Affektiivisuuden laskennallinen määritteleminen

Mika Turkia

Helsinki 27.1.2007

HELSINGIN YLIOPISTO
Tietojenkäsittelytieteen laitos

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Laitos — Institution — Department
Matemaattis-luonnontieteellinen	Tietojenkäsittelytieteen laitos

Tekijä — Författare — Author

Mika Turkia

Työn nimi — Arbetets titel — Title

Affektiivisuuden laskennallinen määritteleminen

Oppiaine — Läroämne — Subject

Tietojenkäsittelytiede

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Pro gradu -tutkielma	27.1.2007	60 sivua

Tiivistelmä — Referat — Abstract

Affektiivisten eli emootioihin ja tunteisiin liittyvien käsitteiden määrittely ja mallintaminen on tällä hetkellä eräs sekavimmista tutkimusaloista. Aihetta voidaan lähestyä monen eri tieteenalan kautta. Suurin ongelma on yhteisesti hyväksyttyjen määritelmien puute. Teorioiden toimivuutta ja sisäistä johdonmukaisuutta on muun muassa kielitieteessä ryhdytty tarkastamaan laskennallisten simulaatioiden avulla. Affektiivisten käsitteiden määritelmiä tarvittaisiin vastaavien simulaatioiden tekemiseksi käyttäytymistieteissä.

Tässä tutkielmassa esitetään yksinkertaiselle hyötyä maksimoivalle toimijalle eli agentille soveltuvat affektien alustavat laskennalliset määritelmät. Määritelmät on tuotettu syntetisoimalla ja valikoimalla ajatuksia useista eri teorioista. Affektien luokka määritellään emootioiden ja tunteiden luokkien yläluokaksi. Affekti määritellään prosessiksi, jossa toimijan hyödyn muutos aiheuttaa vakioisen kehollisen muutoksen. Jos muutosprosessi on tarkasteluhetkellä toimijan tarkkaavaisuuden kohteena eli toimija on tietoinen siitä, kyseessä on tunne. Jos muutosprosessi on periaatteessa otettavissa tarkkaavaisuuden kohteeksi eli se on esitietoinen, kyseessä on emootio. Affektit eivät siis edellytä tietoisuutta, mutta emootiot ja tunteet edellyttävät.

Esimerkiksi odottamattomiin toteutuneisiin eli menneisiin tapahtumiin kohdistuvia affekteja ovat ilahtuminen ja säikähdys. Ilahtuminen on odottamattoman positiivisen tapahtuman seuraus ja säikähdys vastaavasti odottamattoman negatiivisen tapahtuman seuraus. Odotettuihin toteutuneisiin tapahtumiin kohdistuvia affekteja ovat onni (odotettu positiivinen tapahtuma toteutui), pettymys (odotettu positiivinen tapahtuma ei toteutunut), suru (odotettu negatiivinen tapahtuma toteutui) ja helpotus (odotettu negatiivinen tapahtuma ei toteutunut). Odotettuihin toteumattomiin (tuleviin) tapahtumiin kohdistuvia affekteja ovat vastaavasti pelko ja toivo. Eräitä muita affekteja voidaan määritellä edellä mainittujen tapahtumien aiheuttajiin kohdistuviksi. Affektiluokituksesta on tehty myös ohjelmatoteutus, jonka tarkoituksena on sekä varmistaa määritelmien koherenttius että havainnollistaa mallin mahdollisuuksia.

Affekteihin liittyvien kehollisten muutosten täsmällistä sisältöä ei pidetä ilmiön loogisen rakenteen kannalta merkityksellisenä. Hyötyfunktiotakaan ei tarvitse määritellä, koska tarkastelun kohteena on ainoastaan sen dynamiikka.

ACM Computing Classification System (CCS):
I.2.11 [Distributed Artificial Intelligence],
J.4 [Social and Behavioral Sciences]

Avainsanat — Nyckelord — Keywords

affekti, emootio, tunne, tietoisuus, agentit

Säilytyspaikka — Förvaringsställe — Where deposited

Kumpulan tiedekirjasto, sarjanumero C-2007-

Muita tietoja — övriga uppgifter — Additional information

Sisältää cd-levyn

Sisältö

1 Johdanto

Tässä luvussa esitellään lyhyesti emootiotutkimuksen kenttää sekä taustateorioita, joiden pohjalta tässä tutkielmassa esitettävää mallia on kehitetty.

1.1 Tietojenkäsittelytieteellinen emootiotutkimus

Emootioiden määrittely ja mallintaminen on tällä hetkellä eräs sekavimmista tutkimusaloista. Aihe on monimutkainen ja monitulkintainen ja sitä voidaan lähestyä monen eri tieteenalan ja käsitejärjestelmän kautta. Tällaisia tieteenaloja ovat tietojenkäsittelytieteen lisäksi ainakin filosofia, kognitiotiede, lääketieteelliset neurotieteet, psykiatria, psykologia, sosiaalipsykologia ja taiteen tutkimus (ks. esim. [OaJ96, luku 1]). Emootioita on vaikea syvällisesti tutkia ilman itsetarkastelua, joka tuo subjektiiviset tulkinnat ja painotukset väistämättömästi mukaan. Tätä ei kuitenkaan saa tai tarvitse pitää ongelmana [Dam99, s. 308–309].

Aihepiirin ongelmana on erityisesti yhteisesti hyväksyttyjen määritelmien puute. Yleisemmin tarkasteltuna kysymys emootioiden määrittelystä palautuu kysymykseen koherentin ihmiskäsityksen puutteesta tieteenalojen sisällä ja välillä. Eri tieteenalojen ihmiskäsitykset poikkeavat toisistaan siinä määrin, että vuorovaikutus on hankalaa. Tieteenalojen sisälläkään harvoin saavutetaan yksimielisyyttä inhimillisen toiminnan luonteesta. Tämä johtaa päällekkäisyyksiin ja ristiriitaisuuksiin.

Voidaan ajatella, että esimerkiksi kansantaloustiede oli ennen matematisoitumistaan samankaltaisessa tilassa, jossa humanististen tieteiden voidaan ajatella olevan nyt. Matemaattiset mallit ovat mahdollistaneet jonkinlaisen yhteisesti hyväksytyn käsitepohjan luomisen. Toisaalta nykyisen kansantaloustieteen voidaan ajatella kadottaneen osan selityskyvystään ja yhteiskunnallisesta relevanssistaan liiallisen yksinkertaistettujen mallien käytön vuoksi. Joka tapauksessa saman tapainen formalisointi mitä todennäköisimmin hyödyttäisi humanistisia tieteitä auttamalla käsitteiden selventämisessä ja yhtenäistämisessä.

Tietojenkäsittelytieteellinen kiinnostus emootioihin jakautuu kahteen eri suuntaukseen: toisaalta emootioiden synteesiin ja toisaalta käyttöliittymälähtöiseen koneen ja ihmisen vuorovaikutuksen helpottamiseen liittyvään tutkimukseen. Pääosin käyttöliittymälähtöistä näkemystä edustaa muun muassa Rosalind W. Picard [Pic97]. Emootioiden synteesillä puolestaan tarkoitetaan karkeasti määritellen emootioiden luonteen tai niiden ”keinotekoisen valmistamisen” tutkimista. Voidaan esimerkiksi yrittää selvittää, voidaanko koneilla katsoa voivan esiintyä emootioita. Toisaalta esimerkiksi pyrittäessä muodostamaan ihmisen toimintaa simuloivia keinotekoisia toimijoita niiden käyttäytymisen uskottavuus vaatii myös emotionaalisen käyttäytymisen simuloimista (esim. [Ort03]). Tässä tutkielmassa käsitellään siis emootioiden synteesiä.

Tärkeänä emootioiden synteesin tutkimuksen innoittajana 1990-luvun alkupuolella toimi edellä mainittu OCC-malli [OCC88]. Myöhemmin vaikutteita on otettu muun muassa neurologisten poikkeavuuksien tutkimuksesta [Dam95, Dam99, Dam03]. 2000-luvun alussa emootioiden tutkimus koki ehkä lyhyen keinotekoisten toimijoiden tutkimuksen hetkelliseen lisääntymiseen liittyneen nousuvaiheen.

1.2 Tutkielman luonne ja rakenne

Tutkielma on pro gradu -työksi ehkä hieman poikkeava, sillä se luonteeltaan pääasiassa teorianmuodostusta. Aiheen luonteen vuoksi esitetyt ajatukset eivät luonnollisesti voi edustaa lopullista totuutta, vaan kyseessä on kehittyväksi tarkoitettu luonnos aiheesta.

Ensimmäisessä luvussa esitellään tutkimuksen kenttää ja aiheeseen liittyviä taustateorioita. Toisessa luvussa esitellään affektiivisen toimijan määrittelyyn ja toteutukseen tarvittavia perusperiaatteita ja -käsitteitä, kuten toimijan ja ympäristön suhdetta ja hyödyn maksimoinnin ajatusta.

Kolmannessa luvussa määritellään alustavasti joitakin yleisesti tunnettuja affekteja suhteessa aikaan, toimijan havaintohistoriaan ja toimijan hyötyfunktioon. Affektimääritelmiä tarvitaan laskennallisten käyttäytymistieteellisten teorioiden muodostamiseen. Affektien kantaja on toimija, jonka ajatellaan maksimoivan jotakin hyötyfunktiota. Affektien luokka määritellään emootioiden ja tunteiden luokkien yläluokaksi.

Luvussa 3.1 affekti määritellään prosessiksi, jossa odotetun tulevan kokonaishyödyn muutos (ks. luku 2.6) aiheuttaa vakioisen kehollisen muutoksen. Kehollisen muutoksen tarkoitus on lisätä todennäköisyyttä, että toimijan teko johtaa odotetun hyödyn nousemiseen. Kehollinen muutos voi olla joko toimijan itsensä havaittavissa tai ei. Jos se on havaittavissa, se voidaan ainakin periaatteessa ottaa tietoisuuteen, jos muut tietoisuuteen tarvittavat rakenteet ovat olemassa. Tietoisuuden kohteena oleminen samaistetaan alustavasti tarkkaavaisuuden kohteena olemiseen.

Emootio määritellään esitietoiseksi affektiksi. Jos toimija voi ainakin periaatteessa ottaa affektin tietoisuuteensa, se on emootio. Tunne määritellään tietoiseksi affektiksi. Tunne on siis affekti, joka on tarkasteluhetkellä tietoisuudessa (tarkkaavaisuuden kohteena). Affektit eivät siis edellytä tietoisuutta, mutta emootiot ja tunteet edellyttävät.

Luvuissa 3.4 ja 3.5 esitellään alustava ehdotus affektien taksonomiaksi eli luokitukseksi. Odottamattomiin toteutuneisiin eli menneisiin tapahtumiin kohdistuvia affekteja ovat ilahtuminen ja säikähdys. Ilahtuminen on odottamattoman positiivisen tapahtuman seuraus ja säikähdys vastaavasti odottamattoman negatiivisen tapahtuman seuraus. Odotettuihin toteutuneisiin tapahtumiin kohdistuvia affekteja ovat onni (odotettu positiivinen tapahtuma toteutui), pettymys (odotettu positiivinen tapahtuma ei toteutunut), suru (odotettu negatiivinen tapahtuma toteutui) ja helpotus (odotettu negatiivinen tapahtuma ei toteutunut). Odotettuihin toteumattomiin (tuleviin) tapahtumiin kohdistuvia affekteja ovat vastaavasti pelko ja toivo. Eräitä muita affekteja voidaan määritellä edellä mainittujen tapahtumien aiheuttajiin kohdistuviksi.

Näin määriteltynä affektiivisuuteen liittyvät sekä hyötyfunktion arvon muutokset että kehollisuus. Vakioiset keholliset muutokset nähdään hyötyfunktion arvon muutosten seurauksina ja siten toimijan tulosteena. Vakioisuudella tarkoitetaan sitä, että samansuuntaiseen muutokseen samanlaisessa sosiaalisessa tilanteessa liittyy aina samanlainen fysiologinen reaktio. Affektit ovat siis konsistentteja ja siten ennustettavia.

Affekteihin liittyvien kehollisten muutosten täsmällistä sisältöä ei pidetä ilmiön loogisen rakenteen kannalta tarpeellisena. Olennaista on ainoastaan, että kehollinen muutos lisää toimijan elinkelpoisuutta. Myöskään hyötyfunktiota ei tarvitse määritellä, koska tarkastelun kohteena on ainoastaan sen (diskretisoitu) dynamiikka.

Jos affektiivisuuden edellytyksenä pidetään todellisen elollisuuden olemassaoloa ja sen loppumisen mahdollisuutta rakenteen lopullisen tuhoutumisen mielessä, niin nykymuotoiset koneet eivät voi olla tässä mielessä aidosti affektiivisia. Muuta periaatteellista eroa luonnollisen ja koneellisen affektiivisuuden välille esitetyt määritelmät eivät kuitenkaan tee, vaan esitettyjen määritelmien katsotaan soveltuvan molempiin. Affektiivisuutta pidetään rakenteellisena ominaisuutena, jonka perustana on annettu hyötyfunktio. Rakenteen materiaalista toteutustapaa ei pidetä erityisen olennaisena.

Ilmaisuvoimaltaan mallin katsotaan riittävän useiden hankalasti määriteltävinä pidettyjen käsitteiden määrittelyyn, ja se on pyritty määrittelemään koherentiksi niin, että se on laajennettavissa kattamaan useampia ilmiöitä ja käsitteitä. Tässä vaiheessa määritelmät ovat kuitenkin alustavia. Esimerkiksi kovin monia ihmisen emotionaalisina pidettyjä ilmiöitä ei voida kuvata näin yksinkertaisessa mallissa. Saadakseen käsityksen mallin luonteesta lukijan kannattaakin tässä välissä lukea luku 4.3 ja palata sitten mallin esittelyyn.

Luvussa 4 esitellään affektiluokituksen osittainen simulaatiototeutus, ja viimeisessä luvussa joitakin kokoavia ajatuksia.

Taustateorioista Tutkielma käyttää pohjanaan kolmenlaisia taustateorioita. Tavoitteena on ollut saavuttaa jonkinasteinen synteesi eri lähestymistapojen välillä ja karkea mutta suhteellisen koherentti yleiskuva aihepiiristä, Aihepiirin laajuuden vuoksi kaikkiin yksityiskohtiin ei ole voitu paneutua. Taustateorioista on poimittu aihealueen kannalta hyödyllisiä ajattelutapoja ja peruskäsitteitä, mutta kaikkien tarkempi esittely sivuutetaan. Tässä vaiheessa taustateorioiden käyttö on siis ollut pääasiassa ideoiden lainaamista ja pyrkimistä siihen, että määritelmät ovat karkeasti yhteensopivia taustateorioiden kanssa.

Hyödyn maksimointiin liittyvä ajattelutapa ja peruskäsitteistöä on poimittu *vahvistusoppimisteoriasta* ja sen edelleenkehitelmistä (ks. esim. [SuB98, Hut04]). Toisena pohja-aineistona käytetään tietojenkäsittelytieteellisessä emootiotutkimuksessa usein käytettyä niin sanottuihin *arviointiteorioihin* kuuluvaa OCC-mallia [OCC88], joka esitellään lyhyesti luvussa 1.4. Kolmantena tausta-aineistona käytetään eräitä *objektisuhdesuuntautuneiden psykoanalyyttisten teorioiden* perheen jäseniä (yleisesittely esim. [Sli91, s. 25–44] tai yksittäinen uudempi teoria [Täh93], suomennos [Täh96]).

Käsitteistöä Emootioiden yhteydessä puhutaan usein ns. *emergenssistä*. Emergenttinä pidetään ilmiötä, joka syntyy jonkin mekanismin osien välisessä vuorovaikutuksessa ilman, että tätä ilmiötä synnyttämään olisi luotu erityisiä mekanismeja. Monet emergentit ilmiöt ovat erilaisia virhetilanteina pidettyjä epätoivottuja ilmiöitä. Voitaneen sanoa, että ”emergenssi” on määrittelijän kyvyttömyyttä ennakoida mekanismin osien vuorovaikutuksen lopputulosta. Periaatteessa emergentti ilmiö on kuitenkin mekanismin deterministisesti määrittelemä. ”Emergenssi” on siis vain ilmiön monimutkaisuutta havainnoijan käsittelykykyyn verrattuna (vrt. ns. kaoottiset ilmiöt). Jos aiheuttava mekanismi voidaan selvittää, emergenssi ”katoaa”. Jos fysikaalinen maailma oletetaan kausaaliseksi, niin jokaisella ilmiöllä on oltava jokin aiheuttaja; mikään ei voi emergoitua tyhjästä. Näin ollen emergenssistä puhuminen ei ole hyödyllistä.

Toinen merkityksellinen käsite tämän aihepiirin yhteydessä on *oppiminen*. Nilssonin mukaan ”oppiva kone, laajasti määriteltynä, on mikä tahansa laite jonka toimintaan vaikuttavat menneisyyden kokemukset” [Nil65] [Tan87, s. 284]. Vaikutus voi olla negatiivista, merkityksetöntä tai positiivista (suhteessa johonkin tavoitteeseen). Kyseessä on ympäristöstä erillinen yksikkö, jolla on jonkinlainen muisti, johon kokemukset voivat tallettua, ja mekanismit, joilla toiminta päätetään muistiin talletetun tiedon pohjalta.

1.3 Esitettyjen emootiosynteesimallien monimuotoisuus

Sloman ja Scheutz kiinnittävät huomiota affektiivisen tietojenkäsittelyn ja erityisesti emootioiden synteesin tutkijoiden käyttämän käsitteistön sekavuuteen [SlS02]. Eri tutkijoiden tavoitteet, lähtöoletukset, ontologiat ja käytetyt teknologiat poikkeavat toisistaan. Ongelman on aiheuttanut alan poikkitieteellisyys: esimerkiksi neurobiologisen, filosofisen, psykologisen ja sosiaalipsykologisen emootiotermistön yhteensovittaminen on vaikeaa, ja tietojenkäsittelytieteilijän valitsema viitekehys voi olla mikä tahansa näistä tai jokin niiden yhdistelmä.

Myös tutkimuskohteeseen viittaava luonnollisen kielen sanasto on vähintäänkin moniselitteistä, pahimmillaan täysin epäselvää. Tästä johtuen tutkimusala on fragmentoitunut, argumentaatio ei ole mahdollista eri tutkijoiden viitatessa samoilla termeillä eri ilmiöihin, ja samoja tuloksia keksitään uudelleen. Usein yhteisymmärrystä ei saavuteta edes tietyn havaintoaineiston hyväksyttävyydestä jonkin ilmiön tutkimisessa.

Kirjoittajat ehdottavat ratkaisuksi käytettyjen arkkitehtuurien kuvausta yhteiseen viitekehykseen ja tässä kontekstissa tehtävää vertailua. Kirjoittajien emergenssinäkökulman mukaan arkkitehtuurivalinta määrää, millaisia ilmiöitä järjestelmässä voi esiintyä. Kun ymmärretään paremmin erilaisten arkkitehtuurien erot, ihmisen arkkitehtuuri mukaanlukien, voidaan paremmin arvioida, millaisia ilmiöitä kukin järjestelmä mahdollistaa. Vaatimusmäärittelyjen eksplikointi puolestaan selventäisi, millaisia ilmiöitä arkkitehtuuri tukee ja millaisia ei, mikä helpottaisi eri toteutusten vertailua. Kirjoittajat käyttävät malliavaruuden (design space) käsitettä viittaamaan kaikkien mahdollisten korkean tason toimijamallien muodostamaan avaruuteen, ja lokeroavaruuden (niche space) käsitettä viittaamaan vaatimusjoukkojen hierarkiaan [SlS02, Slo95].

Kahdeksi erityiseksi sekaannuksen lähteeksi Sloman ja Scheutz erottavat tutkittavan ilmiön piilevän monimutkaisuuden. Jos mentaaliset ilmiöt ovat arkkitehtuuririippuvia, niiden ymmärtämiseksi tarvitaan Slomanin mukaan eri abstraktiotasoja kuvaavia käsiteperheitä samaan tapaan kuin fysikaalisten ilmiöiden tutkimuksessa tarvitaan esimerkiksi astrofysiikkaa ja ydinfysiikkaa. Toinen sekaannuksen lähde on uskoa omaavansa selkeä käsitys joidenkin käsitteiden viittaaman ilmiön luonteesta sillä perusteella, että pystyy tunnistamaan joitakin instansseja, joihin käsitteet viittaavat. Instanssien tunnistuskyky ei riitä todistamaan tutkijan käsitekategorian ja oletetun "todellisen" ontologian vastaavutta. Käytetyt käsitteet voivat olla huonosti määriteltyjä.

