

Теория на Разума и Вселената на Тодор Арнаудов
Универсален Изкуствен Разум
Математическа теория на разума



* версия
2001 – 2004 с
допълнителни
бележки и
уточнения
-- ЧЕРНОВА --

<http://research.twenkid.com> Twenkid Research Todor Arnaudov

<http://research.twenkid.com>

Забележка

- Препоръчително е четенето на оригиналите.
- Презентацията на български е много съкратена спрямо английската, може да я потърсите в папката при тази:
- http://research.twenkid.com/agi_2011/Todor_Arnaudov_Theory_of_Hierarchical_Universal_Si
- http://research.twenkid.com/agi_english/
- http://research.twenkid.com/agi_english/Todor_Arnaudov_Theory_of_Hierarchical_Universal_

Съдържание

- Водачът на човешкото поведение, учене с подсилване.
- Вселената сметач
- Управляващи устройства и условни УУ
- Йерархичност във Вселената. Въображаеми вселени, йерархични въображаеми вселени и универсални симулатори на въображаеми вселени
- Еволюция на вселената към предсказване и управление
- Предсказването и управлението с нарастваща точност и разделителна способност, и вероятност клоняща към 1 като принципи на разума
- Абстрактна теория за изключенията от правилата в изчислителните машини и връзката и с йерархичността във Вселената.

Условия за разум?

- Въображава вселена (среда, свят) с физични закони
- Входни устройства (сетива, възприятия – **права връзка**)
- Изходни устройства (действия, „мускули“, причиняващи сили – **обратна връзка**)
- **Управляващи устройства + „Съдници“**

Съдници – измерват успеваемостта на УУ и го наказват или поощряват

УУ – запомнят входните, изходните данни и успеваемостта и търсят зависимости между тях

УУ – стремят се да максимизират успеваемостта като променят по подходящ начин изходните данни (действията си), с което могат да променят и входните данни

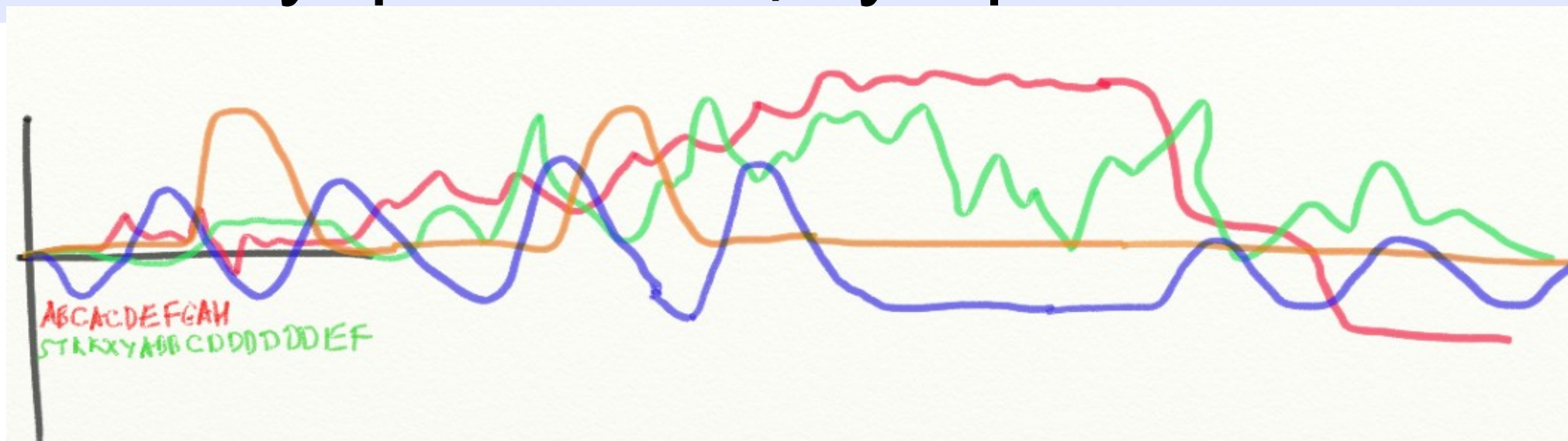
Управляващи устройства

- Устройство, единица, същност, „нещо“, което има **причиняваща сила – може да върши целенасочени промени.**
- Устройство, което прехвърля информация от една памет в друга, от своята памет в паметта на друга въображаема вселена по предсказуем начин (в подчинено управляващо устройство).
Устройство с права и обратна връзка.
- Истинско управление – УУ записва информацията с максималната разделителна способност на управление и възприятие на управляваната вселена.
- Условни УУ – информацията се записва с по-ниска разд. Способност от тази на управляваната вселена. Управлението не е пълно, има фактори които са извън контрола на условното УУ.
- **Всички УУ, освен Вселената, описана най-ниско ниво и разгледана в цялост, са Условни УУ**

