



# **Механизмы сознания и самосознания в управлении целенаправленным поведением**

Кузнецова Ю.В.

Осипов Г.С.

Панов А.В.

Чудова Н.В.

Институт системного анализа РАН  
gos@isa.ru

# Поведение интеллектуальных систем

Главная задача - синтез плана целенаправленного поведения:

- в условиях прогнозируемой среды;
- в условиях непрогнозируемой среды.

Эта задача имеет комбинаторный характер, а основные усилия при её решении направлены на борьбу с вычислительной сложностью.



# Планирование поведения

Существующие подходы основаны на поиске в пространстве состояний, поиске в пространстве планов, методах иерархического планирования, планирования на основе прецедентов и удовлетворения ограничений.

# Цели и планы

- Цель или множество целей считаются заданными. На самом деле задачи планирования поведения и целеполагания неразрывно связаны и для их решения используются иные механизмы, нежели в существующих интеллектуальных планировщиках.
- К таким механизмам относятся механизмы знакового опосредования и мотивационно-смысловой регуляции:
  - знаковое опосредование лежит в основе картины мира субъекта;
  - мотив обладает смыслообразующей функцией; благодаря ему предметы обретают смысл с точки зрения текущей деятельности.



# Методологические основы

Основой предлагаемой модели являются базовые положения культурно-исторического подхода в психологии, «тартусского» подхода в семиотике, идеи прикладной семиотики Д.А.Поспелова.

# Методологические основы

- А.Н.Леонтьев: *«ключ к морфологии сознания лежит в морфологии деятельности».*
- Л.С.Выготский: *«... употребление знаков, сущность которого состоит в том, что человек воздействует на поведение через знаки ... ».*
- Ю.М.Лотман и Б.А.Успенский: *«презумпция языка и осмысленности»*

# Задачи

Построение:

- базовых процедур осознания
- модели знакового опосредования
- механизмов формирования знака
- процедур синтеза поведения

# Элементы сознания

- Элементы сознания по А.Н.Леонтьеву:
  - образ,
  - значение,
  - личностные смыслы.
- Значение объекта, его образ и личностные смыслы могут не связываться в единое целое – не происходит осознания.
- Тогда образ носит характер перцепта, личностный смысл становится биологическим смыслом, значение – функциональным назначением.

# Связывание

- Осознавание, т.е. включение объекта в поле сознания связано с возможностью рассмотрения объекта как целостного и существующего независимо от текущего состояния действующего субъекта, обеспечивается введением имени объекта, связывающего упомянутые три компоненты в единое целое.
- Только благодаря связыванию все компоненты закрепляются, происходит «стабилизация».
- При этом, образ, личностные смыслы и значения объекта входят в сознание как элементы не просто знания, а как компоненты знака.

# Базовые процедуры осознавания

Обозначим:

- $A$  – множество личностных смыслов;
- $M$  - множество значений;
- $P$  - множество процедур распознавания образов;

тогда

- $a \subseteq A$  подмножество множества личностных смыслов (возможно пустое);
- $m \subseteq M$  подмножество множества значений;
- $p \subseteq P$  подмножество множества образов.

# Базовые процедуры осознавания

Рассмотрим отображения:

$$\Psi_p^m : 2^P \rightarrow 2^M \quad \Psi_m^a : 2^M \rightarrow 2^A \quad \Psi_a^p : 2^A \rightarrow 2^P$$

и обратные им:

$$\Psi_a^m : 2^A \rightarrow 2^M \quad \Psi_m^a : 2^A \rightarrow 2^M \quad \Psi_p^a : 2^P \rightarrow 2^A$$

соответственно, такие, что

$$\Psi_p^m(p) = m \quad \Psi_m^a(m) = a \quad \Psi_a^p(a) = p$$

аналогично для обратных отображений.

# Образование протознака

Результатом работы БПО является протознак  
если оператор

$$\Psi_a^r \Psi_t^a \Psi_r^t$$

является тождественным оператором.