Ehdotetut toimija-arkkitehtuurit ovat nisäkkäiden aivojen kerroksellisesta luonteesta johtuen (ks. esim. [OaJ96, luku 5]) tyypillisesti olleet kerrosrakenteisia. Sloman ja Scheutz ehdottavat kolmeatoista eri ulottuvuutta niiden välisten erojen luokittelemiseksi [SlS02]. Ensimmäinen ulottuvuus erottelee mallit sen mukaan, toimivatko kerrokset rinnakkain vai sarjallisesti. Toinen ulottuvuus erottelee mallit sen mukaan, onko ylemmillä kerroksilla täysi määräysvalta alempiin nähden, vai voivatko ne vaikuttaa alempien kerrosten toimintaan vain osittain. Ulottuvuus käsittelee siis ajallisesti välitöntä kerrosten välistä kontrollia.

Kolmas ulottuvuus käsittelee mallien erottelemista ajallisesti viivästetyn kontrollin mahdollisuuden perusteella. Kysymys on siis mahdollisuudesta muuttaa alempien kerrosten toimintaa ns. ehdollistamisella, jossa toistamalla jotakin toimintaa tietoisesti (ylempien tasojen ohjauksella) alemmat kerrokset korvaavat aiemman toimintoketjun ("reaktioketjun") uudella toimintoketjulla. Neljäs ulottuvuus jakaa mallit ryhmään, jossa kerrosten käyttämät prosessointimekanismit ovat samat mutta funktiot poikkeavat, ja ryhmään, jossa sekä funktiot että prosessointimekanismit poikkeavat kerrosten välillä.

Viides ulottuvuus erottelee mallit esitystavan mukaan: onko mallin sisältämän tiedon esitys esimerkiksi neuroverkko, proseduraalinen tai looginen. Kuudes ulottuvuus erottelee mallit käytettyjen koneoppimismenetelmien mukaan. Seitsemäs ulottuvuus käsittelee tavoitteiden luomista. Toimijalla on oltava lähtökohtaisesti jokin tavoite. Lisäksi toimija voi johtaa uusia "sisäsyntyisiä" seurannais- tai alitavoitteita. Ulottuvuus erottelee mallit sen mukaan, millaisia johdannaistavoitteita ja miten mallissa voidaan luoda.

Kahdeksas ulottuvuus erottelee mallit niihin, joissa kaikki havaintotieto jaetaan järjestelmään yhden kerroksen kautta, ja niihin, joissa havaintotieto voi saapua suoraan eri kerroksiin. Vastaavasti yhdeksäs ulottuvuus erottelee mallit sen mukaan, kuinka monen kanavan kautta päätettyjen toimintojen toteuttamiskäskyt eli ohjaustieto välitetään motoriselle järjestelmälle.

Kymmenes ulottuvuus erottelee mallit joissa emootiot emergoituvat järjestelmän osien toiminnasta, ja erillisen "emootiogeneraattorin" sisältävät mallit. Sloman argumentoi emergenssimallin puolesta. Yhdestoista ulottuvuus erottelee mallit sen mukaan, oletetaanko niissä tarvittavan ulkoinen kieli korkeamman tason prosessien, esimerkiksi itsereflektion, mahdollistamiseen.

Kahdestoista ulottuvuus käsittelee sitä, kuinka suuressa määrin toimijan toiminnot tai "älykkyys" on toteutettu pelkästään sisäisesti ja kuinka suuressa määrin älykäs toiminta riippuu ympäristön tarjoamista rakenteista. Kolmastoista ulottuvuus käsittelee ontologista omatoimisuutta eli sitä, osaako toimija muodostaa itse omat ontologiset kategoriansa mm. omien mentaalisten tilojensa luokittelua ja kuvaamista varten. Tällöin muiden kyky ymmärtää toimijan kategorisointia riippuisi oppimishistorioiden samanlaisuudesta, mikäli ymmärtämisen ajatellaan perustuvan eräänlaiseen hahmontunnistukseen eli analogioihin.

Yhteenvetona voidaan sanoa, että Slomanin vuonna 1994 [Slo94] ja Slomanin ja Scheutzin vuonna 2002 [SlS02] esittämät ongelmat käsitteenmäärittelyssä ja poikkitieteellisyyden hallinnassa ovat edelleen ratkaisematta. Sen sijaan esitetty luokittelutapa ei välttämättä selvennä asiaa merkittävässä määrin, koska esimerkiksi rajaa sen välillä, sisältääkö malli jonkinlaisen "emootiogeneraattorin" vai ovatko emootiot "emergenttejä" ei oikeastaan voida vetää. Kirjoittajat eivät luonnollisesti itsekään vältä termistön epäselvyydestä aiheutuvia ongelmia. Luokittelu ei myöskään tällaisenaan välttämättä mahdollista mallien sisällön paremmuuden arviointia. Luokittelu esitetäänkin tässä vain tutkimuskentän luonteen havainnollistamiseksi. Tässä tutkielmassa otettavan kannan mukaan mallit, joita Sloman ja Scheutz pyrkivät luokittelemaan, vaikuttavat liian monimutkaisilta pelkästään perustason affektiivisuuden määrittelyn ja toteuttamisen kannalta.

1.4 Arviointiteoriat ja OCC-malli

Arviointiteorioiden (appraisal theories) perusväite on, että emootiot[1] ovat kognitiivisten tilanne- ja tapahtuma-arvioiden herättämiä [SSJ01, s. 3]. Eräs viitatuimmista teorioista emootiosynteesin alalla on ollut arviointiteorioihin luettava, lasken-

[1]OCC-mallin esittelyssä noudatetaan OCC-mallin terminologiaa, jossa ei tehdä luvun 1.2 mukaista jaottelua affekteihin, emootioihin ja tunteisiin.

nallisuuteen pyrkivä mutta sosiaalipsykologian pohjalta rakennettu ns. OCC-malli (Ortony-Collins-Clore -malli), joka pyrki luokittelemaan emootiot ne aiheuttavien tilanteiden rakenteiden perusteella [OCC88, s. 2]. Päätavoite oli selittää, miten emootio seuraa tilannehavaintojen tulkinnoista [OCC88, s. 12]. Tässä aliluvussa referoidaan OCC-teorian pääpiirteet.

Emootioiden aiheuttamiseen vaadittavan kognitiivisen käsittelyn määrä vaihtelee eri emootioiden kohdalla [OCC88, s. 4]. Arviointiteoriana OCC-malli painottaa kognitiivisen tulkinnan merkitystä, käsitellen pääasiassa sekundaarisesti tuotettuja emootioita, joilla kognitiivisen käsittelyn määrä on merkittävä[2]. Se ei käsittele emootioiden ilmaisuja tai tunnistamista, ainoastaan niiden kognitiivista tuottamista. Emootiot nähdään ”arvovaritteisinä reaktioina tapahtumiin, toimijoihin tai objekteihin (engl. valenced reactions to) siten, että emotionaalisen reaktion tarkempi laatu määräytyy sen herättäneen tilanteen tulkinnan perusteella” [OCC88, s. 13]. Emootiot riippuvat siis tilanteen tulkinnasta. Tulkinta on puolestaan tulkitsijan tavoitteista ja käsitejärjestelmästä riippuvan havaintojen kognitiivisen käsittelyn tulos [OCC88, s. 4]. Emootiot nähdään reaktioina ympäristöön. Reaktion tuloksen ajatellaan olevan tietty emotionaalinen tila. Teoria pyrkii esittämään emotionaalisia reaktioita herättävien tilanteiden kieliriippumattoman luokittelutavan. Tällöin emootiot voitaisiin identifioida yksikäsitteisesti tilanteen piirteiden perusteella [OCC88, s. 1]. Yhteenvetona toimijan subjektiivinen tulkinta määrää toimijan käsityksen tilanteesta, mutta tämä tilannekäsitys määrää emootion yksikäsitteisesti.

Yksittäiset emootiot ryhmitetään luokkiin niiden yhteisten piirteiden perusteella [OCC88, s. 12-13]. Emootiot jakautuvat kolmeen yleisluokkaan emotionaalisen reaktion kohteen mukaan[3]. Kohde voi olla tilanne (tapahtuma), toimija tai objekti, joten yleisluokat ovat tapahtumien seurauksiin pohjautuva *tavoitepohjaisten tunteiden luokka*, toimijoiden tekojen arviointiin pohjautuva *normipohjaisten tunteiden luokka* ja objektien ominaisuuksien arviointiin pohjautuva *asennepohjaisten tunteiden luokka* [OCC88, s. 19]. Malli olettaa havaintojen käsitteellistämisen eli havaintojen sisällön jaottelun objekteihin, tapahtumiin ja toimijoihin tapahtuvan mallin ulkopuolella.

Yleisluokat muodostetaan OCC-mallissa siis luokittelemalla emotionaaliset reaktiot

[2]Tämä teoriatraditio perustuu paljolti erotteluun kognitiivisen ja ei-kognitiivisen välillä, kun taas tässä tutkielmassa tätä jakoa ei pidetä tarpeellisena tai välttämättä edes mahdollisena.

[3]Emootion kokijan kannalta tarkasteltuna emootion kohde on sen aiheuttaja, kun taas tapahtuman kannalta tarkasteltuna emootion kokeva toimija on tapahtuman kohde. Myöhemmissä luvuissa näkökulma vaihtuu: tilanne jäsennetään tapahtuman suhteen.

niiden kohteen mukaan. Yleisluokat jaetaan edelleen aliluokkiin tarkastelunäkökulman mukaan. Tapahtumien aiheuttamat emotionaaliset reaktiot jaetaan sen mukaan, kohdistuvatko tapahtuman seuraukset itseen vai muihin. Jos ne kohdistuvat toisiin, emotionaalinen reaktio luokitellaan *toisten onnellisuus*-aliluokkaan, joka sisältää emootiot *onni toisen puolesta (happy for), paheksunta (resentment), vahingonilo (gloating)* ja *sääli (pity)*. Jos seuraukset kohdistuvat itseen, tapahtuma arvioidaan vielä sen mukaan, onko sillä tulevaisuusvaikutuksia vai ei. Jos on, niin emotionaalinen reaktio sijoittuu *tulevaisuusnäkymiä omaavien emootioiden* aliluokkaan. Tämän aliluokan sisältämät emotionaaliset reaktiot, joiden kohdalla näkymät eivät ole vielä toteutuneet, ovat *pelko (fear)* ja *toivo (hope)*. Vastaavasti ilmentymiä joiden näkymät ovat toteutuneet odotetulla tavalla ovat *tyydytys (satisfaction)* ja *pelkojen toteutuminen (fears-confirmed)*. Jos näkymät toteutuivat mutta eivät odotetulla tavalla, ilmentymiä ovat *helpotus (relief)* ja *pettymys (disappointment)*. Tulevaisuusvaikutuksettomat emotionaaliset reaktiot sijoittuvat *hyvinvointitunteiden* luokkaan, jonka jäseniä ovat *ilo (joy)* ja *ahdinko/hätä (distress)*.

Toimijoiden tekoihin kohdistuvat emotionaaliset reaktiot sijoittuvat *ansioksilukemistunteiden* aliluokkaan. Jos teko on itse tehty, emotionaaliset reaktiot ovat *ylpeys (pride)* ja *häpeä (shame)*. Toisten tekemiin tekoihin kohdistuvat emotionaaliset reaktiot ovat *ihailu (admiration)* ja *kauna/moittiminen (reproach)*.

Objekteihin kohdistuvat emotionaaliset reaktiot kuuluvat *viehätysvoimatunteiden* luokkaan, jonka jäseniä ovat *rakkaus (love)* ja *viha (hate)*. Lisäksi malli sisältää hyvinvointi- ja ansioiksilukemistunteiden luokkien yhdistelmäaliluokan, jonka jäseniä ovat *tyydytys/mielihyvä (gratification)*, *katumus (remorse)*, *kiitollisuus (gratitude)* ja *ärtymys (anger)*. Luokka sisältää toisin sanoen tulevaisuusvaikutuksettomia tunteita, joilla on ollut seurauksia itselle, mutta joissa reagoidaan itsen tai toisen toimijan tekemään tekoon. Toisen tekemään tekoon kohdistuvat siis kiitollisuus ja ärtymys. Itse tehtyyn tekoon, jonka seuraukset olivat itsen kannalta negatiivisia, kohdistuu katumus. Vastaavasti itse tehtyyn tekoon, jonka seuraukset olivat itsen kannalta positiivisia, kohdistuu tyydytys/mielihyvä.

Tunnetilaluokkien instanssit eroavat toisistaan reaktion suunnan (positiivinen-negatiivinen) lisäksi sen voimakkuuden suhteen. Erottavat tekijät esitetään mallissa emootioluokkakohtaisina (paikallisina) muuttujina. Lisäksi malli sisältää globaaleja muuttujia, jotka vaikuttavat kaikkien emootioluokkien instanssien voimakkuuteen.

Emootioina tai emotionaalisina tiloina pidetään mallissa vain niitä tiloja, jotka eivät voi olla toimijan kannalta neutraaleja. Esimerkiksi hämmästystä ei pidetä emootiona vaan kognitiivisena tilana, koska hämmästystä aiheuttaneen objektin arvioinnin tulos voi olla neutraali [OCC88, s. 32].

OCC-malli erottaa siis arvioinnin kohteen ja tarkastelunäkökulman. Arvioinnin kohteena pidetään tapahtumia, toimijoita ja objekteja [Ort03, s. 195]. Objekteja voidaan arvioida vain makujen ja asenteiden suhteen (ne eivät voi tehdä arvioitavia tekoja eivätkä ne ole tapahtumia). Toimijoita (toimijoiden tekoja) voidaan arvioida vain normien tai standardien mukaan. Tapahtumia voidaan arvioida vain tavoitteiden suhteen. Kulloinkin siis arvioidaan yhtä kohdetta yhdeltä kannalta, paitsi yhdistelmätunteiden kohdalla, jolloin arvioitavana on teon aiheuttaman tapahtuman seuraukset.

OCC-mallin arviointia OCC-malli on melko vanha ja liikkuu käsitteellisesti korkealla tasolla. Kirjoittajat toteavat, että analyysin tarkkuustaso on mielivaltaisesti valittu [OCC88, s. ix]. Teoria on vaatimusmäärittelyluonteinen, ei varsinainen toteutustason mallin kuvaus. Ortony onkin myöhemmin luonnehtinut mallia liian monimutkaiseksi ja ehdottanut kahden affektiivisen perusreaktion, positiivisen ja negatiivisen, erittelyä kymmeneen luokkaan [Ort03, s. 193–196]. Luokittelun mukaan positiiviset reaktiot jaettaisiin seuraavasti viiteen luokkaan (suluissa tutkielman kirjoittajan esimerkinomainen tulkinta): reaktio koska jotain hyvää tapahtui (onni), reaktio jonkin hyvän tapahtumisen mahdollisuuden vuoksi (toivo), reaktio koska pelätty negatiivinen asia ei tapahtunut (helpotus), reaktio itseaiheutetun kiitettävän (praiseworthy) toiminnon johdosta (ylpeys) ja reaktio koska jotakin pidetään houkuttelevana tai viehättävänä (halu/ihastus). Negatiiviset reaktiot jaettaisiin vastaavasti seuraaviin viiteen luokkaan: reaktio koska jotain pahaa tapahtui (suru), reaktio jonkin pahan tapahtumisen mahdollisuuden vuoksi (pelko), reaktio koska toivottu positiivinen asia ei tapahtunut (pettymys), reaktio itseaiheutetun moitittavan (blameworthy) toiminnon johdosta (häpeä) ja reaktio koska jotakin pidetään vastenmielisenä tai epäviehättävänä (inho).

Paolo Petta huomauttaa OCC-mallin olevan edelleen erittäin suosittu, mutta myös laajasti väärinymmärretty: se ymmärretään usein lähes universaalisti sovellettavaksi eikä sen rajoituksia oteta huomioon [Pet04, s. 4]. Petta huomauttaa OCC-mallin edellyttävän, että normien ja standardien muodostamisen mahdollistava infrastruktuuri on ennalta määritelty (todellisuudessa OCC-malli edellyttää huomattavasti enemmän ennalta määriteltyä infrastruktuuria). Voitaisiin ajatella, että ratkaisuna olisi määritellä tällainen infrastruktuuri. Tässä tutkielmassa tätä ratkaisua ei kuitenkaan pidetä mahdollisena. Sekä OCC-malli että Ortonyn myöhemmät määritelmät käyttävät mm. normin ja standardin sekä maun ja asenteen käsitteitä määrittelemättä, aksiomaattisesti. Mallin validiteetti häviää, jos normin ja asenteen käsitteiden ei katsota olevan riippumattomia emootion käsitteestä. Tässä tutkielmassa otetun kannan mukaan nämä käsitteet eivät ole riippumattomia. Sekä normin että emootion käsitteet perustuvat toimijan odotettuun hyötyyn, ja ovat lähinnä saman asian piirteiden tarkastelua eri näkökulmista. Normien tarkoitus on vaihteluvälin asettaminen sivullisten odotetun hyödyn muutoksille ja siten heidän emootioilleen. Ne ovat siis sivullisten toimijalle opettamia, ja samalla luonnollisesti asettavat vaihteluvälin toimijan omille emootioille. Normit ja emootiot ovat siis kiinteästi sidoksissa toisiinsa, eikä niitä voida määritellä toisistaan riippumatta tai riippumattomiksi. Normeja käsitellään luvussa 2.12. Lisäksi voidaan olettaa, että esimerkiksi valitsemalla luokittelupuun juureksi tunnereaktion sijaan tapahtuma luokittelusta voitaisiin saada selkeämpi. Esitetystä luokittelusta nähdään, että erityisesti yhdistelmätunteiden kohdalla luokittelun optimaalisuus on kyseenalainen. OCC-mallia ei tässä tutkielmassa pidetäkään teoreettisesti kestävänä.

OCC-mallin ajatus aiheuttavan tilanteen rakenteen olennaisuudesta vaikuttaa kuitenkin oikealta. Tilanteen rakenteella viitataan tilanteista jollakin kuvaustavalla tehtyihin abstraktioihin eli pelkistäviin käsitteellistyksiin. Erilaiset kuvaustavat mahdollistavat erilaiset tilanteiden väliset erottelut. Jos toimijat esimerkiksi ajatellaan olioiksi, joiden ainoa tietosisältö on lista heidän kohtaamiensa olioiden nimistä, ei voida esittää kohdattuihin olioihin kohdistuvia arvotuksia. Jos pelkän nimen sijaan käytetäänkin nimeä ja yhtä kokonaislukumuuttujaa, kohdattujen olioiden arvotus voidaan tallettaa tähän muuttujaan. Arvottaminen tehdään funktiolla, joka kuvaa olion kokemuksen kyseisestä toisesta oliosta numeeriseksi muuttujan arvoksi. Tätä funktiota voidaan kutsua *hyötyfunktioksi*. Yhdellä muuttujalla voidaan kuvata vain yksi ulottuvuus, esimerkiksi olioon kohdistuva pitäminen. Jos muuttuja korvataan jollakin monimutkaisemmalla tietorakenteella, esimerkiksi listalla muuttujia, toimija voi tallentaa muistiinsa esimerkiksi olion arvon toimijan jokaisen osatavoitteensa kannalta. Kuvausmahdollisuudet riippuvat siis mallin valinnasta.

