона Милля, — ту особенность, что положительные ценности человека (удовольствие, полезность и т.п.) не обязаны быть узко эгоистическими, а вполне могут быть отражением альтруизма их носителя. Так в нашем конкретном случае, представленном на рис. 2, тот факт, что исход, происходящий от выбора действия P_2 и R_1 , приносит Робинзону максимальный уровень полезности: 100 единиц, — этот факт совершенно нейтрален по отношению к вопросу о том, является ли Робинзон эгоистом или альтруистом или какой-либо смесью того и другого, потому что численный узор ситуации, отраженный в матрице рис. 2, не несет никакой информации о том, чем именно Робинзону так дорог этот исход, — тем ли, что он сможет сам, единолично, потребить какую-либо особенно изысканную корзину благ (эгоизм) или же тем, что данный исход спасет жизнь и здоровье близкого (или даже дальнего!) ему человека, ухудшив одновременно здоровье и материальное благосостояние самого Робинзона (альтруизм). Так что в субоптимальной ситуации могут оказаться и завязтые альтруисты.

Ответ на крик 2: Дилемма заключенных и ее роль в жизни общества

Интерактивные ситуации типа той, что представлена на рис. 2, настолько центральны для математической теории игр и ее экономических и социологических приложений, что сам этот

тип ситуации удостоился имени собственного. Он был обнаружен и описан американцем Мериллом Фладом в 1952 году и впоследствии стал известен в литературе под именем «Дилемма заключенных».

Имя происходит от популярной иллюстрации, кочующей из учебника в учебник. Двое заключенных подозреваются в совместном совершении тяжкого преступления. Прямых улик нет, но есть веские косвенные. Следователь предлагает каждому из двоих (они сидят в разных одиночных камерах) одну и ту же сделку: «Если расскажешь все сам, а твой подельник будет молчать, сразу выйдешь из тюрьмы за помощь следствию, а тот получит 10 лет. Расскажете оба — каждому по девять лет за совершение тяжкого преступления. Будете оба заперты — что делать, прямых улик нет, получите по году каждый за незаконное ношение оружия».

Рациональность повелевает каждому чистосердечно сотрудничать со следователем. Оба делают это и получают по девять лет. А могли бы сидеть по году, если бы оба молчали. Историю делает выпуклой правдоподобность психологических мотивов предательства, но совершенно не отражает колоссальную роль, которую ситуации типа Дилеммы заключенных — субоптимальные ситуации вообще — играют в жизни общества.

Паскалево пари

Я хочу теперь поговорить об одном средстве, упоминаний о котором я не встречал в литературе, — во всяком случае, не встречал упоминаний о нем в связи с Дилеммой заключенных и с проблемой субоптимальности вообще.

Сама идея средства, которое я имею в виду, структурно очень проста: если именно рациональность — идеальная, совершенная, непогрешимая рациональность (инструментальная и когнитивная, взятые вместе) — толкает нас в ДЗ-ситуации на путь предательства и неблагополучия (горе от ума!), то чтобы избежать несчастья, надо, стало быть, поступиться рациональностью или хотя бы какой-то ее частью или совершенством ее качества и т.д.

Подобные идеи уже звучали в истории человеческой мысли. Самый яркий исторический пример — знаменитое пари Паскаля.

Предположим, что человеку надо решить вопрос о своем отношении к вере в Бога.

Один из возможных подходов — попытаться решить этот вопрос на основе как можно более совершенной когнитивной рациональности. Самая совершенная из доступных человеку парадигм когнитивной рациональности — наука и ее методы. Однако в вопросе о существовании Бога наука не способна дать окончательного ответа: она не способна построить методологически непогрешимое доказательство ни существования Бога, ни несуществования Бога. Более того, она не способна дать нам непогрешимые резоны даже для того, чтобы прагматически склонить нас в сторону положительного или отрицательного ответа. Таким образом, самая совершенная парадигма когнитивной рациональности в вопросе о Боге рекомендует абсолютно симметричный скептицизм.

