Лекция 6. Результаты о невозможности: теоремы Эрроу, Гиббарда-Саттертуэйта, Гурвица, Уильямса.

Сергей Николенко*

16 июня 2008 г.

1 Теорема Гиббарда–Саттертуэйта

1.1 Введение и определения

О чём всё это. Мы уже рассмотрели примеры, в которых были получены правдивые механизмы, успешно реализующие социальную функцию в доминантных стратегиях. Однако резонно задуматься, всегда ли это возможно. Сейчас мы рассмотрим один из самых больших подвохов этой теории.

Суть теоремы. Оказывается, что всё-таки не любые механизмы существуют. Мы сформулируем определение довольно узкого и «нечестного» класса социальных функций — *диктаторских*, т.е. таких, которые выгодны одному конкретному участнику. А потом докажем, что никаких других реализовать в доминантных стратегиях нельзя...

Диктаторские функции социального выбора.

Определение 1 *Функция социального выбора f называется* диктаторской, *если существует такой агент i, что для всех $\theta = (\theta_1, \ldots, \theta_N) \in \times$*

$$f(\theta) = \{x \in X \mid u_i(x, \theta_i) \geq u_i(y, \theta_i) \text{ для всех } y \in X\}.$$

Проще говоря, функция социального выбора всегда выбирает один из вариантов, оптимальных для i-го агента.

*Законспектировали Владимир Кулёв и Михаил Дворкин.

Монотонные функции социального выбора. Вспомним определение: множество нижнего контура возможного исхода o при агенте i типа θ_i — это

$$L_i(o, \theta_i) = \{o' \in \mathcal{O} : u_i(o, \theta_i) \geq u_i(o', \theta_i)\}.$$

Определение 2 *Функция социального выбора f называется* монотонной, *если для каждого θ, если θ' таково, что $L_i(f(\theta), \theta_i) \subseteq L_i(f(\theta'), \theta'_i)$ для всех i, то $f(\theta) = f(\theta')$.*

То есть если $f(\theta) = x$, и при переходе к θ' ни у одного агента ни один исход, который раньше был хуже x, не стал строго лучше x, то x должен остаться его социальным выбором.

Порядки предпочтений. Важным для нас понятием будут порядки на возможных исходах $\mathcal{O}$, которые для каждого агента задают, что ему больше нравится. Нам не так важно, сколько именно агент получит (u_i), сколько то, что он исход o_1 ценит выше, чем o_2, но ниже, чем o_3. Обозначим через $\mathcal{P}$ множество всех линейных порядков на $\mathcal{O}$. Через $\mathcal{R}_i$ — множество порядков, которые может реализовывать агент i.

1.2 Формулировка и доказательство

Теорема Гиббарда–Саттертуэйта.

Теорема 1 *Предположим, что множество возможных исходов $\mathcal{O}$ конечно и состоит не менее чем из трёх элементов, все исходы реализуются: $f(\theta) = \mathcal{O}$, и каждый агент может реализовывать любое рациональное множество предпочтений: $\mathcal{R}_i = \mathcal{P}$. Тогда функция социального выбора f правдиво реализуема в доминантных стратегиях тогда и только тогда, когда она диктаторская.*

Справа налево. Очевидно, что диктаторская f правдиво реализуема в доминантных стратегиях. Дальше будем доказывать слева направо.

Структура доказательства. Доказывать будем так:

1. Если $\mathcal{R}_i = \mathcal{P}$ для всех i, и f правдиво реализуема в доминантных стратегиях, то f монотонна.

2. Если $\mathcal{R}_i = \mathcal{P}$ для всех i, f монотонна, и $f(\theta) = \mathcal{O}$, то f эффективна ex post.

3. Если f монотонна и эффективна ex post, то она диктаторская.

Это будут наши три леммы.

Доказательство леммы 1.

- Если $\mathcal{R}_i = \mathcal{P}$ для всех i, и f правдиво реализуема в доминантных стратегиях, то f монотонна. Рассмотрим два профиля типов θ и θ', для которых $L_i(f(\theta), \theta_i) \subseteq L_i(f(\theta'), \theta'_i)$. Хотим показать, что $f(\theta) = f(\theta')$.