Управляващи устройства

- Търсят целевото си състояние
- Имат **воля**, стремят се към целевото състояние и се стремят към все по-истинско управление
- Умът се състои (може да се представи като) от много на брой УУ взаимодействащи си хоризонтално и в йерархия, които се борят да поемат властта над **изпълнителните устройства (най-ниското ниво на обратна връзка)**
- Съвкупностите от виртуални УУ, които вземат надмощие в даден момент, причиняват наблюдаемото поведение.
(Тази дефиниция е вид мултиагентна система)

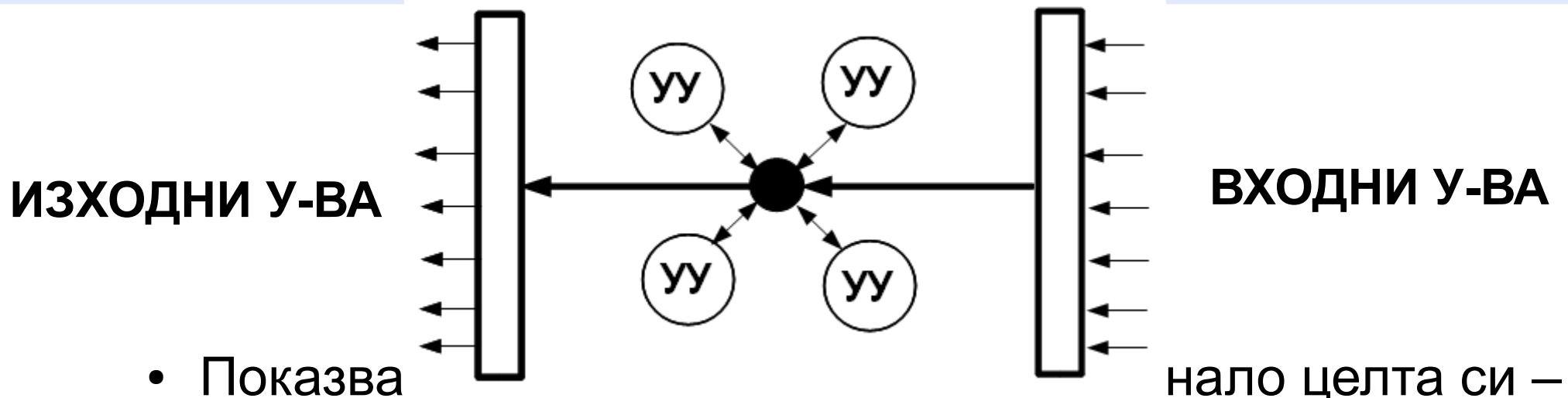
Функция на удоволствието на управляващи устройства



- Ниво на награда/удоволствие/успеваемост във времето на всяко УУ от съвкупност от УУ
- Паралелно – вход и изход на УУ (възприятия и действия)
- Всяко УУ строи модели на възприятията и действията си и изчислява **бъдещи траектории на действията с максимална очаквана сборна награда за избран период напред**

*подобно на модела на М. Хутер и Ш. Лег, но за съвкупност от УУ

„СЪДНИК“



- Показва **нало целта си** – ниво на **успеваемост, „удоволствие“**.
- Разпределя **влиятието** което всяко УУ може да има върху изходния сигнал към **изпълнителните устройства (мускули, двигатели или записвачи на данни във въображаеми вселени)**
- Съдник – вграден в УУ или външен?

Условия за разум?

