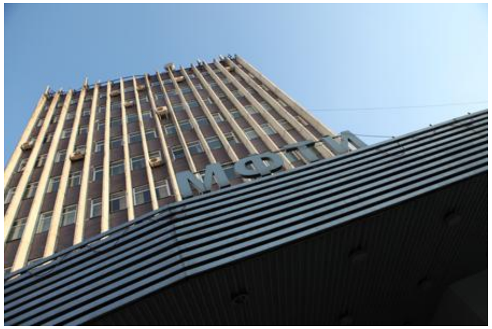
ИПУ РАН



МФТИ

РЕФЛЕКСИЯ И УПРАВЛЕНИЕ

Новиков Дмитрий Александрович

dan@ipu.ru, www.ipu.ru, www.mtas.ru

ПЛАН

1) Рефлексия. Управление.
Информационное и рефлексивное управление
1) Теория: рефлексивные игры
2) Примеры:
- Активный прогноз
- Выборы
- Экспертиза. Рефлексивная неманипулируемость механизмов планирования
- Эвакуация
- Управление толпой
4) Литература

РЕФЛЕКСИЯ

«РЕФЛЕКСИЯ (лат. reflexio – обращение назад). Термин, означающий отражение, а также исследование познавательного акта».

Рефлексия первого рода (авторефлексия) и второго рода.

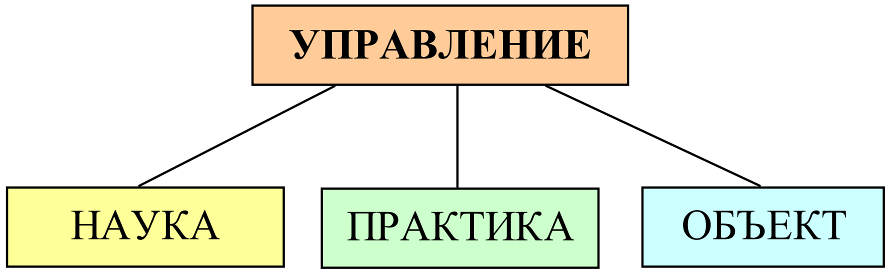
Стратегическая и информационная рефлексия.

Примеры: «Задача о скоординированной атаке», «Electronic Mail Game», «Задача о двух брокерах», … .

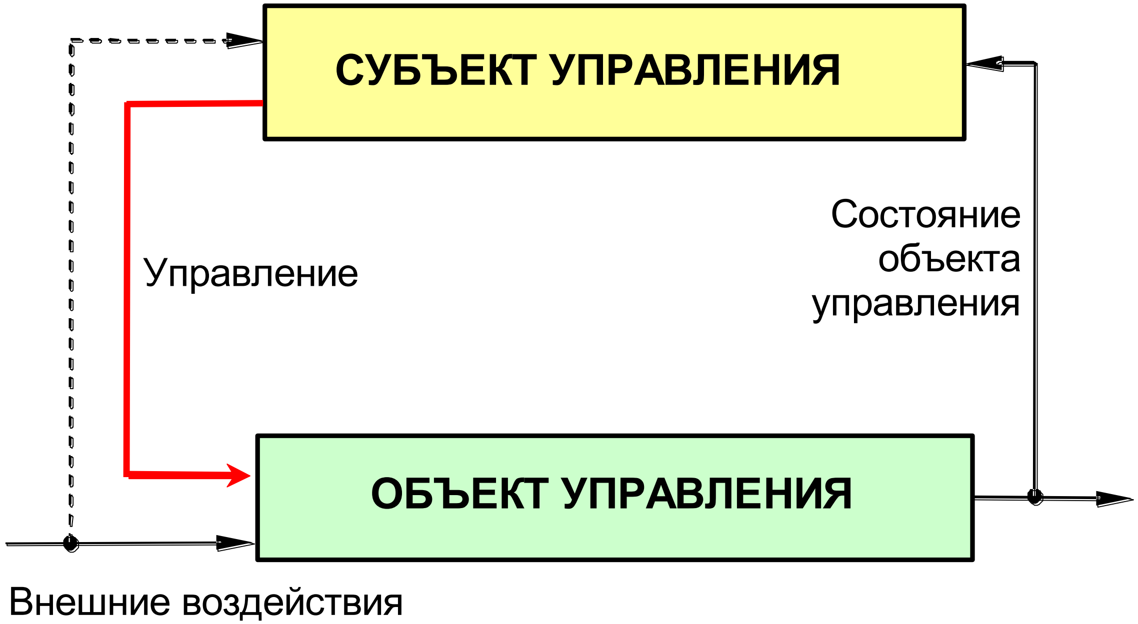
Максимальный ранг рефлексии, который следует иметь агенту для того, чтобы охватить все многообразие исходов игры (упуская из виду некоторые стратегии оппонента, агент рискует уменьшить свой выигрыш), назовем *максимальным целесообразным рангом рефлексии*.

УПРАВЛЕНИЕ

- **Управление** – «элемент, функция организованных систем различной природы: биологических, социальных, технических, обеспечивающая сохранение их определенной структуры, поддержание режима деятельности, реализацию программы, цели деятельности.» [Большой энциклопедический словарь].
- **Управление** – «направление движением кого/чего-нибудь, руководство действиями кого-нибудь» [Словарь русского языка].
- **Управление** – «воздействие на управляемую систему с целью обеспечения требуемого ее поведения» [Теория управления орг. системами].
- Webster's dictionary предлагает следующие значения терминов «control» и «management», которые соответствуют (в зависимости от объекта управления) русскому термину «**управление**»: «control – an activity or organization that directs or regulates an activity»; «management – art or science of directing, conducting and administering the work of others to achieve defined objectives».