- Т.к. f правдиво реализуема, то $f(\theta'_1, \theta_2, \dots, \theta_N) \in L_1(f(\theta), \theta'_1)$ и $f(\theta) \in L_1(f(\theta'_1, \theta_2, \dots, \theta_N), \theta'_1)$. Т.к. порядки линейные (всё сравнимо), из этого следует, что $f(\theta'_1, \theta_2, \dots, \theta_N) = f(\theta)$.

- Далее, $f(\theta'_1, \theta'_2, \theta_3, \dots, \theta_N) = f(\theta'_1, \theta_2, \dots, \theta_N) = f(\theta)$. И т.д., в общем, $f(\theta) = f(\theta')$.

Доказательство леммы 2.

- Если $\mathcal{R}_i = \mathcal{P}$ для всех i, f монотонна, и $f(\theta) = \mathcal{O}$, то f эффективна ex post. Напомним, что «эффективна ex post» означает, что уже после того, как агенты сыграют по своим стратегиям, для каждого возможного значения θ нельзя сместить равновесие туда, где всем будет лучше.

- Предположим противное. Пусть существует $\theta \in \times$ и $y \in X$ такие, что
$$u_i(y, \theta_i) > u_i(f(\theta), \theta_i)$$
(не равно, т.к. нет несравнимых исходов).
$$u_i(y, \theta_i) > u_i(f(\theta), \theta_i)$$
Воспользуемся теперь тем, что $f(\theta) = \mathcal{O}$. Это значит, что есть такой $\theta' \in \times$, что $f(\theta') = y$.

- А теперь воспользуемся тем, что все предпочтения в $\mathcal{P}$ возможны. Выберем такой вектор $\theta'' \in \times$, что
$$\forall i \forall x \neq f(\theta), y \quad u_i(y, \theta''_i) > u_i(f(\theta), \theta''_i) > u_i(z, \theta''_i).$$
$$\forall i \forall x \neq f(\theta), y \quad u_i(y, \theta''_i) > u_i(f(\theta), \theta''_i) > u_i(z, \theta''_i).$$

- Поскольку $L_i(y, \theta'_i) \subset L_i(y, \theta''_i)$ для всех i, по монотонности $f(\theta'') = f(\theta)$. Противоречие, т.к. $y \neq f(\theta)$.

Доказательство леммы 3. Если f монотонна и эффективна ex post, то она диктаторская. Эта лемма следует из теоремы Эрроу о невозможности (Arrow's Impossibility Theorem). Сейчас мы её сформулируем и докажем.

1.3 Теорема Эрроу

Парадокс Кондорсе. Начнём с примера: пусть у нас три участника, у них есть свои предпочтения на трёх исходах, и мы хотим решить дело голосованием. Предпочтения таковы:

$$x \succ_1 y \succ_1 z,$$
$$z \succ_2 x \succ_2 y,$$
$$y \succ_3 z \succ_3 x.$$

Получается, что нарушилась транзитивность.

Формулировка.

Теорема 2 *Пусть в множестве альтернатив* ≥ 3 *элемента, и возможны все рациональные профили* ($\mathcal{R}$) *или вообще все профили, в которых любые две альтернативы различимы* ($\mathcal{P}$)*. Тогда всякая функция социального выбора* F*, которая оптимальна по Парето и удовлетворяет условию попарной независимости, является диктаторской, т.е.* $\exists$ *агент* h *такой, что* $\forall \{x, y\} \subset \mathcal{O}$ *и любого профиля* $(\succeq_1, \ldots, \succeq_I)$ x *социально предпочтителен* y *тогда и только тогда, когда* $x \succ_h y$.

Пояснения.

- Оптимальность по Парето: если для всех профилей $x \succeq_j y$, то F предпочтёт x перед y.
- Попарная независимость: отношения между двумя возможностями x и y зависят только от предпочтений на них и не зависят от других возможных исходов.

Определяющие наборы агентов. Для данного F будем говорить, что набор агентов $S \subset I$:

- *определяющий для* x *перед* y, если когда каждый агент в S предпочитает $x \succ y$ *и* каждый агент в $I \setminus S$ предпочитает $y \succ x$, F выбирает x;

- *определяющим*, если он определяющий для любой пары $\{x, y\}$;
- *полностью определяющим*, если когда каждый агент из S предпочитает $x \succ y$, F тоже выбирает x.

