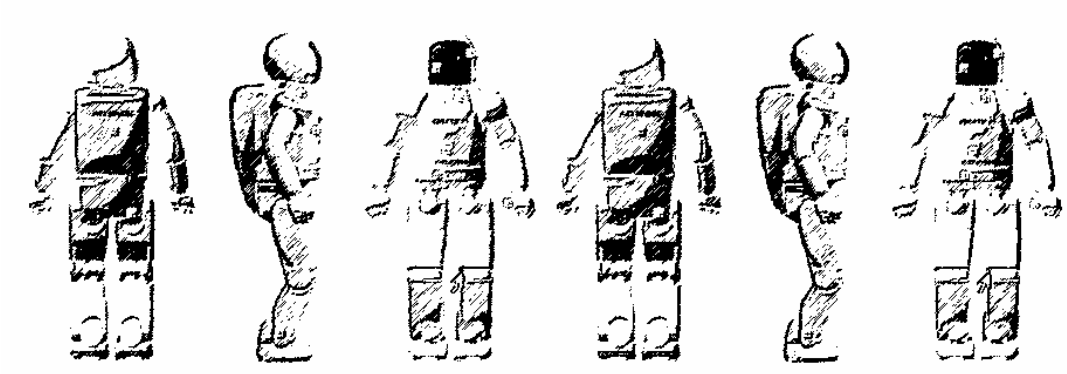
Luonnonfilosofian seura 4.3.2008

Tietoiset Robotit



Pentti O A Haikonen, TkT

pentti.haikonen@nokia.com

Tel. +358 (0) 40 844 2848

Ja ihminen loi robotin kuvakseen

Voidaanko vielä puhaltaa henki?

Repliee Q2
Prof.
Ishiguro,
Osaka
University

DER2 fembot,
Sanrio, Japani

Kauniita, mutta eivät tietoisia!

Miksi robottien pitäisi olla tietoisia?

- Haluammeko lähellemme itsestään liikkuvia mahdollisesti vaarallisia peltipömpeleitä?
- Pitäisikö robottien tietää mitä ne ovat tekemässä ja pystyä keskustelemaan tästä isäntiensä kanssa?
- Jos näin, niin silloin robotin täytyisi olla **tietoinen**.

Service Robot(Fujitsu)

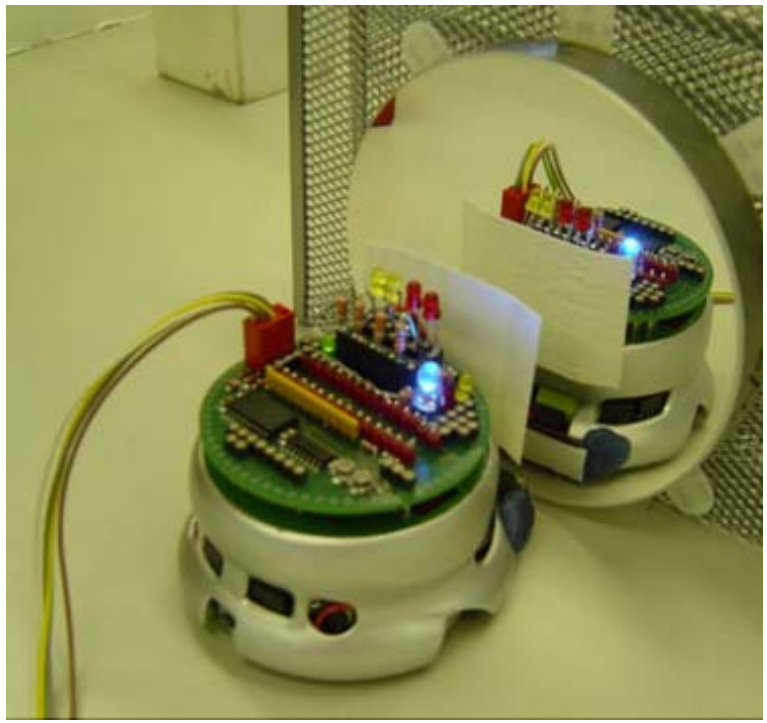

Mental Commit Robot “PARO”

Lemmikit ilman **omaa mieltä** ovat tylsiä.

Kohti itsensä tiedostavia robotteja?

- Tämä robotti pystyy havaitsemaan oman peilikuvansa ja samannäköisen toisen robotin välisen eron. Robotti pystyy myös matkimaan toisen robotin liikettä. LED-valoja on käytetty ilmaisemaan robotin erilaisia sisäisiä tiloja.

- Robotin on kehittänyt prof. Junichi Takenon tutkijaryhmä Meiji Universityssä Japanissa.



Tämäkään vekotin ei kyllä oikeasti ole tietoinen.

Millainen olisi tietoinen robotti?

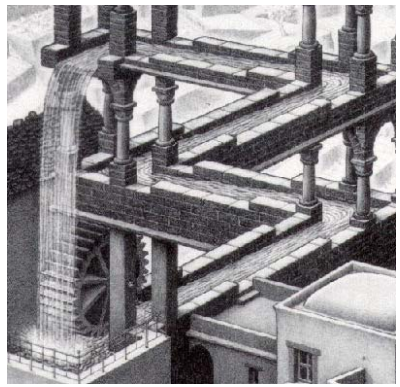
- Robotti **tietää**, mitä on tekemässä, mitä on jo tehnyt ja mitä sen pitää tehdä seuraavaksi
- Robotti **havaitsee** itsensä ja ympäristönsä **suoralla tavalla** samoin kuin ihminen
- Havainnoilla ja asioilla on robotille **merkitys**
- Robotti **ajattelee**; sillä on **mielikuvitus** ja **sisäinen puhe** kuten ihmisellä
- Robotilla on **tuntemuksia**
- Robotilla on **oma tahto**

Ongelmalliset avainkäsitteet

- Mielen ja ruumiin ongelma; dualismi
- Tieto ja tietäminen
- Merkitys, intentionaalisuus
- Ymmärtäminen, käsittäminen
- Havaitseminen, havainto, representaatio
- Tuntemukset, laadullisuus, qualia
- Ajattelu
- Tietoisuus

Dualismi konetietoisuuden esteenä?

► Ikiliikkujan mahdottomuus voidaan johtaa termodynamiikan pääsäännöistä; mikä tahansa ikiliikkuja, sen rakenteesta riippumatta, on mahdoton.

