

# Теория конвенций и понятие коррелированного равновесия<sup>1</sup>

## 1 Конвенции

**Юм:** «Соглашение это не носит характера обещания; мы увидим впоследствии, что сами обещания проистекают из соглашений между людьми. Оно не что иное, как общее чувство общественного интереса; все члены общества выражают это чувство друг перед другом, и оно заставляет их подчинять свое поведение известным правилам. Я замечаю, что мне выгодно предоставлять другому человеку владение его собственностью при условии, что он будет действовать так же по отношению ко мне. Он чувствует, что, подчиняя свое поведение такому же правилу, тоже служит своим интересам. Когда мы это общее чувство взаимной выгоды выражаем друг перед другом и оно становится известно нам обоим, оно влечет за собой соответствующее решение и поведение; и это может по праву быть названо соглашением, или уговором, между нами, хотя и заключенным без посредства обещания, ведь поступки каждого из нас находятся в зависимости от поступков другого и совершаются нами в предположении, что и другой стороной должно быть нечто сделано. Когда два человека гребут веслами в одной лодке, они тоже делают это по взаимному соглашению, или уговору, хотя они никогда не обменивались взаимными обещаниями. Тот факт, что правило, устанавливающее стабильность владения, возникает лишь постепенно и приобретает силу только путем медленного прогресса, а также благодаря тому, что мы постоянно испытываем на опыте неудобства от его нарушения, — этот факт не противоречит происхождению данного правила из соглашения между людьми. Наоборот, опыт еще больше убеждает нас, что чувство взаимного интереса стало общим для всех наших близких, и дает нам уверенность в том, что в будущем их поведение будет регулироваться [этим чувством]; только этим ожиданием и обосновывается наша умеренность, наша воздержанность. Таким же образом, т. е. путем соглашений между людьми, но без посредства обещания, мало-помалу образуются языки. Точно так же золото и серебро становятся общими средствами обмена и признаются достаточной платой за вещи, в сотни раз превышающие их по ценности»<sup>2</sup>.

<sup>1</sup>НУГ «Формальная философия», проект «Динамический поворот в логической семантике» (2015–2016, 15-05-0005)

<sup>2</sup>Юм Д. Сочинения в 2 томах. Том 1. М.: Мысль, 1996. С. 530–531.

## Теория Д. Льюиса

**Рассел** «[w]e can hardly suppose a parliament of hitherto speechless elders meeting together and agreeing to call a cow a cow and a wolf a wolf».<sup>3</sup>

**Куайн:** «What is convention when there can be no thought of convening?»<sup>4</sup>

**Льюис:** «Language is only one among many activities governed by conventions that we did not create by agreeing and that we cannot describe»<sup>5</sup>

**1 (Конвенции)** *A regularity R in the behavior of members of a population P when they are agents in a recurrent situation S is a convention if and only if it is true that, and it is common knowledge in P that, in any instance of A among members of P:*

1. *everyone conforms to R;*
2. *everyone expects everyone else to conform to R;*
3. *everyone has approximately the same preferences regarding all possible combinations of actions;*
4. *everyone prefers that everyone conform to R, on condition that at least all but one conform to R;*
5. *everyone would prefer that everyone conform to R', on condition that at least all but one conform to R' (where R' is some possible regularity in the behavior of members of P in S, such that no one in any instance of S among members of P could conform both to R' and to R.)*

**2 (Координационное равновесие, coordination equilibrium)** «Let me define a coordination equilibrium as a combination in which no one would have been better off had any agent alone acted otherwise, either himself or someone else» [Lewis, p.14]

### Пример 1 Гребцы Юма

	A1	A2	A3
A1	1, 1	0, 0	0, 0
A2	0, 0	2, 2	0, 0
A3	0, 0	0, 0	3, 3

### Пример 2 PD

	L	R
T	2, 2	0, 6
B	6, 0	1, 1

<sup>3</sup>Russell B. The Analysis of Mind, London: Unwin Brothers Ltd, 1921. P. 190.

<sup>4</sup>Quine W.V.O. Foreword // Lewis D. Convention. Cambridge: Harvard University Press, 1969. P. xi.

<sup>5</sup>Lewis D. Convention. Cambridge: Harvard University Press, 1969. P.3.