Доказательство.

1. Если для некоторых x и y $S \subset I$ определяющий для x перед y, то $\forall z \neq x$ S определяющий для x перед z и $\forall z \neq y$ S определяющий для z перед y.

   Если $z = y$, доказывать нечего. Если $z \neq y$, рассмотрим профиль $(\succeq_1, \dots, \succeq_I)$ такой, что

   $$x \succ_i y \succ_i z \quad \forall i \in S,$$

   $$y \succ_i z \succ_i x \quad \forall i \in I \setminus S.$$

   Тогда, значит, по свойству определяющего набора F должна предпочесть x перед y. А по оптимальности по Парето, F предпочитает y перед z. Значит, F предпочитает x перед z. Осталось сослаться на попарную независимость.

2. Если для некоторых x и y $S \subset I$ определяющий для x перед y, и z — третья альтернатива, то S определяюща для z перед w и для w перед z для всех $w \neq z \in \mathcal{O}$.

   По шагу 1, S определяющий для z перед y и для x перед z. Применим снова шаг 1 для пары $\{x, z\}$ и альтернативы w; значит, S определяющий для w перед z. Аналогично для пары $\{z, y\}$.

3. Если для некоторых $\{x, y\} \subset \mathcal{O}$ S определяющий для x перед y, то S определяющий. Доказательство сразу следует из шага 2 и из того, что третья альтернатива существует (это важно!).

4. Если S определяющий и T определяющий, то $S \cap T$ тоже определяющий. Рассмотрим тройку альтернатив $\{x, y, z\} \subset \mathcal{O}$ и профиль $(\succeq_1, \dots, \succeq_I)$ такой, что

   $$\begin{aligned}
   &z \succ_i y \succ_i x && \forall i \in S \setminus (S \cap T),\\
   &x \succ_i z \succ_i y && \forall i \in S \cap T,\\
   &y \succ_i x \succ_i z && \forall i \in T \setminus (S \cap T),\\
   &y \succ_i z \succ_i x && \forall i \in I \setminus (S \cup T).
   \end{aligned}$$

Тогда zFy, потому что $S = (S \cap T) \cup (S \setminus (S \cap T))$ — определяющий, и xFz, потому что T — определяющий. Значит, xFy, и по попарной независимости $S \cap T$ определяющий для x перед y. Значит, он вообще определяющий.

5. Для любого $S \subset I$ либо S определяющий, либо $I \setminus S$ определяющий. Рассмотрим тройку альтернатив $\{x, y, z\} \subset \mathcal{O}$ и профиль $(\succeq_1, \dots, \succeq_I)$ такой, что

$$
\begin{aligned}
&x \succ_i z && \succ_i y \quad \forall i \in S, \\
&y \succ_i x && \succ_i z \quad \forall i \in I \setminus S.
\end{aligned}
$$

   Тогда либо xFy, и S определяющий для x перед y, либо yFx. Если yFx, то по Парето xFz, и, значит, yFz; значит, $I \setminus S$ определяющий для y перед z.

6. Если S определяющий и $S \subset T$, то T определяющий. По Парето пустой набор не может быть определяющим. Значит, $I \setminus T$ не может быть определяющим, потому что тогда и $\emptyset = S \cap (I \setminus T)$ будет определяющим.

7. Если $S \subset I$ определяющий, и $|S| > 1$, то есть определяющее строгое подмножество $S' \subsetneq S$. Рассмотрим $h \in S$. Если $S \setminus \{h\}$ определяющий, то всё. Если нет, то $I \setminus (S \setminus \{h\})$ определяющий, и $\{h\} = S \cap (I \setminus (S \setminus \{h\}))$ определяющий.

8. Для некоторого $h \in I$ $\{h\}$ определяющий. Нужно просто итерировать шаг 7.

9. Если $S \subset I$ определяющий, то для всех x и y S полностью определяющий для x перед y. Нужно получить, что для всех $T \subset I \setminus S$ xFy, если все агенты из S предпочитают $x \succ y$, из T — $x \succeq y$, остальные — $y \succ x$.