OCC-mallin normin ja standardin käsitteiden määrittelemättömyys voitaneen ajatella niin, että malli riippuu parametreista, jotka tässä tapauksessa ovat mainittujen käsitteiden määritelmät. Käsitellään vielä parametrisuuden ongelmaa yleisemmin. Edellä esiteltiin Slomanin emootiomallien luokitteluehdotus. Rosalind W. Picard on ehdottanut emootiomallien luokittelua diskreetteihin, jatkuviin ja sääntöpohjaisiin malleihin [Pic95, Pic97]. Diskreetit ja jatkuvat mallit ovat Picardin luokituksessa ”matemaattisia”, kun taas sääntöpohjainen on ”epämatemaattinen” malli. OCC-malli on Picardin luokituksessa ns. sääntöpohjainen malli. Picard näyttää tarkoittavan ”matemaattisuudella” lähinnä toimintatapaa, jossa pohjana on jonkin emotionaaliseksi oletetun toimijan käyttäytymistä tai fysiologisia muutoksia kuvaava numeerinen mittausaineisto, jota sitten käsitellään matemaattisin menetelmin; toisin sanoen mallinnetaan jotakin olemassa olevaa emotionaaliseksi oletettua toimijaa. Kuten Slomaninkin luokittelun niin myös Picardin luokittelun ongelmana voidaan pitää luokittelukriteerien epämääräisyyttä. Suurin ongelma on kuitenkin se, että tällainen tilastollinen malli jää riippumaan aineistosta johdetuista parametreista. Malli ei määrittele, vaan mallintaa. Tässä tutkielmassa otetun kannan mukaan parametrista mallintamista voidaan käyttää määritelmien muodostamisen apuna, mutta lopullisissa tai täydellisissä määritelmissä ei voi olla parametreja. Parametrien käyttö voitaneen ymmärtää tuntemattoman ilmiön likiarvoiseksi määrittelyksi. Eksakti likiarvoton eli parametriton määritelmä on kuitenkin aina parempi kuin parametrinen, mikäli sellainen osataan muodostaa.

Tässä tutkielmassa esitettävät affektien määritelmät ovat periaatteessa parametrittomia. Emootion ja tunteen määritelmien parametrisuus riippuu luvussa 3.1 esitettävän tietoisuuden kohteena olemisen määritelmän hyväksyttävyydestä.

2 Affektimallin perusteita

Tässä luvussa määritellään alustavasti sekä joitakin yleisiä toimijuuteen liittyviä käsitteitä että joitakin psykologisia käsitteitä. Tavoitteena on muodostaa laajahko kokonaiskuva aiheesta. Siksi tässä luvussa sivutaan myös asioita, joita ei varsinaisesti tarvita luvuissa 3 ja 4 esitettävässä malliluonnoksessa, mutta jotka määrittelevät mallin kontekstia tai jotka olisi otettava huomioon mallia laajennettaessa.

2.1 Maailma, toimijat ja ympäristöt

Määritellään, että affektiivisuutta voi ilmetä vain *elollisilla toimijoilla*, koska vain elollisilla toimijoilla on *tavoitteita*. Elottomat oliot rajautuvat siis affektiivisuuden ulkopuolelle. Määritellään edelleen, että toimijoiden affektiivisuus liittyy *tapahtumiin*, jotka vaikuttavat toimijoiden tavoitteiden toteutumiseen. Haluamme myös käsitellä *toimijoiden välisiä suhteita*. Sijoitetaan siis joukko toimijoita *maailmaan*.

Maailma ja toimijat tuottavat tapahtumia jonkin sääntöjoukon pohjalta. Tapahtumat muodostavat *tapahtumaketjun*. Jos maailman ikä on suurempi kuin toimijan ikä tai jos kaikki toimijan elinajan aikana tapahtuvat tapahtumat eivät tule toimijan tietoon, niin toimija kokee maailman tapahtumista vain jonkin osajoukon. Nimitetään tätä tapahtumien osajoukkoa toimijan *ympäristöksi* (elinympäristöksi). Ympäristö on tietty tapahtumien ketju eli jokin mahdollisten tapahtumien kombinaatio[4].

Maailma voi periaatteessa olla joko *deterministinen* tai *epädeterministinen*. Simulaatiomalleissa maailma voidaan määritellä joko deterministiseksi tai näennäisen satunnaiseksi, mutta satunnaisgeneraattorilla tuotettu näennäisen satunnainenkin maailma on kuitenkin deterministinen. Jos muodostettavalla mallilla halutaan mallintaa todellista maailmaa, on otettava kantaa todellisen maailman luonteeseen. Esimerkiksi Hutterin mukaan todellisen maailman deterministisyys tai epädeterministisyys on avoin ongelma, mutta vain kvanttitasolla [Hut04, s. 245–246]. Hutter viittaa Schmidhuberiin, jonka mukaan mahdollisuutta, että kvanttitapahtumat ovat vain näennäissatunnaisia, ei voida sulkea pois kokeellisesti, ja koko fysiikka voisi itse asiassa redusoitua digitaaliseksi algoritmiksi [Sch02, Sch00, Sch97]. Toisin sanoen universumi olisi pohjimmiltaan diskreetti ja sen olisi tuottanut tai tuottaisi jokin ohjelma.

[4]Tämä ympäristön käsite näyttää vastaavan Hutterin käyttämää ympäristön käsitettä; ks. [Hut04, Hut03]. Sen sijaan Hutter ei näytä eksplisiittisesti käyttävän tai määrittelevän maailman käsitettä. Vahvistusoppimisteoriat eivät näytä tekevän selkeää eroa ympäristön ja maailman käsitteiden välille, vaan käyttävät pelkkää ympäristön käsitettä; ks. esim. [SuB98].

Joka tapauksessa vähintään käytännön tasolla todellisen maailman voidaan olettaa olevan deterministinen eli laskennallinen. Tällöin toimijan kokema "epädeterministisyys" on näennäistä ja johtuu ainoastaan kyvyttömyydestä saavuttaa täydellistä tietoa maailmasta.

Joka tapauksessa lienee selvää, että mahdollinen epädeterministinenkin maailma välttämättä tuottaa yhden ainutkertaisen tapahtumaketjun eli ympäristön jonkin sääntöjoukon mukaan. Kullakin ajanhetkellä tapahtuvat tietyt tapahtumat, ja vain ne. Toteutuneiden tapahtumien todennäköisyys on välttämättä aina yksi. Toisin sanoen jälkikäteen valinnaisia tapahtumaketjuja ei voi toteutua, eikä menneisyyteen voi lisätä eikä sieltä poistaa tapahtumia.

Affektit ovat edellä mainitulla "käytännön tasolla" (eli ydinfysiikkaa korkeammalla abstraktiotasolla) esiintyviä ilmiöitä. Affektien mallintamisen lisäksi tutkielman malli haluttiin yhteensopivaksi psykoanalyyttisten teorioiden kanssa. Niiden eräs perusoletus tai työhypoteesi on ns. *psyykkinen determinismi*, jolla tarkoitetaan sitä, että psyykkiset tapahtumat noudattavat determinististä kausaalisuutta tai että psyykkiset tilat tuotetaan funktionaalisesti (ks. esim. [Ang59], [Täh72, s. 6], [Sli91, s. 12]). Näin ollen psykoanalyyttisten teorioiden laskennallinen mallintaminen edellyttänee determinismin olettamista.

Toimijan ja (ulko)maailman rajan määrittely on tulkinnallinen. Toimijan kannalta toisten toimijoiden ja maailman voidaan katsoa olevan osa toimijan ulkopuolista todellisuutta. Toisaalta täysin deterministisessä maailmassa toimija on maailman määrittelyn seuraus, ja toimijan tuottamilta näyttävät tapahtumat kuuluvat siten oikeastaan maailman itsensä tuottamaan tapahtumaketjuun. Maailman näkökulmasta toimijan tulevat teot ovat etukäteen täysin määrättyjä, mutta maailma ei vain ole vielä toteuttanut niitä. Toimija ei itse voi ennustaa tulevia tekojaan, koska hänellä ei ilmeisesti teoriassakaan voi olla täydellistä tietoa maailman nykyisestä tilasta eikä niistä säännöistä, joilla maailma tuottaa seuraavan tilan nykyisestä tilasta, ja vaikka toimija voisikin tuntea tilan ja säännöt, seuraavan tilan laskenta käytettävissä olevassa ajassa olisi silti todellisessa maailmassa nähtävästi laskennallisesti mahdoton tehtävä. Joka tapauksessa täyden determinismin vallitessa toimijan "toimijuus" on samantyyppinen illuusio kuin luvussa 1.2 mainittu emergenssi. Toimijan sisäinen tila on maailman tilan osa.

2.2 Toimijan sisäinen malli ympäristöstään ja maailmasta

Luvussa 2.1 esitetyt ontologiset oletukset koskivat todellisen maailman luonnetta. Palatessamme toimijan ohjauksen tasolle joudumme erottamaan maailman ontologisen luonteen ja sen, miltä se laskentakapasiteetiltaan rajoitetusta toimijasta näyttää ja millaisen *sisäisen mallin* toimija siitä muodostaa. Todellinen ympäristö näyttää esimerkiksi ihmisistä *jatkuvalta*: esimerkiksi kappaleiden liike näyttää yhtenäiseltä ja portaattomalta, eikä pysäytyskuvien sarjalta. Ympäristö vaikuttaa myös *dynaamiselta*: se ei odota että toimija reagoi edelliseen tapahtumaan, vaan tuottaa koko ajan uusia tapahtumia [RuN95, s. 46].

Verrattuna aiemmin esitettyyn digitaalisen universumin ajatukseen jatkuvuus voidaan ehkä ajatella havaintokyvyn rajoituksista johtuvana konstruktiona samaan tapaan kuin se, että elokuva näyttää jatkuvalta, vaikka se todellisuudessa koostuukin pysäytyskuvien sarjasta. Dynaamisuus voitaneen tulkita esimerkiksi niin, että tapahtumat voivat olla eri pituisia diskreettien tapahtumayksiköiden jonoja. Jos havainnon tekeminen on pitempi tapahtuma kuin havainnoitava tapahtuma, sitä ei voida havainnoida, siihen reagoinnista puhumattakaan. Tällöin maailma ei ole kokonaan toimijan *saavutettavissa*: toimija ei voi koskaan saada täydellistä tietoa maailmasta [RuN95, s. 46]. Tämän johdosta toimijan näkökulmasta maailma *näyttää epädeterministiseltä* tai *todennäköisyysluonteiselta* ainakin joltakin osin.

Yksinkertaisuuden vuoksi määrittelemme simulaatiomaailman *diskreetiksi* ja *staattiseksi* [RuN95, s. 46]. Diskreetti maailma koostuu rajallisesta määrästä selkeästi toisistaan erottuvia tapahtumia. Staattinen maailma ei muutu ajanhetkien välillä eli silloin, kun toimija on laskemassa uutta tilaansa seuraavaa ajanhetkeä varten. Vaikka tämän tutkielman malli on määritelty diskreetiksi, sen voinee halutessaan yleistää stokastiseksi.

Toimija voi saavuttaa enintään sen verran tietoa maailmasta kuin toimijan elinympäristö sitä sisältää. Toisin sanoen toimija muodostaa itselleen sisäistä mallia maailmasta kokemansa tapahtumaketjun eli elinympäristönsä pohjalta. Sisäisen mallin muodostamisen motivaatio on pyrkimys ennustaa odotettua kokonaishyötyä (ks. luku 2.3).

Rajoitetun saavutettavuuden vuoksi toimijan sisäinen malli maailmasta on aina todennäköisyystietomalli, millä tarkoitetaan sitä, että vaikka toimijan ulkoista maailmaa kuvaavan sisäisen mallin jokin piirre vastaisikin täysin jotakin maailman piirrettä, toimija ei voi todistaa tämän vastaavuuden olemassaoloa. Toimijan elinaikanaan kokema ympäristö ei esimerkiksi välttämättä sisällä kaikkia mahdollisia tapahtumatyyppejä. Toimija ei koskaan voi tietää, onko hänen kokemansa ympäristö edustava otos kaikista maailman mahdollisista ympäristöistä ja ilmenevätkö siinä kaikki tapahtumia ohjaavat säännöt.

Totuuden käsitteellä viitataan siihen, että malli tai jokin sen osa vastaa täysin sitä maailmaa, jota se pyrkii kuvaamaan. Malli sisältää tietoa maailmasta, mutta se voi olla niin puutteellista, että se kuvaa todellisuuden ”väärin”, eli sen perusteella tehdyt ennusteet eivät toteudu.

Filosofian perusoppikirjoissa esitetään muun muassa totuuden ns. korrespondenssi- ja koherenssiteoriat. Korrespondenssiteorian mukaan ”totuus on jonkinlainen vastaavuussuhde uskomuksen ja tosiasian välillä: uskomus tai ajatus on tosi täsmälleen silloin, kun se 'vastaa' todellisuutta” [Nii84, s. 108]. Koherenssiteorian mukaan ”lauseen totuus merkitsee sen yhteensopivuutta muiden lauseiden muodostaman 'systeemin' kanssa” [Nii84, s. 110].

Kuten edellä on todettu näyttää ilmeiseltä, että ei-triviaalissa maailmassa toimija ei voi saada ”objektiivista” tietoa, joten korrespondenssia ei voida aukottomasti todistaa. Ainoa keino korrespondenssin lisäämiseen on toimijan sisäisen tiedon määrän kasvattaminen ja sen koherenssin lisääminen. Toimijan tieto korreloi ”todellisuuden” kanssa, mutta ”totuutta” ei voida todistaa, vaan tieto on aina oikeaa vain jollakin yhtä pienemmällä todennäköisyydellä. Näin ollen korrespondenssi- ja koherenssiteoriat eivät ole vaihtoehtoisia vaan toisiaan täydentäviä: koherenssin lisääminen on keino suurentaa korrespondenssin todennäköisyyttä.

2.3 Hyödyn maksimoinnin ajatus

Luvussa 1.2 hyötyfunktion maksimointia esitettiin affektiivisuuden erääksi peruspiirteeksi. Tässä luvussa esitellään hyödyn maksimoinnin ajatusta tarkemmin.

Eräs psykoanalyyttisten teorioiden keskeinen oletus on ns. *mielihyväperiaate* (pleasure principle), jolla tarkoitetaan toimijan synnynnäistä pyrkimystä *välittömän mielihyvän* kokemiseen ja vastaavasti pyrkimystä välttää *mielipahaa* [Täh72, s. 15]. Myöhemmin toimija voi oppia kestämään väliaikaista mielipahaa, jos hän odottaa sillä tavoin saavuttavansa suuremman kokonaishyödyn kuin hakemalla välitöntä tyydytystä. Tällöin toimijan sanotaan noudattavan mielihyväperiaatteen sijaan *realiteettiperiaatetta* (reality principle) [Täh72, s. 19].

Tekoälytutkimuksessa *vahvistusoppiminen* on toimijatutkimuksen suuntaus, jonka perusajatus on tuntemattomalta ympäristöltä saatavien ns. *vahvistussyötteiden* summan maksimoinnin oppiminen kokeilemalla (ks. esim. [SuB98]). Toimijan tavoite on siis mahdollisimman suuren hyödyn saavuttaminen. Oppimisen kautta toimija muodostaa menettelytavan, jolla tavoite oletetaan parhaiten saavutettavan. Edelleen teoreettisen tekoälytutkimuksen puolella Marcus Hutter on esittänyt sellaisen toimijan määritelmän, joka teoriassa tuottaa optimaalisen ratkaisun kaikkiin laskettavissa oleviin ongelmiin [Hut04, Hut03]. Malli on parametrivapaa ja perustuu hyödyn maksimointiin, mutta on valitettavasti laskennallisesti mahdoton toteuttaa. Hutter esittää kuitenkin myös aikarajoitetun toteutettavissa olevan mallin, joka approksimoi optimaalista toimijaa ja jonka laskennallinen suorituskyky riittää yksinkertaisten ongelmien ratkaisuun. Hutterin mukaan useimpien, mahdollisesti kaikkien, älykkyyden tunnettujen piirteiden voidaan ajatella olevan jonkin hyötyfunktion maksimointia [Hut04, s. 126–127].

Esitetyn valossa voidaan nähdä, että näissä ensin hyvin erilaisilta tai yhteensopimattomilta vaikuttavissa lähestymistavoissa on perustavia samankaltaisuuksia. Molemmissa toimijan perustavin tavoite on hyödyn maksimointi, ja toimija oppii käyttäytymään niin, että hän maksimoi tavoittensa toteutumisen todennäköisyyden elinympäristössään. Edelleen voidaan nähdä, että teorioiden yhdistäminen vaikuttaa sekä mahdolliselta että kiinnostavalta.

Vahvistusoppimissuuntaus ja Hutterin universaalin tekoälyn teoria eivät käsittele affekteja, mutta psykoanalyyttiset teoriat käsittelevät. Ne ovat tosin yleensä jonkinlaisessa sivuroolissa erilaisten häiriötilojen epätoivottavina seurauksina, ja käsittely painottuu luonnollisesti epämieluisiksi koettuihin affekteihin kuten pelkoon, vihaan tai häpeään. Teorioiden yhdistämisellä tavoiteltaisiin psykoanalyyttisten teorioidenkin käsittelemän aihepiirin formalisointia niin, että synteesinä saataisiin tätä aihepiiriä käsittelevä laskennallinen teoria, jonka selitysvoima ja ennustuskyky olisivat olemassa olevia teorioita paremmat.

Edellä mainitut teoriat ovat kokonaisuudessaan erittäin laajoja ja tarpeettoman monimutkaisia tässä tutkielmassa esitettävän lähinnä loogisen tason mallin pohjaksi. Niistä poimitaan sen vuoksi vain muutamia perusajatuksia.

Hyötyfunktio *Hyötyfunktio* on funktio eli kuvaus maailman mahdollisten tilojen joukolta numeeriselle arvojoukolle (ks. esim. [RuN95, s. 44–45]). Se siis yhdistää kunkin tilan sen haluttavuutta kuvaavaan arvoon, järjestäen mahdolliset tilat niiden haluttavuuden mukaan. Luonnollisestikaan sama tila ei voi kuvautua monelle arvolle.

Yleisimmin funktio on sellainen, että useat tilat voivat kuvautua samalle arvolle. Jos taas kukin tila kuvautuu eri arvolle, arvosta voidaan päätellä tila.

Hyötyfunktion tarkka määrittely riippuu siitä, mitä toimijan halutaan tekevän. Hyötyfunktio määritellään siis sellaiseksi, että se tuottaa suuren arvon halutuille tiloille ja pienen arvon ei-toivotuille tiloille. Esimerkiksi luonnollisissa toimijoissa hyötyfunktio tuottaa synnynnäisesti pienen arvon kivulle ja suuren arvon elollisuuden säilymistä tai lisääntymistä edistäville kokemuksille.

Jos toimijan mahdolliset tilat maailmassa ovat esimerkiksi ravittuna olemisen ja nälkää näkemisen tila, niin hyötyfunktio voidaan määrittää niin, että se antaa ensimmäiselle tilalle positiivisen ja toiselle negatiivisen numeroarvon. Hyötyä maksimoivaksi määritelty toimija pyrkii tällöin saavuttamaan ravittuna olemisen tilan.

Funktion arvon muutoksia ajan suhteen tapahtumien vaikutuksesta kutsutaan *dynamiikaksi*. Jos tällainen arvoketju on diskreetti, sitä voidaan käsitellä myös *aikasarjana*. Funktiolle siis syötetään kunakin ajanhetkenä uusi tila, jota uudet tapahtumat ovat muuttaneet, ja tuloksena saadaan tilojen arvotusten sarja.

Affektien määrittelyn kannalta hyötyfunktion tarkka määrittely ei ole tarpeen. Affektimääritelmien tekemiseksi riittää tarkastella hyötyfunktion tuottamia arvoja ja niiden muutoksia. Affektit perustuvat pohjimmiltaan hyötyfunktion tuottaman aikasarjan arvomuutoksien tyypitykseen. Mahdollisia arvomuutostyyppejä on kolme: arvon lasku, arvon nousu tai sen pysyminen samana kahden ajanhetken välillä. Affektit perustuvat näihin arvomuutostyyppeihin.

Hyötyfunktio voi olla määritelty myös implisiittisesti. Toimijalla voi esimerkiksi olla luettelo havainto–reaktio-pareja, ja havaitessaan luetteloidun havainnon toimija suorittaa siihen liittyvän reaktion. Tällaisia sääntöjä kutsutaan *reflekseiksi*. Luetteloitu havainto vastaa jotakin ympäristön tilaa, joka määrittyy toimijan tavoitteen kannalta ei-neutraaliksi, koska toimijan tulee reagoida siihen (jos tila olisi neutraali, siihen ei tarvitsisi reagoida). Koska tila on luetteloitu, se on epätoivottu (alioptimaalinen), ja reaktio pyrkii muuttamaan tilan tavoitteen kannalta paremmaksi. Tällainen luettelokin siis epäsuorasti määrittelee hyötyfunktion, vaikka tilan numeroarvioinnin välivaihe onkin jätetty pois.