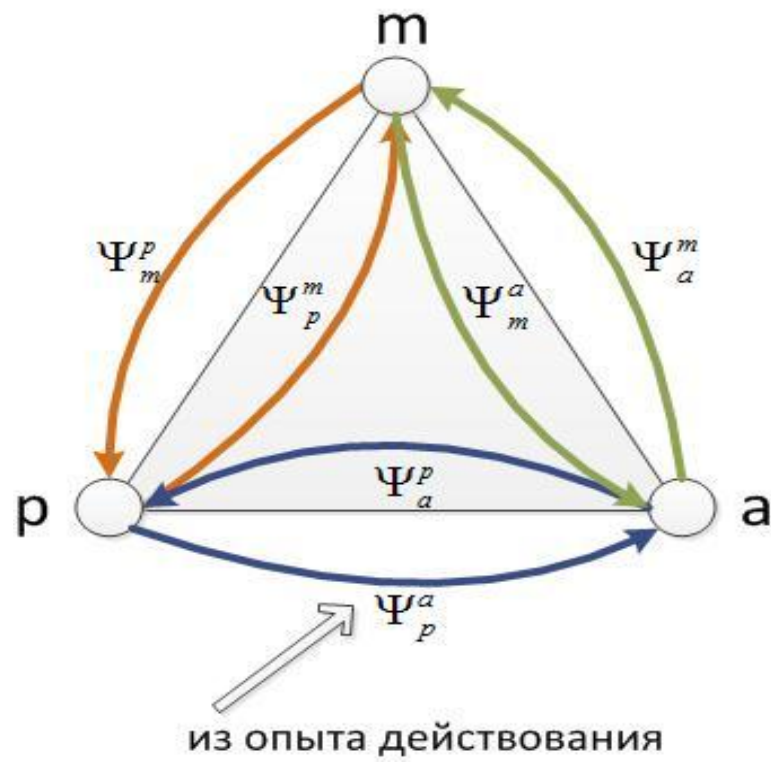
**Утверждение 1.** Для всякого протознака операторы

$$\Psi_t^a \Psi_r^t \Psi_a^r$$

$$\Psi_r^t \Psi_a^r \Psi_t^a$$

суть тождественные операторы.

# Протознак



# Интерпретация БПО

- Переход от биологического смысла к перцепту обеспечивает выбор в конкретной ситуации объекта, наилучшим образом удовлетворяющего потребность;
- переход от перцепта объекта к его функциональному значению в конкретной ситуации реализует выбор способа его использования;
- переход от функционального значения объекта к его биологическому смыслу создаёт условия для выбора нового мотива деятельности;

# Интерпретация БПО

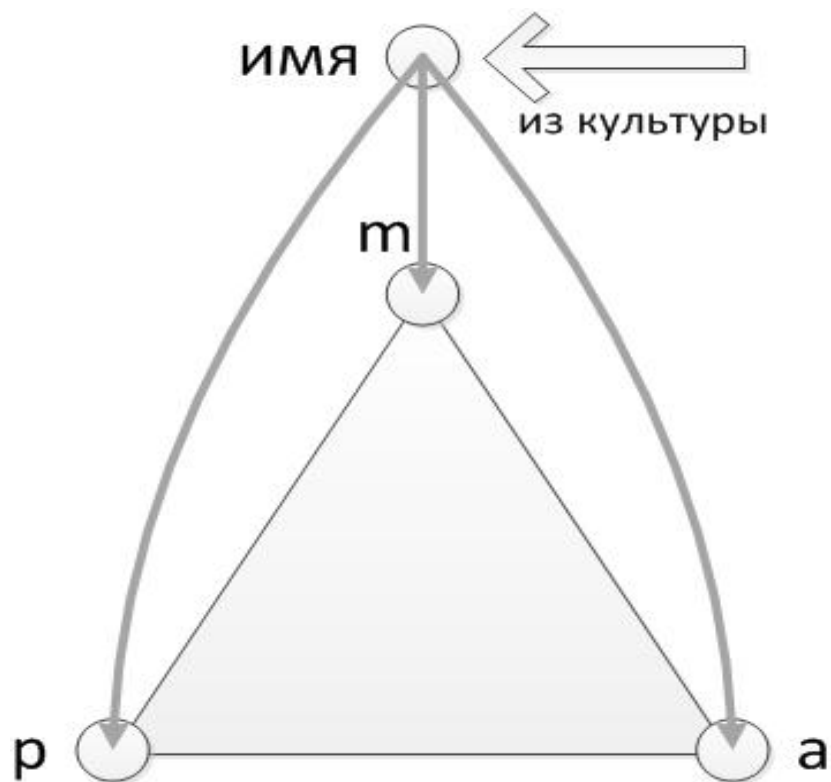
- переход от значения к перцепту соответствует выбору средства;
- переход от перцепта к биологическому смыслу соответствует определению ценности объекта;
- переход от биологического смысла к значению соответствует возникновению неопредмеченой потребности.