БАЗОВАЯ СТРУКТУРА СИСТЕМЫ УПРАВЛЕНИЯ

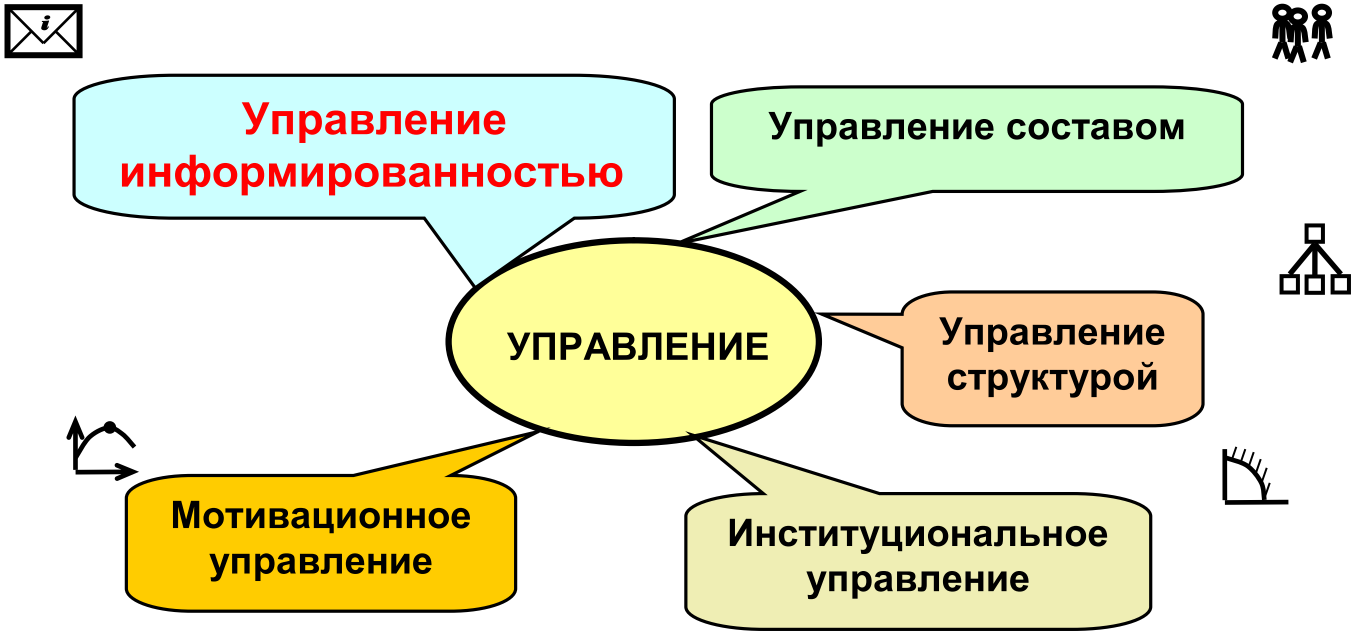


Структура системы управления

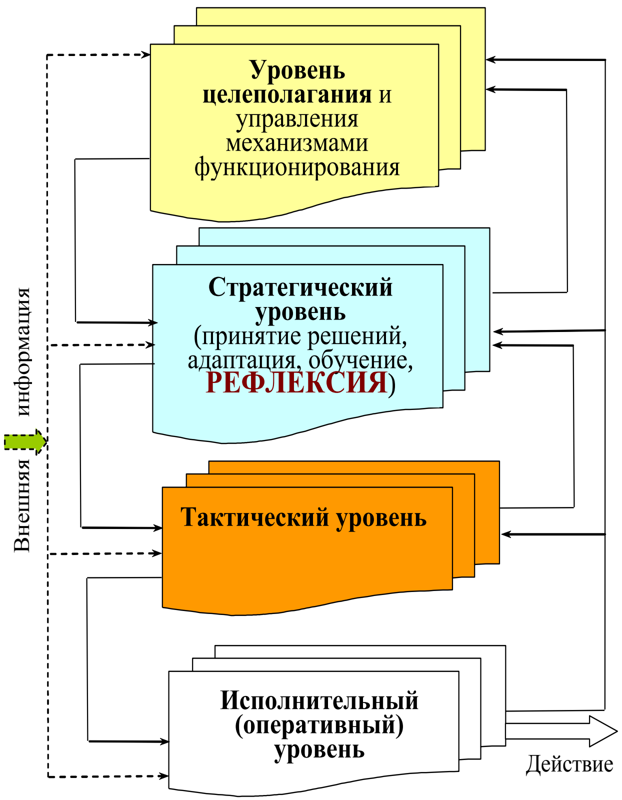
МЕТОДЫ (ВИДЫ) УПРАВЛЕНИЯ

Модель организационной системы определяется заданием:
• структуры ОС (совокупности информационных, управляющих, технологических и других связей между участниками ОС);
• множеств допустимых стратегий (ограничений и норм деятельности) участников ОС, отражающих, в том числе, институциональные, технологические и другие ограничения и нормы их совместной деятельности;
• предпочтений участников ОС;
• состава ОС (участников, входящих в ОС, то есть ее элементов);
• информированности – той информации о существенных параметрах, которой обладают участники ОС на момент принятия решений о выбираемых стратегиях;
• порядка функционирования (последовательности получения информации и выбора стратегий участниками ОС).

Управление ОС, понимаемое как воздействие на управляемую систему с целью обеспечения требуемого ее поведения, может затрагивать каждый из шести перечисленных параметров ее модели. Обычно порядок функционирования тесно связан со структурой, поэтому получаем пять классов задач управления.



ПЕРСПЕКТИВЫ. МУЛЬТИАГЕНТНЫЕ СИСТЕМЫ



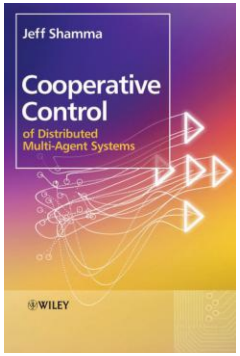
Противоборство, иерархические структуры	Confrontation Hierarchies
Коллективное принятие решений	Collective Decision-making
Кооперативное управление	Cooperative Control
Взаимодействие	Distributed Optimization (e.g. Task Assignment)
Планирование	Mission Planning
Реализация	Formation Control
Стабилизация	Consensus Problem

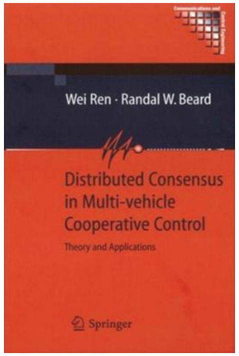
Теория игр

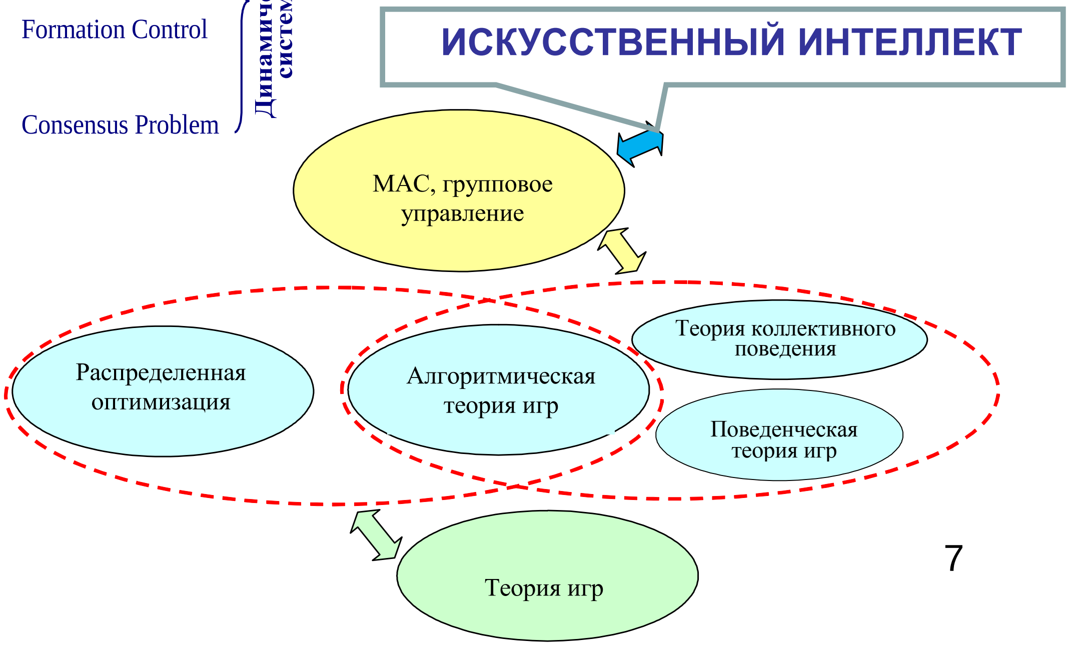
Модели коллективного поведения

Искусственный интеллект

Динамические системы







ТЕНДЕНЦИИ:

1) Интеграция теории МАС с теориями игр и искусственного интеллекта.
2) Стратегическое поведение (принятие решений).
3) Возрастающая роль теории игр и логик коммуницирующих и мобильных систем.
4) Тестовые задачи и сценарии.

СТРАТЕГИЧЕСКАЯ РЕФЛЕКСИЯ (БИМАТРИЧНЫЕ ИГРЫ)

Матрицы выигрышей $\boldsymbol{A} = ||a_{ij}||$ и $\boldsymbol{B} = ||b_{ij}||$ размерности $n \times m$ соответственно.

$I = \{1, 2, ..., n\}$ – множество действий первого агента (выбирающего строку),

$J = \{1, 2, ..., m\}$ – множество действий второго агента (выбирающего столбец).

Гарантирующие стратегии агентов следующие:

$$i_0 \in \operatorname{Arg} \max_{i \in I} \min_{j \in J} a_{ij}, \quad j_0 \in \operatorname{Arg} \max_{j \in J} \min_{i \in I} b_{ij}.$$

Пример («*Игра в прятки*»)

Ранг рефлексии агентов и соответствующие действия по выбору комнат

Ранг рефлексии агента	**0**	**1**	**2**	**3**	**4**
Комната, выбираемая прячущимся	Самая темная	Любая, кроме самой светлой	Любая, кроме самой темной	Самая светлая	Самая темная
Комната, выбираемая ищущим	Самая светлая	Самая темная	Любая, кроме самой светлой	Любая, кроме самой темной	Самая светлая

ЭКСПЕРИМЕНТЫ

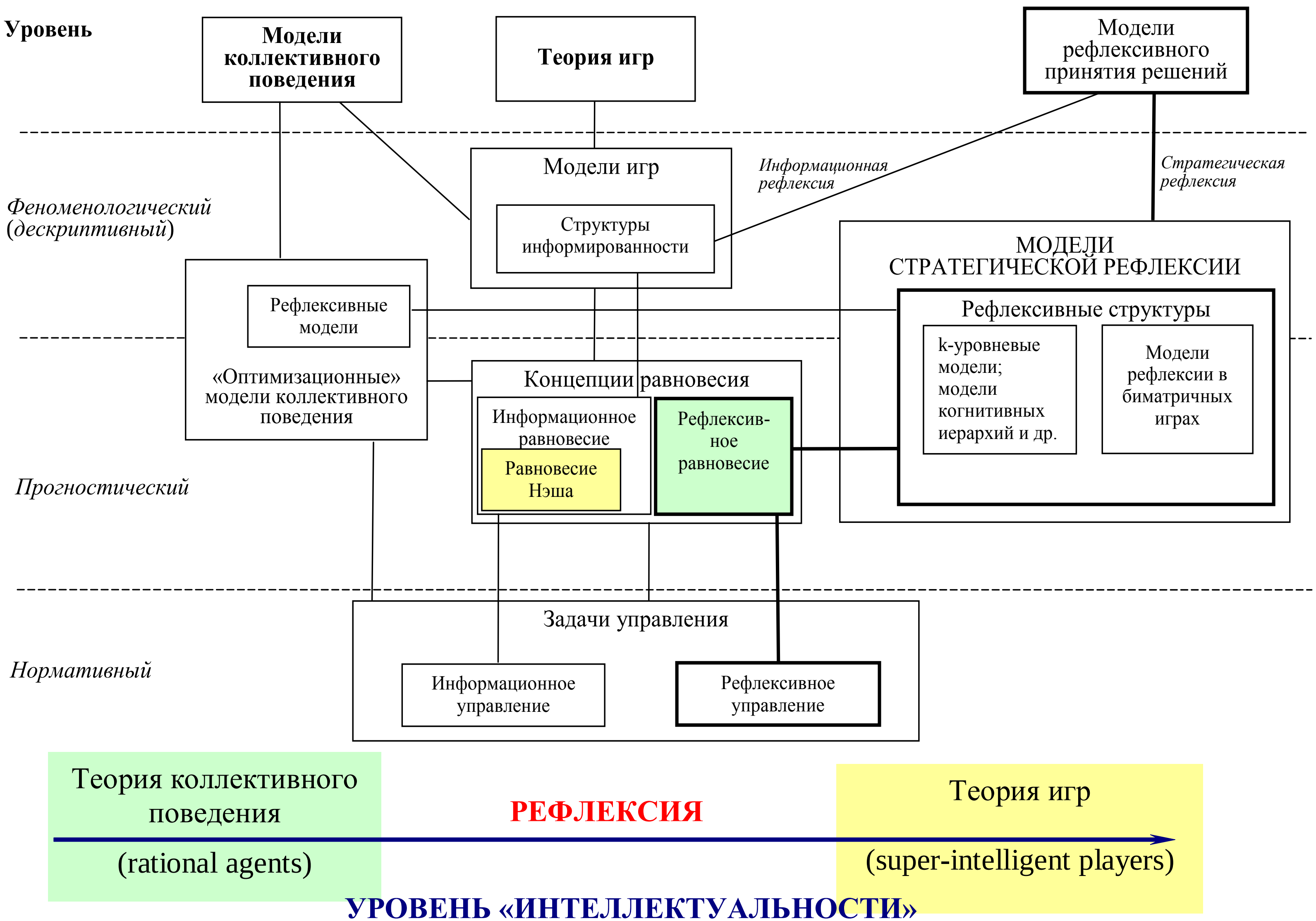
Для **игры «11-20»** (для 108 участников) получено следующее распределение агентов по рангам рефлексии:

Ранг	0	1	2	3	4	5	6	7	8	9
Действие	20	19	18	17	16	15	14	13	12	11
Доля агентов (%)	6	12	30	32	6	1	6	3	0	4

Известны результаты экспериментов, в соответствии с которыми от 40 % до 60 % агентов имеют ненулевой ранг рефлексии (выбирают действия, отличные и от действий агентов нулевого ранга, и от равновесия Нэша). Сводка распределений (по результатам трех исследований) агентов по рангам рефлексии:

Ранг	**Доля агентов**	**Доля агентов**	**Доля агентов**
0	0,25	0,42	0,21
1	0,12	0,44	0,21
2	0,12	0,11	0,27
3	0,12	0,03	0,19
4	0,12	0,01	0,09
5	0,12	0	0,03
6	0,12	0	0,01
7 и выше	0	0	0

ПРИНЯТИЕ РЕШЕНИЙ (ИНФОРМАЦИОННАЯ И СТРАТЕГИЧЕСКАЯ РЕФЛЕКСИЯ)



РОЛЬ ИНФОРМИРОВАННОСТИ. ОБЩЕЕ ЗНАНИЕ

Игра в нормальной форме:

$\Gamma_0 = \{N, (X_i)_{i \in N}, (f_i(\cdot))_{i \in N}\}$

N - множество игроков (агентов),

$(X_i)_{i \in N}$ - множества допустимых действий

$(f_i(\cdot))_{i \in N}$, f_i: $X \to \Re^1$ - целевые функции, $i \in N$.

Равновесие Нэша:

$$\forall i \in N \qquad x_i^* \in \operatorname*{Argmax}_{x_i \in X_i} f_i(x_1^*, \ldots, x_{i-1}^*, x_i, x_{i+1}^*, \ldots, x_n^*).$$

Общее знание - факт, который:

i) известен всем агентам

ii) всем агентам известно i)

iii) всем агентам известно ii)

и т.д. до бесконечности.

МОДЕЛИ ИНФОРМАЦИОННОЙ РЕФЛЕКСИИ

$\Gamma_I = \{N, (X_i)_{i \in N}, (f_i(\cdot))_{i \in N}, \Omega, I\}$ – **рефлексивная игра**

Ω – множество возможных состояний природы

I – структура информированности

N – множество элементов (игроков)

f_i: $\Omega \times X' \to \Re^1$ – целевая функция i-го элемента

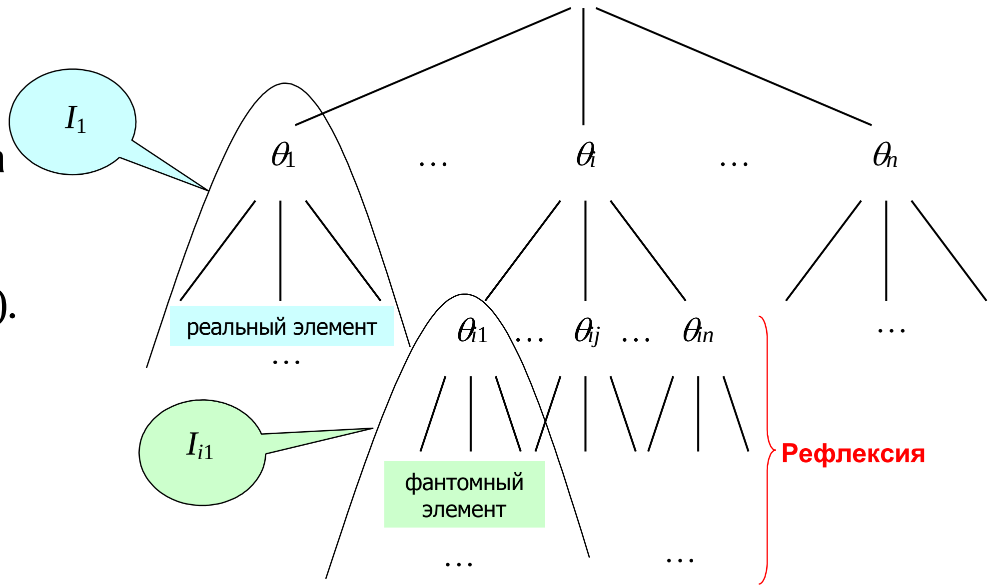
Информационное равновесие

$$x^*_{\sigma i} \in \operatorname*{Arg\,max}_{x_i \in X_i} f_i(\theta_{\sigma i}, x^*_{\sigma i1}, \dots, x^*_{\sigma i,i-1}, x_i, x^*_{\sigma i,i+1}, \dots, x^*_{\sigma i,n}).$$

Задача информационного управления

$$\min_{x \in \Psi_X(I)} \Phi(x, I) \xrightarrow[I \in \Im]{} \max.$$

Структура информированности



$\Psi_X(I) \subseteq X'$ – множество векторов действий реальных элементов, являющихся равновесными при структуре информированности I;

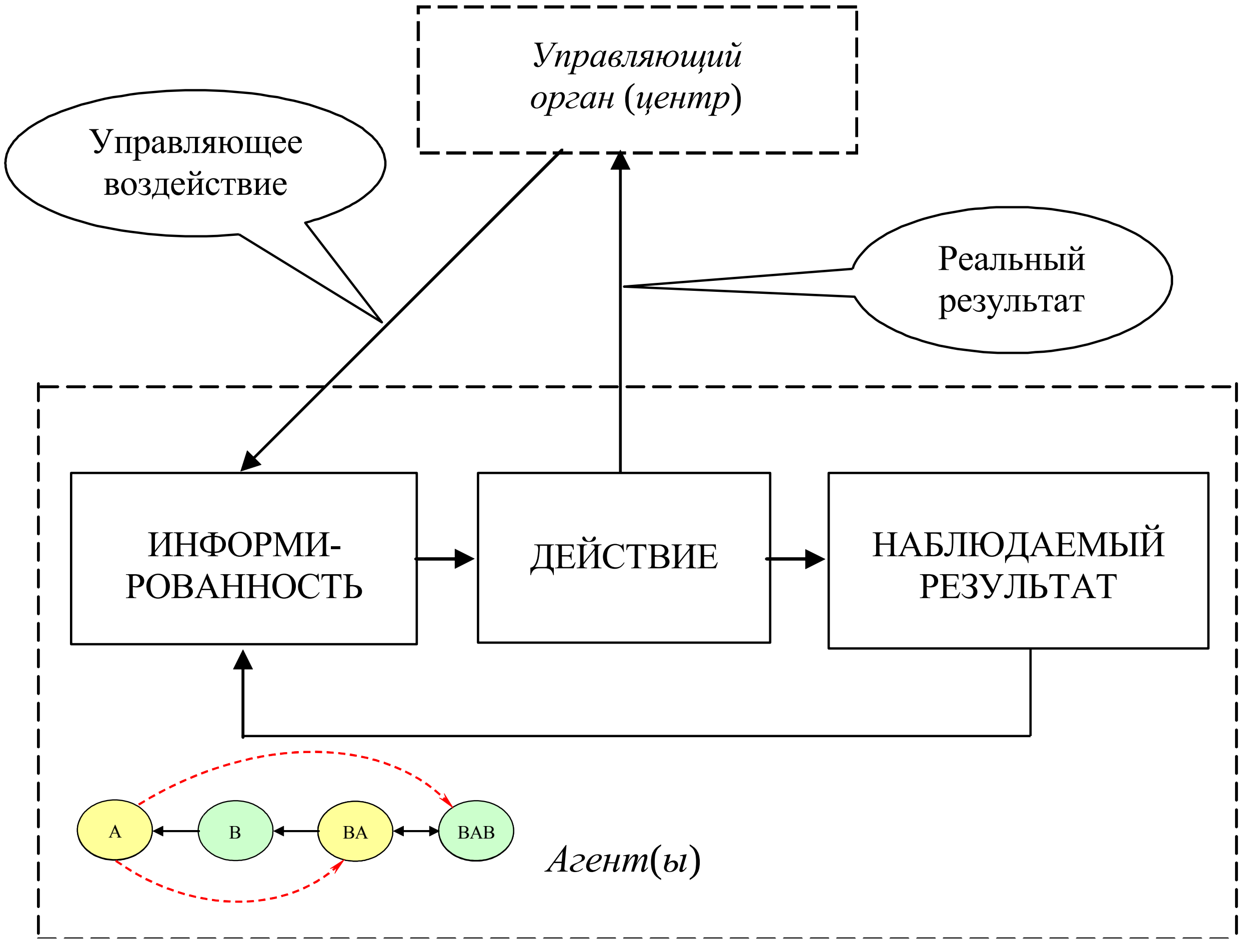
$\Phi(x, I)$ – целевая функция центра;