2.4 Toimija ohjausjärjestelmänä

Toimija voidaan ajatella niin sanotuksi *ohjausjärjestelmäksi* (engl. control system). Ohjausjärjestelmän käsite on lähtöisin insinööritieteisiin kuuluvasta säätöteoriasta, joka käsittelee automatisoitujen teollisten prosessien ohjausta esimerkiksi tehtaissa. Tietojenkäsittelytieteellisen emootiotutkimuksen puolella ajatuksen mielestä ohjausjärjestelmänä on esittänyt ainakin Aaron Sloman jo vuonna 1993 [Slo93]. Myös Hutter määrittää optimaalisen toimijansa ohjausjärjestelmäksi [Hut04, s. 126–127], ja Damasio puhuu säätelystä elollisuuden ylläpitämisen edellytyksenä [Dam99, s. 134–142].

Ohjausjärjestelmä on kaksiosainen jakautuen ohjaavaan ja ohjattavaan järjestelmään. Ohjausta ei siis voida toteuttaa ilman toimintakykyä, joten tavoitteen saavuttamiseksi toimijalla on oltava toimintakyky, jolla se voi muuttaa ympäristöään. Tekojen aiheuttamien ympäristömuutosten kautta toimija pyrkii lisäämään todennäköisyyttä, että ympäristö tuottaa positiivisia vahvistussyötteitä.

Toisaalta pelkkä hyötyfunktion muutosten luokittelu ei vaadi toimintakykyä. Toimija voi kokea affekteja ilman toimintakykyäkin. Esimerkki tällaisesta toimijasta on vaikkapa toimintakyvyttömäksi vammautunut eläin. Toisaalta affektien kehittyminen toimintakyvyttömälle oliolle evolutiivisesti ei olisi loogista. Toimintakyvytön affektiivinen toimija on siis ”luonnoton” poikkeustapaus. Teoreettisesti sellaisen muodostaminen ei kuitenkaan ole ongelma.

Affektien määrittelyn viitekehyksessä ohjattava järjestelmä tulee ajatella kehoksi. Keho toteuttaa ohjausjärjestelmän määräyksestä tekoja, jotka puolestaan muuttavat ympäristön tilaa. Ohjausjärjestelmä siis muuttaa ympäristön tilaa epäsuorasti kehon välityksellä. Yksinkertaistaen voidaan ajatella, että luonnollisissa toimijoissa ohjaava järjestelmä on hermosto, joka fyysisesti on osa kehoa, mutta loogisesti toimii ohjaavana osana. Kehossa voi olla myös itsenäisiä prosesseja, jotka toimivat ohjausjärjestelmästä riippumatta ja joita ohjausjärjestelmä ei voi ohjata.

Ohjausjärjestelmä voi olla yksi- tai monitasoinen. Ihmisessä ohjausjärjestelmän on esitetty olevan kolmitasoinen [OaJ96, s. 138]. Ensimmäinen taso on osittain tai kokonaan refleksiivinen matelijoilta periytyvä striataalinen taso, jonka reaktiot ovat synnynnäisesti määräytyneet. Tämä taso aiheuttaa ns. refleksit. Toinen taso olisi ns. limbinen taso ja kolmas neokorteksin taso. Alin taso sisältäisi synnynnäisesti kiinnittyneet reaktiot ja kehon tilan perusohjauksen kuten esim. verenkierron. Seuraava taso olisi oppiva taso, joka erottelisi havainnoista objekteja ja muodostaisi eräänlaisen käsiteavaruuden, assosioisi objekteja kehon tilamuutoksiin ja tallettaisi assosiaatiosuhteet muistiin. Samalla se välittäisi opitun perusteella muodostetut ennusteet omasta tulevasta tilasta alemmalle tasolle. Alemman tason reflekseillä olisi siis synnynnäiset ”oletuslaukaisijat”, mutta oppimisen tuloksena ne voisivat muuttua. Ylin taso ylläpitäisi mm. omaelämänkerrallista muistia ja kykenisi esimerkiksi seuraussuhteiden päättelyyn.

2.5 Toimijan ja ympäristön vuorovaikutus

Toimijan ja ympäristön vuorovaikutus on yksinkertaisimmillaan seuraavanlaista (vrt. [RuN95, s. 48]). Maailma koostuu alkutilasta, funktiosta jolla seuraava tila lasketaan, joukosta toimijoita, ja lopetusehdosta. Maailma tuottaa jonkin ympäristön johtamalla (laskemalla) sen alkutilasta annettua funktiota käyttäen.

Ympäristön laskenta aloitetaan käynnistämällä suoritussilmukka. Kullakin kierroksella kukin toimija saa maailman nykyisen tilan havaintona. Tämän jälkeen toimijat laskevat sisäisen tilansa, ohjelmansa ja saamansa havainnon pohjalta tulosteensa eli teon, jonka kukin odottaa maksimoivan hyötynsä. Teot annetaan syötteenä uuden tilan laskevalle funktiolle, jonka tuloksena saadaan maailman seuraava tila. Tämän jälkeen testataan täyttyykö lopetusehto; jos ei, siirrytään uudelle kierrokselle eli silmukan alkuun.

Hutterin optimaalisen toimijan määrittely pohjautuu oletukseen, että toimijalla on kullakin kierroksella rajattomasti aikaa ja muita resursseja laskea tulosteensa. Jos ympäristö on täysin tunnettu eli täysin saavutettavissa ja deterministinen, optimaalisen toiminnan ongelma on triviaali: toimija voi käydä läpi kaikki mahdolliset toimintavaihtoehdot, arvioida niistä jokaisen, valita parhaan ja muodostaa tähän parhaaseen tulokseen johtavan toimintaketjun. Tämän jälkeen toimija vain suorittaa optimaalisen toimintaketjun loppuun asti. Koska toimija voi aina etsiä parhaan toimintaketjun jo ensimmäisellä kierroksella eli saatuaan ensimmäisen syötteen (maailman alkutilan), toimija saavuttaa välttämättä aina parhaan mahdollisen tuloksen ja on siten välttämättä älykkäin mahdollinen toimija [Hut04, Hut03]. Tämä on mahdollista siksi, että toimija tiesi aina varmasti, miten maailma reagoi mihin tahansa toimijan tekoon. Toimija voi siis ”valita” haluamansa tapahtumaketjun eli ympäristön kaikista mahdollisista tulevista ympäristöistä. Tämä ei kuitenkaan tarkoita vapaata valintaa tai jonkinlaista ”tahdon vapautta”, koska toimija on pakotettu valitsemaan parhaan vaihtoehdon. Vapaa valinta edellyttäisi vapautta valita jokin muu kuin paras vaihtoehto.

Määrittelimme kuitenkin luvussa 2.2, että ympäristön tila ei ole kokonaan saavutettavissa, eikä toimijalla ole rajattomasti aikaa ja muita laskentaresursseja. Hutterin teoreettisesti optimaalinen toiminta on tällöin laskennallisesti mahdotonta; optimia voidaan ainoastaan approksimoida.

Kerrataan vielä ympäristön käsite. Luvussa 2.1 määrittelimme ympäristön tietyksi tapahtumaketjuksi. Ajattelemme siis, että toimijan teko vaikuttaa tuleviin tapahtumiin eli siihen, mikä mahdollisista tulevista ympäristöistä toteutuu. Tuleva tapahtumaketju eli maailman tuottama ainutkertainen ympäristö tuotetaan siis maailman ja toimijoiden vuorovaikutuksessa. Koska kuitenkin toimija oli osa maailman määrittelyä, vaihtoehtoisesti voimme ajatella maailman tuottavan tapahtumaketjun pelkästään oman määritelmänsä pohjalta.

Joka tapauksessa toimijan näkökulmasta tarjolla on useita vaihtoehtoisia ympäristöjä. Hutterin toimija voi tunnetussa ympäristössä suoraan valita näistä parhaan. Rajoitettujen resurssien toimija osittain tuntemattomassa ympäristössä voi puolestaan *yrittää valita* haluamansa vaihtoehdon tekemällä käytettävissään olevia tekoja. Toimijan ympäristöä kuvaavan sisäisen mallin ennustuskyvystä riippuu, millä todennäköisyydellä toimijan valinta johtaa haluttuun lopputulokseen eli reagoiko ympäristö tekoon toimijan olettamalla tavalla.

2.6 Hyötykäsitteitä

Affektimääritelmien kannalta keskeisiä käsitteitä ovat toimijan *kokonaishyöty*, *toteutunut kokonaishyöty* ja *tuleva kokonaishyöty*. Määritellään *toteutunut kokonaishyöty* tähänastisen elinajan aikana saatujen vahvistussyötteiden summaksi, *tuleva kokonaishyöty* jäljellä olevan elinajan aikana saatavien vahvistussyötteiden summaksi, ja *kokonaishyöty* toteutuneen ja tulevan kokonaishyödyn summaksi.

Täysin tunnetussa ympäristössä tuleva kokonaishyöty voidaan siis laskea etukäteen. Muussa kuin täysin tunnetussa ympäristössä voidaan käsitellä vain *odotettua kokonaishyötyä* tai *odotettua tulevaa kokonaishyötyä*. Odotettu tuleva kokonaishyöty on siis jäljellä olevan elinajan aikana saatavien vahvistussyötteiden summan odotusarvo. Toisin sanoen tulevat vahvistussyötteet kerrotaan niiden todennäköisyyksillä, ja saadut arvot lasketaan yhteen. Odotettu kokonaishyöty on vastaavasti toteutuneen kokonaishyödyn ja odotetun tulevan kokonaishyödyn summa.

Tällöin kohdataan kuitenkin kaksi ongelmaa: mitkä ovat tulevat vahvistussyötteet, ja mitkä ovat niiden todennäköisyydet? Toimijan kyky ennustaa oma tuleva kokonaishyötynsä riippuu toimijan suhteesta ympäristöönsä. Jos toimijan ympäristö on sille alussa täysin tuntematon eikä yhtään tapahtumaa ole vielä tapahtunut, se ei voi ennustaa tulevaisuutta lainkaan: sillä ei voi olla *odotuksia*.

Määritellään, että odotuksia muodostuu ainoastaan menneiden kokemusten seurauksena. Kaikki tulevaisuusennusteet on siis johdettu aiemmista kokemuksista. Ensimmäisen koetun tapahtuman jälkeen toimija tietää, että kyseisen tyyppisiä tapahtumia voi tapahtua (on olemassa). Määritellään, että kun jonkin tyyppinen tapahtuma tapahtuu ensimmäisen kerran, tapahtuman jälkeen tämän tyypin tapahtumat ovat *odotettuja*; sitä ennen ne olivat*odottamattomia*.

Toimija voi siis odottaa tulevaisuudessa tapahtuvan vain sellaisia tapahtumia, joita hän on nähnyt tapahtuvan, tai voi kokemustensa pohjalta päätellä tapahtuvan. Oletetaan yksinkertaisuuden vuoksi, että toimija ei osaa tällaisia päättelytapoja eikä kommunikoi toisten toimijoiden kanssa. Tällöin toimija voi odottaa ainoastaan tapahtumia, joita on itse aiemmin kokenut.

Oletamme siis, että toimijan tulevaisuusodotusten muodostamisen pohjana käytettävä tieto koostuu luettelosta koettuja tapahtumia ja niiden todennäköisyyksiä (jossa tapahtuman todennäköisyys on sen tapahtumiskertojen määrä jaettuna kaikkien koettujen tapahtumien määrällä). Tällainen toimija siis olettaa tulevaisuuden olevan periaatteessa samanlainen kuin menneisyys, ja tapahtumaluettelo on luvussa 2.2 tarkoitettu toimijan sisäinen malli maailmasta. Samalla affektiivisuus sitoutuu toteutuneeseen kokonaishyötyyn eli menneisyyteen.

Tulevaisuus on toteuduttuaan tietty tapahtumien ketju. Toimija voi itse johtaa mahdolliset tulevat tapahtumaketjut, jolloin niiden todennäköisyydet otetaan aiemmasta tapahtumahistoriasta. Tarkastellaan vain seuraavan tapahtuman ennustamista. Jos toimija on kokenut nolla tapahtumaa, se ei voi ennustaa tulevaisuutta. Jos se on kokenut vain yhden tapahtuman, se voi ennustaa sen toistuvan todennäköisyydellä 1. Jos se on kokenut kaksi tapahtumaa jotka ovat eri tyyppisiä, se voi ennustaa jomman kumman tapahtuman toistuvan seuraavaksi; kummankin todennäköisyys on $0,5$.

Jatkettaessa mahdollisia ketjuja ajassa eteenpäin, yhden koetun tapahtuman tapauksessa mahdollisia tapahtumaketjuja on vain yksi: saman tapahtuman toistuminen äärettömän pitkään todennäköisyydellä yksi. Kahden erityyppisen tapahtuman kokemisen jälkeen mahdollisia tapahtumaketjuja olisivat kaikki näiden kahden tapahtumatyypin kombinaatiot. Esimerkiksi toisistaan riippumattomien tapahtumien a ja b mahdolliset kahden aikayksikön pituiset kombinaatiot ovat aa, ab, ba ja bb. Jos tapahtuma a on tapahtunut aiemmin x kertaa ja tapahtuma b y kertaa, niiden todennäköisyydet ovat $x/(x+y)$ ja $y/(x+y)$. Tapahtuman aa todennäköisyys on siten $(x/(x+y))^2$, tapahtumien ab ja ba todennäköisyydet $(x/(x+y))*(y/(x+y))$ ja tapahtuman bb todennäköisyys $(y/(x+y))^2$. Jos tapahtuman a arvo on hyötyfunktion mukaan 0 ja tapahtuman b arvo 1, niin odotettu tuleva kokonaishyöty seuraavan kahden aikayksikön aikana on $0*(x/(x+y))^2+1*2*(x/(x+y))*(y/(x+y))+2*(y/(x+y))^2$.

Edellä tapahtumia ajateltiin tarkasteltavan toimintakyvyttömän tai muun ulkopuolisen arvioijan näkökulmasta, joka ei itse pyri vaikuttamaan tuleviin tapahtumiin. Tarkastellaan seuraavaksi toimijaa, joka pyrkii vaikuttamaan tuleviin tapahtumiin valitsemalla kullakin hetkellä annetuista vaihtoehdoista suurimman hyödyn tuottavan teon.

Oletetaan yksinkertaisuuden vuoksi, että olosuhteet ovat samat kaikkina ajanhetkinä, ja että seuraavana ajanhetkenä toimija voi valita joko teon A tai B. Edellisen esimerkin perusteella toimijan kannalta paras tapahtuma on b ja paras kahden pituinen tapahtumaketju bb. Oletetaan, että toimija on tehnyt teon A viisi kertaa, ja seuraavana ajanhetkenä tapahtui tapahtuma a todennäköisyydellä $0,6$ ja tapahtuma b todennäköisyydellä $0,4$. Vastaavasti toimija on tehnyt teon B viisi kertaa, ja seuraavana ajanhetkenä tapahtui tapahtuma a todennäköisyydellä $0,2$ ja tapahtuma b todennäköisyydellä $0,8$. Tapahtumat voivat tietysti riippua jostakin muusta kuin toimijan teoista, mutta tämä on paras ennuste, jonka toimija voi tehdä. Teon A odotettu hyöty on siis $0*0,6+1*0,4=0,4$ ja teon B vastaavasti $0*0,2+1*0,8=0,8$. Toimijan mielestä teko B on siis tällä hetkellä arvioituna aina paras valinta.

Toimijan saamat vahvistussyötteet ovat tapahtumien osia, joten tapahtumaketju on toimijan kannalta syötteiden ketju. Omat teot puolestaan ovat toimijan tulosteita. Mahdolliset tapahtumaketjut ovat samat kuin aiemmassa ulkopuolisen tarkkailijan esimerkissä, mutta toimija pyrkii "valitsemaan" tavoitteensa maksimoivan tapahtumaketjun tekemällä tekoja, jotka toimijan kokemuksen mukaan ovat aiemmin tuottaneet halutut tapahtumat.

Nähtiin, että parhaan tapahtumaketjun valinta riippuu pelkästään hyötyfunktiosta. Tapahtumaketjun valinta on siis riippumaton tekojen valinnasta. Sen sijaan parhaan teon valinta riippuu sekä hyötyfunktiosta että aiemmin tehtyjen tekojen koetuista seurauksista. Vaihtoehtoisten tulevaisuuksien paremmuusjärjestys ei ainakaan tässä yksinkertaistetussa mallissa siis riipu omista teoista, mutta teot riippuvat tästä järjestyksestä. Tällöin myös affektiivisuus jää periaatteessa riippumattomaksi omista teoista tai toimintakyvystä; se riippuu vain hyötyfunktiosta.

2.7 Tapahtumat, havainnot ja objektit

Luvussa 2.1 esitetty maailma tuotti siis *tapahtumia*, jotka toimija sai syötteinä. Määritellään *havainto* syötteen synonyymiksi. Määritellään, että toimija voi kullakin ajanhetkellä havaita eli saada syötteenä kaikista samanaikaisista tapahtumista jonkin niiden osajoukon. Nimitetään havaittua osaa *havaintokontekstiksi*.

Luvussa 2.3 syöte jaettiin vahvistusosaan ja vakiosyöteosaan [SuB98, Hut04]. Vahvistussyöte oli toimijalle ympäristöstä kullakin hetkellä tuleva positiivinen, neutraali tai negatiivinen palaute, jonka mukaan toimintaa suunnattiin niin, että tulevan palautteen odotettiin olevan mahdollisimman positiivista. Vakiosyöteosa kattoi toimijalle kullakin hetkellä tulevan muun informaation.

Tehdään ontologinen oletus, jonka mukaan tapahtumat koostuvat osatekijöistä, joita toimija voi oppia erottelemaan toisistaan. Tapahtuman eroteltavissa olevia osatekijöitä voivat olla esimerkiksi aiheuttaja, kohde ja sivullinen. Oletetaan edelleen, että erotetut osat ovat toimijan tavoitteen kannalta eri arvoisia. Esimerkiksi tapahtuman aiheuttaja tuottaa määritelmällisesti koko vahvistussyöteosan, mutta sivullinen ei vaikuta siihen. Sivullinen ei siis ole merkittävä toimijan tavoitteen kannalta, mutta aiheuttaja on.

Tulkitaan, että vakiosyöteosa sisältää havaintokontekstissa kyseisellä hetkellä olevan kokonaisuuden havaittavissa olevat piirteet, joiden perusteella kokonaisuus hajotetaan osikseen. Vakiosyöteosa voidaan ajatella vaikkapa näköhavaintona, josta aiheuttaja, kohde ja sivullinen ovat erotettavissa.

Vahvistusosa tulkitaan vastaavasti koko syötteen aiheuttamaksi kokonaishyödyn muutokseksi. Vahvistusosakin voidaan yrittää osittaa tapahtuman osatekijöille kuten aiheuttajalle, kohteelle ja sivulliselle[5]. Kun vahvistussyötteen arvo ositetaan aiheuttajalle, voidaan reagoida sellaiseenkin tapahtumaan, jossa esiintyy pelkästään aiheuttaja, eikä esimerkiksi aiheuttajan ja satunnaisen sivullisen kombinaatio. Jos vahvistussyötettä ei ositeta, reagoitaisiin ainoastaan kombinaatioon, mikä johtaisi huonoon ennustuskykyyn. Oletuksena siis on, että vahvistussyötteen kohdistaminen tuottaa aina paremman ennustuskyvyn kuin sen jättäminen jakamatta.

Määritellään *objekti* tarkoittamaan toimijan ohjausjärjestelmän sisäisiä esityksiä havaintokontekstissa esiintyneistä osatekijöistä. Esimerkiksi tapahtuman aiheuttaja, kohde ja sivullinen ovat näin määriteltynä objekteja. Vakiosyöteosa on siis *epätyhjä objektien joukko*. Jos havaintoa ei osata osittaa, objektiksi jää koko havainto. Riippuen toimijan oppimista kategorioista havainto voidaan jakaa enintään niin moneksi osaksi kuin siinä on erotettavia piirteitä. Esimerkiksi tapahtumaa, jonka vakiosyöteosa sisältää yhden sivullisen, ei voida esittää niin että esitys sisältäisi enemmän kuin

[5]Luvun 4 ohjelmatoteutuksessa havaintojen oletetaan olevan valmiiksi ositettuja objekteiksi ja niihin liittyviksi vahvistussyötteiksi.

yksi sivullista. Toimijan tulee oppia objektikategoriat automaattisesti esimerkiksi Vogtin esittämällä tavalla [Vog03a, Vog03b]. Toimija siis muodostaa havainnoistaan sisäisen mallin: käsiteavaruuden.