# Двухфазное решение задач

- Переход от биологического смысла к значению соответствует возникновению неопредмеченной потребности;
- переход от значения к перцепту соответствует выбору средства

Последовательное выполнение этих двух переходов дает решение задачи

# Формирование знака



# Модель знака

Знак образуется в результате связывания имени с компонентами протознака:

- $\hat{A}: N \rightarrow 2^A$ , так что  $\hat{A}(n) = a$ ,
- $\hat{Y}: N \rightarrow 2^M$ , так что  $\hat{Y}(n) = m$ ,
- $\pi: N \rightarrow 2^P$ , так что  $\pi(n) = p$ ,

Легко видеть, что, например,

- $\Psi_a^m(\hat{A}) = \hat{Y}$
- $\Psi_p^a(\pi) = \hat{A}$
- $\Psi_p^m(\pi) = \hat{Y}$

# Картины мира субъекта

Взаимодействие компонент знаков приводит к формированию индивидуальных картин мира субъектов:

- картины мира, построенной на образах,
- картины мира, построенной на личностных смыслах,
- картины мира, построенной на значениях.

Имена знаков могут наследовать одну из трех возникающих структур, а фиксация одной из них, порожденной либо образами, либо смыслами, либо значениями, порождает *точку зрения* субъекта.

# Объектная картина мира

- Картина мира, построенная на образах , возникает как результат взаимодействия знаков посредством их *образов*.
- Это взаимодействие проявляется в формировании семейства бинарных отношений на множестве образов.
- Среди формирующихся на множестве образов отношений - отношения сходства, противопоставления и вхождения

# Нарративная картина мира

Картина мира на множестве личностных смыслов возникает как результат взаимодействия множеств личностных смыслов различных знаков. Для её формирования используются следующие отношения:

- поглощения  $a_1 \subseteq a_2$  (смысл  $a_2$  поглощает смысл  $a_1$ ),
- объединения (агглютинации)  $a_1 \sqcup a_2$  и
- противопоставления по смыслам  $a_1 \perp a_2$
- предпочтения  $a_1 \blacktriangleright a_2$ .

Выражение  $a_1 \subseteq a_2$  отражает то обстоятельство, что множество действий, соответствующих смыслу  $a_1$  есть подмножество множества действий, соответствующих смыслу  $a_2$ .

Выражению  $a_1 \sqcup a_2$  соответствует объединение множеств действий, соответствующих смыслам  $a_1$  и  $a_2$ .

Выражение  $a_1 \perp a_2$  означает, что действия, соответствующие смыслу  $a_1$  исключают действия, соответствующие смыслу  $a_2$ .

# Сюжетно-ролевая картина мира

- При формировании картины мира, образованной значениями, строятся ситуативные классификации, роли и сценарии.

# Семиотическая сеть

- Каждая из картин мира представима в виде семантической сети (например, неоднородной семантической сети).
- Базовые операции осознания обеспечивают переходы между сетями.
- Тройку семантических сетей вместе с базовыми операциями осознания уместно назвать семиотической сетью.

# Семиотическая сеть

Таким образом, семиотическая сеть

$SN = \langle S, R, \Psi \rangle$ , где

- $S$  - множество знаков,
- $R$  – семейство отношений на знаках,
- $\Psi$  – семейство базовых операций осознания.

# Семиотическая сеть

Заметим, что имена знаков играют особую роль – они являются элементами системы языка, что создаёт новые связи между знаками – связи, определяемые языковой нормой, и связи, определяемые узусом (например, профессиональным). Таким образом, система языка выступает как эвристическая машина для формирования новых знаков.

# Сознание

- Образование новых знаков, пополнение новыми элементами отношений на множествах их компонент, появление новых отношений приводят к изменениям  $S$  и  $R$  во времени или, иначе говоря, к изменениям картин мира субъекта.
- Сознанием субъекта будем называть процесс порождения и трансформаций его картин мира.