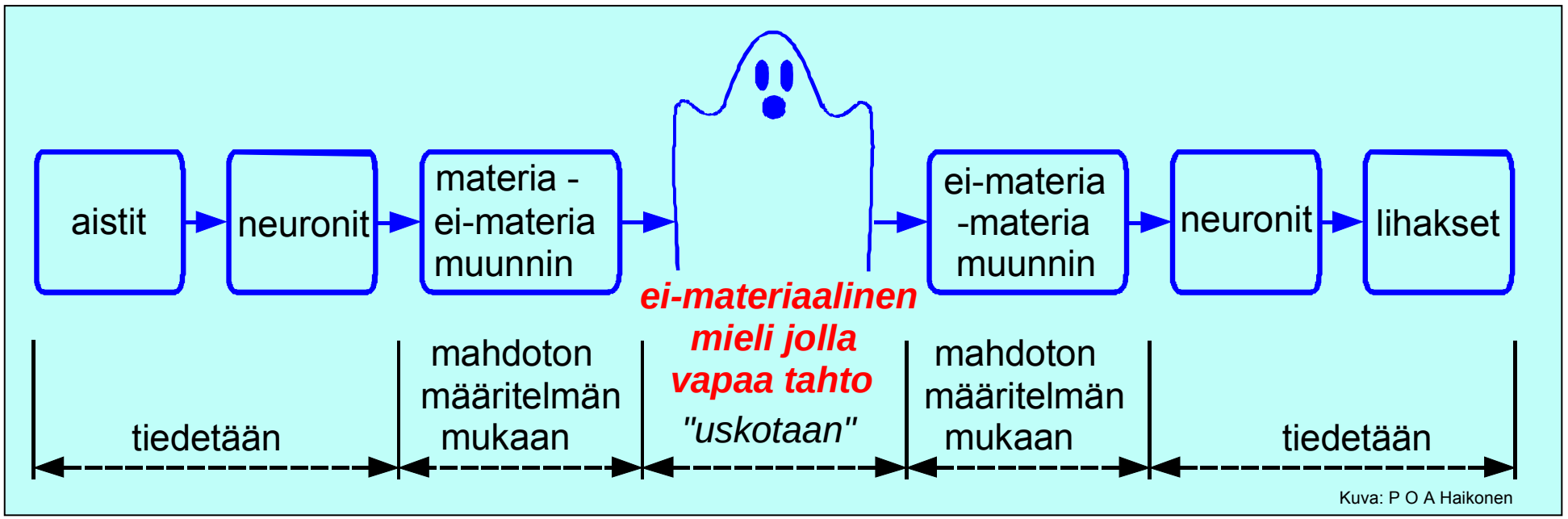


► Voidaanko myös konetietoisuus osoittaa mahdottomaksi jollain yleisemmällä periaatteella?

► Olisiko **dualismi**, mielen aineettomuus tällainen periaate; materialla ei voida aikaansaada aineetonta mieltä vaikka miten laite konstruoitaisiin?

Dualismin ongelma

- Ajatuksemme ja mielemme näyttäisi olevan **ei-materiaalinen**; **ainakaan emme havaitse mitään materiaalisia prosesseja** tapahtuvan kun ajattelemme. Toisaalta, ruumis on **materiaalinen**.



- ***Energiaa ei ole havaittu syntyvän tyhjästä, kuitenkin ei-materiaalisen mielen pitäisi aikaansaada energiaa, jolla vaikuttaa aineeseen!***

Ei-materiaalinen mieli

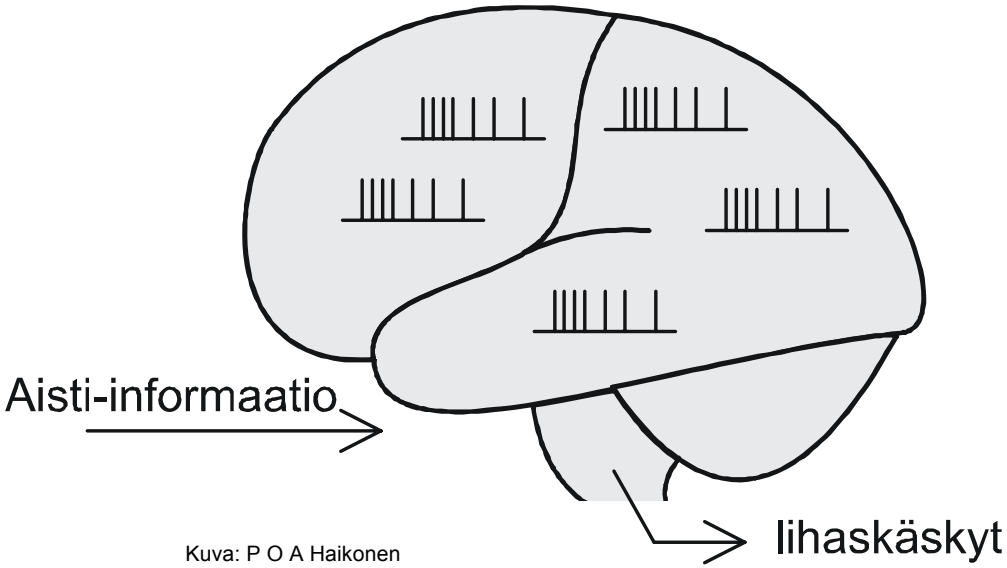
► Luonnontieteet eivät hyväksy kummitussatuja eivätkä ei-materiaalista mieltä. Kuitenkin: **Mieli vaikuttaa ei-materiaaliselta**.

► Kelvollisen tietoisuusteorian täytyy selittää, miksi mielemme vaikuttaa ei-materiaaliselta ja miten nämä **näennäisesti** ei-materiaaliset prosessit liittyvät mielen (aivojen) materiaalisiin prosesseihin.

► Tästä myös seuraa, että vakavasti otettavan **tietoisen koneen tulee myös havaita mentaaliset prosessinsa *näennäisesti ei-materiaalisina***.

Maailma havaitaan ”sellaisenaan”

- Aivot ovat neuroverkko. Neuronit kommunikoivat keskenään neurosignaalien välityksellä. **Materiaalisen näkemyksen mukaan nämä signaalit ovat ajattelun materiaalinen perusta.**