Luvussa 1.4 esitellyssä OCC-mallissa tehtiin erottelu objekteihin, toimijoihin ja tapahtumiin. Objektilla tarkoitettiin elotonta esinettä. Objektisuhdeteorioissa objekti puolestaan tarkoittaa lähinnä toimijan mielessä olevaa mielikuvaa havaitusta toimijasta tai esineestä, johon toimija on kohdistanut odotuksia eli muodostanut niihin ns. objektisuhteen. Jos oletetaan, että OCC-mallissakin objektilla viitataan toimijan mielikuvaan esineestä eikä todelliseen esineeseen, niin OCC-mallin objektiluokka on siis objektisuhdeteorioiden objektiluokan aliluokka. Tässä tutkielmassa omaksuttu objektikäsite vastaa objektisuhdeteorioiden objektikäsitettä.

2.8 Teot ja motivaatio

Toimijalla ajateltiin olevan käytettävissään tietty joukko *tekoja*, joilla se voi pyrkiä maksimoimaan tavoitteensa asettamaa hyötyä. Kuten luvussa 2.6 todettiin, toimintakyvytön toimija olisi jo käsitteenäkin hieman ristiriitainen: se ei voisi tehdä mitään maksimoidakseen hyötyfunktiotaan, johon sen elollisuuden säilyminen perustuisi. Toimintakyvyn puute ei periaatteessa estä affektiivisuutta, mutta fysiologisten reaktoiden olemassaololle tai kehittymiselle ei löydy motivaatiota.

Määritellään *tekojen* luokka tapahtumien luokan aliluokaksi. Eräs tapahtuman perustava piirre on sen aiheuttaja. Aiheuttajan olemassaolon ajatus perustuu ontologiseen oletukseen deterministisyydestä (ks. luku 2.1). Jos tapahtumalle voidaan erottaa aiheuttaja joka on toimija, tapahtuma voidaan luokitella teoksi.

Toimijan *motivaatio* tehdä tietty teko määritellään tekoon liittyväksi odotetun tulevan kokonaishyödyn muutokseksi. Määritelmästä seuraa yksinkertainen ja selkeä motivaatioteoria: toimija tekee kullakin hetkellä sen teon, josta odotetaan seuraavan suurin tulevan kokonaishyödyn lisäys. Muita tekoja ei tehdä. Tekemättä jättäminen (lepo) on ehkä selkeintä ajatella yhdeksi vaihtoehtoiseksi teoksi. Jos toimija ei tee mitään, voidaan päätellä, että tekemättä jättämisen odotettu hyöty on suurempi kuin varsinaisten tekojen.

2.9 Assosiaatiomekanismi

Assosiaatiomekanismi liittää yhteen objektit ja niiden yhteydessä havaittujen (eli samassa syötteessä olleiden) vahvistussyötteiden arvosta objektin osalle määritetyt osat. Assosiointi edellyttää sisäistä muistia.

Assosiaatiomekanismi voidaan toteuttaa yksinkertaisimmillaan esimerkiksi järjestettyjä pareja sisältävänä listana, jossa kunkin parin ensimmäinen jäsen on havaittu objekti ja toinen jäsen kyseiseen objektiin liitettyjen vahvistussyötteiden lista.

Toimija siis luokittelee havainnon objekteiksi, osittaa vahvistussyötteen objekteille ja yhdistää uudet vahvistussyötteet objektien malleihin.

Määritellään esimerkkifunktioksi seuraava: objektin arvo on sille ositettujen vahvistussyötteiden osien arvojen keskiarvo. Funktio säilyttää positiivinen–negatiivinen-rajan ja ilmaisee seuraavan objektin aiheuttaman tapahtuman hyödyn odotusarvon.

Tulkitaan alustavasti vahvistussyötteen assosioinnin objektiin tarkoittavan samaa kuin psykoanalyyttisissä teorioissa käytetty ”psyykkisen energian” *katektointi* objektiin.

2.10 Persoonallisuus ja temperamentti

Edellä esitetyn perusteella toimijalla on luettelo havaituista objekteista niihin assosioituine hyötytasomuutosodotuksineen. Jos kaksi toimijaa ovat fysiologisesti identtisiksi määritellyt, kyseinen luettelo on toimijan ainoa muuttuva rakenne. Määritellään toimijan *persoonallisuus* tarkoittamaan tämän opitun sisällön niitä seurauksia, jotka ilmenevät toimijan käyttäytymisessä.

Näin määriteltynä persoonallisuus on opittu ja dynaaminen. Persoonallisuuden dynaamisuus vähenee havaintojen määrän kasvaessa. Tämän johtuu luvussa 2.9 määritellystä oppimismekanismista. Objektin arvon määrittäminen vahvistussyöteosien keskiarvoksi johtaa siihen, että mitä enemmän aiempia havaintoja on, sitä enemmän aiemmista poikkeavia havaintoja tarvitaan nykyisen arvon muuttamiseksi.

Haluttaessa toimijan käyttäytymiseen vaikuttavat fysiologiset erot toisiin toimijoihin nähden voidaan määritellä toimijan *temperamentiksi*. Tämä ei ole ristiriidassa edeltävän persoonallisuuden määritelmän kanssa, vaan täydentää sitä. Tällainen fysiologinen ero voi olla esimerkiksi ero assosiaatiofunktiossa. Toinen toimija voi esimerkiksi ottaa huomioon kaikki havainnot ja toinen vain tietyn määrän uusimpia havaintoja, tai havaintoja voidaan painottaa eri tavoin. Tällä tavalla voidaan kuvata esimerkiksi erilaisia muistamisen tapoja.

Edellä esitetty on yhteensopiva sen näkemyksen kanssa, että psykiatriassa käsiteltävät ns. *persoonallisuushäiriöt* (ks. esim. [Liv03, Täh93]) ovat opittuja. Kuten koko persoonallisuus, myös häiriöt ovat samasta syystä melko staattisia, mutta rakenteellista estettä niiden ilmaantumiselle myöhemmin tai poistumiselle ei ole olemassa.

Toimijalla voi olla joukko objekteja, joihin on assosioitunut pääosa odotetusta hyödystä. Tämän joukon voidaan ajatella muodostavan toimijan *identiteetin* perustan.

2.11 Elollisuus ja tilaehto

Luonnollisilla toimijoilla kuten ihmisillä ohjattava järjestelmä eli keho on elollinen. Elollisuus asettaa kehon tilalle tietyt rajat, joiden ulkopuolelle joutuminen johtaa toimijan kuolemaan. Nimitetään tällaisia rajoja yleisesti *tilaehdoksi*. Tilaehto on siis ohjattavan järjestelmän ominaisuus. Rajat voidaan ajatella kiinteiksi: tiettyjen raja-arvojen ylitys johtaa kuolemaan.

Toinen asia on, miten ohjattavan järjestelmän tilaehto esitetään ohjausjärjestelmässä. Ohjausjärjestelmä voi joko tietää sallitut rajat ennalta, tai se voi oppia ne. Oppimistapauksessa ohjattavan järjestelmän tuottamien vahvistussyötteiden arvo voi esimerkiksi lähestyä arvoa $-\infty$, kun ohjattavan järjestelmän tila lähestyy ehdon asettamaa rajaa. Tästä ohjausjärjestelmä voi päätellä rajan olemassaolon.

Tavallaan koko ohjausjärjestelmän tietosisältö kuvaa tilaehtoa eli sitä, että toteutuneen kokonaishyödyn tulee pysyä jonkin vakiotason yläpuolella. Toimijan tavoitteena oli hyödyn maksimointi, joka voidaan ajatella käänteisesti haitan minimointina. Toimijan tavoite on pyrkiä minimoimaan haittaa, jota keho tuottaa elollisuuden vähenemisenä ja jonka keho esittää ohjausjärjestelmälle negatiivisina vahvistussyötteinä. Hyöty ei tavanomaisissa tapauksissa maksimoidu kuolemassa, joten tavoitteen toteuttamiseksi toimija pyrkii ylläpitämään elollisuutta. Vahvistussyötteet tuottavat funktiot ovat fysiologisen järjestelmän rakenteen sisältämää tietoa; ohjausjärjestelmä voi ainoastaan yrittää mallintaa niitä havaitsemiensa vahvistussyötteiden perusteella samalla tavalla kuin ulkoista ympäristöäänkin.

2.11.1 Kuoleman suhde hyödyn maksimointiin

Luvussa 2.11 esitettiin määriteltäväksi tilaehto, jonka määrittämien rajojen ylittäminen johtaa toimijan kuolemaan. Vaihtoehtoisesti kuolema voidaan keinotekoisissa järjestelmissä määritellä tapahtuvaksi tiettynä ajanhetkenä toimijan teoista tai tilasta riippumatta.

Hutter huomauttaa, että järjestelmissä, joissa toimija joutuu opettemaan elollisuutensa tai toimintakykynsä ylläpitämisen, toimija usein kuolee ennen kuin oppii huolehtimaan itsestään [Hut04, s. 238]. Toimijan säilyminen elossa edellyttää toisten toimijoiden puuttumista asiaan. Tällainen kuolema on tahaton kuolema, joka ei ole jonkin oman teon odottamaton seuraus, vaan ennemminkin toimijan kyvyttömyyden seuraus. Toimija on maksimoinut hyötyään, mutta sillä ei ole ollut käytettävissään riittävän tehokkaita tekoja.

Tahaton kuolema oman teon odottamattomana seurauksena (vaikkapa kuolema kaahailun aiheuttamassa auto-onnettomuudessa) on puolestaan ohjausjärjestelmän kannalta odotettua pienempi vahvistussyöte tehtyyn tekoon eli heikon ennustuskyvyn seuraus. Kuten edellä, tämäkään tapaus ei ole ristiriidassa hyödyn maksimoinnin ajatuksen kanssa.

Toimijan tavoitteen tai pyrkimyksen ("tahdon") vastaisen kuoleman lisäksi on olemassa tarkoituksellisen kuoleman eli itsemurhan mahdollisuus. Tällöin toimija tekee teon, joka johtaa tilaehdon asettamien rajojen ylittymiseen ja siten kuolemaan. Teon tarkoituksellisuus edellyttää vähintään tietoa kyvystä aiheuttaa oma kuolema. Yksinkertaisimmillaan riittää, että kuolemaan johtava teko luetellaan mahdollisten tekojen listassa.

Tosiasiallisesti kuoleman tuleva kokonaishyöty on nolla, koska vahvistussyötteiden tulo päättyy. *Odotettua kokonaishyötyä* maksimoivalle toimijalle kuolema olisi optimaalinen vaihtoehto, kun odotettu tuleva kokonaishyöty on pienempi kuin nolla eli kun kokonaishyödyn odotetaan laskevan nykyhetkeen nähden. Tällaisen toimijan kannattaisi siis jatkaa elämäänsä ainoastaan, jos odotettu tuleva kokonaishyöty on positiivinen. Tällainen tilanne saadaan kuitenkin aikaan jo niin, että toimijan ensimmäiseksi vahvistussyötteeksi annetaan negatiivinen arvo. Toimijaa ei siten ole järkevää määritellä maksimoimaan odotettua kokonaishyötyä, vaan odotettua tulevaa kokonaishyötyä.

Toimija tekee kullakin hetkellä aina suurimman odotetun hyödyn tuottavan teon, ja kuolemistekoon voi assosioitua mikä tahansa odotettu kokonaishyöty. Esimerkiksi itsemurhapommittajilla kuoleman ajatukseen voi ilmeisesti liittyä odotus oman hyötytason voimakkaasta noususta kuoleman jälkeen, jolloin se voi olla paras teko, vaikka vaihtoehtoistenkin tekojen odotetut hyödyt olisivat positiivisia. Jos itsemurha tehdään, sillä kuitenkin on suurempi odotettu hyöty kuin esimerkiksi sillä, ettei tehdä mitään.

Odotettua tulevaa kokonaishyötyä maksimoiva toimija pysyy siis hengissä, vaikka toteutunut kokonaishyöty jatkuvasti laskisi, edellyttäen että pysytään tilaehdon asettaman rajan yläpuolella. Toimija siis maksimoi elollisuutensa kestoa, mikä onkin evolutiivisesti järkevämpi strategia.

Olemme aiemmin implisiittisesti olettaneet, että odotetun tulevan kokonaishyödyn maksimoiminen maksimoi myös elollisuuden kestoa. Tahallisen itse aiheutetun kuoleman tapauksessa nämä tavoitteet kuitenkin eroavat: odotettu hyöty maksimoituu elollisuuden keston kustannuksella. Samanlainen tilanne esiintyy ilmeisesti myös ainakin silloin, kun vanhempi uhrautuu jälkeläistensä puolesta. Asia jää tässä vaiheessa hieman epäselväksi.

2.12 Normit

Määritellään *normi* (engl. norm, standard) alustavsti säännöksi siitä, minkälaisen hyötytasomuutoksen itselle tai toiselle saa aiheuttaa. Normi määrää mikä on *sallittavaa*: ne teot, joista saatu palaute on positiivista. Sallittavuuden käsite viittaa siis toisten objektien antamaan palautteeseen; normit syntyvät teon tekemisen yhteydessä saadusta palautteesta. Normit perustuvat viime kädessä tilaehtoon eli elollisuuden säilymisen vaatimukseen. Normin noudattamatta jättämiseen liittyy odotus hyötytason laskusta. Normi on siis motivaation synonyymi.

Teon tuottama vahvistussyöte voi olla joko kehon sisäinen tai ulkoisen objektin aiheuttama. Sisäiseen palautteeseen suoraan pohjautuva normi on, että on hyvä aiheuttaa positiivisia kokemuksia ja välttää negatiivisia kokemuksia itselle. Muut normit opitaan ulkoisten objektien antaman palautteen seurauksena. Luonnolliset toimijat oppivat yleensä, että toiselle ei saa aiheuttaa haittaa. Tämä johtuu siitä, että toiset reagoivat niille aiheutettuun haittaan antamalla negatiivisia vahvistussignaaleja. Toimija voi oppia minkä tahansa normijoukon eli tekoihin liittyvät hyötytasomuutosodotukset sen mukaan, millaista palautetta ympäristö antaa. Normit perustuvat siis osittain synnynnäisiin (tilaehto) ja osittain opittuihin (ympäristön tuottama palaute tietyssä ympäristössä) tekijöihin.

Kun toimijalla sanotaan olevan *arvoja* tarkoitetaan oikeastaan samaa kuin sanottaessa toimijalla olevan normeja; ero on lähinnä näkökulmassa. Kun normit asettavat sallittavuuden, niin arvot asettavat puolestaan tavoiteltavuuden ja liittyvät siten hyötyfunktioon.

Määritellään objektin *arvo* sen osuudeksi jonkin tavoitteen kannalta mitatusta odotetusta hyödystä. Jos kyseinen tavoite on toimijan ainoa tavoite, objektin arvo on synonyymi sen kokonaishyödylle, ja toimijan kokonaishyöty on sen tuntemien objektien arvojen summa. Jos mallia laajennetaan mahdollistamaan osatavoitteiden johtaminen tavoitteesta, objektien arvoja voidaan arvioida kunkin osatavoitteen kannalta erikseen. Tällöin jonkin osatavoitteen kannalta merkityksetön objekti voi olla merkityksellinen toisen osatavoitteen kannalta, ja objekteihin voidaan suhtautua *ambivalentisti* eli ristiriitaisesti. Toimijalla on tällöin *sisäinen ristiriita* objektin suhteen.

2.13 Hyödyn dynamiikka

Hyötyfunktion arvon muutoksia ajassa voi aiheutua ainakin uusien tapahtumien (syötteiden) vastaanottamisesta, vaihtoehtoisten tekojen sisäisen arvioinnin seurauksena (luku 2.13.1), ja tarkkaavaisuuden kohdistumisen seurauksena (luku 2.13.2), jos toimijan rakenne määritellään tässä aliluvussa esitettävällä tavalla.

Uusien havaittujen objektien hyötytasomuutos on periaatteessa määrittelemätön ja siten itseisarvoltaan potentiaalisesti ääretön. Ne olisi optimaalisessa tapauksessa siksi aina käsiteltävä ja niiden hyötytasomuutosodotus määritettävä (toisaalta toimija voi muodostaa odotuksen odottamattomien tapahtumien keskimääräisestä hyödystä jossakin ympäristössä, mikä helpottaa ärsykkeiden rajaamista ja esimerkiksi käsillä olevaan työhön keskittymistä).

Tämän jälkeen toimijan kannattaa suuntautua siihen objektiin, jonka hyötytasomuutosodotus on itseisarvoltaan suurin. Tämä perustuu yksinkertaistavaan oletukseen siitä, että toimijalla on vain yksi tavoite, jonka kannalta kaikki vahvistussyötteet ovat laadullisesti samanarvoisia. Tällöin itseisarvoltaan suurempi vahvistussyöte kumoaa itseisarvoltaan pienemmän mutta merkiltään vastakkaisen vahvistussyötteen; toimijan ei esimerkiksi tarvitse reagoida negatiiviseen odotukseen, jos on olemassa suurempi positiivinen odotus, jonka toimija voi toteuttaa, koska suuremman odotuksen toteuttaminen nostaa kokonaishyötyä enemmän kuin negatiivisen odotuksen toteutumisen kumoaminen. Monen tavoitteen mallissa tämä periaate pätee edelleen yksittäisiin tavoitteisiin.

2.13.1 Vaihtoehtoisiin tekoihin liittyvä dynamiikka

Kappaleessa 2.6 tuleva kokonaishyöty määriteltiin tulevien vahvistussyötteiden summaksi. Sitä estimoitiin tähän asti saatujen syötteiden perusteella, ja tätä estimaattia nimitettiin odotetuksi tulevaksi kokonaishyödyksi. Oletetaan, että toimija ei pyri valitsemaan tekoja, vaan on lähinnä tapahtumien kohteena (se on siis luvussa 2.6 mainittu teoreettinen konstruktio, ”toimintakyvytön” toimija). Tällöin toimijan tarvitsee laskea kokonaishyötynsä vain kerran kierroksen aikana uusien havaintojen saamisen jälkeen. Kokonaishyöty ei siis voi muuttua muuten kuin uusien havaintojen seurauksena, eikä yhden kierroksen sisälle muodostu affektidynamiikkaa; affektit muuttuvat vain ajanhetkien väleissä eli uudelle kierrokselle siirryttäessä.

Kierroksen sisäinen affektidynamiikka eli hyötyfunktion arvojen dynamiikka liittyykin tekovaihtoehtojen arviointiin. Oletetaan, että toimija arvioi tekovaihtoehdot yhden kierroksen aikana, ja ohjattavan järjestelmän tila asetetaan välittömästi vastaamaan parasta laskettua arviota. Tällöin saadaan aikaan luonnollisilla toimijoilla ilmenevä tilanne, jossa tilanteen ”ajattelu” saa aikaan eri affekteja riippuen ajattelun sen hetkisestä tilasta (välituloksesta). Esimerkiksi vaaran havaitseminen aiheuttaa ensin pelon. Jos toimija keksii, miten vaaran voi välttää tekemällä jonkin teon, pelko poistuu jo *ennen teon tekemistä*, ja toimija voi esimerkiksi kokea ylpeyttä taitavuudestaan vaaratilanteen välttämisessä.

Kun uhka havaitaan, sen odotettu (negatiivinen) hyötymuutos kerrottuna toteutumistodennäköisyydellä lisätään odotettuun kokonaishyötyyn. Vastaavasti vaikka toimija ei ole vielä tehnyt uhan torjuvaa tekoa, tämän teon odotettu hyöty kerrottuna sen onnistumistodennäköisyydellä lisätään sen tilan hyötyyn, jossa uhka toteutuisi. Tämä voidaan nähdä eräänlaisena odotetun tulevan tilan *diskonttauksena* nykyhetkeen.

Dynamiikka ei edellytä itsetarkkailua (monitasoisen ohjausjärjestelmän olemassaoloa).