# Модель сознания

Динамическая система

$$C: SN \times T \rightarrow SN$$

# Самосознание

В результате решения возникающих задач у субъекта накапливаются не только знания о мире, но и знания о самом себе. Этот опыт фиксируется в его самосознании.

Самосознание содержит зафиксированный опыт субъекта действующего в рамках культурной традиции:

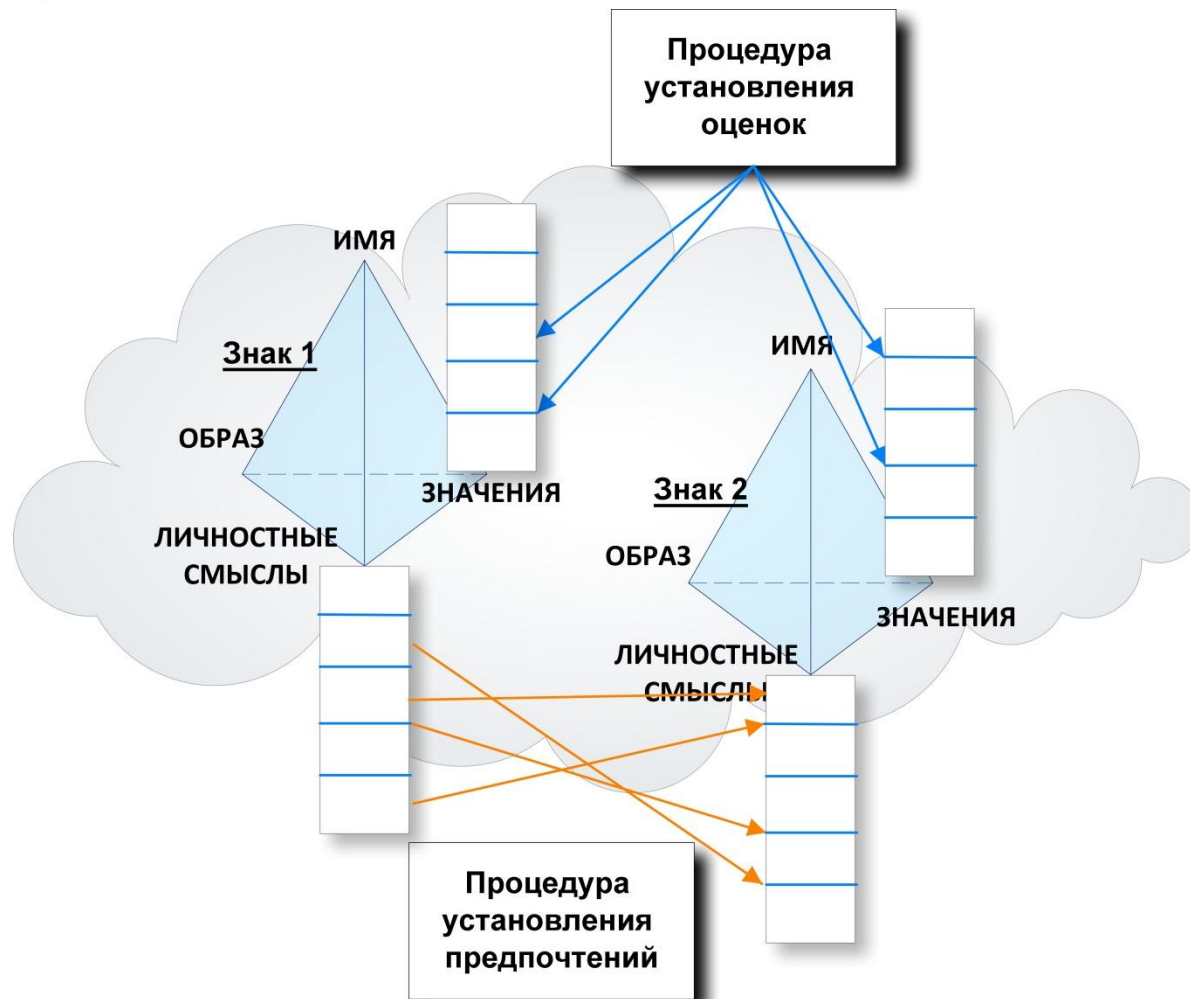
- для чего предназначен предмет (предпочитаемые значения),
- какие признаки оказались более надёжными при распознавании предмета (предпочитаемые правила построения образа),
- что даёт субъекту оперирование данным предметом (предпочитаемые смыслы).