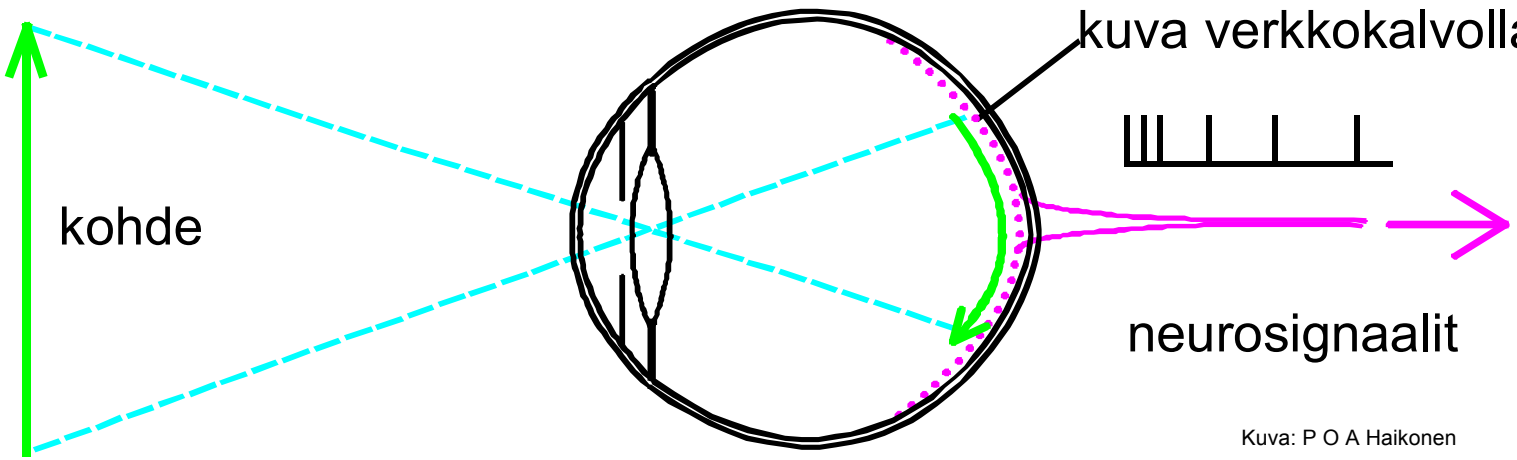
Kuva: P O A Haikonen

- **MUTTA,** miksi kuitenkin havaitsemme kohteet; esineet, äänet, jne. **näennäisesti** **sellaisenaan** emmekä niiden sijaan jotain neurosignaaleja, jotka jotenkin liittyisivät ulkomaailman kohteisiin?

Jos kerran emme havaitse neurosignaaleja, niin miten ne muka voisivat mitenkään liittyä ajatteluun ja mieleen? ***Jospa mieli ja aivot ovatkin eri asia?*** *”Aivojen tehtävä: aineenvaihdunta?”*

Neurosignaalien näkymättömyys on fakta

► **Silmä:** Emme havaitse ympäristöämme verkkokalvolla olevana kuvana emmekä neurosignaaleina jotka lähtevät silmän verkkokalvolta. Ovatko nämä siis näkemisen kannalta turhia?

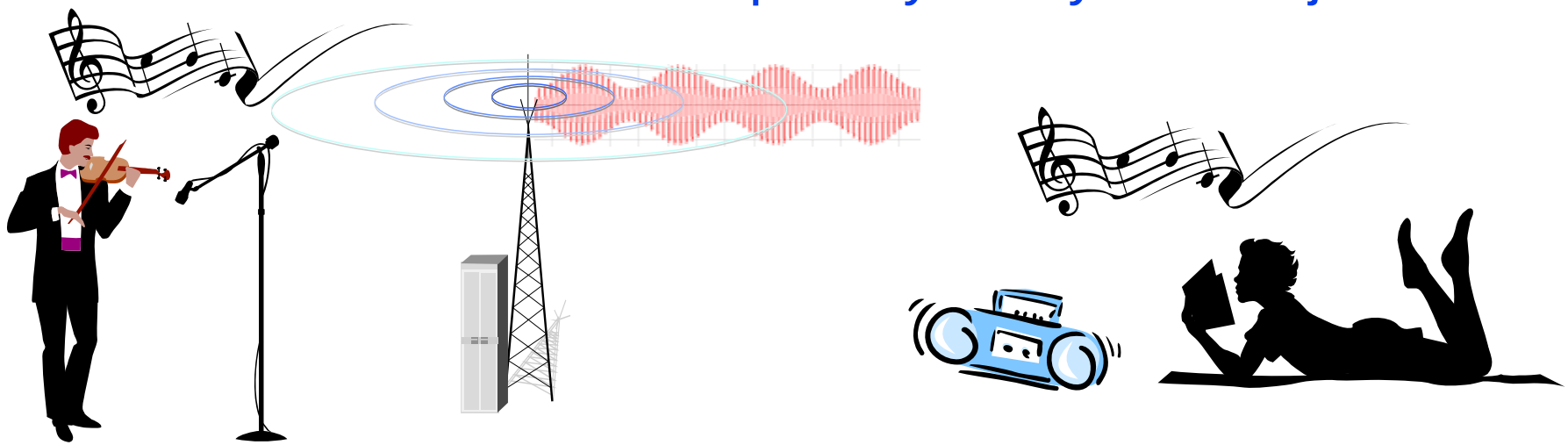


Kuva: P O A Haikonen

► **SIIS: On mahdollista havaita pelkkä informaatio eikä niitä mekanismeja jotka sitä välittävät. Välittävä mekanismi jää läpinäkyväksi!**

Modulaatio ja systeemin läpinäkyvyys

Tekniikka tuntee läpinäkyviä systeemejä

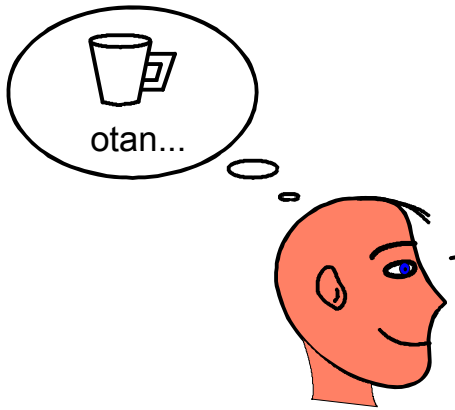


- FM vai AM? Putkia vai transistoreja? Mitä väliä, musiikkiahan tässä kuunnellaan.

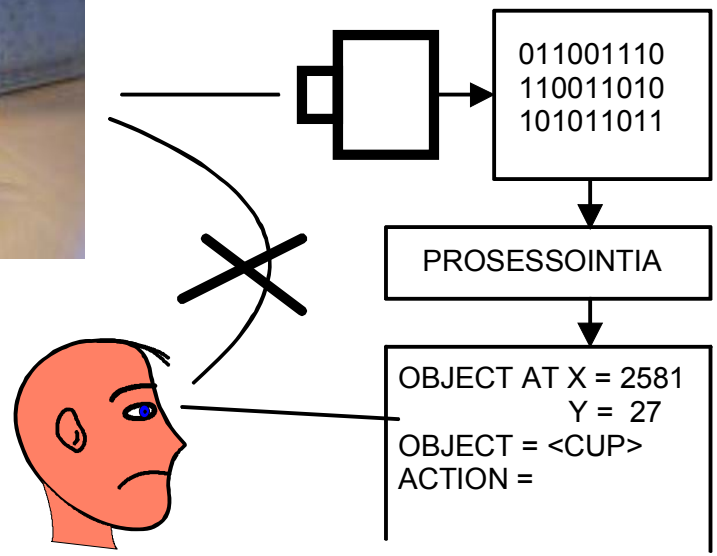
- **Modulaatio;** kannettu informaatio ratkaisee, ei radioaalto eikä laitteiden rakenne. Nämä jäävät läpinäkyviksi ja niitä voidaan havaita vain muilla keinoin. Mutta ilman näitä ei olisi musiikkia.

informaation prosessointi;

Suora ja läpinäkyvä vs. epäsuora



Kuva: P O A Haikonen



Olemme näkevinämme kohteet **sellaisena kuin ne ovat;** kohteen havainnon perusteella voimme toimia suoraan. Helppo kohdistaa huomio yksityiskohtiin.