$\Im$ – множество допустимых структур информированности.

Учет взаимной информированности элементов дает возможность:

1. **(с нормативной точки зрения) расширить множество исходов их игры, что, в свою очередь, увеличивает эффективность управления;**
2. **(с дескриптивной точки зрения) многие наблюдаемые на практике ситуации, которые не могут быть интерпретированы как «обычные» равновесия Нэша в условиях общего знания, являются информационным равновесием.**

МОДЕЛЬ ИНФОРМАЦИОННОГО УПРАВЛЕНИЯ



АКТИВНЫЙ ПРГНОЗ

Активный прогноз – сообщение информации о будущих значениях параметров, зависящих от состояния природы и действий агентов.

$\Gamma_I = \{N, (X_i)_{i \in N}, (f_i(\cdot))_{i \in N}, \Omega, I\}$ – рефлексивная игра

$z(\cdot): \Omega \times X' \to Z, \ i \in N$

Z – множество прогнозов

Центр сообщает агентам прогноз $z \in Z$.

Агенты решают систему соотношений

$x_i \in BR_i(\theta, x_{-i}), \ i \in N,$

$z(\theta, x) = z$.

«Вечером 6 января 1981 года Джозеф Гранвилл, известный советник по капиталовложениям во Флориде, отправил своим клиентам телеграмму: «Цены на акции резко упадут; продавайте завтра». Очень скоро все узнали о совете Гранвилла, и 7 января стало самым черным днем во всей истории Нью-йоркской фондовой биржи. По общему мнению, акции потеряли в цене около 40 миллиардов долларов».

АКТИВНЫЙ ПРОГНОЗ: «ПРИНЦИП ДЕФИЦИТА» (пример)

Чалдини Р. Психология влияния.

Три типа клиентов (закупщиков импортной говядины для супермаркетов):

1. услышали предложение, сделанное в стандартной форме;
2. дополнительно была предоставлена информация о том, что поставки импортной говядины будут сокращены;
3. те же сведения, что и вторая группа, а также информацию о том, что мало кто узнает о предстоящем сокращении поставок.

Результат информационного управления:

клиенты типа 2 заказали в 2 раза больше, чем клиенты типа 1;

клиенты типа 3 заказали в 6 раза больше, чем клиенты типа 1.

«ПРИНЦИП ДЕФИЦИТА». ВЗАИМОДЕЙСТВИЕ АГЕНТОВ

n агентов (клиентов) с целевыми функциями

$$f_i(x_1,\dots,x_n) = (Q - \sum_{j \in N} x_j)\, x_i - c x_i,$$

где $x_i \geq 0$, $i \in N = \{1, \dots, n\}$, $c \geq 0$.

x_i – объем продаж агента за период времени

$(Q - \sum_{j \in N} x_j)$ – цена, которая при этом устанавливается на рынке

c – оптовая цена, по которой агенты закупают товар

$x_i = \dfrac{Q - c}{n + 1}$ – действие агентов типа 1 (равновесие в обычных условиях)

$2x_i$ – действие агентов типа 2, считающих прекращение поставок общим знанием

$6x_i$ – действие агентов типа 3, считающих себя инсайдерами (а долю инсайдеров равной пятой части от общего числа агентов)

БИПОЛЯРНЫЙ ВЫБОР: СТОРОННИКИ, КОЛЕБЛЮЩИЕСЯ, ПРОТИВНИКИ

Агенты из бесконечно большой «популяции» выбирают между позитивным и негативным полюсами.

Сторонники (их доля в популяции α_1): $x_1 = 1$.

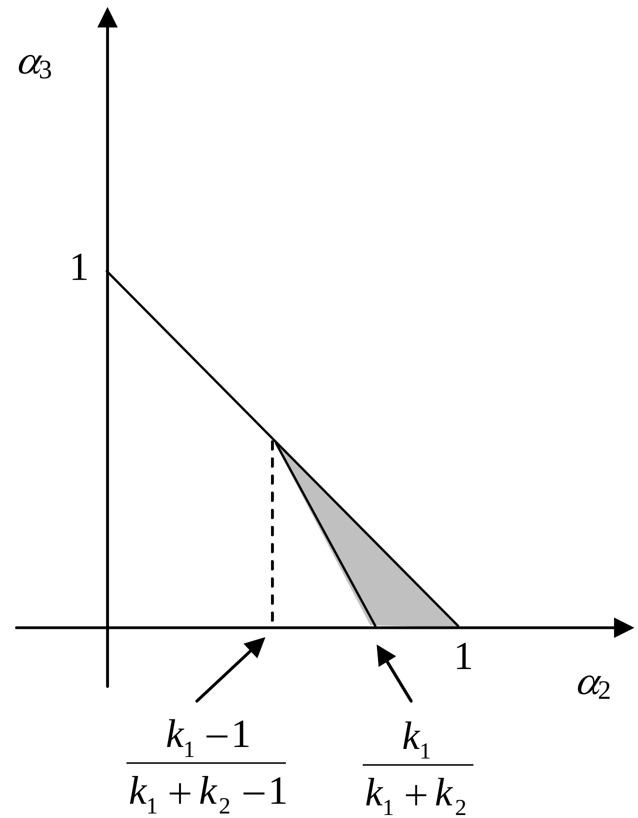
Колеблющиеся (их доля в популяции α_2) поступают так, как ожидают от «среднестатистического» члена популяции: $x_2 = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3$.

Противники (их доля в популяции α_3): $x_3 = 0$.