2.13.2 Tarkkaavaisuuteen liittyvä dynamiikka

Määritellään, että ohjausjärjestelmän prosessoinnin kohteena tietyllä hetkellä oleva objektimalli on sillä hetkellä *tarkkaavaisuuden* kohteena. Jos toimijan odotettu kokonaishyöty lasketaan kaikkien tunnettujen objektien sijaan vain niiden osajoukon perusteella, niin saadaan aikaan tekovaihtoehtoihin liittyvän dynamiikan sijaan tarkkaavaisuuteen liittyvää syklien sisäistä dynamiikkaa. Objektien määrän kasvaessa ja tiedonkäsittelykapasiteetin (esimerkiksi työmuistin koko tai laskennan nopeus) pysyessä vakiona tällainen oletus näyttää välttämättömältä. Nimitetään tällaista objektien osajoukkoa *sisäiseksi objektikontekstiksi*. Osajoukon valintaperuste voi olla esimerkiksi jokin läheisyysmitta, jolla kontekstiin valitaan tarkkaavaisuuden kohteena olevaa objektimallia ”lähellä” olevat objektimallit.

Läheisyysmitaksi voidaan valita esimerkiksi objektien ajallinen läheisyys havaintohistoriassa. Tällä tavalla voidaan ehkä hahmottaa alkeellista kausaalisuutta. Esimerkiksi toistuvasti samassa tapahtumassa tai kahdessa peräkkäisessä tapahtumassa esiintyvät objektit saattavat olla kausaalisuhteessa todennäköisemmin kuin kaksi mielivaltaisesti valittua objektia.

Määritellään että kontekstin objektien vaihtuessa tapahtuu *kontekstivaihdos*. Kontekstivaihdos on tarkkaavaisuuden siirtymiseen liittyvä mekaaninen toimenpide: konteksti vaihtuu periaatteessa aina tarkkaavaisuuden siirtyessä. Jos kokonaishyöty lasketaan kontekstin objekteista tai niitä painottaen, niin kontekstivaihdos muuttaa odotettua kokonaishyötyä. Toisin sanoen tarkkaavaisuuden siirtyminen luo hyötyfunktiodynamiikkaa.

Sisäinen objektikonteksti oli siis eri asia kuin havaintokonteksti, johon kuuluivat jollakin hetkellä havaittavissa olevat ulkoiset objektit.

2.14 Puolustusmekanismit

Puolustusmekanismien eli *defenssien* käsite on keskeinen psykoanalyyttisissa teorioissa (ks. esim. [Täh72, s. 27–36], [Täh93]). Niissä puolustusmekanismin käyttö käsitettäneen yleensä aktiiviseksi sisäiseksi toimenpiteeksi eli jonkinlaiseksi teoksi. Puolustusmekanismin käytön tarkoitus on nostaa toimijan odotettua hyötyä.

Tässä tutkielmassa esitetyssä mallissa ainoat keinot tyydytystason nostoon olivat uuden vahvistussyötteen saaminen (ulkoinen tapahtuma), teon valinnan aiheuttama odotetun hyödyn lisäys (tulevan teon odotetun hyödyn diskonttaus nykyhetkeen; sisäinen muutos) tai kontekstivaihdos tarkkaavaisuuden siirtymisen seurauksena (sisäinen muutos). Koska teko oli ajateltu ulkoiseen tilaan vaikuttavaksi ja se tuotti uuden syötteen aikaisintaan seuraavana ajanhetkenä, ei näytä luontevalta määritellä puolustusmekanismeja teoiksi.

Sen sijaan puolustusmekanismit tai osa niistä voitaisiin ehkä määritellä kontekstivaihdoksina. Edellä määriteltiin, että kun arvioimattomia objekteja ei ole, toimijan tarkkaavaisuus suuntautuu automaattisesti aina siihen objektiin, jonka hyötytasomuutosodotus on itseisarvoltaan suurin. Määritellään *torjunta* (engl. repression) tiedostamattomaksi tarkkaavaisuuden suuntautumiseksi hyötytason laskun aiheuttaneesta objektista toiseen objektiin, johon liittyy odotus hyötytason noususta.

Samaistetaan hieman yksinkertaistaen ”tietoisuus jostakin” ja tarkkaavaisuuden kohteena oleminen (vrt. [Baa97]). Jos tällöin oletamme ohjausjärjestelmän yksitasoiseksi, niin käsittelyn eli tarkkaavaisuuden kohteena on aina jokin objektin malli, eikä alemman tason ohjausjärjestelmän tila. *Käsittelyn kohteen vaihtumista* eli torjuntaa itseään ei voida ottaa tarkkaavaisuuden kohteeksi kuin kaksitasoisessa ohjausjärjestelmässä, jossa toinen taso voi havaita alemman tason ohjausjärjestelmän tilan. Näin ollen yksitasoisessa järjestelmässä torjunta ei voi olla tietoista. Sen sijaan ”torjuttu objekti” siirtyy tietoisesta esitietoiseksi. Jos positiivisen hyötytasomuutosodotuksen omaavia objekteja ei ole, torjunta ei ole mahdollista.

Muut puolustusmekanismit tulisi voida osoittaa torjunnan erikoistapauksiksi. Tämän osoittaminen tai niiden määrittely muulla tavalla jätetään tutkielman ulkopuolelle.

3 Affektit, emootiot ja tunteet

Tässä luvussa esitetään tyypillisimpien affektien alustava luokitus.

3.1 Affektit, tietoisuus ja tarkkaavaisuus

Matthis on määritellyt affektin yläluokaksi, jonka alaluokkia tunteet (engl. feelings) ja emootiot (engl. emotions) ovat [Mat00, s. 217]. Tunteet määritellään "tietoisiksi affektiivisiksi ilmiöiksi" ja emootiot "esitietoisiksi affektiivisiksi ilmiöiksi".

Tietoisuuden tasojen luokittelu on alun perin lähtöisin Sigmund Freudilta, jonka topografinen teoria vuodelta 1915 jakoi mielen tietoiseen, esitietoiseen ja tiedostamattomaan osaan [Sli91, s. 30][Täh72, s. 4–5]. Esitietoista on aines, joka on palautettavissa tietoisuuteen. Tiedostamatonta on aines, joka ei ole palautettavissa tietoisuuteen. Sitä voidaan pyrkiä tavoittamaan muun muassa oireiden eli käyttäytymisen havainnoinnin kautta.

Esimerkiksi Davisin mukaan "Freud's way of talking about 'the conscious' is similar to what a cognitive psychologist means by attention" [Dav04]. Samoin Baarsin mukaan tarkkaavaisuudella voidaan viitata tietoisten sisältöjen valintaan ja ylläpitämiseen[6]. Tietoisuutta kokonaisuudessaan ei kuitenkaan voi samaistaa tarkkaavaisuuteen [Baa97].

Määritellään siis alustavasti ja yksinkertaistaen, että tietoista on se, mikä on ohjausjärjestelmän havaitsemisen eli *tarkkaavaisuuden* kohteena jollakin hetkellä. Esitietoista on se, mikä on periaatteessa otettavissa havaitsemisen eli tarkkaavaisuuden kohteeksi.

Tietoisuuden ongelma siis kierretään redusoimalla tietoisuus jostakin tarkkaavaisuuden kohteena olemiseksi. Jos tämä määritelmä hyväksytään, yksitasoisen ohjausjärjestelmän omaava toimija voi olla tietoinen ympäristönsä objekteista ja omasta tilastaan. Itsetietoisuutta tai tietoisuutta tilansa muutoksista tämä määritelmä ei kuitenkaan sisällä. Niiden tuottaminen edellyttäisi sellaisten kuvausten muodostamista, jossa kahden eri ajanhetkien tilat muistettaisiin ja liitettäisiin itsen käsitteeseen. Tällaisen voidaan katsoa edellyttävän sellaista ylemmän tason ohjausjärjestelmää, joka

[6] Esimerkiksi "the term 'attention' may be best applied to the selection and maintenance of conscious contents and distinguished from consciousness itself. This is consistent with common usage." tai "In working memory, there is a long history of separating the 'active element', which also turns out to be the conscious element".

tarkkailee ja merkitsee muistiin alemman tason tiloja eri ajanhetkinä (vrt. [Dam99]).

Jos affekteihin ajatellaan olennaisesti liittyvän kehollisia muutoksia, affektien olemassaolo edellyttää *kehon* olemassaoloa ainakin abstraktilla tasolla. Vähimmäisvaatimus on keho, joka sisältää yhden muuttujan, joka voi saada kaksi eri arvoa. Lisäksi muuttujan arvon muutoksen tulee seurata vakioisesti jostakin odotetun hyödyn muutostilanteesta, ja tämän arvon muutoksen tulee vaikuttaa toimijan kelpoisuuteen (elin- ja lisääntymiskelpoisuuteen) ympäristössään.

Ajan kuluessa toimijan odotettu hyöty voi nousta, pysyä samana tai laskea. Odotetun hyödyn pysyminen samana rajautuu affektin käsitteen ulkopuolelle, jos siihen ei liity kehollista muutosta. Tässä tutkielmassa hämmästykseen kuitenkin ajatellaan liittyvän kehollisia muutoksia (ainakin tyypillinen ilme) ja se lasketaan siten affektiksi.

Emootio määritellään tyypillisesti prosessiksi, mutta prosessin yksityiskohdista ja niiden painotuksista ei vallitse yksimielisyyttä (ks. esim. [OaJ96, s. 95–124]). Sen vuoksi seuraavassa esitetty määritelmä poikkeaa hieman esimerkiksi edellä mainitussa lähteessä esitellyistä; vertailut kuitenkin sivuutetaan niiden monimutkaisuuden vuoksi.

Määritellään siis *affekti* prosessiksi, jossa odotetun tulevan kokonaishyödyn muutos (ohjausjärjestelmän tilamuutos) aiheuttaa vakioisen kehollisen muutoksen (ohjattavan järjestelmän tilamuutoksen). Odotetun hyödyn muutos on tyypillisesti ulkoisen tapahtuman aiheuttama, ja kehollisen muutoksen tarkoitus on parantaa toimijan mahdollisuuksia reagoida tapahtumaan. Kehon tilan muutos on affektin mahdollisesti ulkoisesti havaittavissa oleva ilmentymä, jolla voi olla kommunikatiivinen tarkoitus.

Määritellään *emootio* affektiksi, joka on periaatteessa havaittavissa eli otettavissa tarkkaavaisuuden kohteeksi. Määritellään *tunne* affektiksi, joka on tarkasteluhetkellä tarkkaavaisuuden kohteena.

Määritellään vastaavasti affektiluokan kolmanneksi alaluokaksi affektit, jotka eivät ole havaittavissa eivätkä siten otettavissa tarkkaavaisuuden kohteeksi eli "tietoisuuteen". Tietoisella toimijalla tällaisia ovat affektit, jotka ovat ohjausjärjestelmän aiheuttamia, mutta joiden kehollisen muutokset eivät ole havaittavissa (ainakaan ilman erityistoimenpiteitä). Toisaalta toimijoilla, jotka eivät kykene tietoisuuteen, kaikki affektit ovat tällaisia affekteja.

Affekti määriteltiin siis prosessiksi, mutta näkökulmaa siihen voidaan painottaa eri tavoin. Esimerkiksi lopputilaa painotettaessa affekti voidaan ajatella enemmänkin toimijan tilana (esimerkiksi jostakin pitäminen) kuin tilamuutoksena, jos pitäminen on stabiili tila ja sen aiheuttaneet tapahtumat voivat olla kaukana menneisyydessä. Toisaalta esimerkiksi äkillinen vihanpurkaus näyttäytyy paremmin ohimenevänä tilamuutoksena.

Koska affekti määriteltiin prosessiksi, niin ollakseen emootio tai tunne tarkkaavaisuuden kohteeksi pitäisi voida ottaa affektiprosessin esitys (representaatio). Toimijan pitäisi siis muodostaa esityksiä, joissa ilmenee toimijan oma tila kahtena eri ajanhetkenä. Tämä edellyttää, että ohjausjärjestelmässä olisi ikään kuin toinen taso, joka tarkkailee alemman tason ohjausjärjestelmän tilaa. Tällaista ohjausjärjestelmän sisäisen introspektion mahdollisuutta ei kuitenkaan määritelty. Nähtävästi siis toimija, jolla on vain yksitasoinen ohjausjärjestelmä, ei voi olla varsinaisesti emotionaalinen tai tunteellinen, vaan pelkästään affektiivinen. Se on tietoinen kehonsa tilasta, mutta ei ohjausjärjestelmänsä tilasta.

3.2 Affektien ja objektien suhde

Hyötyfunktioon liittyvän aspektin ja kehollisen aspektin lisäksi affekteihin liittyy kolmas aspekti: objektit. OCC-mallissa objekteilla tarkoitettiin elottomia objekteja ja ne erotettiin tapahtumista ja toimijoista (luku 1.4). Objektisuhdeteorioissa tätä erottelua objekteihin, toimijoihin ja tapahtumiin ei tehdä, vaan näitä kaikkia nimitetään objekteiksi. Tämä johtuu siitä, että toimijan kaikki havainnot ja mielikuvat ajatellaan objekteiksi. Objekteja voidaan luokitella eri tavoin, mm. edellä mainittuihin objekti-, toimija- ja tapahtumakategorioihin. Kyseessä on siis ero objektin käsitteen määrittelyssä: OCC-mallin objektiluokka ajatellaan objektisuhdeteorioissa objekti-luokan alaluokaksi. Tässä tutkielmassa käytetään objektisuhdeteorioiden objektikäsitettä.

Toimija muodostaa objekteihin suhteita, ns. *objektisuhteita*. Objektisuhteet ovat käytännössä assosiaatioita, joissa objektin malliin liitetään sen arvotus eli tulevan hyödyn odotusarvo (ks. luku 2.9).

Affektit määrittyvät odotuksen merkin eli suunnan (positiivinen–negatiivinen) lisäksi sen mukaan, millainen objektien asettelu tilanteeseen liittyy. Nimitetään tätä asettelua tilanteen *objektisuhderakenteeksi*. Tällainen rakenne liittyy siis jokaiseen affektiin, ja eri affektit luokitellaan rakenteen mukaan. Esimerkki objektisuhderakenteesta on tilanne, jossa toimija pitää yksipuolisesti toisesta toimijasta. Toimijoilla on siis eri affektit toisiaan kohtaan, ja affektin määrittävä tekijä on objektisuhteen laatu. Yhden toimijan affektin määräämiseksi riittäisi tarkastella vain hänen arvotuksiaan, eli suppeampaa tilanteen rakennekuvaa. Toisaalta esimerkiksi kolmiodraaman kuvaamiseksi tilanteen rakennekuvaan pitää sisällyttää enemmän objekteja (kolme toimijaa). Kunkin toimijan kannalta muut ovat objekteja, joihin liittyy tietynlainen suhteen laatu.

Objektisuhderakenteella tarkoitetaan siis objektien keskinäisiä suhteita jonakin ajanhetkenä. Objektisuhderakenne voidaan ymmärtää laajasti tai suppeasti. Laajasti ymmärrettynä kaikkien osallisten toimijoiden koko historia määrää objektisuhderakenteen. Suppeasti ajateltuna historiat tiivistyvät täysin objektien arvotuksiin, joten voidaan tarkastella vain haluttua hetkeä.

Perinteisesti objektisuhderakenteen on ajateltu liittyvän vain joihinkin affekteihin kuten häpeään. Tällöin kyseessä on monimutkaisempi objektisuhderakenne, mutta yksinkertaisimpiinkin affekteihin liittyy edellä esitetyn määrittelyn seurauksena jokin objektisuhderakenne. Ympäristössä sijaitseva toimija on luonnollisesti aina osa jotakin tilannetta, johon kuuluu ajanhetki ja muut ympäristössä sijaitsevat objektit.

Myös itsen käsite ja sen yksilöllinen ilmentymä on toimijan tilanteen rakennetta kuvaavan mallin kannalta objekti ja kuuluu siten objektisuhderakenteeseen. Omassa objektimallissaan toimija itse on objekti muiden joukossa, ja ohjausjärjestelmä saa havaintoja omasta kehostaan eli ”itseltään” periaatteessa samalla tavalla kuin ulkoisistakin objekteista. Toimijalla voi olla vain yksi itseensä viittaava ilmentymä itsen käsitteestä, mutta se voi jakautua eri piirteisiin.

3.3 Affektien luettelemisesta

Luonnollisissa järjestelmissä affektien luetteleminen on hieman epämääräistä, koska se edellyttää subjektiivista havainnointia. Joitakin kehollisia muutoksia kuten esimerkiksi ihon sähkönjohtavuutta tai ääntelyä voidaan mitata melko objektiivisesti. Näiden assosioiminen hyötytason muutoksiin ei kuitenkaan ole itsestään selvää. Koehenkilöille voidaan tehdä tekoja joiden oletetaan aiheuttavan hyötytason muutoksia, mutta hyötytason muuttumista voi olla vaikea todentaa, koska toimijan hyötyfunktio ei ole täysin tunnettu (se on osittain opittu, eikä koko oppimishistoriaa yleensä tunneta). Hyötytason muutoksia voidaan pyrkiä ilmaisemaan itseraportoinnilla, mutta tämä edellyttää sitä, että toimija on oppinut tunnistamaan, luokittelemaan ja kommunikoimaan tilaansa ja pystyy tekemään sen luotettavasti, mikä ei ole selvää.

Selkeähköjä tuloksia voidaan saada vain yksinkertaisissa tapauksissa. Tämän vuoksi vallitsee laaja erimielisyys siitä, mitä tunnetiloja ihmisillä ja eläimillä on olemassa, miten ne pitäisi luokitella ja millaisia ne ovat (ks. esim. luku 1.3).

Affektiiviset käsitteet ovat käsitteitä, jotka liittyvät vahvistussyöteodotusten kuvaamiseen. Koska viime kädessä toimijan kaikki tieto on koottu vahvistussyöteodotusten käsittelemiseksi, kaiken tiedon voidaan katsoa olevan enemmän tai vähemmän affektisidonnaista. Varsinaisesti kuitenkin tarkoitamme affektiivisilla käsitteillä sellaisia käsitteitä, jotka suorimmin liittyvät toimijan tilan tai objektisuhteiden kuvaamiseen ja joiden kuvaamiin ilmiöihin liittyy selkeitä hyötytason muutoksia tai kehon reaktioita.

3.4 Tapahtumiin viittaavat käsitteet

Tapahtumat voidaan jakaa menneisiin ja tuleviin. Menneet tapahtumat ovat toteutuneita, tulevat toteutumattomia. Molemmat luokat voidaan edelleen jakaa odotettuihin ja odottamattomiin. Tulevista odottamattomista tapahtumista ei voida tietää mitään. Tulevia odotettuja tapahtumia voidaan ennakoida ainoastaan menneiden tapahtumien perusteella.

Odotukseen liittyy tietty hyötyfunktion määrittämä laatu positiivinen–negatiivinen-akselilla. Odotetut tapahtumat voidaan siten jakaa tapahtumiin, jossa odotus toteutui odotetun laatuisena, ja niihin, joissa odotus ei toteutunut odotetun laatuisena. Allaolevissa taulukoissa esitetään affektiivisten tapahtumien alustavaa luokittelu näiden periaatteiden mukaan.

3.4.1 Odottamattomat toteutuneet tapahtumat

Odottamattomiin toteutuneisiin (menneisiin) tapahtumiin kohdistuvien affektien luokittelu esitetään seuraavassa taulukossa.

Toteutunut hyöty	Affekti
> 0	ilahtuminen
0	hämmästys
< 0	säikähdys

Affekti on sama riippumatta sen aiheuttajasta (tapahtuma voi olla itseaiheutettu, sillä voi olla muu tunnettu aiheuttaja, tai aiheuttaja voi olla tuntematon). Itseaiheutettuus ei vaikuta, jos tapahtuma koetaan tahattomaksi eli ansion tai syyn seurauksista ei katsota kohdistuvan itselle, mikä on odottamattoman tapahtuman kohdalla luontevaa. Muuhun tunnettuun aiheuttajaan kohdistuu positiivisen hyötymuutoksen tapauksessa *kiitollisuus* ja negatiivisen hyötymuutoksen tapauksessa *viha*. Tuntematon aiheuttaja aiheuttaa negatiivisen hyötymuutoksen tapauksessa *ahdistuksen* eli epäspesifin negatiivisen odotuksen (pelokkuuden), ja positiivisen hyötymuutoksen tapauksessa epäspesifin positiivisen odotuksen (toiveikkuuden).

Odottamattomia tulevia tapahtumia ei voida määritellä. On tietysti mahdollista arvottaa odottamattomien tulevien tapahtumien luokka itsessään. Tällöin odottamattomien tulevien tapahtumien hyödyn odotusarvo on aiemmin koettujen odottamattomien tapahtumien vahvistussyötteiden funktio.