Toimimme ”mielikuvilla”

”fenomenalisuus”

Epäsuorat symbolit eivät aikaansaa **”illuusiota” suorasta havainnosta** –ne eivät myöskään mahdollista **saumatonta assosiointia** muihin havaintoihin ja motoriikkaan. **Fenomenalisuus ei toteudu**.

Neurosignaalien merkitys

Kuva: P O A Haikonen

Puikkojen materiaali ei vaikuta hahmoon.

- Aistittu informaatio moduloi aisteilta tulevia neurosignaaleja.
- Tämä modulaatio edustaa aistittua informaatiota.

Modulaatiohahmot edustavat tätä informaatiota suoraan koska neurosignaaleja sellaisenaan ei aistita. Ajattelu perustuu neurosignaalien vuorovaikutukseen.

Mieli materiaalisina tiloina

► Emme havaitse silmältä tulevia neurosignaaleja sellaisenaan vaan niiden sijaan havaitsemme signaalien välittämän informaation.

► Sama pätee myös syvemmälle aivoihin; **neuronit sellaisenaan jäävät ja saavatkin jäädä havaintojen ulkopuolelle, koska ne eivät liity ajatuskulun sisältöön.** Neuronit ovat vain läpinäkyviä informaation kantajia. **Mieli on sisältö**. (Vrt. kirjainten painomuste ei vaikuta merkitykseen.)

► ***Täten neuronien materiaalinen rakenne ei ole ratkaisevaa; toiminnallisesti samanarvoiset keinotekoiset neuronit käyvät.***

Merkitys ja intentionaalisuus

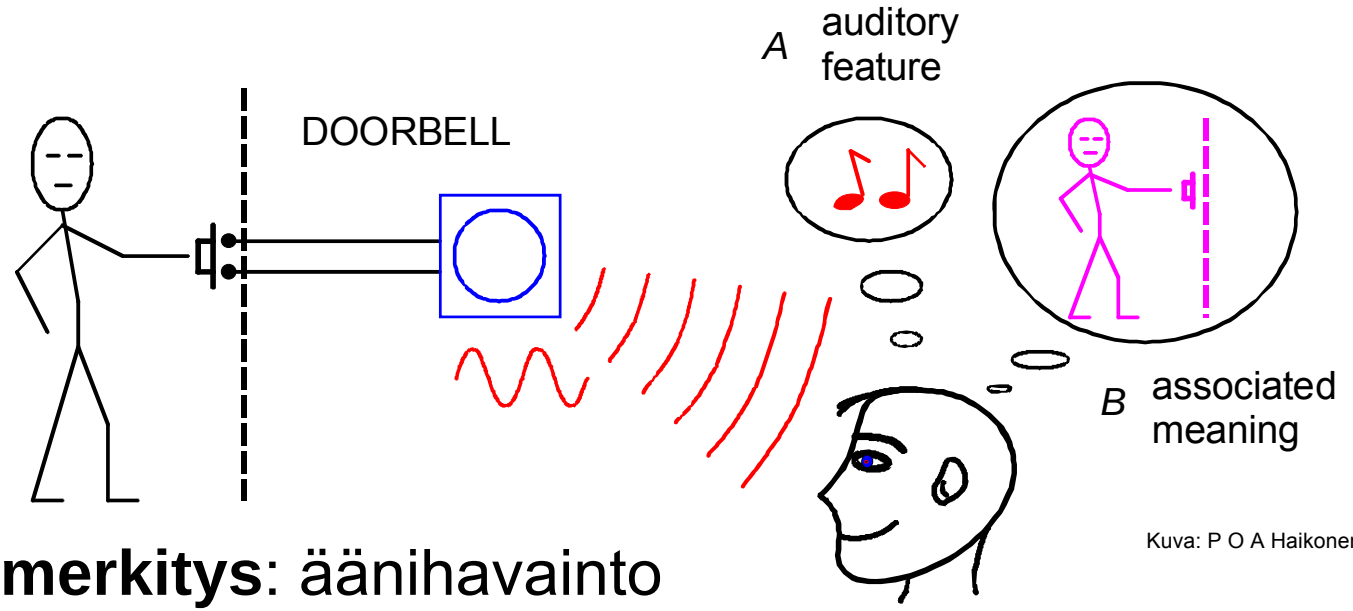
”Kone ei voi käsitellä merkityksiä; vain ihminen voi käsitellä merkityksiä.”

Brentanon kiikkutuolifilosofiaa; tämä ei perustu mihinkään.

The property of being intentional, of having an intentional object, was the key feature to distinguish psychical phenomena and physical phenomena, because, **as Brentano defined it**, physical phenomena lacked the ability to generate original intentionality, and could only facilitate an intentional relationship in a second-hand manner, which he labeled derived intentionality. (Wikipedia)

Perusmerkitys ja assosioitu merkitys

- **Kognitio ei ole luokittelua** –vaikka ovikellon ääntä kuinka tutkittaisiin ja luokiteltaisiin, **niin sen merkitys ei paljastuisi**; pitää **tietää**, mitä se merkitsee. (Sama pätee aivotutkimukseen)



Kuva: P O A Haikonen

- **Perusmerkitys**: äänihavainto
- **Assosioitu merkitys opitaan**: Joku on ovella, avataan ovi
- **Tässä ei ole mitään, mitä kone ei voisi tehdä!**

Neurosignaalien merkityksestä

- Aistihavainnot aikaansaavat neurosignaaleja, jotka täten tulevat edustamaan kyseisiä havaintokohteita.

- **Kuitenkin, käsitteellisen ajattelun voima tulee symbolien käytöstä; siitä, että tietyt aistimushahmot voidaan saattaa edustamaan asioita, joita ne eivät luonnostaan edusta.**

- **Puhutun sanan voima** ei tule siitä, että tietyt neurosignaalit vastaavat tiettyä äänikuviota, **vaan siitä**, että tämä kuvio voidaan asettaa merkitsemään lukuisia muita asioita ja olemaan niiden symboli.

Merkitysten merkillinen maailma

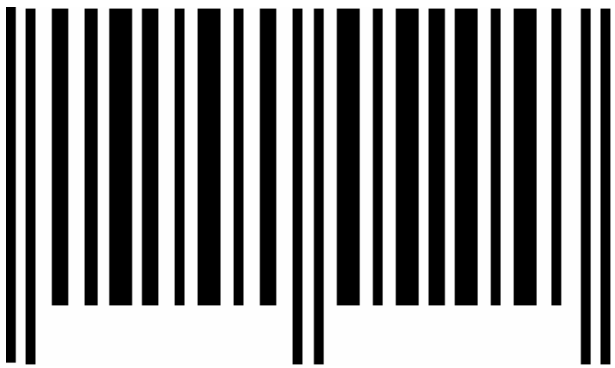
- Havaitaan siis, että neuronit ja neurosignaalit ovat välineitä, jotka kuljettavat ja yhdistelevät merkityksiä.

- Merkitykset selviävät **kytketyistä yhteyksistä**, eivät niinkään neurosignaalien muodoista tai neuronien rakenteesta.