Центр стремится максимизировать вероятность позитивного выбора. Он может

1. повлиять на третью группу, переведя долю y ее членов во вторую и затратив ресурс $C_2 y$;
2. повлиять на вторую группу, изменив представления ее членов об α_3 и затратив ресурс $C_1 x$.

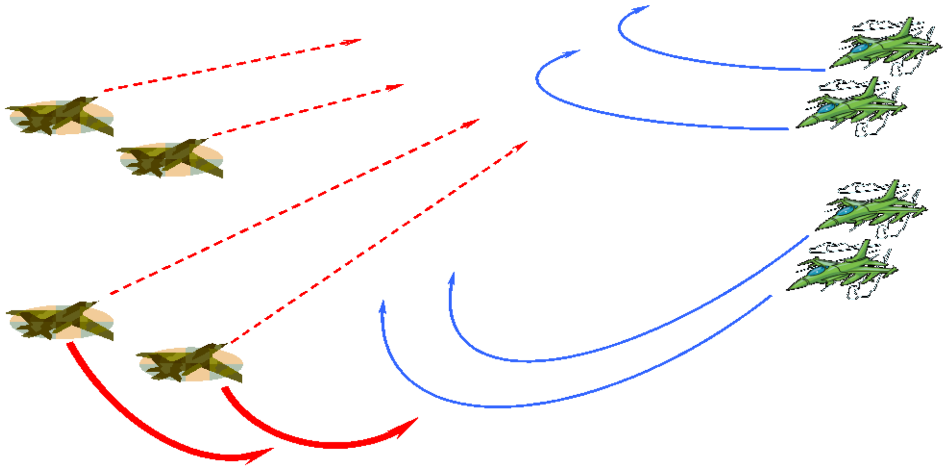
Совокупный ресурс (бюджет) центра составляет C.



На рисунке затемнена область, где оптимально весь ресурс направить на изменение представлений агентов второй группы.

$k_i = C_i / C > 1$, $i = 1, 2$

РЕФЛЕКСИВНАЯ МОДЕЛЬ ГРУППОВОГО ВОЗДУШНОГО БОЯ; ИНФОРМАЦИОННОЕ УПРАВЛЕНИЕ В ЭКСПЕРТИЗЕ



Манипулирование результатами экспертизы. Предположим, что центр заинтересован в том, чтобы результат экспертизы был как можно ближе к значению $x_0 \in [d;D]$. Пусть центру известны мнения n экспертов $\{r_i \in [d;D]\}_{i \in N}$, но никому из них не известны достоверно мнения остальных. Принимаемое решение $x = \pi(s)$, где $s \in [d;D]^n$ – вектор сообщений экспертов. Обозначим: $x_{0i}(a,r_i)$ – решение уравнения $\pi(a, \ldots, x_0, \ \ldots, a) = r_i$,

$$d_i(r_i) = \max \{d;\ x_{0i}(D,\ r_i)\},\ D_i(r_i) = \min \{D;\ x_{0i}(d,\ r_i)\},$$

$$d(r) = (d_1(r_1),\ d_2(r_2),\ \ldots,\ d_n(r_n)),\ \ D(r) = (D_1(r_1),\ D_2(r_2),\ \ldots,\ D_n(r_n)).$$

Утверждение. Если мнение каждого эксперта известно организатору экспертизы, но неизвестно другим экспертам, то за счет рефлексивного управления любой результат $x_0 \in [\pi(d(r)); \pi(D(r))]$ может быть реализован как коллективное решение. При этом достаточно ограничиться вторым рангом рефлексии экспертов.

РЕФЛЕКСИВНАЯ НЕМАНИПУЛИРУЕМОСТЬ МЕХАНИЗМОВ ПЛАНИРОВАНИЯ

Механизм: i-й агент ($i \in N$) выбирает сообщение центру $s_i \in S_i$.

Центр назначает агентам планы $x_i = h_i\,(s) = h_i(s_1, \ldots, s_n)$.

Агенты получают выигрыши $f_i(x_i, r_i)$.

1. «Классическая» неманипулируемость (типы агентов являются общим знанием): для любого типа r_i равновесной стратегией i-го агента является сообщение $s_i^* = r_i$.
2. Рефлексивная неманипулируемость: для любого типа r_i существует подструктура информированности $(r_{i\sigma})$, $i \in N$, $\sigma \in \Sigma_+$, такая, что $s_i^* = r_i$.

РЕФЛЕКСИВНАЯ НЕМАНИПУЛИРУЕМОСТЬ МЕХАНИЗМОВ ПЛАНИРОВАНИЯ

E_N – множество наборов типов, для которых сообщение каждым агентом своего истинного типа является равновесием Нэша:

$$E_N = \{ r \in \Re^n \mid \forall i \in N \ \forall s_i \in \Re^1 \ \ f_i(h_i(r), r_i) \geq f_i(h_i(r_1, \ldots, r_{i-1}, s_i, r_{i+1}, \ldots, r_n), r_i) \}.$$

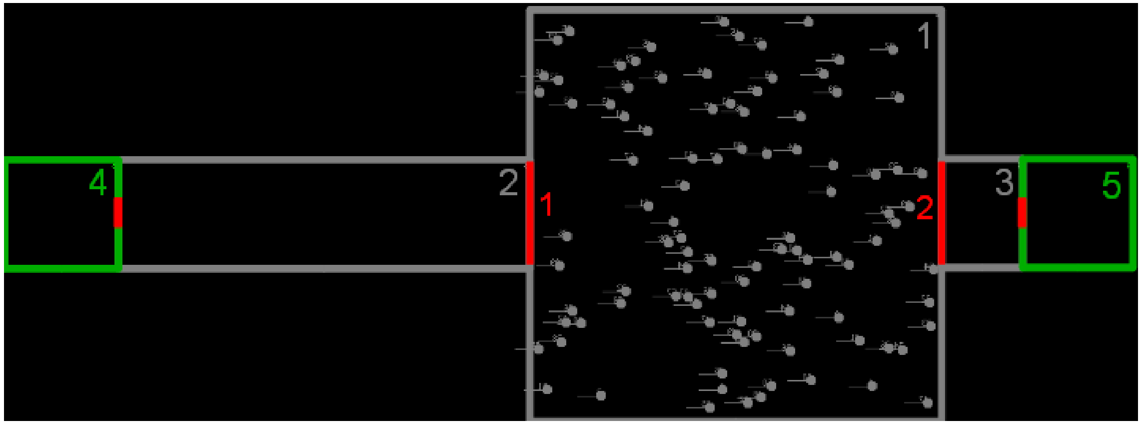
<u>Утверждение.</u> Для того, чтобы механизм планирования являлся рефлексивно неманипулируемым, достаточно, чтобы для любого i-го агента, $i \in N$, существовал набор типов $r' = (r'_1, \ldots, r'_n) \in E_N$ такой, что выполнено

$$r_i \in \operatorname*{Arg\,max}_{s_i \in \Re^1} f_i(h_i(r'_1, \ldots, r'_{i-1}, s_i, r'_{i+1}, \ldots, r'_n), r_i).$$

При этом для реализации рефлексивной неманипулируемости достаточно ограничиться не более чем вторым рангом рефлексии агентов.