- Не бива всички УУ да могат да получат максимално удоволствие едновременно - за да има сложно поведение
- Не бива едно УУ да може да задържа управлението над изходните устройства неограничено дълго (зацикляне).
С времето:
 - **потискане** на влиянието на УУ, което е „завзело управлението“ в момента.
 - **подсилване** на влиянието на УУ, които не управляват в момента.
- **Йерархично управление** – по-лесно справяне с неопределени ситуации. (Виж Абстрактна теория за изключенията от правилата)

Висши процеси

- „Сдружаване“ на УУ – взаимодействие и синхронизация, подсилване, между различни УУ, взаимовръзка между удоволствията
- „Раздор“ между УУ – противопоставяне/потискане на УУ, удоволствието на едното намалява удоволствието на друго
- Групиране на УУ така че сборното удоволствие на всички групирани УУ да бъде максимално (общества, държави)
- “...бъдещи траектории на действията с максимална очаквана сборна награда за **избран период напред...**“ - трябва да се уточни как се избира периода (йерархия – по-високо ниво, по-дълъг период)

Период за изчисление на бъдещо натрупаното удоволствие

- “...бъдещи траектории на действията с максимална очаквана сборна награда за **избран период напред...**”
- Универсални симулатори на въображаеми вселени (УСВВ, състоят се от УУ)
- Йерархични УСВВ
- Най-долното ниво работи с реално време - **тактове** на разума
- По-горните нива работят с абстрактно (компресирано) време, и времето се определя от съвпадения между събития/образци

Основни понятия на разума

При човека, повечето обекти (от всякакъв вид) не се запомнят "фотографски", а се "преразказват" в мозъка, записват се най-характерните особености на информационните обекти, входната информация се компресира. Във "фотографска", "фонографска", "текстографска", "стереографска" (пространствена) и пр. памет вероятно се съхраняват **само основните понятия**. (...) Хитро е новопостъпилият информационен обект да се обясни с наличните информационни обекти. Просто му се дава етикет, а същността му се описва с известните понятия, като се използват техните етикети - връзки към значението им, съдържащи само "адрес". Ние хората наричаме такова запомняне "разбиране" и "осмисляне". -

- Човекът и Мислещата машина, 2001

Основни понятия на разума?

- Зависят от модалността – **вид входно или изходно устройство** (сетиво или актуатор/мускул/писач на данни във въображаема вселена)
- **Слух:**
 - интензитет на звук с дадена честота (спомнете си как работи човешкото ухо и кората на мозъка)
 - „контраст“ - относителен интензитет между различно отдалечени честоти в пространството
- **Зрение:**
 - интензитет на пиксел (цвят, яркост)
 - контраст/разлика между съседни пиксели във всяко направление
 - ...

Удоволствие?

- *Показатели за изпълнението на целта на поведението.*

Елементарна цел/

Елементарно удоволствие:

Съвпадение между две стойности:

А) очаквано (желано/целево)

Б) действително усещане

Елементарно усещане – число, променлива:

IF (Input == Target_Input) Feels = NIRVANA;

IF (Input != Target_Input) Feels = HELL;

Степенувано просто удоволствие

- Разстояние, Разлика, Сравнение

- При животните и хората - удоволствие/болка – показват на хипотетични УУ от модел на поведението ни, колко близо се намираме до осъществяване на първично заложеното целево състояние: храна, вода, секс (размножаване) и нецелево: нараняване (болка) и пр.

Feels = *Difference(Input, Target_Input)*

Очаквано ~ предвиждано ~ желано (цел)

Разликата между очакваното и

действителността е “грешка” - неудоволствие.

Умът се стреми да намали грешката.

Йерархия на нуждите на Маслоу



Maslow's Hierarchy of Needs

Abraham Maslow, Motivation and Personality, 2nd ed., Harper & Row, 1970

Познавателно и физическо удоволствие

- **Познавателно/когнитивно удоволствие** - компресиране, предсказване, съвпадение, оптимизиране (разбиране, усъвършенстване ...)

Успешното предсказване е удоволствие.

Feels = *Difference*(Input, Predicted_Input)

- **Физическо удоволствие** – първични нужди за оцеляване. Първични „награди“ и цели на поведение (съхранение (болка), храна, секс ...)

Желаното усещане е удоволствие.

Feels = *Difference*(Input, Desired_Input)

Познавателният инстинкт за самосъхранение е научен

- **Самоубийци** – смятат (предсказват), че това което ще чувстват като се самоубият ще бъде по-малко неприятно от това, което чувстват в момента на вземане на решението да се самоубият.
- **Смърт** – виж по-нататък.
- **Физическото и познавателно удоволствие трябва да се съгласуват.**
- Познавателната йерархия (Б. Казаченко) или Йерархичните подражатели на въображаеми вселени (Т. А.) получават обратна връзка за физическото удоволствие в средата в която съществува умът и търсят зависимости между познавателното и физическото удоволствие.