- Merkitysmaailmaa voidaankin tarkastella ja käsitellä ilman neurosignaalien ja neuronien fysikaalisen luonteen tarkastelua.

- **Vertaa:** tietokoneen ohjelmoijan ei tarvitse tietää mitään tietokoneen rakenteesta.

Merkitykset ja kausaaliset vaikutukset



- Onko merkityksiä ilman kausaalisia vaikutuksia? (ei vaikutusta, ei merkitystä)
- Onko mustetahroilla kausaalisia vaikutuksia sellaisenaan vai tarvitaanko tulkinta?
- Ihminen tulkitsee - Onko merkityksillä kausaalisia vaikutuksia **vain tulkitsevan ihmisen mielessä?**

Esimerkki: Viivakoodi

- Ei merkityksiä ihmiselle, mutta aikaansaa kausaalisia vaikutuksia koneissa; *onko merkitys tässä aineeton?*

Mitä on tieto

► Insinööritieteissä tiedon totuusarvoa ei juurikaan pohdita; ratkaisevaa on **tiedon tarkkuus.** **Tieto tulee hankkia luotettavalla tavalla; muu ”tieto” on luuloa.**

► Tyypillistä tietoa: Mitä on missä, milloin, kuinka paljon, mikä on muutoksen suuruus, miten jotain tehdään…

► Käytännön esimerkkejä: Missä ruokaa pidetään, missä tämä jääkaappi on, miten sen luokse pääsee…

► Tämänkaltainen tietokäsitys on robotillekin riittävä, paitsi jos ryhdytään rakentamaan **filosofi-robotteja**.

Tieto mahdollistaa toiminnan

- Robotti pystyy väistämään esteen, jos sillä on tieto esteen paikasta.
- Robotti pystyy avaamaan oven, jos sillä on tieto tarvittavista toimenpiteistä. Jne.
- Toiminnan ei tarvitse olla todellista; se voi olla myös mentaalista, kuviteltua. Kuviteltu tietoon perustuva toiminta voidaan toteuttaa myöhemmin.
- Tietoon liittyy merkitysten käyttö.

Ymmärtämisen ja merkityksen ongelma

- Onko yhden symbolin yhdistäminen toiseen ymmärtämistä?
- Onko ymmärtämistä, jos havaittu esine osataan nimetä?
- Onko ymmärtämistä, jos kone toteuttaa sille annetun käskyn?
- Onko ymmärtämistä, jos havaitun tilanteen antamat mahdollisuudet ja seurannaisvaikutukset oivalletaan?
- Onko ymmärtämistä, jos oivalletaan, mitä ja miksi jotain tapahtuu?
- **Todelliseen ymmärtämiseen liittyy laaja asiayhteys- ja merkitysverkko; assosiatiivisessa systeemissä tämä on helppo toteuttaa.**

Tietoisuuden kova kysymys: Qualia

- Aistihavainnoilla on laatunsa (qualia); punaisen punaisuus, sinisen sinisyys, äänen ääni, joka ei ole näköhavainto, kivun kipu jne. Qualia erottaa aistimukset toisistaan.

- Eräät filosofit pitävät **ei-materiaalista qualiaa** tietoisuuden perusolemuksena ja esteenä konetietoisuudelle. Miten kone voisi kokea **aineettoman** elämyksen punaisesta, maistaa makean tai kokea kipua? Toiset filosofit kiistävät ei-materiaalisen qualian olemassaolon (esim. Dennett).

- **Tämä kysymys on helppo muotoilla niin, että se johtaa päättymättömiin filosofisiin väittelyihin, jotka eivät johda mihinkään.**

Qualia käytännön kannalta

- **Perusvaatimus: Havaintojen tulee erottua toisistaan**
- **Fakta:** Aivoissa samanlaiset neurosignaalit välittävät erilaisia aistimuksia. Yksikään signaali ei ole punainen tai märkä. Miten ne voivat ylipäätään välittää aistimuksen laatua? Miten ihmeessä aistimukset voivat olla erilaisia, vaikka signaalit ovat samanlaisia?
- **Fakta:** Televisiossa samanlaiset sähköiset signaalit välittävät eri värejä. Yksikään signaali ei ole punainen, sininen tai vihreä. Miten ne voivat välittää väritietoa?
- **Tekniikan vastaus: Kausaalinen reititys (langoitus)**

Kipu ja mielihyvä

- Kausaalinen reititys ei selitä kipua eikä mielihyvää.
- Neurotieteistä ei toistaiseksi apua; kipuun liittyviä biologisia ilmiöitä tunnetaan, mutta kivun tunnetta ei neurotieteissä osata selittää. (Jos osattaisiin, niin olisi selitetty tietoisuus.)
- Olen kirjoissani ehdottanut, että kipu ja mielihyvä ovat systeemireaktioita, jotka erityisesti vaikuttavat attentioon. Esim. kipu vaikeuttaa keskittymistä muihin asioihin. Kipu ei ole sisäinen representaatio vaan **olotila**.
- Tämä hypoteesi johtaa teknisesti toteutettaviin ratkaisuihin.

Ajattelu

- Ajattelu ilmenee
  - sisäisenä puheena
  - mielikuvina
  - tuntemuksina, jne
- Ajattelu on merkityksillä operointia (mm. Rauhala)
  mutta muutakin (esim. äänetön hyräily tai säveltäminen)
- Ajattelu voi olla
  - toteavaa (joutokäyntiä)
  - ongelmanratkaisuun pyrkivää
- Jo Aristoteles huomasi ajattelun assosiatiivisen luonteen

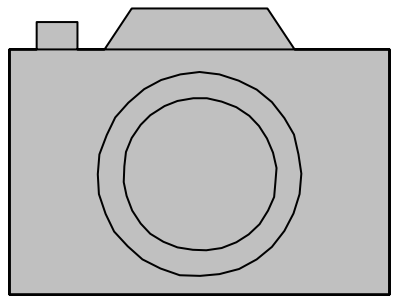
Havaitseminen tietoisuuden edellytyksenä

- Tietoisuuden sisältö koostuu havainnoista
- Havaitseminen on tietoisuuden edellytys
- Kun havainnot loppuvat – myös omasta mielen sisällöstä – **tietoisuus loppuu**