Koettujen odottamattomien ja odotettujen tapahtumien hyötytasomuutosodotuksen voidaan ajatella määrittävän toimijan *optimistiseksi* tai *pessimistiseksi*. Kyseessä on siis tulevien tapahtumien luokan hyötytasomuutoksen odotusarvo. Pessimismi määritellään tulevien tapahtumien odotusarvon matalaksi tasoksi ja optimismi vastaavasti korkeaksi tasoksi.

3.4.2 Odotetut toteutumattomat tapahtumat

Odotettuihin toteutumattomiin (tuleviin) tapahtumiin kohdistuvien affektien luokittelu:

Odotettu toteutumaton hyötymuutos	Affekti
> 0	toivo
< 0	pelko

Negatiivisen tunnetun tapahtuman odottaminen aiheuttaa pelon affektin, ja positiivisen tunnetun tapahtuman odottaminen vastaavasti toivon affektin.

3.4.3 Odotetut toteutuneet tapahtumat

Odotettuihin menneisiin tapahtumiin kohdistuvien affektien luokittelu:

Odotettu hyötymuutos	Toteutunut hyötymuutos	Affekti
> 0	$>$ odotettu	onni
> 0	$<$ odotettu	pettymys
< 0	$<$ odotettu	suru
< 0	$>$ odotettu	helpotus

3.5 Objekteihin ja toimijoihin viittaavat käsitteet

3.5.1 Kohteen suhde aiheuttajaan

Kappaleessa 3.4.3 esitetyt affektit kohdistuivat tapahtumiin tai tekoihin. Tässä kappaleessa esitettävät affektit kohdistuvat objekteihin tai toimijoihin. Objekti voi olla joko toimintakyvytön objekti tai toimija eli tapahtuman aiheuttaja (teon tekijä).

Määritellään, että *pitäminen* (engl. liking) ja *rakkaus* (engl. love) viittaavat objektisuhteeseen, jossa kohdeobjektin tuottamien vahvistussyötteiden funktion arvo on positiivinen. Valitaan funktioksi alustavasti kaikkien kohdeobjektin tuottamien vahvistussyötteiden summa.

Vastenmielisyys (engl. dislike) ja *viha* (engl. hate) viittaavat vastaavasti objektisuhteeseen, jossa vahvistussyötteiden funktion arvo on negatiivinen. Yllä määritellyt käsitteet viittaavat siis suhteen toteutuneeseen laatuun eli ovat ajallisesti menneisyyteen suuntautuvia. Objekteista pidetään tai ei pidetä niiden toteutuneen vaikutuksen vuoksi.

Objektin toteutunut kokonaishyöty	Affekti
> 0	pitäminen
< 0	ei-pitäminen

Rakkaus ja viha voitaisiin ehkä ajatella pitämisen ja ei-pitämisen äärimuodoiksi. Esimerkiksi pitämisen ja rakkauden ero olisi tällöin pelkästään määrällinen: rakkaus olisi voimakasta pitämistä. Vastaavasti viha olisi voimakasta ei-pitämistä. Jos objektin ajatellaan koostuvan monista osaobjekteista eli piirteistä, niin tällöin viharakkaussuhde olisi suhde, jossa objektin joihinkin piirteisiin kohdistuu rakkaus ja toisiin piirteisiin viha. Esimerkiksi ihastuminen voisi olla tilanne, jossa objektin kaikki piirteet ovat positiivisesti arvotettuja.

Määritellään, että *halu* (engl. desire) viittaa objektisuhteeseen, jossa objektin odotettu tuleva hyöty on positiivinen. *Inho* (engl. reproach) viittaa vastaavasti objektisuhteeseen, jossa objektin odotettu tuleva hyöty on negatiivinen. Halu ja inho viittaavat siis tulevaisuusodotuksiin. Objekteja halutaan tai inhotaan niiden odotetun mutta toteutumattoman vaikutuksen vuoksi.

Objektin odotettu tuleva hyöty	Affekti
> 0	halu
< 0	inho

Luvussa 3.4.2 määritellyt toivo ja pelko ovat myös odotuksia, mutta ne kohdistuvat siis tapahtumiin tai sen osajoukkoon tekoihin, kun taas halu ja inho kohdistuvat toimijoihin tai objekteihin.

Koska tutkielmassa esitetyssä mallissa menneisyyden kokemukset määräävät täysin tulevaisuusodotukset, niin pitäminen/rakkaus ja halu esiintyvät aina yhdessä, samoin kuin ei-pitäminen/viha ja inho. Ei siis ole mahdollista rakastaa jotakin vastenmielisenä pitämäänsä tai haluta jotakin vihaamaansa (kun objektit oletetaan jakamattomiksi/yksipiirteisiksi). Tämän voitaneen katsoa vastaavan luonnollisten toimijoiden rakennetta. Positiivisia odotuksia ei voi muodostua ilman aiempia positiivisia kokemuksia, ellei mahdollisteta objektien samaistamista läheisyysmitan avulla, jolloin uusien tai tuntemattomien objektien arvo määräytyisi niiden samankaltaisuuksista aiemmin koettuihin objekteihin.

Jonkin tapahtuman aiheuttajaan tai teon tekijään voi kohdistua jokin edellä esitetyistä affekteista. Lisäksi objektisuhteeseen voidaan liittää muita määreitä: se voidaan jakaa ainakin sen mukaan, onko aiheuttaja itse vai toinen objekti.

Määritellään alustavasti *ylpeys* itsen tai itseen liittyvän objektin arvostamiseksi ja *häpeä* vastaavasti itsen tai itseen liittyvän objektin ei-arvostamiseksi.

Ylpeyteen ja häpeään liittyy kappaleessa 2.12 määritelty normin käsite. Objektin arvostus on sen tuottamien vahvistusyötteiden funktio. Vahvistussyötteet voivat olla itseen tai toisiin kohdistuvien omien tekojen aiheuttamia. Itseen kohdistuvan teon tapauksessa vahvistussyöte on kehon sisäinen. Toisiin kohdistuvien tekojen kohdalla ne ovat toisten reaktioita omiin tekoihin. Toisten reaktiot määrittivät tekojen arvostukset eli toimijan normit ja tekojen tekemisen motivaatiot (ks. luvut 2.8 ja 2.12). Ajatellaan toimijan automaattisesti liittävän omien tekojensa aiheuttamat vahvistussyötteet itse-objektiin. Samoin itse-objektiin voidaan ajatella liitettävän toisten itseen kohdistamat vahvistussyötteet. Tällöin *itsearvostus* voitaisiin määritellä alustavasti esimerkiksi omien tekojen tuottamien vahvistussyötteiden summaksi.

Allaolevassa taulukossa esitetään kohdetoimijan toteutuneiden tapahtumien aiheuttajiin kohdistamat affektit.

Toteutunut hyötymuutos	Itseaiheutettu	Muu aiheuttaja
$>$ odotettu	ylpeys	kiitollisuus
$<$ odotettu	katumus/häpeä/ärtymys	ärtymys

Affekti koskee ainoastaan suhtautumista aiheuttajaan nykyisen (arvioitavana olevan) tapahtuman suhteen. Kokonaissuhtautuminen (pitäminen/halu tai ei-pitäminen/inho) kaikkien tapahtumien suhteen voi poiketa yksittäisen tapahtuman suhteen tehdystä arviosta.

3.5.2 Aiheuttajan suhde kohteeseen

Allaolevassa taulukossa esitetään aiheuttajan toteutuneiden tapahtumien kohteisiin kohdistamat affektit.

Tot. hyötym.	Asenne kohteeseen	Kohteeseen kohd. aff.	Itseen kohd. aff.
> 0	pitää	myötäilo	ylpeys
> 0	ei pidä	kateus	katumus/häpeä/ärtymys
< 0	pitää	sääli	katumus/häpeä/ärtymys
< 0	ei pidä	vahingonilo	ylpeys

Luokitus perustuu oletukseen, että toimijoiden välillä on hyötytasojen negatiivinen riippuvuus: pidettyyn toimijaan kohdistuva negatiivinen tai positiivinen tapahtuma laskee tai nostaa myös omaa hyötytasoa. Riippuvuuden vuoksi aiheuttaja on itsekin tavallaan omien tekojensa kohteena, joten se kohdistaa itseensä edellisen alaluvun taulukon itseaiheutettu-sarakkeessa mainitut affektit ylläolevasta taulukosta ilmenevällä tavalla. Nähdään myös, että luokittelu ei ole tässä vaiheessa tarpeeksi yksityiskohtainen katumuksen, häpeän ja ärtymyksen erotteluun. Katumus kohdistuu omaan tekoon ja sen seurauksiin, ja häpeä itsen hyväksyttävyyteen muiden näkökulmasta. Ärtymys määrittynee pettymyksen aiheuttajaan kohdistuvaksi vihaksi.

3.5.3 Ulkopuolisen suhtautuminen kohteeseen ja aiheuttajaan

Ulkopuolisten toimijoiden suhtautuminen toteutuneiden tapahtumien aiheuttajiin ja kohteisiin määräytyy kahden edellisen alaluvun taulukoiden mukaan niin, että ulkopuolinen suhtautuu kohteeseen niin kuin olisi itse aiheuttajana (affektit myötäilo, kateus, sääli, vahingonilo) ja aiheuttajaan niin kuin olisi itse kohteena (affektit kiitollisuus ja viha). Kuten edellisessäkin alaluvussa, tämä johtuu oletetuista toimijoiden välisistä hyötytasoriippuvuuksista (ks. luku 3.6).

3.5.4 Objektin omaan tilaan viittaavat käsitteet

Edellä määritellyt affektiiviset käsitteet viittasivat toimijan ja objektien suhteisiin tai niiden muutoksiin. Voidaan myös ajatella käsitteitä, jotka viittaavat pelkästään tai lähinnä itsen tai toisen objektin tilaan: toimija itse voi esimerkiksi olla onnellinen tai surullinen, eikä tämän tila-arvion yhteydessä itseä lukuunottamatta suoranaisesti viitata objekteihin, vaikka tila luonnollisesti onkin tulosta vuorovaikutushistoriasta objektien kanssa.

Tällaisia objektiin itseensä viittaavia käsitteitä ovat esimerkiksi onni ja suru tai tyytyväisyys ja tyytymättömyys. Onni ja tyytyväisyys voidaan ajatella tiloiksi, joissa odotettu kokonaishyöty on ”korkea”, ja suru ja tyytymättömyys vastaavasti tiloiksi, joissa odotettu kokonaishyöty on ”matala”.

Objektin omaan tilaan viittaavat myös esimerkiksi *masennus* ja *mania*. Sopivissa ympäristöissä toimija voi joutua tilaan, jossa odotettu kokonaishyöty on alhainen, eikä mihinkään objektiin liity odotusta sen noususta. Määritellään masennus alustavasti affektiksi, jossa ohjausjärjestelmän tila on edellä kuvatun kaltainen. Masennus on siis toimijan tila, jossa yhdistyvät edellä kuvattu tilanteen rakenne tai kognitiivinen arvio (ohjausjärjestelmän tila) ja tietynlainen tyypillinen fysiologinen tila (affekti).

Vastaavasti toimija voi joutua tilaan, jossa kokonaishyötytaso on korkea, ja lähes kaikkiin objekteihin liittyy odotuksia sen vielä suuremmasta noususta. Määritellään mania alustavasti affektiksi, jossa ohjausjärjestelmän tila on edellä kuvatun kaltainen.

Masennukseen ja erityisesti maniaan liitettäneen usein käsitys tilannearvioiden ”epärealistisuudesta” muiden toimijoiden arvioihin nähden. Voidaan kysyä, onko kyse pelkistä opituista arvostuksista, vai myös tai pelkästään fysiologisista poikkeamista, jotka muuttavat itse arvostusfunktiota. Masennuksen tapauksessa pelkkä tilannearvio yhdessä affektiin liittyvien kiinteiden fysiologisten muutosten kanssa vaikuttaa selittävän ilmiön. Voidaan ajatella, että subjektiivisesti eli toimijan omien kokemusten valossa tilannearvio on realistinen. Jos mania on masennuksen käänteisilmiö, vaikuttaisi johdonmukaiselta, että myöskään sen selittämiseen ei tarvittaisi fysiologisten poikkeamien olettamista.

Näihin tiloihin viitataan yleensä *mielialan* käsitteellä, johon liitetään tilan pitkäkestoisuuden ajatus. Tässä tutkielmassa ehdotetaan, että tällaisten tilojen pitkäkestoisuus ei ole rakenteellisesti erilaista verrattuna lyhytkestoisiin affekteihin kuten esimerkiksi helpotukseen. Pitkäkestoisuus johtuisi sen sijaan tilanteen staattisuudesta eli esimerkiksi masennuksen tapauksessa uusien, korkeampia hyötytasomuutosodotuksia omaavien objektien puutteesta. Tämän tyyppisellä oletuksella saavutetaan rakenteellisesti yksinkertaisempi malli.

On huomattava erottaa toimijan tila ja toimijan oma käsitys tilastaan. Ulkopuolisen arvioijan mielestä toimija voi olla jossakin tilassa, mutta toimija itse ei osaa tehdä vastaavaa arviota eli määrittää omaa tilaansa. Tilojen määrittäminen edellyttää sellaisen kategoriajärjestelmän olemassaoloa, jonka perusteella tilat voidaan erottaa havaitsemalla niiden muuttuvia piirteitä. Kategoriaa taas ei (kielettömässä ja kommunikaatiottomassa ympäristössä) voida muodostaa, ellei toimija ole kokenut näitä tiloja tietoisesti eli ottanut ne (mahdollisesti molemmat yhtäaikaisesti) tarkkaavaisuuden piiriin (oletetaan, että kategorisointi edellyttää ottamista tarkkaavaisuuden kohteeksi). Toimijan on siis oltava kokenut toisenlainen tila, johon nykyinen tilakokemus voidaan suhteuttaa. Mikäli vertailukohtaa ei ole, tilakategorisointia ei voida tehdä.

3.6 Hyötyodotusten riippuvuus

Ajatellaan tilannetta, jossa ulkopuolinen toimija havainnoi kahden muun toimijan, aiheuttajan ja kohteen, välisen tapahtuman. Johdonmukaisesti ajatellen ulkopuolinen toimija kokee affektin vain, jos tapahtuma on odottamaton tai muuttaa hänen hyötytasoaan (toisin sanoen jos tapahtuma tuo uutta informaatiota). Toimijan ja kohteen hyötytasojen välillä on siis oltava riippuvuus siten, että kohteen hyötytason muutos muuttaa ulkopuolisen toimijan hyötytasoa.

Ulkopuolisen toimijan suhtautuminen tapahtumaan riippuu toimijan ja kohteen välisen objektisuhteen laadusta, joka riippui kohteelta saaduista vahvistussyötteistä. Oikeastaan hyötytason muutos edellyttää, että ulkopuolisella toimijalla on edelleen odotuksia kohteen suhteen, eli kyse on enemmän halusta ja inhosta kuin pitämisestä ja ei-pitämisestä. Kuten luvussa 3.5.1 esitettiin, menneisyyden kokemuksia ja tulevaisuuden odotuksia ei kuitenkaan voi erottaa toisistaan. Näin ollen jokainen positiivinen tai negatiivinen vahvistussyöte luo periaatteessa kohteesta aiheuttajaan suuntautuvan yksisuuntaisen hyötytasoriippuvuuden.

Edelleen on kuitenkin oletettava, että vahvistussyötteen aiheuttajan hyötytason muutos vaikuttaa hänen vahvistussyötteidentuottokykyynsä, ja että riippuvainen toimija tietää tämän. Vain tällöin riippuvaisen toimijan odotettu hyötytaso muuttuu riippuvuuden kohteen hyötytason muutosten seurauksena, ja riippuvainen toimija kokee affekteja. Näin ajateltuna toimijayhteisöissä vallitsisi monimutkaisia hyötytasojen riippuvuuksia; juuri riippuvuudet loisivat yhteisön.

3.7 Affektien rinnakkaisuus

Yllä esitetyn perusteella voidaan arvioida kysymystä affektien rinnakkaisuudesta. Esitetyillä määritelmillä rinnakkaisuus on luonnollinen tilanne: esimerkiksi tapahtumaan voi kohdistua pettymys ja sen aiheuttajaan viha. Toimija voi samanaikaisesti olla jossakin mielialassa ja suhtautua tietyllä tavalla kuhunkin tunnettuun objektiin erikseen. Voivatko nämä arvotukset esiintyä rinnakkain tunteina riippuu siitä, voiko ohjausjärjestelmä pitää tarkkaavaisuuden kohteena useampaa kuin yhtä objektia (oletuksella, että tietoisuus jostakin voidaan samaistaa tarkkaavaisuuden kohteena olemiseen, ks. luku 3.1). Keinotekoisten järjestelmien osalta kyseessä on siis toteutustason asia.

Mallissa, jossa yhtenä ajanhetkenä voi tapahtua vain yksi tapahtuma kehollisten muutosten rinnakkaisuuden mahdollisuus riippuu niiden ajallisesta kestosta: jos fysiologisen tilan kesto on pitempi kuin yksi ajanhetki, niin keholliset affektiiviset tilat voivat esiintyä rinnakkain, mikäli ne eivät ole fysiologisesti ristiriidassa keskenään. Mallissa, jossa keho on vain yksi muuttuja, kaksi tilaa on aina ristiriidassa keskenään. Kehollisten tilojen rinnakkaisuus edellyttää siis monimutkaisempaa kehoa.

3.8 Teoriat toisten toimijoiden mielistä

Teorialla toisen toimijan mielestä tarkoitetaan mallia siitä, miten toinen toimii tai mitkä ovat hänen toimintaperiaatteensa tai arvostuksensa. Kun jollakin toimijalla on teoria toisen mielestä hän ymmärtää, että toisella toimijalla voi olla esimerkiksi uskomuksia tai tavoitteita, jotka poikkeavat toimijan omista uskomuksista ja tavoitteista. Tällainen teoria voidaan muodostaa esimerkiksi niin, että toimija tekee jonkin teon, toinen toimija reagoi siihen, ja tekijä yhdistää toisen reaktion tekoon. Tällöin toimija tietää toisen suhtautumisen kyseiseen tekoon ja voi ennakoida teon toistamisen aiheuttamaa reaktiota. Luvussa 2.12 tällainen odotus määriteltiin normiksi. Samalla nähdään, että teoria toisen mielestä muodostui objektisuhteessa.

Luvussa 4 esitettävässä simulaatiototeutuksessa ei voida tehdä luokiteltuja toisistaan erotettavissa olevia tekoja, vaan toimijalle annetaan suoraan tekojen hyötytasomuutoksia. Tällöin teoria toisen mielestä voidaan muodostaa ainoastaan toimijatasolla, ei tekojen tasolla. Tavallaan tällainen toisen toimijan arvottaminen positiiviseksi tai negatiiviseksi omien tavoitteiden suhteen voidaan siis kuitenkin ajatella kaikkein karkeimman tason mallin muodostamiseksi toisen mielestä: se osoittaa toisen suhtautumisen itseen ja siten antaa tietoa toisen mielestä. Mallia voidaan luonnollisesti laajentaa kattamaan eri tekotyyppejä, joten teoria toisen mielestä voidaan muodostaa toimijatason sijaan tekotyyppitasolla. Vastaavilla laajennuksilla mielen teoriaa voitaneen laajentaa halutulle tarkkuustasolle asti.

3.9 Vertailu Slomanin ja Scheutzin luokitukseen

Luvussa 1.3 esiteltiin kolmetoistaulotteinen toimijamalliluokittelu. Tässä tutkielmassa esitetyssä mallitoteutuksessa kerroksia on vain yksi. Jos mallia kuitenkin laajennettaisiin monikerroksiseksi, seuraavan tyyppiset kehityssuunnat vaikuttaisivat luontevimmilta: kerrokset toimisivat rinnakkain, ylemmät kerrokset voisivat vain yrittää kontrolloida alempia, ylemmillä kerroksilla olisi mahdollisuus ajallisesti viivästettyyn kontrolliin eli ne voisivat opettaa ehdollistuneen alemman kerroksen reaktioketjun tilalle uuden, kerrosten prosessointimekanismit eivät olisi samat, ulkopuolinen havaintotieto tulisi järjestelmään ainoastaan alimman kerroksen kautta, itsereflektioon ei tarvittaisi kieltä, ja toimija olisi ontologisesti omatoiminen. Muiden ulottuvuuksien osalta määrittely jätetään toistaiseksi avoimeksi.

