

Карпов Валерий Эдуардович

К чему приводит наличие эмоций и темперамента у роботов.

Биологически инспирированные подходы в робототехнике

НИЦ Курчатовский институт
Отделение нейрокогнитивных наук и интеллектуальных систем,
Лаборатория робототехники
Karpov_ve@mail.ru

Биологически инспирированные архитектуры, модели и методы

Два аспекта заимствования из биологии:

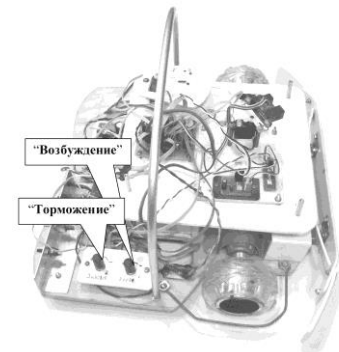
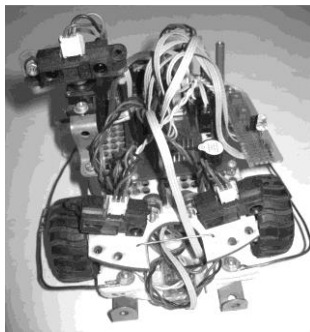
1. «Черпание вдохновения», попытки создания те или иных аналогов биологических объектов
2. Проверка общности уже разработанных моделей и методов, попытка применить их для более широкого круга задач, т.е., верификация.

Здесь рассматривается именно верификационный аспект биоинспирированности.

- Продолжение исследований по реализации механизмов эмоций и темперамента в СУ роботов. В основе модели эмоциональной СУ - информационно-потребностная теория П.В.Симонова.

$$E=f(N, p(I_{need}, I_{has}))$$

- Развитие этого подхода - создание **психологического** уровня управления поведением робота – **темперамента**.
- Особенности поведения роботов, обладающих эмоциональной-темпераментной компонентой. На этой базе реализуются такие феномены, как поведение **зомбированного паразитами индивида** и **шизофрения** (раздвоение личности).
- Имитационные модели и реальные робототехнические устройства.



Архитектура эмоционального робота

1. Набор поведенческих правил представлен в MYCIN-подобной форме:

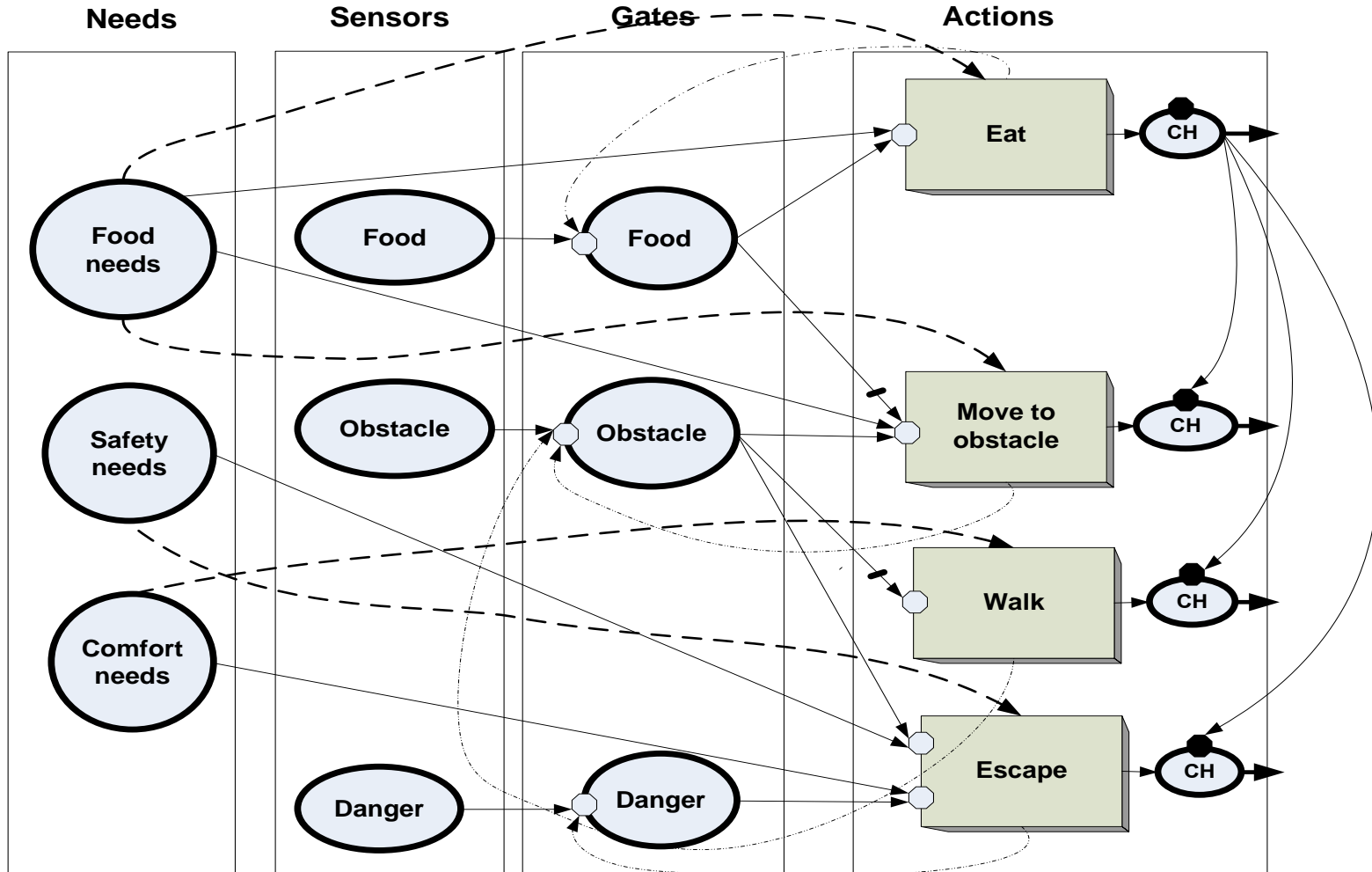
$$P_n = \text{Cond}_1 \& \dots \text{Cond}_i (a_n)$$