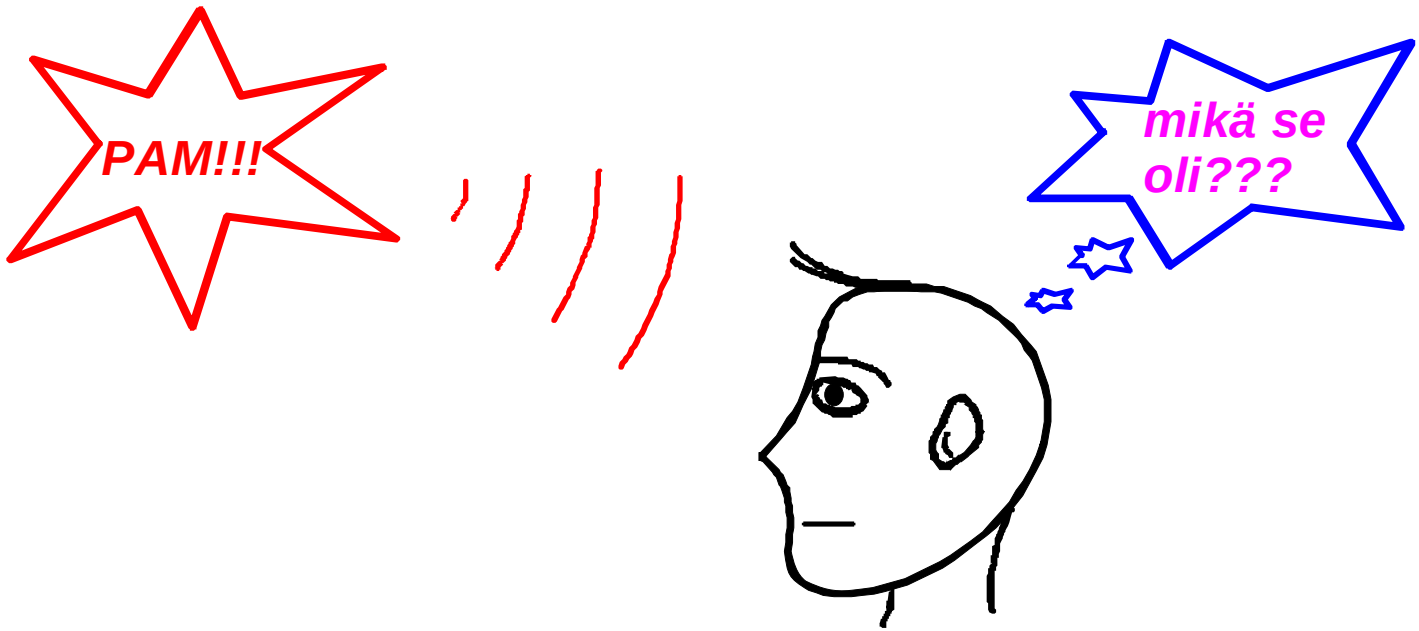
Tietoiset havainnot

- Tietoinen havaitseminen ei ole pelkkää aisti-informaation vastaanottoa eikä myöskään hahmontunnistusta.

- Kamera ei ole tietoinen ympäristöstään vaikka muodostaakin kuvan siitä. Se tekee kuvan toisia varten, ei itselleen.



Tietoiset havainnot



Kuva: P O A Haikonen

► **Tietoiselle havainnolle** on ominaista huomion (attention) kiinnittyminen, raportointimahdollisuus ja ainakin lyhytaikainen muistijälki, tapahtuma voidaan palauttaa mieleen.

Tietoinen vs. tiedostamaton

- Teemme monia asioita tiedostamattomasti; käveleminen, liikerutiinit, jne.
- Joskus tietoisen huomion keskittäminen tiedostamattomasti suoritettavaan rutiiniin voi vaikeuttaa suoritusta; Esim. solmion solmiminen, pillerin nieleminen
- Monimodaalisessa systeemissä nämä ilmiöt on helppo selittää.

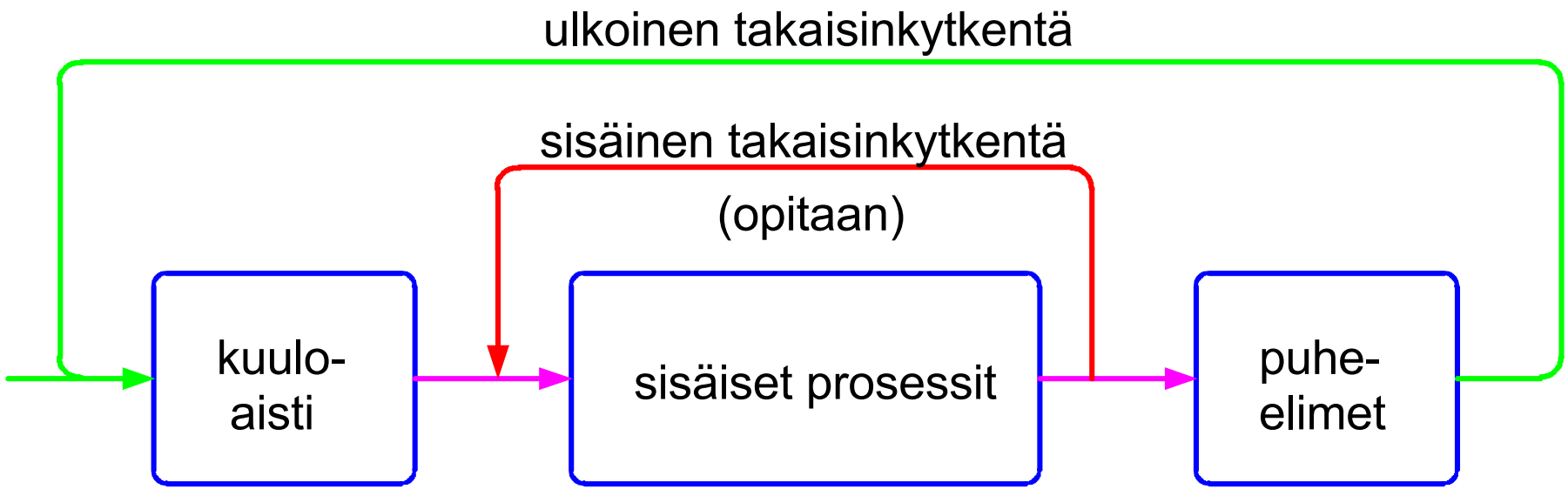
Tietoisuus ja sisäinen puhe

Sisäinen puhe: Kuka puhuu, kuka kuuntelee? Itse puhumme itsellemme – kuuntelija on sama kuin puhuja. Puhuja tietää mitä puhuu. Jos kuuntelija on sama, niin kuuntelijakin tietää. ***Miksi siis puhua ollenkaan kun asia on jo kuulijan tiedossa?***

Tietoisuus syntyy havaintojen kautta. Emme voi suoraan havaita neurosignaaleja, niinpä nämä on palautettava vastaamaan aistihavaintoja; ikään kuin kuulisimme puhetta. Tämä voi tapahtua sisäisen ***takaisinkytkennän*** kautta. Ajatus on tietoisuudessamme vasta kun olemme ajatelleet (puhuneet) sen.

Tämä takaisinkytkentä on ilmeisesti opittava; pienten lasten lienee pakko puhua ääneen **tietääkseen** mitä ajattelevat.

Sisäinen puhe ja takaisinkytkentä



Kuva: P O A Haikonen

Tietoisuus ja sisäiset mielikuvat

- Sisäiset mielikuvat vastaavat sisäistä puhetta visuaalisessa tasossa. Tässä sisäinen takaisinkytkentä palauttaa neurosignaalit näköhavaintoa vastaavaksi.
- Ilmeisesti esineiden käsittely ja piirtäminen myötävaikuttavat takaisinkytkennän syntyyn tässä ja visuaalisen tietoisuuden syntyyn omasta mielensisällöstä.