### Пример 3 Координация

	L	R
T	1, 1	0, 0
B	0, 0	1, 1

	L	R
T	2, 2	0, 0
B	0, 0	1, 1

### Пример 4 Фокальные точки

	A1	B	A2	A3
A1	1, 1	0, 0	0, 0	0, 0
B	0, 0	1, 1	0, 0	0, 0
A2	0, 0	0, 0	1, 1	0, 0
A3	0, 0	0, 0	0, 0	1, 1

**Schelling** «focal point[s] for each person's expectation of what the other expects him to expect to be expected to do»<sup>6</sup>

**Vanderschraf:** «I define a convention as a correlated equilibrium»<sup>7</sup>

## 2 Коррелированное равновесие

### 2.1 Идея

Понятие предложено Р.Ауманном<sup>8</sup>. Коррелированное равновесие предполагает, что игроки действуют согласованно в соответствии с распределением вероятностей на множестве исходов. Каждый из игроков получает сигнал, который советует ему, что делать (конкретный сигнал может быть частным, но типы сигналов и вероятности их использования являются общим знанием).

### R. Myerson:

«if there is intelligent life on other planets, in a majority of them, they would have discovered correlated equilibrium before Nash equilibrium»

### Пример 5 Компромисс и "Семейный спор"

<sup>6</sup>Schelling T. The strategy of conflict. Cambridge: Harvard University Press, 1960. P.57

<sup>7</sup>Vanderschraaf P. Convention as correlated equilibrium // Erkenntnis. 1995. Vol. 42, № 1. P. 65–87; Vanderschraaf P. Knowledge, Equilibrium and Convention // Erkenntnis. 1998. Vol. 49, № 3. P. 337–369.

<sup>8</sup>Aumann R.J.. Correlated Equilibrium as an Expression of Bayesian Rationality // Econometrica. 1987. Vol. 55, № 1. P. 1–18.

	L	R
T	1, 2	0, 0
B	0, 0	2, 1

	L	R
T	1/2	0
B	0	1/2

### Пример 6 Светофор и "Chicken"

	L	R
T	6, 6	2, 7
B	7, 2	0, 0

	L	R
T	1/3	1/3
B	1/3	0

## 2.2 Определение

**3** Для любого распределения вероятностей  $p \in \Delta(S)$  определим игру  $\Gamma^*(p)$

- внешний наблюдатель ("доброжелательный диктатор") выбирает профиль стратегий в соответствии с распределением  $p$
- каждый игрок  $i \in N$  узнает о своей рекомендованной стратегии  $s_i$ , но может не знать о рекомендации другим игрокам  $s_{-i}$
- каждый игрок  $i$  выбирает стратегию  $s'_i \in S_i$  (то есть, игрок может не следовать рекомендации)
- для каждого игрока функция полезности определяется через реальные действия, а не рекомендованные, т.е. как  $u_i(s'_1, \dots, s'_n)$

**4** Стратегией игрока  $i$  в игре  $\Gamma^*(p)$  будем называть функцию  $\tau_i : S_i \mapsto S_i$ , которая сопоставляет каждой рекомендации реальное действие  $\tau_i(s_i) \in S_i$

**5**  $\tau_i^*(s_i) = s_i$  – стратегия игрока, в которой он следует рекомендации;  $\tau^*$  – профиль стратегий, в котором все игроки следуют рекомендациям

$$\mathbf{6} \quad p(s_{-i} | s_i) = \frac{p(s_i, s_{-i})}{\sum_{s'_{-i} \in S_{-i}} p(s_i, s'_{-i})}$$

**7** Распределение вероятностей  $p \in \Delta(S)$  является **коррелированным равновесием** е.т.е. профиль стратегий  $\tau^*$  в игре  $\Gamma^*(p)$  является равновесием Нэша. Т.е.:

$$\forall i \in N \quad \forall s, s' \in S_i : \sum_{s_{-i} \in S_{-i}} p(s_{-i} | s_i) \cdot u_i(s_i, s_{-i}) \geq \sum_{s_{-i} \in S_{-i}} p(s_{-i} | s_i) \cdot u_i(s'_i, s_{-i}) \quad (1)$$