# Модель самосознания

Носитель модели  $CS = \langle Q, PV, PA, Sk \rangle$

- $Q: N \times M \rightarrow O$ ,  $Q(n, m) = q$ ,  $m \in \hat{Y}(n)$ ,  $q \in O$  – ставящее в соответствие каждому значению  $v$  некоторую его субъективную оценку  $q$  из множества  $O$  - оценок значений;
- $PV \subseteq (M \times M)$ , так что, если  $PV(m_1, m_2)$ , то  $m_1$  предпочтительнее  $m_2$  – отношение субъективных предпочтений на множестве значений  $M$ ;
- $PA \subseteq (A \times A)$  так что, если  $PA(a_1, a_2)$ , то  $a_1$  предпочтительнее  $a_2$  – отношение предпочтений на множестве действий;
- $Sk: N \times A \rightarrow A$ ,  $Sk(n, a) = \alpha$ ,  $\alpha \in \hat{A}(n)$  – множество оценок собственных умений из множества  $\hat{A}(n)$ .

# Установление оценок и предпочтений



# Процесс целеобразования

- Для каждой ситуации из множества всех сохранённых в картине мира ситуаций вычисляется мера согласованности/рассогласованности,
- выбирается ситуация с максимальным уровнем рассогласованности,
- среди знаков картины мира проводится поиск знака, включение которого в ситуацию приводит к максимальному уменьшению рассогласованности,
- получившаяся после включения знака целевая ситуация помещается во множество текущих целей.

# Процесс планирования

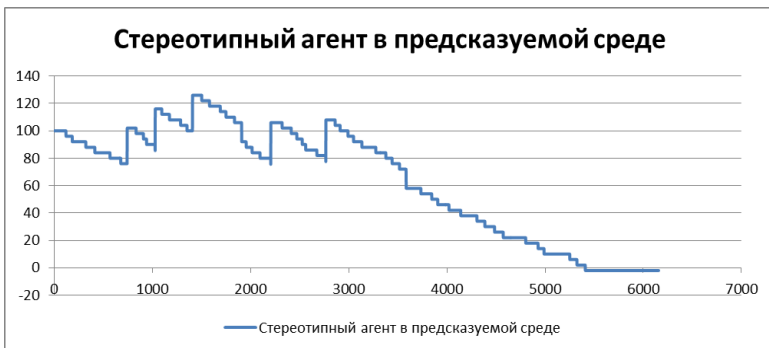
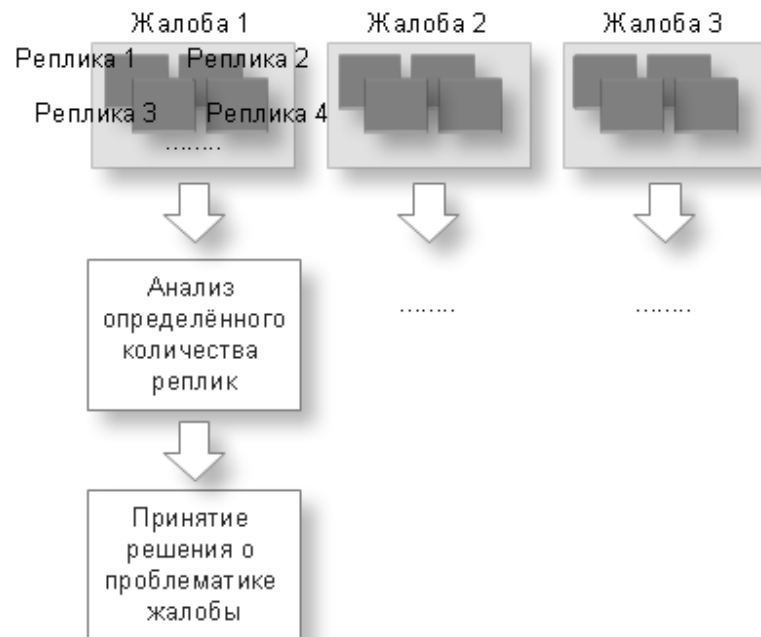
- В качестве стартовой выбирается некоторая ситуацию в картине мира,
- с помощью отношения предпочтения на множестве личностных смыслов определяется множество наиболее предпочтительных действий для каждой ситуации,
- образуется новое множество ситуаций:

$$\mathbf{N}_{sit_i} = (\mathbf{N}_{sit} \setminus \mathbf{N}_{a_i}^{out}) \cup \mathbf{N}_{a_i}^{in}, a_i = (\mathbf{N}_{a_i}^{in}, \mathbf{N}_{a_i}^{out}) \in \mathbf{a}_{act}$$

- если в полученном множестве нет искомой целевой ситуации, то переходим к шагу 2, иначе процесс считается завершенным и построенный план помещается во множество текущих планов.

# Модельная задача

Цель - показать, что поведение программных агентов, основанных на знаковом представлении, моделирует поведение психолога-консультанта в процессе интерпретации жалоб и принятия решения о том, к какой категории относится проблематика поступившей жалобы



Категории жалоб:

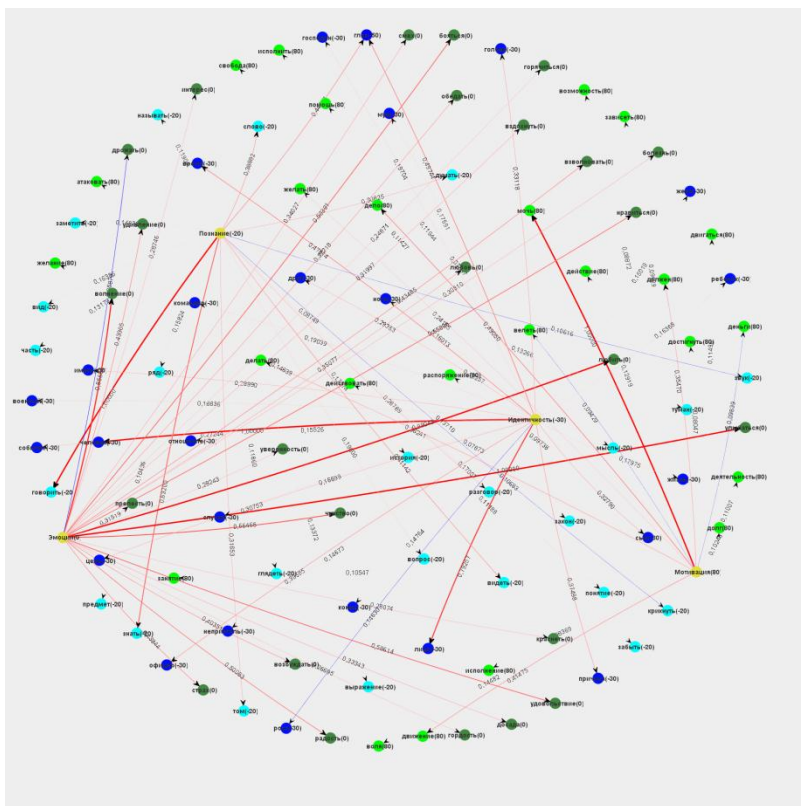
- Мотивация
- Познание
- Эмоции
- Идентичность

Типы агентов:

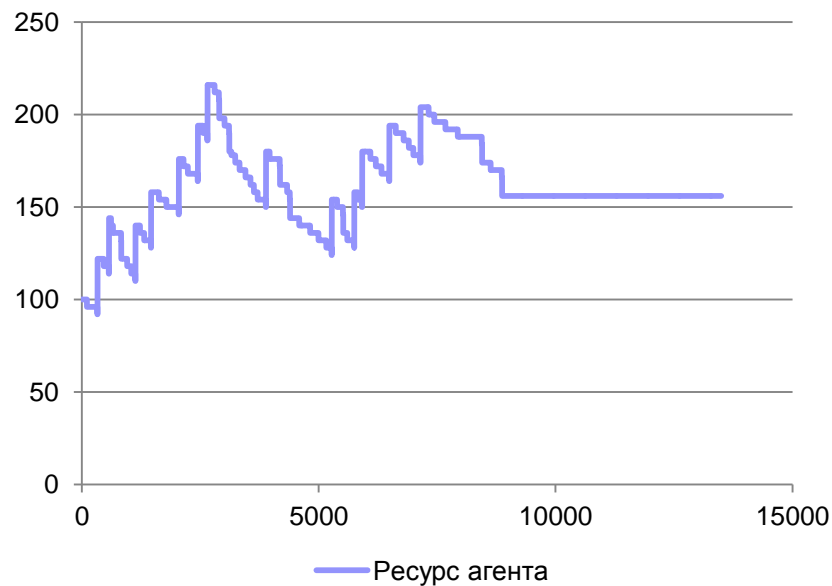
- Активный
- Пассивный
- Хаотичный
- Стереотипные