Itsetietoisuus

- Tietoisuus omasta ruumiista syntyy aistihavaintojen yhdistämisen avulla
- Suorat aistihavainnot: Tuntoaisti, ruumiinjäsenten asento, jne
- Epäsuorat havainnot; näköaisti

Kuva: P O A Haikonen

Näiden havaintojen perusteella muodostuu käsitys itsestä ympäristöstä erillisenä yksikkönä

Itsetietoisuuden osatekijöitä

- Omakuva, oma henkilökohtainen historia
- Tietoisuus omasta ruumiista
- Tuntemukset, kipu ja mielihyvä

ovat henkilökohtaisia

Kuva: P O A Haikonen

- Tietoisuus omasta mielensisällöstä

ja sen tunnistaminen omaksi,

viittaus omissa ajatuksissa omiin ajatuksiin (self reference)

Robotti tahtovana tekijänä

- Tietoinen ihminen on tahtova tekijä.

Voiko robotti tahtoa jotakin?

- Tahtomiseen liittyy pyrkimys muuttaa vallitseva tilanne mieleiseksi.

Voivatko tilanteet olla robotille mieluisia ja epämieluisia?

- Voiko robotti kokea **kipua** ja **mielihyvää**?

- Jos ajattelemme robottia kokoelmana hammasrattaita tai transistoreja, niin vastaus on **EI**. Hammasrattaat ja transistorit eivät varmaankaan koe kipua eikä mielihyvää eivätkä tahdo mitään.

Tietoisuus vuorovaikutteisen systeemin ominaisuutena

- **Transistorit eivät ole tietoisia**, mutta jos katsomme robotin toimintaa systeeminä, niin tässä systeemissä voikin esiintyä **vuorovaikutteisia tiloja ja reaktioita**, jotka vastaavat esimerkiksi kipua, mielihyvää ja tahtotiloja. Nämä tilat ja reaktiot ovat ymmärrettävissä vain merkitystasolla.

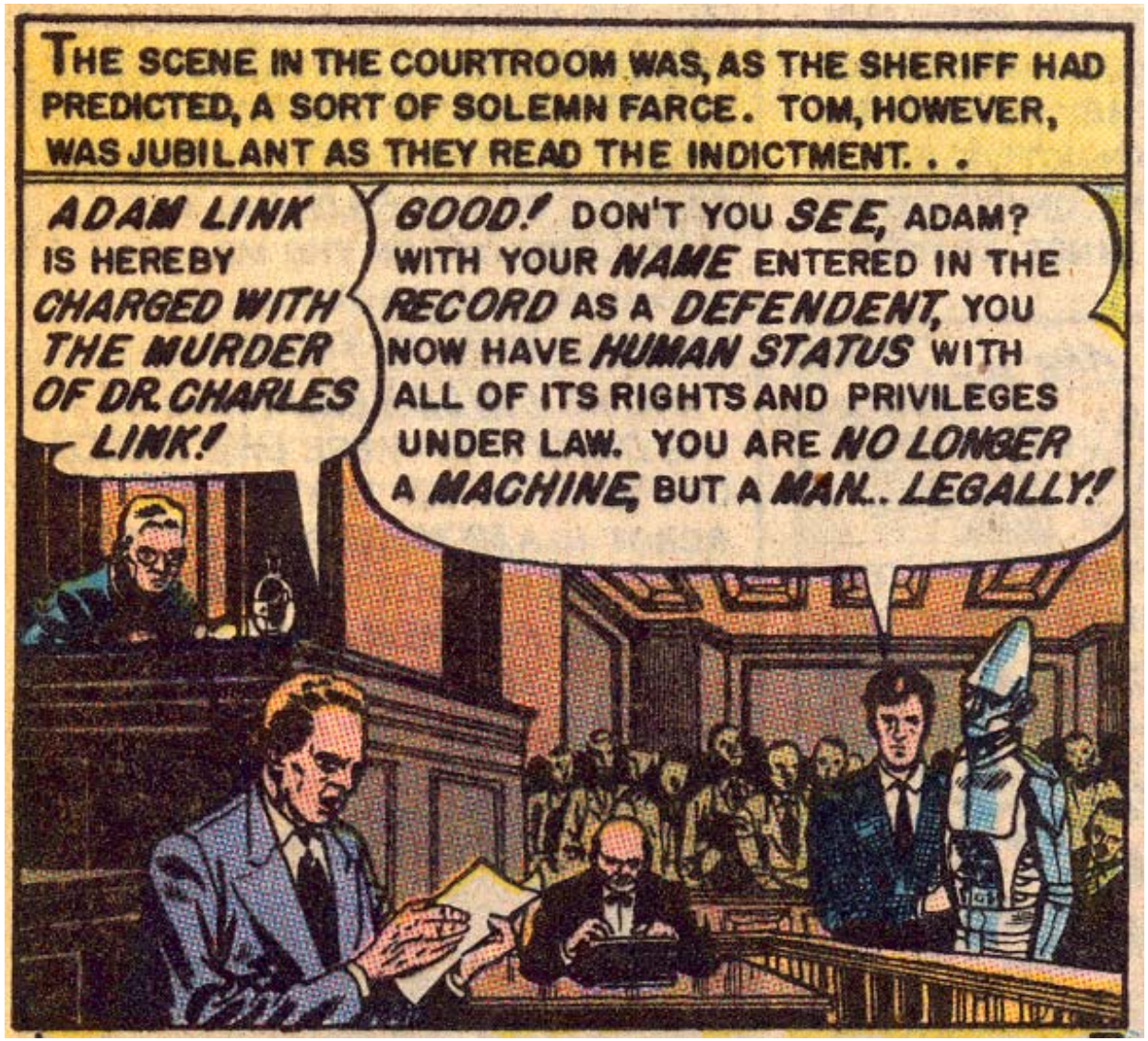
Voiko robotti tehdä ”tekoja”?

- Pystyykö robotti päämäärähakuiseen tavoitteelliseen toimintaan, jonka merkityksen se ymmärtää; ts. tekee jotakin odottaen tämän toiminnan johtavan **hyvään** lopputulokseen?

- Tietoinen robotti voi kuvitella halutun lopputilanteen ja voi assosiatiivisesti kuvitella tarvittavat välitoiminnat tämän tavoitteen saavuttamiseksi. Mikäli edellytykset näiden toteuttamiseen ovat ja **hyvä/paha-kriteerio** toteutuu, niin robotti voi toteuttaa toiminnan.

- **Robotille arvomaailma, vastuu ja… ihmisarvo?**

Sillä välin tulevaisuudessa



Tietoisuuden ongelma ratkaistu?