И вот Паскаль делает совершенно иной заход — он пытается подойти к вопросу о религиозной вере на основе инструментальной (а не только когнитивной!) рациональности⁴. Содержание Паскалева аргумента можно передать с помощью следующей матрицы вероятностей и полезностей:

		Возможные состояния мира	
		Бог существует [вероятность=0.5]	Бог не существует [вероятность=0.5]
Возможные действия	Действие 1 Привести себя в состояние веры в Бога	Вечная блаженная жизнь	Статус кво: ничего не приобретаешь, но и ничего не теряешь
	Действие 2 Остаться как есть — атеистом или скептиком	Потеря вечной блаженной жизни	Статус кво: ничего не приобретаешь, но и ничего не теряешь

Рис. 3

Суть аргумента в том, что численная величина полезности исхода «Вечная блаженная жизнь» столь велика (она, по предположению Паскаля, равна бесконечности), что ожидаемая полезность действия 1 больше ожидаемой полезности действия 2 не только при равновероятности существования и несуществования Бога, но и при сколь угодно малой, но ненулевой веро-

ятности Его существования. Поскольку, как мы помним, инструментальная рациональность состоит в максимизации полезности, для деятеля инструментально рационально предпочесть действие 1 действию 2.

Но, конечно же, все своеобразие Паскалева аргумента в том, что действие 1 совершенно необычно. Оно состоит не в воздействии индивида на внешний мир, а в его воздействии (возможно, не прямым, а косвенным) на свое собственное когнитивное состояние. Предположим, что перед совершением действия 1 индивид обладал полной — насколько это доступно человеку — когнитивной рациональностью. Она, как мы видели, рекомендует человеку в вопросе о Боге оставаться в состоянии абсолютно симметричного скептицизма. Поскольку по совершении действия 1 человек отбросит первоначальный скептицизм и будет верить в Бога, постольку действие 1 вступает в конфликт с когнитивной рациональностью — и тем не менее оно, это действие, инструментально рационально.

Стало быть, чтобы успешно совершить действие 1, индивид должен быть готов пожертвовать какой-то частью или каким-то качеством своей первоначальной когнитивной рациональности. Он должен быть готов к тому, что в процессе совершения действия 1 его когнитивная рациональность окажется урезанной или ущербленной.

Паскаль полностью отдает себе отчет и в вытекающей из выбора действия 1 необходимости ущербления интеллекта, и в том, что даже если человек готов пожертвовать полнотой своего интеллекта, — заставить себя верить вопреки своему разуму нелегко, если вообще возможно. Он говорит своему воображаемому оппоненту-скептику: «Раз человек все равно вынужден играть *в эту игру*, он должен отказаться от разума, чтобы сохранить жизнь.» «В этом я согласен с вами..., — отвечает оппонент. — Но мне от этого не легче — я так уж создан, что не могу верить, что мне в таком случае, по-вашему, делать?». Ответ Паскаля поистине замечателен: «...*Есть люди, которые уже познали тот путь, на который вам следует вступить...* Следуйте их примеру; ведите себя так, как если бы вы уже верили, принимайте освященную воду, участвуйте в богослужениях и т.д. Уже одно это естественным образом приведет вас к вере и притупит остроту вашего ума (курсив мой — А. Б.)»⁵.

«И притупит остроту вашего ума...» — в этом-то все дело: отказаться от полноты интеллекта, чтобы избежать несчастья. Паскаль применил этот прием в ситуации одиночного деятеля

ля — т.е. в контексте теории принятия решений одним лицом. Чтобы создать нужду в столь сильном лекарстве — добровольном самооглувлении, Паскалю пришлось обратиться к исключительной, необыденной, сверхъестественной ситуации, в которой не действие человека, а всего лишь наличие или отсутствие того или иного верования (полагания) у него в голове может причинить тот или иной исход: если Бог существует, то Он даст жизнь вечную тому, кто верит в Него. В отсутствии такого сверхъестественного допущения отказ от полноты когнитивной рациональности не может служить максимизации полезности для одиночного принимающего решений*. А вот для участников интерактивной ситуации, оказывается, может!