   Рассмотрим третью альтернативу и профиль $(\succeq_1, \dots, \succeq_I)$ такой, что

$$
\begin{aligned}
&x \succ_i z \succ_i y && \forall i \in S, \\
&x \succ_i y \succ_i z && \forall i \in T, \\
&y \succ_i z \succ_i x && \forall i \in I \setminus (S \cup T).
\end{aligned}
$$

   Тогда xFz, потому что $S \cup T$ определяющий, и zFy, потому что S определяющий. Значит, xFy, что и требовалось.

10. Если $\{h\}$ определяющий, то h — диктатор. Это в точности следует из определения полностью определяющего набора.

Если в $\mathcal{O}$ два элемента. Мы воспользовались тем, что $|\mathcal{O}| \geq 3$? В самом деле, если $|\mathcal{O}| = 2$, то теорема неверна. Например, функция социального выбора «большинство голосов» в данном случае и недиктаторская, и правдиво реализуемая в доминантных стратегиях.

2 Теорема Гурвица

Об эффективности. Мы уже говорили об эффективных механизмах. В частности, доказали следующую теорему.

Теорема 3 *Эффективный, правдивый и рациональный механизм, у которого сходится баланс, существует тогда и только тогда, когда механизм VCG даёт положительную ожидаемую прибыль аукционеру.*

К результатам о невозможности. Эта теорема сводит вопрос о существовании хорошего механизма к вопросу о свойствах конкретного механизма VCG. В тех ситуациях, где VCG не даёт положительную прибыль, хорошего механизма не будет. Сейчас мы поймём, что в довольно общей ситуации не будет никаких хороших механизмов.

Simple exchange economy. Простая обменная экономика — это такая ситуация, когда на рынке есть продавцы и покупатели, которые торгуют ровно одним товаром.

Теорема. Итого получается довольно печальная ситуация.

Теорема 4 *В простой обменной экономике с квазилинейными преференциями невозможно в доминантных стратегиях реализовать эффективный, правдивый и рациональный механизм, у которого сходится баланс.*

Эту теорему некоторые называют «теорема Гурвица о невозможности» (Hurwicz impossibility theorem). Сейчас мы ее доказывать не будем; возможно, позже, когда будем рассматривать эту ситуацию подробнее. А сейчас нас ждёт очень серьёзная теорема Вильямса.

3 Теорема Вильямса

3.1 Торговля между двумя участниками

Bilateral trade. Начнём с примера bilateral trade. Один хочет продать, другой — купить. У продавца своё распределение себестоимости C; в частности, $c \in [c_0, c_1]$. У покупателя — своё распределение ценности V; в частности, $v \in [v_0, v_1]$.

Постановка задачи. Распределения всем известны, конкретные стоимости — нет. Предположим, что конфликт *может* возникнуть, т.е. $v_0 < c_1$. Можно ли построить механизм так, чтобы торговля происходила тогда и только тогда, когда выгодно обоим?

Формально говоря, механизм должен определить:

- p — сколько покупатель заплатит;
- r — сколько продавец получит.

Эффективен механизм, если объект продан тогда и только тогда, когда $v > c$.

Теорема о невозможности.

Теорема 5 *В вышеописанной задаче не существует механизма, который бы был эффективен, правдив, рационален и у которого в то же время сходился бы бюджет.*

Это называется *теорема Майерсона–Саттертуэйта.*

Доказательство. Рассмотрим механизм VCG. Он работает в данном случае так:

- Покупатель объявляет v, продавец объявляет c.
- Если $v \leq c$, ничего не происходит.
- Если $v > c$, покупатель платит $\max\{C, v_0\}$, а продавец получает $\min\{v, c_1\}$.

Доказательство.

- Механизм правдивый и эффективный (объект продаётся iff $v > c$). Более того, он рационален:
  - у покупателя с ценностью v_0 ожидаемая прибыль равна 0, дальше — больше;

  - у продавца с ценностью c_1 ожидаемая прибыль равна 0, дальше — больше.

- Но есть одна проблема – если $v_0 < c_1$, это значит, что когда вообще есть обмен, $\min\{v, c_1\} > \max\{c, v_0\}$. То есть продавец получает строго больше, чем платит покупатель. Значит, VCG не может сбалансировать бюджет.

- Любой другой хороший механизм, по теореме об эквивалентности доходности, должен на константу отличаться от VCG.