ПРИМЕР: ЭВАКУАЦИЯ ИЗ ПОМЕЩЕНИЯ

- Рассмотрим помещение, в котором находятся n агентов;
- В помещении имеются два выхода, условно назовем их «левым» (L) и «правым» (R). Время выхода определяется моментом времени, когда из данного выхода вышел последний агент, направившийся к нему. Каждый агент однократно принимает решение, из какого выхода он будет выходить. Обозначим n_L (n_R) – число агентов, направившихся к левому (правому) выходу, $n_L + n_R = n$.
- Скорости движения всех агентов в отсутствие пробок примем одинаковыми;
- Пусть известна зависимость $T(k)$ времени выхода в зависимости от числа агентов $k \geq 0$. Зависимость эту будем считать непрерывной, выпуклой (отражение эффекта «пробок») и равной нулю в нуле (когда имеется один агент, пробки отсутствуют, и он покидает помещение без задержек). Обозначим через T_L (T_R) время движения агента до левого (правого) выхода, причем $T_L > T_R$, то есть правый выход расположен ближе левого. Итак, полное время выхода налево равно $T^L(n_L) = T_L + T(n_L)$, направо: $T^R(n_R) = T_R + T(n_R)$.



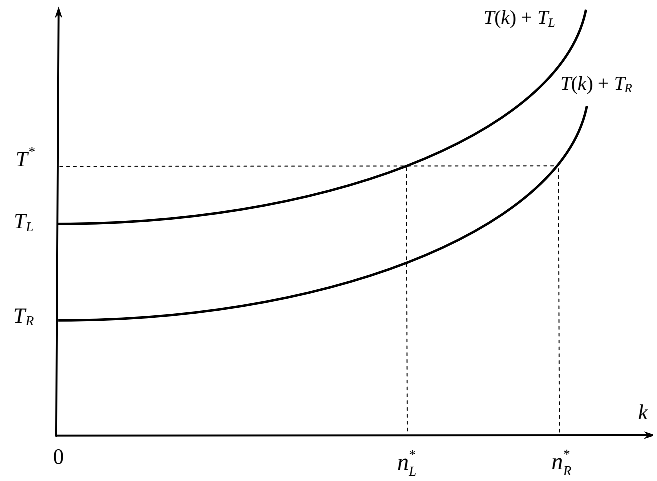
ОПТИМАЛЬНОЕ ВРЕМЯ ЭВАКУАЦИИ

Оптимальное с точки зрения *времени эвакуации* T^* – покидания помещения последним из агентов – распределение агентов по направлениям движения $(n_L^*;\ n_R^*)$ является решением следующей системы уравнений (смотрите также график):

(13) $$\begin{cases} T(n_L^*) + T_L = T(n_R^*) + T_R, \\ n_L^* + n_R^* = n. \end{cases}$$

Минимальное время эвакуации равно

(14) $$T^* = T(n_L^*) + T_L = T(n_R^*) + T_R\,.$$



Зависимость времени эвакуации от числа агентов, выбирающих правый или левый выход

ЭВАКУАЦИЯ РЕФЛЕКСИВНЫХ АГЕНТОВ

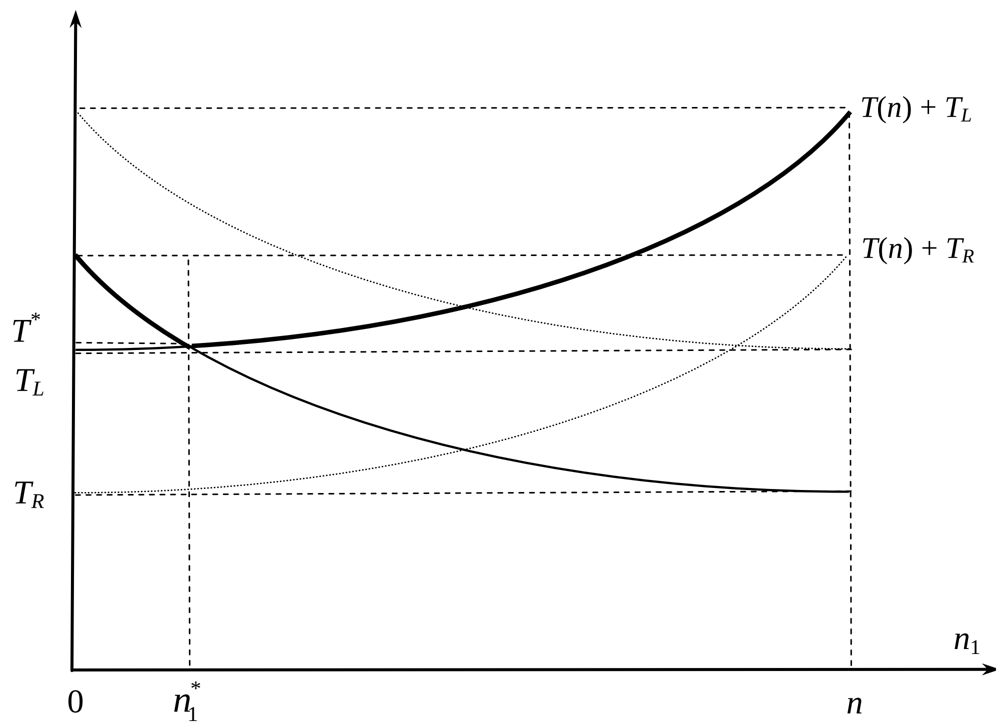
Агенты нулевого ранга рефлексии будут выбирать правый выход, агенты первого ранга рефлексии, прогнозируя, что в правом выходе агенты нулевого ранга создадут пробку, выберут левый выход.

Время выхода в зависимости от числа агентов первого ранга рефлексии равно

(15) $T1(n_1) = \max\{T(n_1) + T_L; T(n - n_1) + T_R\}$

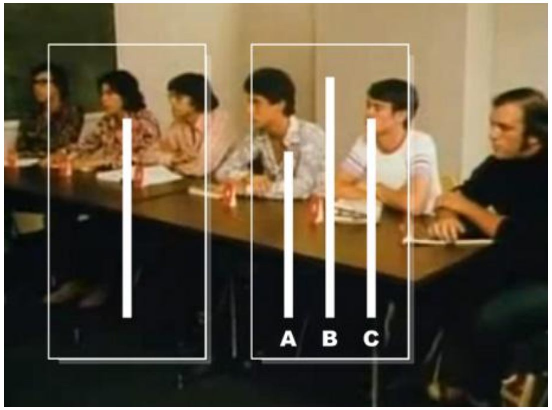
Видно (см. график), что существует оптимальное число агентов 1 ранга, при котором время эвакуации минимально. Обозначим это оптимальное число агентов первого ранга рефлексии n_1^*, его можно найти из свойств (15): $T(n_1^*) + T_L = T(n - n_1^*) + T_R$.