ЕСЛИ "Потребность в пище" (N_{food}) & "Вижу пищу" (S_{food}) ТО "Питаться" (a_{eat})

N_{food} , S_{food} могут интерпретироваться как коэффициенты уверенности.

2. Множество поведенческих процедур
(5-6: защитная реакция, поиск, поедание...)

Архитектура



○ - excitation

CH – Service neuron

---> Communication between action and need

● - inhibition

-> - «NOT»

-.-.-> Emotional feedback

Формализм

$$P=R(N, S)$$

N - множество актуальных потребностей,

S - значения сенсоров

- Введем G - “значимость” для данного правила в текущий момент времени

$$P=R(N, S, G)$$

G отражает влияние эмоционального состояния агента на выполняемое действие

$$G=V(E, N, S)$$

- Эмоция E - разность между тем, какие действия робот мог совершить в зависимости от текущей ситуации (исходя из имеющихся N и S) и вектором фактического выхода Y

$$E = k_{em}(Y - P)$$

$Y_i=1, i=a, 0$ – иначе

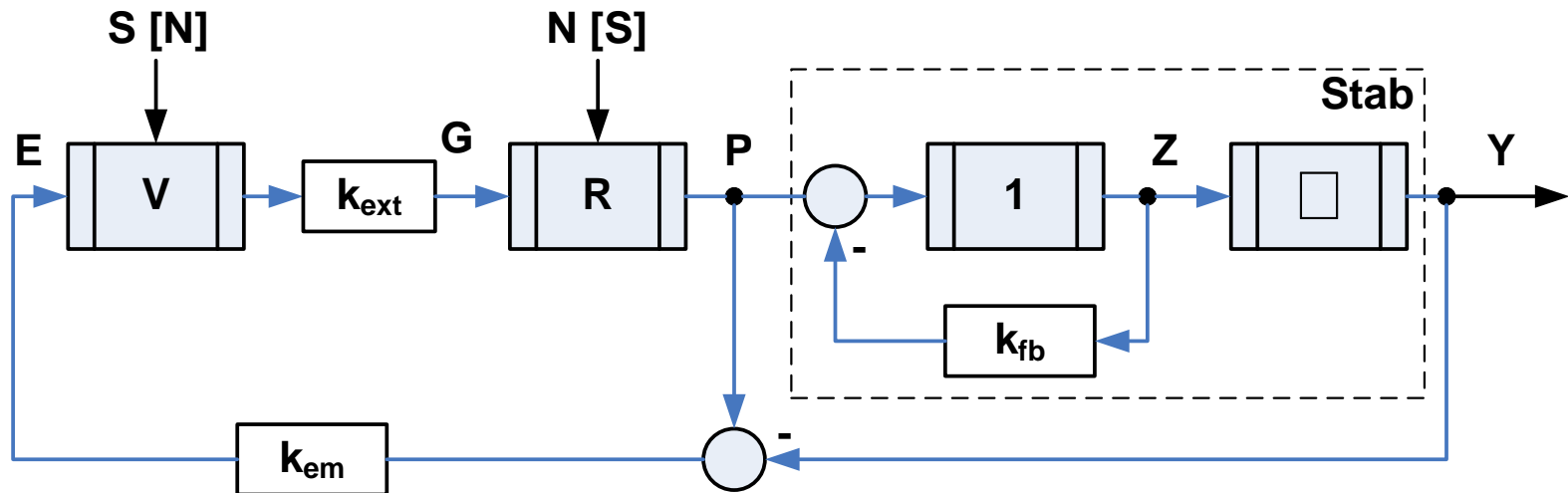
a – номер актуального действия

$a: Z_a = \max Z, Z = P - k_{fb} \sum Z$

Итоговое выражение:

$$P=R(k_{ext}, N, S, G)$$

Структурная схема



Смысл введенного формализма:

Эмоции и темперамент (коэфф. k_{ext} и k_{inh}) – не метафора, а способ описания состояния системы, характера его поведения и способов управления.

Несколько слов о темпераменте

- Экспериментально определенная зависимость между эмоциями робота и параметрами возбуждения/ торможения

Значение возбуждения (1..10)	Значение торможения (1..10)	Эмоции (-100..+100)	Тип
10 (высокое)	1 (низкое)	-60	Холерик
5 (высокое)	10 (высокое)	-24	Сангвиник
1 (низкое)	10 (высокое)	+20	Флегматик

- Соответствие конструкции автомата и темперамента

Уровень возбуждения	Уровень торможение	Темперамент	Автомат
Низкий	Низкий	Меланхолик	Автомат с линейной тактикой
Высокий	Низкий	Холерик	«Неврастеничный» автомат
Высокий	Высокий	Сангвиник	«Доверчивый» автомат
Низкий	Высокий	Флегматик	«Недоверчивый» автомат или «Пессимистический» автомат

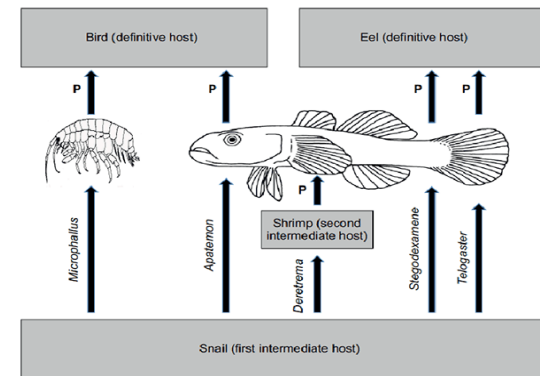
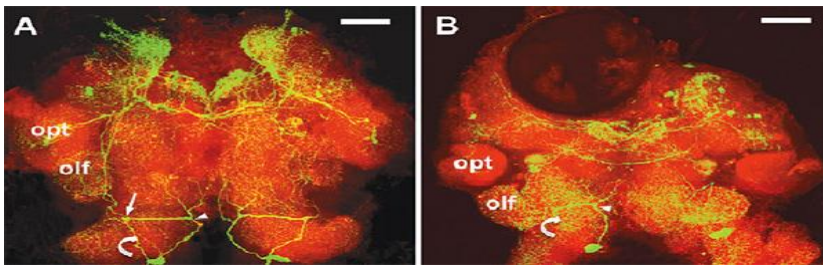
Паразитическое манипулирование (зомбирование)

Стратегии изменения поведения организма-хозяина.

1. Изменение временных последовательностей поведенческих ответов хозяина
2. Изменение направления реакции
3. Расщепление целостного поведения и синтез новых последовательностей действий

Механизмы:

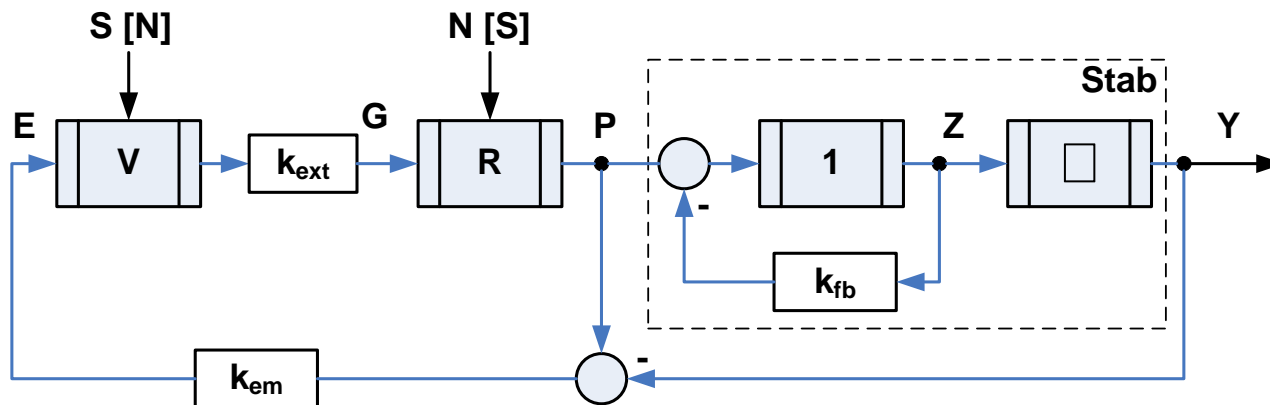
1. Атака иммунной системы хозяина => «болезненное поведение» хозяина: защитная реакция этой системы - выработка нейромодуляторов => перехват управления нейромодуляторами.
2. Поражение генов, ответственных за работу нервной системы.
3. Прямое управление различными участками мозга



- Споровик *Toxoplasma gondii*. Кошка->мышь (пром.хозяин)->кошка. Агрессия мышей и влечение к кошкам.
- Черви скребни. Вынуждают бокоплава (пром.хозяин) не избегать света, не спасаться бегством при обнаружении хищника, не искать партнера для спаривания и т. д. Задача – быть съеденным уткой.

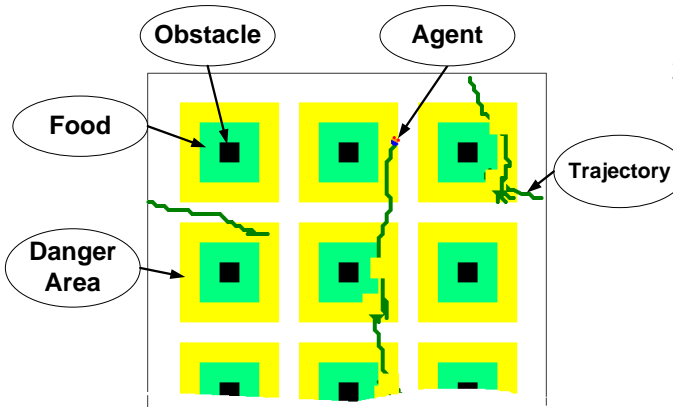
Особенность эмоциональной СУ

- Наличие контуров положительной обратной связи.
- Чувствительность к слабым воздействиям.
- Паразит - *простой* объектом с весьма ограниченным спектром возможностей.
- Изменение поведение агента, воздействуя на потребностные и оценочные каналы СУ (аналог нейромодуляционного манипулирования), а также влияя на параметры контура обратной связи (непосредственное управление участками мозга).

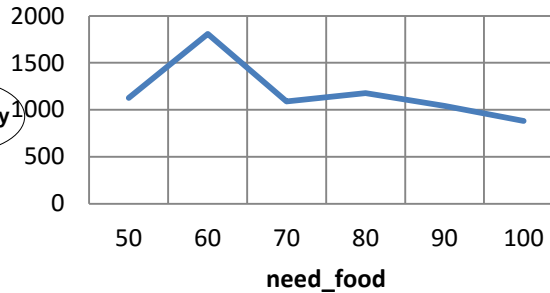


Эксперименты

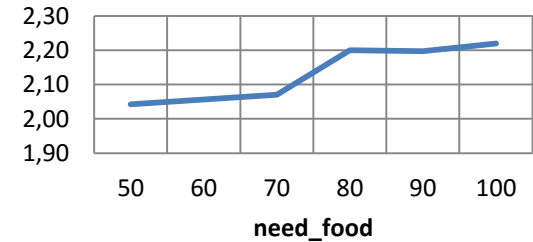
- Задача паразита – заставить агента находится в освещенной области



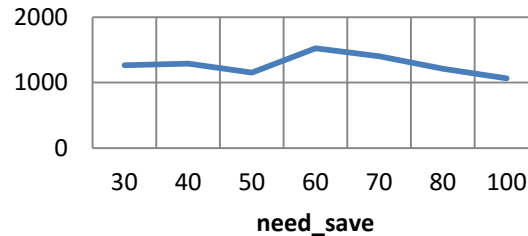
avr(Danger)