  Но в VCG продавец с себестоимостью c_1 получает 0, то есть уменьшить доход продавца, сохранив рациональность, не получится.

  И покупатель с ценностью v_0 получает 0, т.е. увеличить платёж, сохранив рациональность, тоже не получится.

  Итого – теорема доказана.

3.2 Теорема Вильямса: дифференцируемый случай

История вопроса. Про bilateral trade придумали Майерсон и Саттертуэйт (1983). Обобщение, которое сейчас буду рассказывать — это статья Williams (1999), «A characterization of efficient, bayesian incentive compatible mechanisms». Эта невозможность тоже будет следовать из теоремы об эквивалентности.

Вспоминаем определения.

Определение 3 Квазилинейная функция полезности *агента i с типом θ_i имеет вид*

$$u_i(o, \theta_i) = u_i(p_i, a, \theta_i) = v_i(a, \theta_i) - p_i,$$

где исход o определяет выбор $a \in \mathcal{K}$ из дискретного множества $\mathcal{K}$ и выплату p_i*, производимую агентом.*

Агенты с квазилинейными преференциями. У агента с квазилинейными преференциями есть *функция оценки* (valuation function) $v_i(a, \theta_i)$, $a \in \mathcal{K}$. Например, в аукционе, где продаётся одна вещь, $\mathcal{K} = \{0, 1\}$ — агент либо получит эту вещь, либо не получит. p_i в этом случае — выплата агента продавцу.

Постановка задачи.

- Для начала предположим, что тип агента лежит в интервале $\theta_i \in [\underline{\theta}_i, \overline{\theta}_i] \subset \mathbb{R}$.
- Обозначим через θ_i тип агента, а через θ_i^* — тип, который он говорит.
- $U_i(\theta_i^* \mid \theta_i)$ — ожидаемая прибыль (utility) агента i:
$$U_i(\theta_i^* \mid \theta_i) = \mathbf{E}_{\theta_{-i}}[u_i(p_i(\theta^*, \theta_{-i}), a(\theta^*, \theta_{-i}), \theta_i)].$$
- U_i складывается из V_i и P_i:
$$V_i(\theta_i^* \mid \theta_i) = \mathbf{E}_{\theta_{-i}}[v_i(a(\theta^*, \theta_{-i}), \theta_i)],$$
$$P_i(\theta_i^* \mid \theta_i) = \mathbf{E}_{\theta_{-i}}[p_i(\theta^*, \theta_i)],$$
$$U_i(\theta_i^* \mid \theta_i) = V_i(\theta_i^* \mid \theta_i) - P_i(\theta_i^* \mid \theta_i).$$
- Тогда правдивость означает, что
$$U_i(\theta_i) = U_i(\theta_i \mid \theta_i) \geq U_i(\theta_i^* \mid \theta_i) \quad \forall \theta_i^*, \theta_i \in \Theta_i.$$
- Рациональность: $U_i(\theta_i) \geq 0$ для всех θ_i.
- Баланс бюджета (ex ante!) означает, что ожидаемая сумма выплат неотрицательна:
$$\mathbf{E}\left[\sum_i v_i(a(\theta), \theta_i) - U_i(\theta_i)\right] = \mathbf{E}\left[\sum_i p_i(\theta)\right] \geq 0.$$

Теорема об огибающей. Вспомним механизм VCG:
$$M_i^V(\mathbf{x}) = W(\alpha_i, \mathbf{x}_{-i}) - W_{-i}(\mathbf{x}).$$

Есть в мат. анализе такая теорема об огибающей (envelope theorem). Рассмотрим задачу оптимизации $M(a) = \max_x f(x, a)$. Тогда при достаточно хороших условиях дифференцируемости
$$\frac{dM(a)}{da} = \left.\frac{\partial f(x^*, a)}{\partial a}\right|_{x^*=x(a)},$$
где $x(a)$ — точка, в которой достигается максимум.

Иначе говоря, можно продифференцировать f по a и вычислить в точке максимума.