Въображаема вселена

- Среда, система, свят.
- Има **състояние** (памет) и (физични) **закони** – правила, според които се променя състоянието.
- Реалната Вселена, с която разумът взаимодейства може да се представи като „въображаема вселена“ от най-ниско възможно ниво (предсказващ модел на реалната).
- **Реалната вселена няма изрична памет за минало и информацията в нея не е компресирана.**
- В реалната вселена могат да се разпознаят „**подвселени**“ - подсистеми, устойчиви съвкупности от взаимосвързани явления, които могат да се представят със свои състояния и закони. (Модели, patterns)

Разумът симулира възприятията си

- Умът подражава/емулира/симулира възприятията си, за да предсказва.
- ...Разумът работи с "алчни" алгоритми: търси най-преките и най-"правдоподобни" сили - според разбирането му за "правдоподобност" в конкретния случай, които може да нарече **причини**; (припомнете си *Бръснач на Окам*)
- Разумът "интегрира" - от крайните събитията се опитва да извлече функцията, която ги е породила.
- Но това, както знаем, не е еднозначно действие! Ако началната функция липсва, а само нейни производни, не можем да бъдем сигурни че интегрираната функция е съвсем същата като началната: че причините, със сигурност, са били тези, които ние сме предположили.

Симулатор на въображаема вселена

- **Помни минало и настояще**, и на тяхна основа **предсказва бъдещето**. Работи с **компресирани** представяния на законите и състоянията.
- Изгражда се от управляващи устройства.
- **Подражанието** се настройва във взаимодействието на УУ с въображаема вселена – установяват се траектории на поведение на УУ.
- **ВАЖНО:** „Удоволствията“ на УУ в познавателната йерархия са абстрактни и показват успеваемостта им в предсказването на бъдещето. Удоволствие == правилно предсказване на бъдещите възприятия спрямо миналите.
- Траектории, които предсказват правилно в дадени условия, се утвърждават - „самозадоволяват“ ...

Йерархия в природата

кварк

протон/неутрон

електрон

атом

молекула

макромолекула

ДНК

клетъчна органела

клетка

тъкан

орган

система в организъм

организъм (.....)

Висите процеси са вид странични ефекти на най-нисшите?

„Машинен език на вселената“

- **Най-ниско ниво – изход на УУ, инструкции и данни,** които УУ извежда (управление == запис на данни в памет). За някои УУ на най-ниско това са команди към двигатели или мускули. („Моторни команди“ по Д. Хокинс, обратна връзка и промяна на координати по Б. Казаченко)
- **Траекториите на УУ са поредици от инструкции,** „подпрограми“. **Поредици от инструкции, които предсказват бъдещето.**
- **Утвърдените траектории** - тези които познават успешно се взимат като физични закони от по-горно ниво. По-горното ниво прихваща кои УУ са активни. „Функции“, които се викат от по-горното ниво по адрес. Спомнете си НТМ - вектор с причини/вери, които се предават между нивата на йерархията.

Въображаема вселена от по-горно ниво

- Законите от по-долно ниво, изразени като траектории на УУ, които предсказват, и на взаимодействия между УУ, се използват като първични закони на управляващи устройства и въображаеми вселени от по-горно ниво.
- Състоянието на Вселената от по-горно ниво може да се изрази от състоянията на УУ от по-долно ниво, които го описват.
- УУ от по-горно ниво се описват с повече елементарни „инструкции“, по-бавни са и чакат резултата от по-долните нива.

Учене с подсилване

- **Когато УУ намери целево състояние, то се стреми да остане в него (да компенсира отместванията) - пристрастяване.**
- **Опасност от зацикляне и мъртва хватка, ако само едно централизирано УУ владее всички ресурси.**
 - Йерархично учене с подсилване и „надзорници“ с прекъсвания, които извеждат системата от мъртва хватка

Абстрактно управление и предсказване

- Записаното във въображаема вселена на разума да се потвърди от външната среда (обратна и права връзка да съвпаднат) и да се обогати с допълнителни подробности.