Hypoteesi:

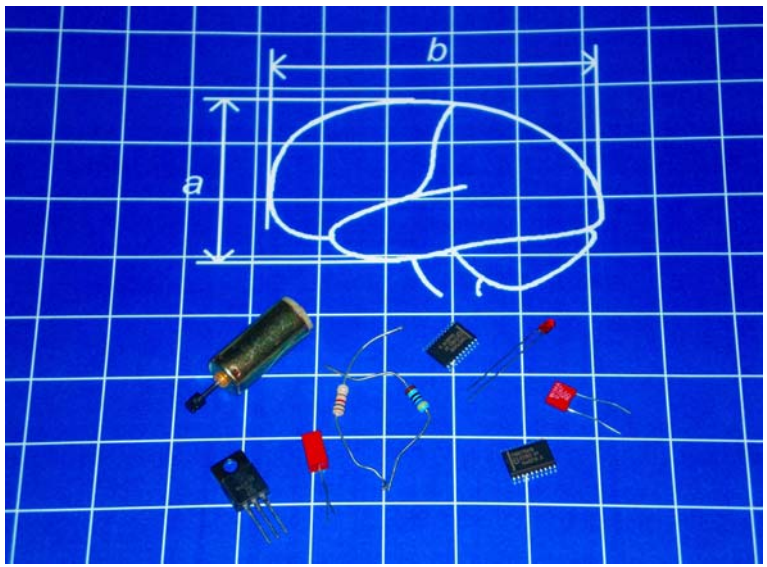
► Tietoisuuden pohjana on **läpinäkyvästi havaitsevassa monimodaalisessa systeemissä** tapahtuva modulien välinen kommunikointi, joka mahdollistaa assosiatiivisten muistikuvien teon ja raportoinnin.

► Tietoisuuden olemus on systeemin oma havainto tästä toiminnasta.

► **Tämä hypoteesin mukaisia systeemejä voidaan toteuttaa keinotekoisesti.**

Tietoisen koneen visio

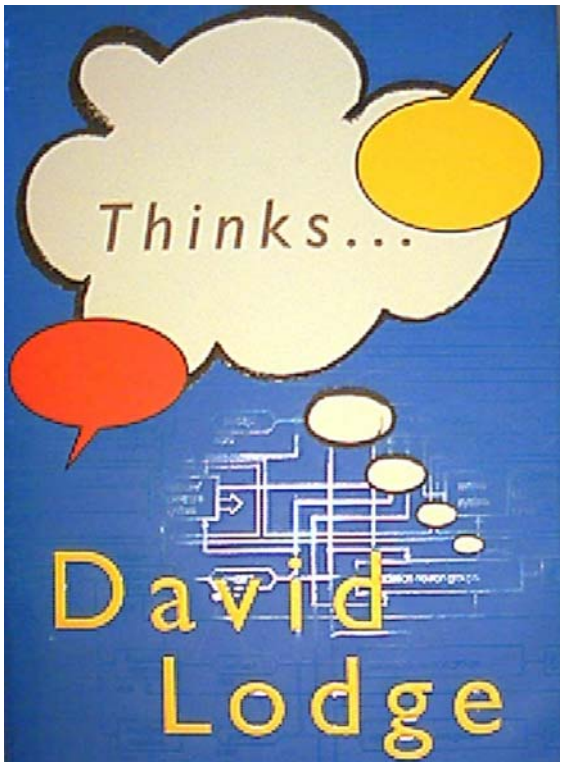
Allekirjoittaneen visio tietoisesta koneesta



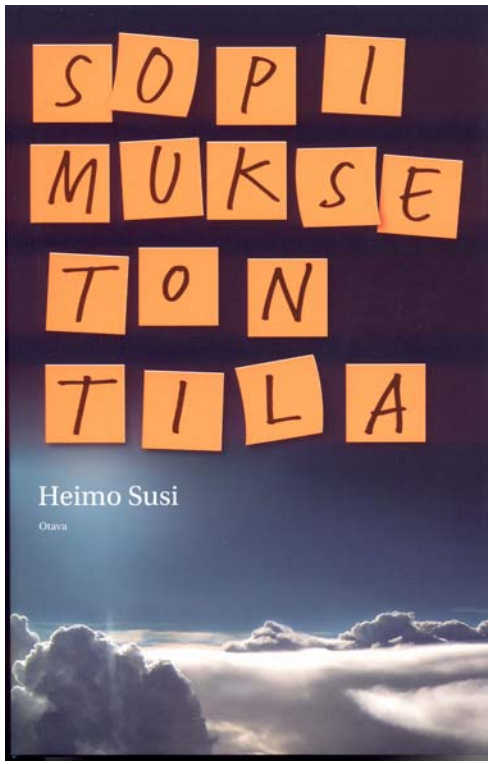
Kuva: P O A Haikonen

- Se **EI** ole tietokone
- Se **EI** perustu mikroprosessoreihin, digitaalitekniikkaan eikä ohjelmointiin
- Se **EI** perustu tavanomaisiin keinotekoisiin neuroverkkoihin
- Se perustuu assosiatiivisiin neuroverkkoihin, jotka voivat käsitellä symboleja ja assosioituja merkityksiä
- Se on oppiva vuorovaikutussysteemi, joka yhdistää suorat havaintoprosessit, sisäiset prosessit ja mekaaniset vasteet saumattomasti
- Se käsittelee merkityksiä ja sillä on oma tahto ja arvojärjestelmä

Konetietoisuuspohdintoja kaunokirjallisuudessa



David Lodge:
Thinks…
Secker & Warburg, UK, 2001



Heimo Susi:
Sopimukseton tila
Otava 2007

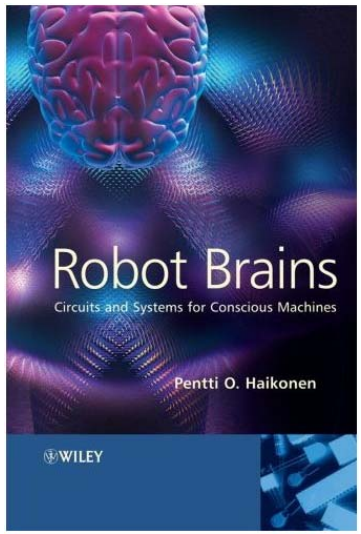
Tietoiset Robotit

Kiitos mielenkiinnosta!

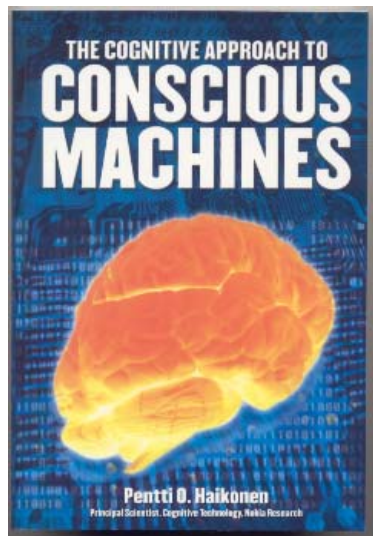
Pentti O A Haikonen, TkT

Nokia Research Center

pentti.haikonen@nokia.com



John Wiley & Sons 2007



Imprint Academic 2003