Применяем теорему об огибающей. Применим её к нашей ситуации:
$$\frac{dU_i(\theta_i)}{d\theta_i} = \left.\frac{\partial U_i(\theta_i^* \mid \theta_i)}{\partial\theta_i}\right|_{\theta_i^*=\theta_i} = \left.\frac{\partial V_i(\theta_i^* \mid \theta_i)}{\partial\theta_i}\right|_{\theta_i^*=\theta_i}.$$
Иначе говоря, получается, что
$$U_i(\theta_i) = U_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} \left.\frac{\partial V_i(\theta_i^* \mid \tau_i)}{\partial\tau_i}\right|_{\theta_i^*=\tau_i} d\tau_i.$$

$U_i(\theta_i) = U_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} \left.\frac{\partial V_i(\theta_i^*|\tau_i)}{\partial\tau_i}\right|_{\theta_i^*=\tau_i} d\tau_i.$

Это и даёт нам результат об эквивалентности всех механизмов, т.к. $V_i(\theta_i^* \mid \tau_i)$ зависит только от правила $a(\theta)$ и ценностей агентов v_i, но не от деталей механизма.

Механизмы Гровса, таким образом, покрывают всё множество «хороших» механизмов. Это и есть основная теорема.

Что такое хорошо. Осталось понять, что такое «хороший» механизм. По идее, в теореме «хороший» должно означать «правдивый и эффективный». Но у нас тут ещё какие-то ограничения на дифференцируемость появлялись.

Вообще говоря, нельзя применять теорему об огибающей к произвольному правдивому и эффективному механизму. Поэтому мы сейчас всё докажем по-другому.

3.3 Теорема Вильямса: общий случай

Формулировка теоремы.

Теорема 6 *Рассмотрим проблему социального выбора с квазилинейными преференциями. Предположим также, что*

- *множества типов Θ_i — связные открытые подмножества $\mathbb{R}^{n_i}$,*
- *ожидаемые interim ценности агентов $V_i(\theta_i^* \mid \theta_i)$ непрерывно дифференцируемы на $\Theta_i \times \Theta_i$ в точках, в которых $\theta_i^* = \theta_i$.*

Тогда механизмы Гровса являются правдивыми и эффективными для этой задачи, и interim ожидаемые ценности агентов $U_i(\theta_i^* \mid \theta_i)$ любого правдивого и эффективного механизма совпадают с ценностями одного из механизмов Гровса.

Что ограничивают ограничения. Важно понять, что именно ограничивают ограничения. Они на V_i. А V_i, как мы уже отмечали, зависит только от свойств *задачи*, но не механизма.

То есть мы немного ограничиваем класс задач, к которым применима теорема. Но при этом класс механизмов остаётся полным — доказываем для *всех* правдивых эффективных механизмов.

Переформулировка теоремы. Как обычно, a good formula stays for ever. Теорема будет следовать из формулы.

Теорема 7 *При вышеозначенных условиях функция доходности любого правдивого эффективного механизма имеет вид (для любых $\theta_i, \theta_i^* \in \Theta_i$)*

$$U_i(\theta_i) = U_i(\theta_i^*) + \int_C D_{\theta_i} V_i(\theta_i^* \mid \theta_i)|_{\theta_i^*=\tau, \theta_i=\tau}\, d\tau,$$

где C — гладкая кривая от θ_i^* к θ_i внутри Θ_i, $\tau \in \mathbb{R}^{n_i}$.

Доказательство.

- Обозначим $\rho \in \mathbb{R}^{n_i}$ — некоторый единичный вектор, $s \in \mathbb{R}$.
- Правдивость гласит, что $\forall \theta_i \in \Theta_i$

$$U_i(\theta_i) \geq U_i(\theta_i + s\rho \mid \theta_i) \text{ и } U_i(\theta_i + s\rho) \geq U_i(\theta_i \mid \theta_i + s\rho).$$

- Суммарно:

$$U_i(\theta_i \mid \theta_i + s\rho) - U_i(\theta_i) \leq U_i(\theta_i + s\rho) - U_i(\theta_i) \leq \\ \leq U_i(\theta_i + s\rho) - U_i(\theta_i + s\rho \mid \theta_i).$$