„Това което си представям и искам да се сбъдне,
да се сбъдне.“

- ОЧАКВАНОТО == РЕАЛНОТО
- Все по-точно предсказване и причиняване на бъдещето

Съзнание и „събитиеен вседържец“

* Вседържец - “операционна система”

- Възникването на съзнанието е създаването на **първия развит (устойчив) събитиеен вседържец** . Събитийният вседържец запомня и обработва поточна информация от сетивата и осигурява запомнянето и възпроизвеждането на събития като "на лента" - можем да изберем "запис" и да го "прослушаме" във въображението си.

- Първият развит събитиеен вседържец е операционна система, способна при усърършенстването и надграждането си да запазва **съвместимостта** с предходните версии. Например имаме спомени и от тригодишна възраст, и от вчера. Мозъкът ни се е променял през годините, но е останал съвместим към най-старите спомени и за всички си спомняме мигновено.

Изпълнителна памет/вседържец и Събитийна памет/вседържец

→ Изпълнителната памет е, в по-голяма степен от събитийната, **хранилище на алгоритми**

→ Изпълнителен вседържец – детето може да говори, да ходи, тича, разпознава предмети и т.н. много да мине теста с огледалото, следователно това е част от по-ниското равнище на Вседържеца на разума - **изпълняващият вседържец (ИВ)**. Той стига до получаване на съвместими отгоре надолу версии много по-рано от Събитийния вседържец.

→ Събитийният вседържец е по-универсален от Изпълнителния (...) по-голям брой формати на данните, съчетание и разширение на форматите на ИВ.

- ...

В МОЗЪКА: ИВ – неокортекс. СВ – хипокампус + неокортекс вж.
<http://eim.hit.bg/3/25/pred-3.htm> т. 61

Изпълнителна памет/вседържец и Събитийна памет/вседържец

Изпълнителната памет - хранилище на алгоритми

Събитийна памет - хранилище на данни, от които след изчисления се получават алгоритми за ИВ

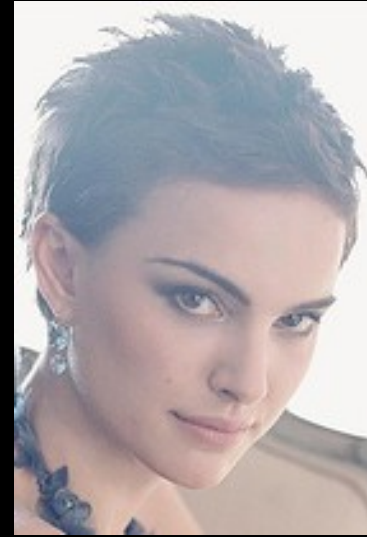
Разумът извлича изпълнителна информация от събитийната. Т.е. разумът може да тълкува (обмисля) събитийната памет с голямо закъснение във времето. Разумът може да преобразува събитийната памет в изпълнителна, във формат за Изпълнителния вседържец, който може да извърши определени действия заради събитие, съхранявано в Събитийната памет.

Танци и музика

- Танците са ритмични движения на тялото - ИЗХОДЕН поток към мускулите, които се намират в някакво съзвучие с ВХОДЕН поток – музика.(...)
- **Ритмичен – предсказуем.**
- Колкото по-голямо СЪЗВУЧИЕ има между ВХОДНИТЕ и ИЗХОДНИТЕ импулси, толкова общото равнище на импулси се УСИЛВА и ни кара да изпитваме наслада.
- Трябва да можем да го обработим (да не е прекалено бързо)

Красота





Красота

- Защо красотата ни носи удоволствие?
- Красотата е удоволствието от определени възприятия?
(в познавателната йерархия)
- Красотата е субективна, но за много неща хората са еднородни в оценката си. **Има нещо изчислимо и неизменно.**
- Какво е общото между красотата на автомобил, самолет, женско лице, рисунка, китара, кон, котка...?

** Красотата може да бъде не само познавателно възприятие. В езика има припокриване между красиво и физически приятно. Нещо, което напомня за физическо удоволствие може да се приеме за красиво, без да има свойствата на познавателна красота, или пък може да я подсили. (еротични снимки)*



Красота

- Съвпадения между желани/търсени **съотношения** (пропорции), които разумът открива между образци (в паметта, известни) и новите, които се оценяват.
- Тези съотношения трябва да се извлекат от данните
- Красивите лица са симетрични*
- **Защо симетрията е красива?** – наслагване, взаимно подсилване; възможно е напасване на две части чрез **прости преобразования** (Спомнете си лекцията за Психология на развитието, как може би е получена думата „мама“)
- Повторението е най-прост начин за потвърждаване на значимост на възприятие.