Благотворные коллективные заблуждения

Действия людей определяет не сама ситуация, а ее преломление в сознании деятелей. Это относится и к субоптимальным ситуациям вообще и к ДЗ-ситуациям в частности.

Вернемся к сказочке о двух заключенных. Представьте себе, что пока они сидели в заключении, в банде, к которой они принадлежат, в их отсутствие и без их ведома произошло знаменательное событие: общее собрание банды приняло решение отменить наказание за сотрудничество с полицией. Наши бедняги, не зная о принятом решении, продолжают думать, что они знают, что если они предадут товарища, их ждет кара. И, что решающе важно, каждый из них знает, что другой думает, что он знает, что за предательством последует наказание; и каждый из них знает, что каждый из них знает, что другой думает, что он знает, что за предательством последует наказание;... И т.д. *ad infinitum*.

По структуре это условие похоже на допущение о взаимном знании, которое мы приводили выше, характеризуя понятие инструментальной рациональности. Разница лишь в том, что в обсуждаемом случае с заключенными самая первая клауза — самая маленькая, центральная кукла этой бесконечно раздутой матрешки — есть не «каждый знает, что п» (назовем это клаузой о знании), а «каждый думает, что знает, что п» (назовем это клаузой о субъективной уверенности).

* Смягчаю категоричность этого отрицания. При желании можно придумать и другие случаи, когда может, но случаи эти все же будут маргинальны для теории принятия решений одного лица и ее приложений.

Какой выбор они сделают — молчание или предательство? Очевидно, молчание — ибо оба ошибочно считают, что находятся в ситуации, где выбор предательства отягощен последующим наказанием, и оба знают о полаганиях друг друга. Между тем на самом деле они находятся в ситуации Дилеммы заключенных, но не знают об этом.

Выходит, что ДЗ-ситуация страшна только для знающих. Заблуждающиеся спасены от ловушки неоптимального исхода.

Спасение обоих происходит, конечно, лишь в том случае, когда (1) заблуждаются оба и (2) оба знают о мнениях (=полаганиях) друг друга. Если один заблуждается, а другой знает о решении собрания, то, конечно, знающий предаст — и немедленно выйдет на свободу; незнающий промолчит — и останется в дураках, получив 10 лет.

Таким образом, благотворными для всех могут быть лишь заблуждения, разделяемые всеми, — *коллективные заблуждения*⁶.

Благотворный коллективный отказ от приобретения знания

Усложним теперь нашу сказочку еще на один вершок. К условиям предыдущего варианта добавим следующую деталь: стражник, обслуживающий камеры наших двух заключенных, воспыпал к ним симпатией и предлагает им бесплатную услугу: он может разузнать и доложить им итоги голосования в банде. Однако он оказывается поборником справедливости: соглашается доложить итоги либо обоим — если этого пожелают оба, либо ни одному из них — в противном случае. Добавим, что оба заключенных полностью доверяют добросовестности стражника и каждый из них знает, что они оба доверяют.

Ситуация, таким образом, превращается в двухходовку: (1) на первом ходу каждый из заключенных должен выразить свое согласие или несогласие с получением информации на условиях стражника; (2) на втором ходу, получив или соответственно не получив информацию о голосовании, каждый из двух должен принять решение по существу дела: молчание или предательство сотоварища.

Можно показать, что (опять-таки при достаточно высокой отрицательной полезности наказания) в этой игре имеется ровно один Нэш-равновесный исход, который и представляет собой решение (solution) игры. Подсказывает ли вам интуиция, какие действия нашей пары заключенных ведут к этому исходу? Обоюдный отказ от предлагаемой информации на первом ходу

и обоюдное молчание на втором. Надо ли говорить, что исход этот одновременно и Парето-оптимален, — ведь обоюдное молчание приводит всего лишь к году заключения для каждого?