- $U_i(\theta_i \mid \theta_i+s\rho)-U_i(\theta_i) \leq U_i(\theta_i+s\rho)-U_i(\theta_i) \leq U_i(\theta_i+s\rho)-U_i(\theta_i+s\rho \mid \theta_i)$.
- Сократим там P_i слева и справа и разделим на s:

$$\frac{V_i(\theta_i \mid \theta_i + s\rho) - V_i(\theta_i)}{s} \leq \\ \leq \frac{U_i(\theta_i + s\rho) - U_i(\theta_i)}{s} \leq \\ \leq \frac{V_i(\theta_i + s\rho) - V_i(\theta_i + s\rho \mid \theta_i)}{s}.$$

- $\frac{V_i(\theta_i|\theta_i+s\rho)-V_i(\theta_i)}{s} \leq \frac{U_i(\theta_i+s\rho)-U_i(\theta_i)}{s} \leq \frac{V_i(\theta_i+s\rho)-V_i(\theta_i+s\rho|\theta_i)}{s}$.

- Устремим $s \to 0$. По условию о дифференцируемости V_i, левая часть сходится к производной $V_i(\tau_i^* \mid \tau_i)$ по направлению ρ в точке $\tau_i^* = \tau_i = \theta_i$.

- Правая часть раскладывается на

$$\frac{V_i(\theta_i + s\rho) - V_i(\theta_i)}{s} - \frac{V_i(\theta_i + s\rho \mid \theta_i) - V_i(\theta_i)}{s}.$$

  Первое слагаемое сходится к производной $V_i(\tau_i)$ по τ_i по направлению ρ в $\tau_i = \theta_i$. Второе слагаемое — к производной $V_i(\tau_i^* \mid \tau_i)$ по τ_i^* по направлению ρ в $\tau_i^* = \tau_i = \theta_i$.

- А вся правая часть — к производной $V_i(\tau_i^* \mid \tau_i)$ по τ_i по направлению ρ в $\tau_i^* = \tau_i = \theta_i$. Значит, $D_{\theta_i} U_i(\theta_i) = D_{\theta_i} V_i(\theta_i^* \mid \theta_i)|_{\theta_i^* = \tau, \theta_i = \tau}$.

- Отсюда следует теорема, т.к. производная по предположению непрерывна.

Это очень показательный метод доказательства. Собственно, это развитие исходной идеи Майерсона в максимальной (или близкой к тому) общности. Видно, что откуда берётся во всех таких теоремах: нужно взять изменение (приращение $s\rho$) и посмотреть, что от него изменится.

3.4 Рациональность

Давайте применим теорему Вильямса. Мы бы хотели создавать рациональные механизмы. Посмотрим, когда это получится.

Теорема 8 *В предположениях теоремы Вильямса минимальная субсидия, которая требуется рациональному, правдивому и эффективному механизму, равна*

$$\min\{0, -(N-1)\mathbf{E}_\theta \left[\sum_{i=1}^{N} v_i(a(\theta), \theta_i)\right] + \sum_{i=1}^{N} U_i(\underline{\theta}_i)\}.$$

Значит, рациональные, правдивые и эффективные механизмы со сбалансированным бюджетом существуют iff

$$(N-1)\mathbf{E}_\theta \left[\sum_{i=1}^{N} v_i(a(\theta), \theta_i)\right] \leq \sum_{i=1}^{N} U_i(\underline{\theta}_i).$$

Доказательство. По теореме Вильямса, достаточно рассмотреть механизмы Гровса. Для них ожидаемая сумма трансферов

$$\mathbf{E}_\theta \left[\sum_{i=1}^{n} p_i(\theta) \right] = -\mathbf{E}_\theta \left[\sum_{i=1}^{n} \sum_{j \neq i} v_j(a(\theta), \theta_j) \right] + \sum_{i=1}^{n} k_i =$$
$$= -(N-1)\mathbf{E}_\theta \left[\sum_{i=1}^{N} v_i(a(\theta), \theta_i) \right] + \sum_{i=1}^{N} k_i.$$

По рациональности, $U_i(\underline{\theta}_i) \geq k_i$ для всех i. Отсюда и получается утверждение теоремы.

Итоги.

- Мы уже фактически прошлись по всей классической теории дизайна механизмов, уж точно по всему тому, за что нобелей давали; не считая, конечно, экономических, т.е. практических приложений (а без этого, конечно, Нобелей не дадут).
- Отныне будем заниматься более узкими вещами: онлайн-аукционами, например.