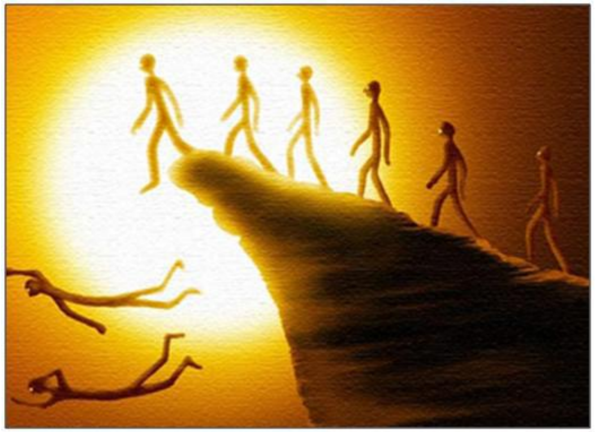
Последнее условие совпадает с условием (14), т. е. $n_1^* = n_L^*$, $T1(n_1^*) \equiv T^*$, значит, первый ранг рефлексии является максимальным целесообразным в рамках рассматриваемой модели.



Зависимость времени эвакуации от числа агентов первого ранга рефлексии

МОДЕЛЬ ТОЛПЫ





Пороговое поведение: $x_i = BR_i(x_{-i}) = \begin{cases} 1, \text{ если } \sum_{j \neq i} t_{ij} x_j \geq \theta_i, \\ 0, \text{ если } \sum_{j \neq i} t_{ij} x_j < \theta_i. \end{cases}$

Модель динамики коллективного поведения: в начальный момент времени все агенты бездействуют, далее в каждый из последующих моментов времени агенты одновременно и независимо действуют в соответствии с пороговой процедурой. Обозначим

$$Q_0 = \{i \in N \mid \theta_i = 0\},\ Q_k = Q_{k-1} \cup \{i \in N \mid \sum_{j \in Q_{k-1},\, j \neq i} t_{ij} \geq \theta_i\},\ k = 1, 2, \ldots, n-1,$$

$$q(T, \theta) = \min \{k = \overline{0, n-1} \mid Q_{k+1} = Q_k\}.$$

Равновесие коллективного поведения (РКП): $x_i^*(T, \theta) = \begin{cases} 1, \text{ если } i \in Q_{q(T,\theta)} \\ 0, \text{ если } i \in N \setminus Q_{q(T,\theta)} \end{cases}$, $i \in N$.

Агрегированным показателем состояния толпы будем считать число действующих агентов:

$$K(T, \theta) = |Q_{q(T, \theta)}|,$$

Задача управления: $H(K(T, \theta)) - C(T, \theta, T^0, \theta^0) \to \max_{T \in \mathrm{T},\, \theta \in \Theta}$.

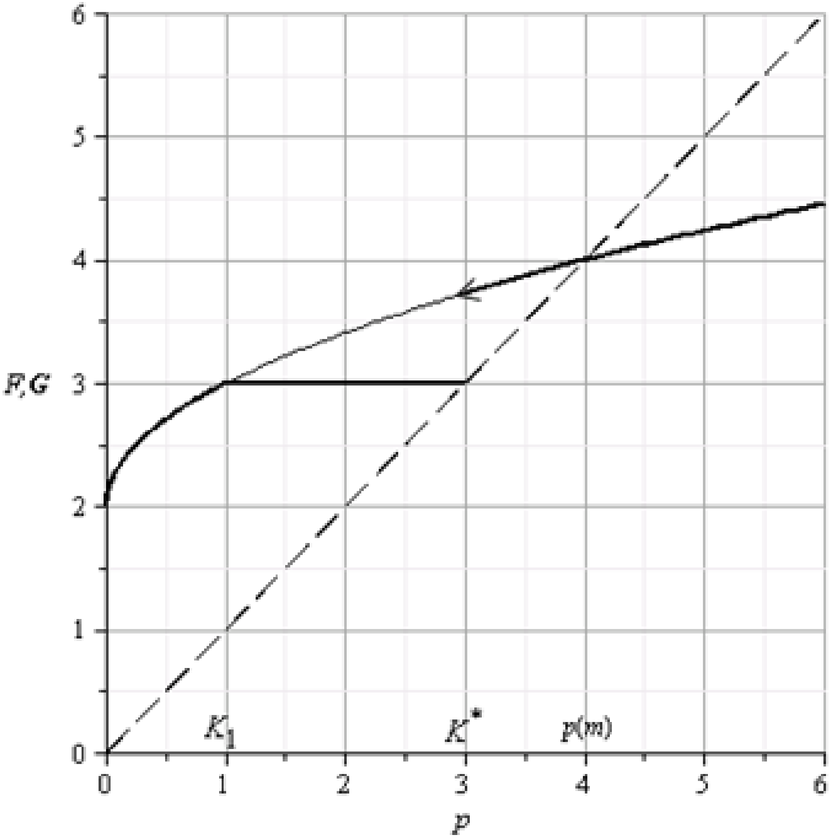
МОДЕЛЬ УПРАВЛЕНИЯ ТОЛПОЙ

Затраты на изменение порогов:

$$C\left(m,m^0\right)=\sum_{i=1}^{n}c_i\left(m_i,m_i^0\right)=\sum_{i=1}^{n}g\left(\left|m_i-m_i^0\right|\right)$$

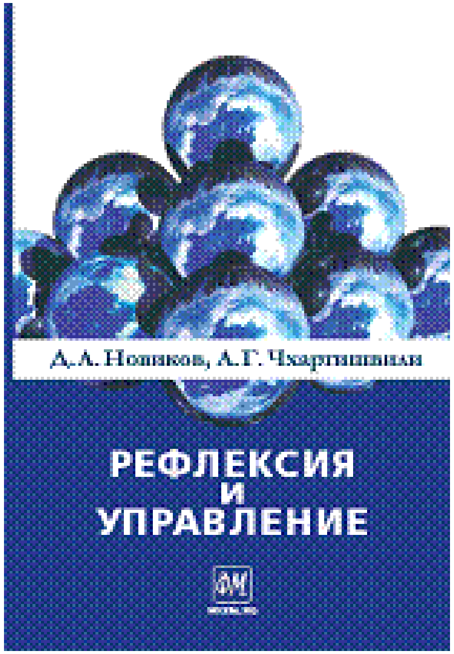
Минимальные затраты перехода в K^*:

$$\int_{F^{-1}\left(K^*\right)}^{K^*} g\left(K^*-t\right)dF\left(t\right)$$



<u>Утверждение</u>. Не существует стабильного информационного равновесия, при котором действуют строго меньшее число агентов, чем в РКП.

ЛИТЕРАТУРА



2003 **2012**

www.mtas.ru