Замечательность этой ситуации в том, что она опровергает глубоко укоренившуюся в нас интуицию о том, что всегда и при всех обстоятельствах знание — сила. В данном случае знание может оказаться слабостью — и именно поэтому рационально (то есть инструментально рационально, но, конечно, не рационально с когнитивной точки зрения!) отказаться от его получения. Должно быть понятно, почему это так. Ведь 50 шансов из 100, что ответ на их запрос будет: «Наказание отменено». И оба будут знать об этом. И оба будут знать, что оба знают, — и т.д. Они, стало быть, узнают, что они находятся в ситуации Дилеммы заключенных, — и тогда рациональность заставит их предать друг друга и отсидеть по 9 лет. Между тем в первоначальном состоянии симметричной неопределенности (неполноты знания) рациональность диктует обоим молчать — ибо шансы, что наказание не отменено, достаточно велики. Они молчат — и выходят на свободу через год. Так разумно ли им рисковать и запрашивать стражника об исходе голосования?

Благотворная коллективная иррациональность

Заключительное и решающее видоизменение обсуждаемой ситуации состоит в следующем. В начальном состоянии два заключенных когнитивно рациональны настолько, насколько это вообще доступно человеку, — предположим, что оба они — гениальные ученые, владеющие всей полнотой современной научной методологии. В результате приложения всей своей великолепной рациональности к доступным им обоим фактам оба приходят к выводу, что имеется 1 шанс из 100, что голосование привело к сохранению наказания, а 99 из 100, что наказание отменено; и оба знают, что их выводы совпадают.

В этой ситуации (если отрицательная величина наказания не сверхвелика) рациональность диктует им выбрать предательство — ибо оба считают, что почти наверняка находятся в ситуации Дилеммы заключенного.

Стражник их, однако, оказывается еще и отменным знахарем; он предлагает им обоим некое снадобье, которое замечательным, строго избирательным образом притупляет остроту ума того, кто его выпьет. Именно: в результате этого «точечного уда-

ра» по своим великолепным интеллектам каждый из них ущербит свою когнитивную рациональность в том и только в том отношении, что будет полагать, что все доступные ему факты (а они — те же что и раньше) заставляют сделать вывод, что имеется 50 шансов из 100, что голосование привело к сохранению наказания, и 50 из 100, что наказание отменено. И опять же стражник берется напоить снадобьем либо обоих, либо ни одного. Оба знают о действии снадобья и оба знают, что оба знают.

Как и в предыдущем случае, математическая сторона этой ситуации репрезентируема в терминах классической теории игр, и решение (solution) соответствующей игры Парето-оптимально (по одному году отсидки каждому) и получается в результате применения каждой стратегии: (1) на первом ходу выпить снадобья (=ущербить свой интеллект); (2) на втором ходу промолчать.

Интерес этой интерактивной ситуации и репрезентирующей ее игры в том, что они (ситуация и игра) показывают, что в принципе возможны положения, в которых локализованная («точечная») коллективная когнитивная иррациональность имеет позитивную ценность, ибо дает возможность рациональным во всех остальных отношениях индивидам избежать соблазна предательства и успешно сотрудничать в интересах всех участников ситуации.

Примечания

- ¹ *Iraham R., HaKen H.* Synergetik — die Lehre vom Zusammen wirkhen, Umschau 6, 1971. P. 191.
- ² См., напр.: *Macnger K.* Thingking in Complexity: The Complex Dynamic of Matter, Mind and Mankind. Berlin, 1997. P. 265—270.
- ³ *Neumann I. Von, Morgenstern O.* Theory of Games and Economic Behavior, 1953. P. 10—11.
- ⁴ *Pascal B.* Pensees, Para. 418.
- ⁵ Ibid.
- ⁶ См.: *Blinov A.* Epistemic rationality at anyprice // Univ of New England. N.SW. Avstralia, 1997. Module 3; *Блинов А.* Общение. Звуки. Смысл: об одной проблеме аналитической философии языка. М., 1996.