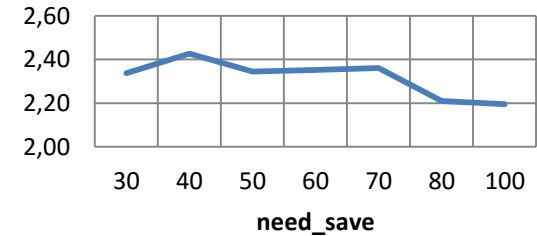
var(Proc)



avr(Danger)

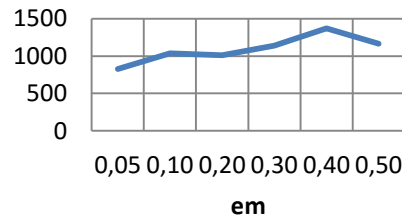


var(Proc)

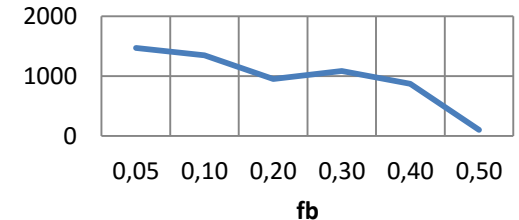


Максимальный эффект – при изменении эмоц. коэффициента k_{em} и коэффициента о.с. в контуре стабилизации k_{fb}

avr(Danger)



avr(Danger)

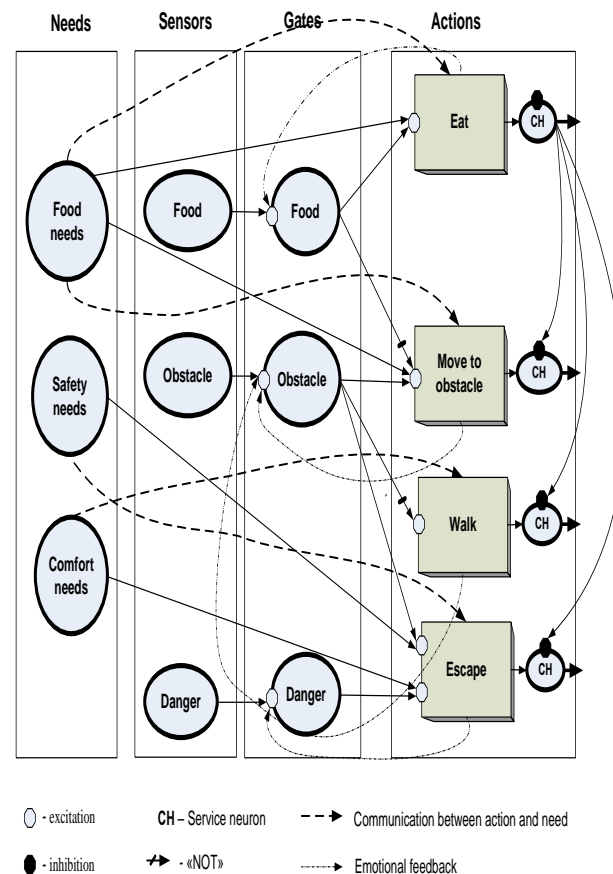


Шизофрения

- Строгое понимание Ш. связано с неадекватностью **эмоциональных** реакций => если робот обладает эмоциональной компонентой, то он вполне может быть подвержен и шизофрении.

Феномен раздвоения сознания.

- Архитектура эмоциональной СУ отражает наличие потребностей и способов их удовлетворения (моторных программ).
- Можно рассматривать СУ как совокупность агентов (G, см.рис.), обладающих своим целеполаганием, потребностями и способами их удовлетворения.
- Это означает наличие конфликтов между агентами, а эмоциональные обратные связи служат средством согласования работы подсистем, стабилизации поведения системы в целом.
- Эффекты дисбаланса, наличие автоколебаний в системе (фантомные сигналы, несуществующие потребности и проч.) можно рассматривать с точки зрения психических расстройств.



Агент

- активность (реализация действий);
- автономность
- общительность;
- целенаправленность (наличие мотивации)

Заключение

- Термины «темперамент робота», «раздвоение личности», «эмоциональное состояние» - не только и не столько просто метафорическое обозначение состояния технической системы.
- Эти обозначения определяют способы организации внешнего управления сложной системой. Например, изменение *характера поведения* робота может быть осуществлено переводом его характера из сангвинистического в холерическое.
- Реализация того или иного метода зомбирования позволяет изменить *целеполагание* робота.
- Несколько сложнее обстоит дело с пользой от шизофренических эффектов. Возможно, такие определения и модели могут помочь в диагностировании состояния сложной системы.