Повторение и съотношения

- **Симетрия** – повторение, компресия $\times 2$
- **Средно лице** - при усредняването на особеностите на лицата на много хора обикновено се получава “красиво” лице

<http://www.faceresearch.org/demos/average>

Изследвайки "красивите съотношения", т.е. информацията, която ни доставя удоволствие, бихме могли да разберем какви съотношения има в устройството, което ги намира.

Съотношенията в образите са между точки, отсечки, петна, ъгли, дъги, площи, цветове, успоредност - равно разстояние между отсечки; равни ъгли, съчетани или не с равни съотношения между дължини на отсечки; равни по размер площи, еднакви форми, т.е. последователности от еднакви насочени отсечки (посочници, вектори) в двуизмерното пространство, или триизмерни, проектирани върху равнина и пр.

<http://research.twenkid.com>

Интересността ...

- Красиви са изображения, които умът успява да компресира ~ предскаже добре чрез своите методи.
- Успешното компресиране и успешното предсказване са “приятни” за разума, но:
- Баланс на предсказуемост/непредсказуемост – прекалена предсказуемост – няма нужда от повече търсене.
- Драматургия
- Лекция на Юрген Шмидхубер <http://vimeo.com/7441291>

Грозота (познавателна)

- Когнитивно претоварване!
- Трудно за компресиране:
непредсказеумост ~ безсмислие,
нелогичност



Красиво и грозно (емоционална)

- Стимули, които се асоциират с приятни/желани и неприятни/нежелани неща, без значение какви точно са стимулите.
- Голота.

Случайност

- Липса/недостиг на информация за събитието
- Наличие на по-малко количество информация от това, което разумът иска или очаква да намери.
- Способността за предсказване е по-слаба, отколкото се приема за необходима за доказване на „закономерност“.
- Свободна воля – неспособност да предскажеш/моделираш поведението с точността с която очакваш, че би трябвало да можеш
- Парадоксът със свободната воля и случайността

Произход на повторенията на събития в различни равнища на Вселенския сметач и на нецеливите, за преживяващия ги разум, съвпадения с ниска вероятност

- **Взаимосвързаност.**
- **Въображаеми вселени и условни подвселени**
- Реални случки
- Днешната сложност е произлязла от миналата – зависимости между частите
- Възпроизводството е страничен ефект на компресията на информацията.
- Висши форми на описание на вселената
- Висши форми на физични закони и градивни частици

„Вселената сметач“ (компютър)

- **Въображаеми вселени и условни подвселени**
- Реални случаи
- Днешната сложност е произлязла от миналата – зависимости между частите
- Възпроизводството (размножаването) е страничен ефект на компресията на информацията.

Еволюция

- **Частите на Вселената се стремят да се групират в подвселени, които имитират принципите на работа на цялата Вселена – затвореност (запазване, отделяне) и способност за предсказване на бъдещето си въз основа на части от самите тях**
- **Стремят се това което искат (вътрешните представления) да се потвърдят от входа**
- **По-сложните подвселени са събрали по-дълга част от „кода на Вселената“ и предвиждат бъдещето все по-точно, с повече ситуации и по-голяма област.**
- **По-сложните подвселени са по-близо до „ядрото на Вселенския сметач**
- **По-сложните системи предсказват все по-добре бъдещето**

„Странни съвпадения“

- Случването на едно събитие с ниска вероятност изисква случване на голям брой събития с висока вероятност (безброй събития с вероятност 1)
- Обичайните явления работят за да се случат “странните” съвпадения
- Произлиза от компресираността и взаимосвързаността.

Предсказуемост в термините на подвселената

- **Държавата** се стреми към предсказуемост спрямо законите, разпоредбите и т.н. - иска да става това, което е предписано там с максимална точност.
- Държавата се стреми да управлява частиците в себе си, но управлението е условно (по-висшето ниво не може да има точен модел на по-нисшето) и от там има разминавания
- Различни форми на подвселени се борят за надмощие.
- **Различни нива на условно управление – по-висшето се стреми да управлява по-нисшето, но това е невъзможно. Противоречия, всяко се стреми да управлява. Различни гледни точки могат да дефинират различни условни управляващи устройства/въображаеми вселени.**

Връзки

- <http://eim.hit.bg/razum>
- <http://research.twenkid.com>
- <http://artificial-mind.blogspot.com>