# Интерпретация результатов

Фрагмент картины мира агента

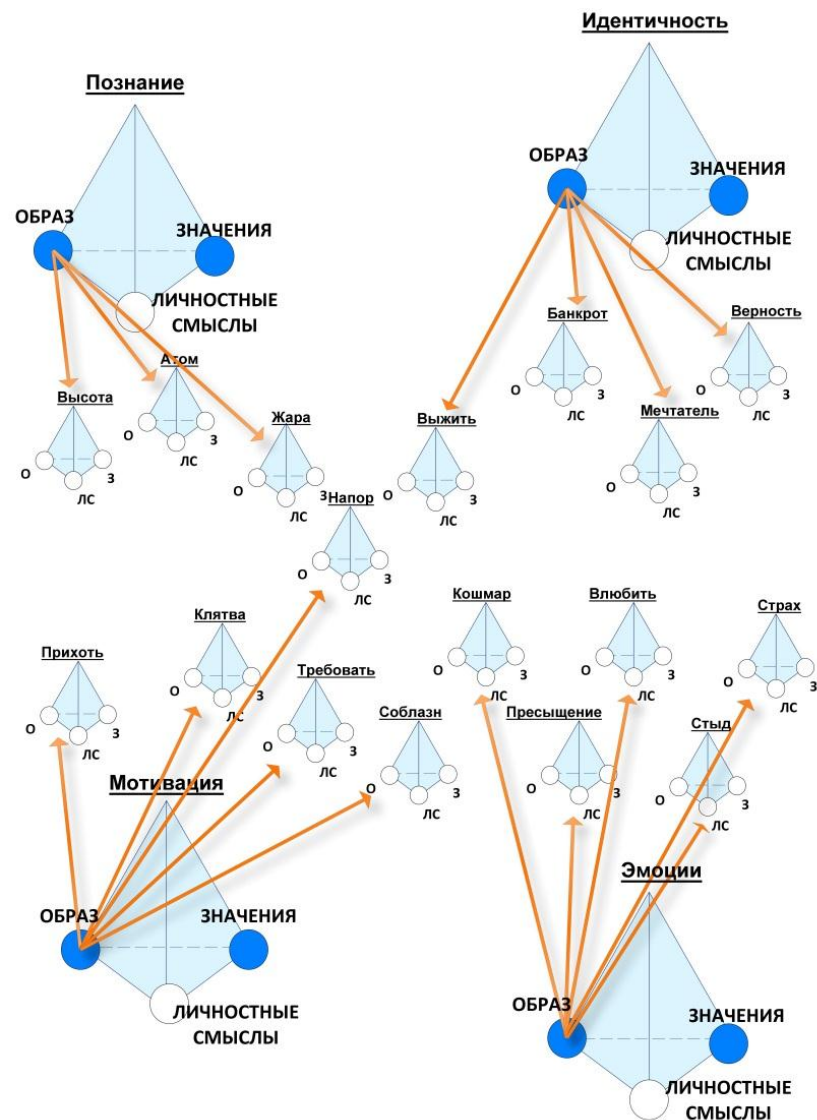


Ресурс агента



# Картина мира агента перед началом работы

- Синие круги – заполненные составляющие знака
- Пустые круги – пустые личные смыслы
- Оранжевые стрелки – отношения между знаками по образу
- О – образ, З – значения, ЛС – личные смыслы



# Картина мира агента после анализа жалоб

- Красные круги – отрицательные личные смыслы
- Зеленые круги – положительные личные смыслы