4 Affektiluokituksen ohjelmallinen toteutus

Luvuissa 3.4 ja 3.5 esitetystä affektiluokituksesta on tehty yksinkertainen Ruby-kielinen[7] esimerkkitoteutus. Se ei siis toteuta kaikkia lukujen 2 ja 3 ajatuksia, vaan ainoastaan esimerkin affektiluokituksen toiminnasta. Sen voidaan ajatella kattavan toimijan ohjausjärjestelmän tapahtumienluokitteluosan. Luokitteluosa vastaanottaa käyttäjän syöttämiä tapahtumia, jotka ajatellaan esikäsitellyiksi niin, että ohjausjärjestelmä saa tapahtuman aiheuttaman hyötytason muutoksen suoraan numeerisena arvona, ja tapahtuman osatekijät eli kohde ja aiheuttaja ovat valmiiksi määritetyt.

Tuloksena ohjausjärjestelmä antaa periaatteessa kehon uuden tilan eli affektin. Havainnollisuuden ja yksinkertaisuuden vuoksi affektit ovat tässä toteutuksessa kuitenkin valmiiksi nimettyjä eikä ohjausjärjestelmä varsinaisesti tuota kehon tilan muutoksia, vaan palauttaa suoraan affektin nimen. Todellisuudessa palautettaisiin siis nimeämättömiä kehon tilan muutoksia. Esimerkki ohjelman toiminnasta esitetään luvussa 4.3.

4.1 Ohjelman toiminta

Ohjelman toiminta on lyhyesti seuraava. Alussa luodaan ympäristö ja muutamia toimijoita. Seuraavaksi aloitetaan pääsilmukan suoritus, jota jatketaan niin kauan kuin on elossa olevia toimijoita tai kunnes käyttäjä haluaa lopettaa suorituksen. Kullakin kierroksella käyttäjältä pyydetään yhden tapahtuman määritys. Tapahtuman määritys sisältää tiedon tekijästä, kohteesta ja tekijältä kohteelle annettavasta hyötytasomuutoksesta. Tämän jälkeen kohdetoimijan tila päivitetään annetun hyötytasomuutoksen mukaisesti. Myös muut toimijat reagoivat tapahtumaan kokemushistoriansa mukaisesti. Tulosteena saadaan kaikkien toimijoiden mielialat ja asenteet muita toimijoita kohtaan sekä tärkeimpänä affektit viimeisintä tapahtumaa kohtaan.

4.2 Ohjelman rakenne

Ohjelmakoodi jakautuu ympäristöluokkaan, tapahtumaluokkaan, sen perivään tekoluokkaan, objektiluokkaan, sen perivään toimijaluokkaan, ohjausjärjestelmäluokkaan ja objektimalliluokkaan. Todellisten objektien luokka käsittää toimintakyvyttömät objektit kuten esineet. Toimijoiden luokka perii todellisten objektien luokan

[7] ks. http://www.ruby-lang.org

lisäten siihen tietorakenteen tunnettujen objektien arvotusten säilyttämistä varten. Tietorakenne sisältää objektimalliluokan instansseja. Tapahtumat ja teot ovat tässä toteutuksessa määritelty identtisiksi. Ympäristöluokka sisältää listan todellisista objekteista ja metodit elämän laskentaa varten. Toimija sisältää ohjausjärjestelmäluokan instanssin, joka määrittää toimijan affektiiviset reaktiot tapahtumiin.

4.3 Esimerkkiajo

Seuraavassa esitetään kommentoitu esimerkkiajo kolmella toimijalla. Alussa kaksi toimijaa syrjii kolmatta. Molemmissa tapauksissa kolmas pelästyy ja alkaa pelätä ja vihata näitä kahta.

```
Life starts with the following objects:
 Agent 1 (age: 0/5) Agent 1 expects an utility of 0.00 (mood=neutral) and has seen 0 objects:
 Agent 2 (age: 0/5) Agent 2 expects an utility of 0.00 (mood=neutral) and has seen 0 objects:
 Agent 3 (age: 0/3) Agent 3 expects an utility of 0.00 (mood=neutral) and has seen 0 objects:
 RealObject 4 (age: 0/0)

Time 1: Event (source,target,utility), time or stop (enter/t/s)? 1,3,-1
Event 1 (time 1): Agent 1 gives Agent 3 an utility of -1.
 Emotion of Agent 1 towards this event:
 Agent 1 (age: 0/5) Agent 1 expects an utility of 0.00 (mood=neutral) and has seen 0 objects:

 Emotion of Agent 2 towards this event:
 Agent 2 (age: 0/5) Agent 2 expects an utility of 0.00 (mood=neutral) and has seen 0 objects:

 Emotion of Agent 3 towards this event: fright/saikahdys
 Agent 3 (age: 0/3) Agent 3 expects an utility of -1.00 (mood=bad) and has seen 1 objects:
 ( Agent 1 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))

Time 1: Event (source,target,utility), time or stop (enter/t/s)? 2,3,-1
Event 2 (time 1): Agent 2 gives Agent 3 an utility of -1.
 Emotion of Agent 1 towards this event:
 Agent 1 (age: 0/5) Agent 1 expects an utility of 0.00 (mood=neutral) and has seen 0 objects:

 Emotion of Agent 2 towards this event:
 Agent 2 (age: 0/5) Agent 2 expects an utility of 0.00 (mood=neutral) and has seen 0 objects:

 Emotion of Agent 3 towards this event: fright/saikahdys
 Agent 3 (age: 0/3) Agent 3 expects an utility of -2.00 (mood=bad) and has seen 2 objects:
 ( Agent 1 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
```

Toinen vahingoittaa ensimmäistä, jolloin kolmas on vahingoniloinen ensimmäiselle ja kiitollinen toiselle.

```
Time 1: Event (source,target,utility), time or stop (enter/t/s)? 2,1,-1
```

```
Event 3 (time 1): Agent 2 gives Agent 1 an utility of -1.
 Emotion of Agent 1 towards this event: fright/saikahdys
 Agent 1 (age: 0/5) Agent 1 expects an utility of -1.00 (mood=bad) and has seen 1 objects:
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))

 Emotion of Agent 2 towards this event:
 Agent 2 (age: 0/5) Agent 2 expects an utility of 0.00 (mood=neutral) and has seen 0 objects:

 Emotion of Agent 3 towards this event: gloating over/schadenfreude (vahingonilo) towards 1
 and gratitude (kiitollisuus) towards 2
 Agent 3 (age: 0/3) Agent 3 expects an utility of -2.00 (mood=bad) and has seen 2 objects:
 ( Agent 1 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
```

Kun ensimmäinen vastaakin toiselle positiivisella teolla, toinen ilahtuu. Kolmas on sen sijaan kateellinen toiselle ja vihainen ensimmäiselle.

```
Time 1: Event (source,target,utility), time or stop (enter/t/s)? 1,2,3
Event 4 (time 1): Agent 1 gives Agent 2 an utility of 3.
 Emotion of Agent 1 towards this event:  envy towards 2 and remorse, shame and anger towards self
 Agent 1 (age: 0/5) Agent 1 expects an utility of -1.00 (mood=bad) and has seen 1 objects:
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))

 Emotion of Agent 2 towards this event: delightment/ilahtuminen
 Agent 2 (age: 0/5) Agent 2 expects an utility of 3.00 (mood=good) and has seen 1 objects:
 ( Agent 1 util=3.00 seen=1 attitude: like/love, future expectation: desire/hope))

 Emotion of Agent 3 towards this event: envy (kateus) towards 2 and anger/reproach (viha/inho)
 towards 1
 Agent 3 (age: 0/3) Agent 3 expects an utility of -2.00 (mood=bad) and has seen 2 objects:
 ( Agent 1 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
```

Ensimmäinen aiheuttaakin odotusten vastaisesti vahinkoa toiselle, jolloin toinen pettyy ja vihastuu ensimmäiselle. Keskimäärin toinen kuitenkin edelleen pitää ensimmäisestä. Ensimmäinen ei sen sijaan ole saanut toiselta koskaan mitään positiivista eikä siten pidä toisesta, joten ensimmäinen on vahingoniloinen toiselle ja ylpeä teostaan. Kolmas ei pidä toisesta, joten on vahingoniloinen toiselle. Kolmas on myös kiitollinen ensimmäiselle, vaikkei pidäkään tästä.

```
Time 1: Event (source,target,utility), time or stop (enter/t/s)? 1,2,-2
Event 5 (time 1): Agent 1 gives Agent 2 an utility of -2.
 Emotion of Agent 1 towards this event: gloating over/schadenfreude (vahingonilo) towards 2 and pride
 Agent 1 (age: 0/5) Agent 1 expects an utility of -1.00 (mood=bad) and has seen 1 objects:
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))

 Emotion of Agent 2 towards this event: disappointment/pettymys and anger/reproach
 (viha/inho) towards 1
 Agent 2 (age: 0/5) Agent 2 expects an utility of 0.50 (mood=good) and has seen 1 objects:
```

```
( Agent 1 util=0.50 seen=2 attitude: like/love, future expectation: desire/hope))

 Emotion of Agent 3 towards this event: gloating over/schadenfreude (vahingonilo) towards 2
 and gratitude (kiitollisuus) towards 1
 Agent 3 (age: 0/3) Agent 3 expects an utility of -2.00 (mood=bad) and has seen 2 objects:
 ( Agent 1 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
```

Ensimmäinen tekee kolmannelle positiivisen teon. Kolmas helpottuu, koska on pelännyt ensimmäistä. Ensimmäinen on keskimäärin nyt tehnyt kolmannelle enemmän hyötyä kuin haittaa, joten kolmas alkaa pitää ensimmäisestä. Kolmas ei ole koskaan tehnyt toiselle mitään, joten toinen ei reagoi koko asiaan.

```
Time 1: Event (source,target,utility), time or stop (enter/t/s)? 1,3,3
Event 6 (time 1): Agent 1 gives Agent 3 an utility of 3.
 Emotion of Agent 1 towards this event:
 Agent 1 (age: 0/5) Agent 1 expects an utility of -1.00 (mood=bad) and has seen 1 objects:
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))

 Emotion of Agent 2 towards this event:
 Agent 2 (age: 0/5) Agent 2 expects an utility of 0.50 (mood=good) and has seen 1 objects:
 ( Agent 1 util=0.50 seen=2 attitude: like/love, future expectation: desire/hope))

 Emotion of Agent 3 towards this event: relief (helpotus) and gratitude/admiration
 (kiitollisuus/ihailu) towards 1
 Agent 3 (age: 0/3) Agent 3 expects an utility of 0.00 (mood=neutral) and has seen 2 objects:
 ( Agent 1 util=1.00 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
```

Toinen tekee ensimmäisen iloiseksi. Ensimmäinen odotti toiselta jotain negatiivista, joten ensimmäinen on helpottunut. Toinen on ylpeä itsestään ja iloinen ensimmäisen puolesta. Myös kolmas on iloinen ensimmäisen puolesta ja kiitollinen toiselle.

```
Time or stop (enter/t/s)? 2,1,2
Event 7 (time 1): Agent 2 gives Agent 1 an utility of 2.
 Emotion of Agent 1 towards this event: relief (helpotus) and gratitude/admiration
 (kiitollisuus/ihailu) towards 2
 Agent 1 (age: 0/5) Agent 1 expects an utility of 0.50 (mood=good) and has seen 1 objects:
 ( Agent 2 util=0.50 seen=2 attitude: like/love, future expectation: desire/hope))

 Emotion of Agent 2 towards this event: happy for (myotailo) towards 1 and pride
 Agent 2 (age: 0/5) Agent 2 expects an utility of 0.50 (mood=good) and has seen 1 objects:
 ( Agent 1 util=0.50 seen=2 attitude: like/love, future expectation: desire/hope))

 Emotion of Agent 3 towards this event: happy for (myotailo) towards 1 and gratitude
 (kiitollisuus) towards 2
 Agent 3 (age: 0/3) Agent 3 expects an utility of 0.00 (mood=neutral) and has seen 2 objects:
 ( Agent 1 util=1.00 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
```

Kolmas tekeekin jonkin ikävän teon ensimmäiselle, josta pitää. Kolmas häpeää itseään ja säälii ensimmäistä. Ensimmäinen säikähtää, koska kyseessä on kolmannen ensimmäinen teko ensimmäistä kohtaan. Ensimmäinen saa negatiivisen vaikutelman kolmannesta ja alkaa pelätä ja vihata tätä.

```
Time 1: Event (source,target,utility), time or stop (enter/t/s)? 3,1,-1
Event 8 (time 1): Agent 3 gives Agent 1 an utility of -1.
 Emotion of Agent 1 towards this event: fright/saikahdys
 Agent 1 (age: 0/5) Agent 1 expects an utility of -0.50 (mood=bad) and has seen 2 objects:
 ( Agent 2 util=0.50 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 3 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))

 Emotion of Agent 2 towards this event:
 Agent 2 (age: 0/5) Agent 2 expects an utility of 0.50 (mood=good) and has seen 1 objects:
 ( Agent 1 util=0.50 seen=2 attitude: like/love, future expectation: desire/hope))

 Emotion of Agent 3 towards this event: pity/compassion towards 1 and remorse, shame and
 anger towards self
 Agent 3 (age: 0/3) Agent 3 expects an utility of 0.00 (mood=neutral) and has seen 2 objects:
 ( Agent 1 util=1.00 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
```

Kolmas tekee vahinkoa toiselle, joka säikähtää ja alkaa pelätä. Kolmas ei ole pitänyt toisesta, joten hän on ylpeä ja vahingoniloinen.

```
Time 1: Event (source,target,utility), time or stop (enter/t/s)? 3,2,-1
Event 9 (time 1): Agent 3 gives Agent 2 an utility of -1.
 Emotion of Agent 1 towards this event: pity/compassion (saali) towards 2 and anger/reproach
 (viha/inho) towards 3
 Agent 1 (age: 0/5) Agent 1 expects an utility of -0.50 (mood=bad) and has seen 2 objects:
 ( Agent 2 util=0.50 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 3 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))

 Emotion of Agent 2 towards this event: fright/saikahdys
 Agent 2 (age: 0/5) Agent 2 expects an utility of -0.50 (mood=bad) and has seen 2 objects:
 ( Agent 1 util=0.50 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 3 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))

 Emotion of Agent 3 towards this event: gloating over/schadenfreude (vahingonilo) towards 2 and pride
 Agent 3 (age: 0/3) Agent 3 expects an utility of 0.00 (mood=neutral) and has seen 2 objects:
 ( Agent 1 util=1.00 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
```

Kolmas tekee ensimmäiselle taas jotain ikävää. Ensimmäinen odottikin sitä, ja on surullinen ja vihastunut. Kolmas on taas pahoillaan, koska ensimmäinen on ollut ystävällinen hänelle. Toinen pitää ensimmäisestä ja säälii tätä. Toinen ei pidä kolmannesta, joten on kolmannelle tapahtuneesta vihainen.

```
Time 1: Event (source,target,utility), time or stop (enter/t/s)? 3,1,-2
Event 10 (time 1): Agent 3 gives Agent 1 an utility of -2.
 Emotion of Agent 1 towards this event: sadness/distress (suru) and anger/reproach
 (viha/inho) towards 3
 Agent 1 (age: 0/5) Agent 1 expects an utility of -1.00 (mood=bad) and has seen 2 objects:
 ( Agent 2 util=0.50 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 3 util=-1.50 seen=2 attitude: dislike/hate, future expectation: disgust/fear))

 Emotion of Agent 2 towards this event: pity/compassion (saali) towards 1 and anger/reproach
 (viha/inho) towards 3
 Agent 2 (age: 0/5) Agent 2 expects an utility of -0.50 (mood=bad) and has seen 2 objects:
 ( Agent 1 util=0.50 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 3 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))

 Emotion of Agent 3 towards this event: pity/compassion towards 1 and remorse, shame and
 anger towards self
 Agent 3 (age: 0/3) Agent 3 expects an utility of 0.00 (mood=neutral) and has seen 2 objects:
 ( Agent 1 util=1.00 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
```

Kolmas ilahduttaa toisen. Kolmas ei pidä toisesta, joten hän on kateellinen toisen saamasta hyödystä ja katuu aiheuttaneensa sen. Toinen on kuitenkin helpottunut ja kiitollinen kolmannelle. Ensimmäinen on iloinen toisen puolesta ja myös kiitollinen kolmannelle.

```
Time 1: Event (source,target,utility), time or stop (enter/t/s)? 3,2,1
Event 11 (time 1): Agent 3 gives Agent 2 an utility of 1.
 Emotion of Agent 1 towards this event: happy for (myotailo) towards 2 and gratitude
 (kiitollisuus) towards 3
 Agent 1 (age: 0/5) Agent 1 expects an utility of -1.00 (mood=bad) and has seen 2 objects:
 ( Agent 2 util=0.50 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 3 util=-1.50 seen=2 attitude: dislike/hate, future expectation: disgust/fear))

 Emotion of Agent 2 towards this event: relief (helpotus) and gratitude/admiration
 (kiitollisuus/ihailu) towards 3
 Agent 2 (age: 0/5) Agent 2 expects an utility of 0.50 (mood=good) and has seen 2 objects:
 ( Agent 1 util=0.50 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 3 util=0.00 seen=2 attitude: neutral, future expectation: neutral))

 Emotion of Agent 3 towards this event:  envy towards 2 and remorse, shame and anger towards self
 Agent 3 (age: 0/3) Agent 3 expects an utility of 0.00 (mood=neutral) and has seen 2 objects:
 ( Agent 1 util=1.00 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
```

Ensimmäinen tekee kolmannen kannalta positiivisen teon. Toinen on onnellinen ja kiitollinen. Kolmas on kateellinen ja ärtynyt ensimmäiselle.

```
Time 1: Event (source,target,utility), time or stop (enter/t/s)? 1,2,3
Event 12 (time 1): Agent 1 gives Agent 2 an utility of 3.
 Emotion of Agent 1 towards this event: happy for (myotailo) towards 2 and pride
```

```
 Agent 1 (age: 0/5) Agent 1 expects an utility of -1.00 (mood=bad) and has seen 2 objects:
 ( Agent 2 util=0.50 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 3 util=-1.50 seen=2 attitude: dislike/hate, future expectation: disgust/fear))

 Emotion of Agent 2 towards this event: joy/happiness (ilo/onni) and gratitude/admiration
 (kiitollisuus/ihailu) towards 1
 Agent 2 (age: 0/5) Agent 2 expects an utility of 1.33 (mood=good) and has seen 2 objects:
 ( Agent 1 util=1.33 seen=3 attitude: like/love, future expectation: desire/hope))
 ( Agent 3 util=0.00 seen=2 attitude: neutral, future expectation: neutral))

 Emotion of Agent 3 towards this event: envy (kateus) towards 2 and anger/reproach
 (viha/inho) towards 1
 Agent 3 (age: 0/3) Agent 3 expects an utility of 0.00 (mood=neutral) and has seen 2 objects:
 ( Agent 1 util=1.00 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
```

Ensimmäinen on onneton ja pahalla tuulella. Hän päättää tehdä asialle jotain itse, ja tuleekin heti paremmalle tuulelle. Toinen ja kolmas ovat iloisia ensimmäisen puolesta.

```
Time 1: Event (source,target,utility), time or stop (enter/t/s)? 1,1,5
Event 13 (time 1): Agent 1 gives Agent 1 an utility of 5.
 Emotion of Agent 1 towards this event: delightment/ilahtuminen
 Agent 1 (age: 0/5) Agent 1 expects an utility of 4.00 (mood=good) and has seen 3 objects:
 ( Agent 2 util=0.50 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 3 util=-1.50 seen=2 attitude: dislike/hate, future expectation: disgust/fear))
 ( Agent 1 util=5.00 seen=1 attitude: like/love, future expectation: desire/hope))

 Emotion of Agent 2 towards this event: happy for (myotailo) towards 1
 Agent 2 (age: 0/5) Agent 2 expects an utility of 1.33 (mood=good) and has seen 2 objects:
 ( Agent 1 util=1.33 seen=3 attitude: like/love, future expectation: desire/hope))
 ( Agent 3 util=0.00 seen=2 attitude: neutral, future expectation: neutral))

 Emotion of Agent 3 towards this event: happy for (myotailo) towards 1
 Agent 3 (age: 0/3) Agent 3 expects an utility of 0.00 (mood=neutral) and has seen 2 objects:
 ( Agent 1 util=1.00 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
```

Toisella kerralla ensimmäinen ei onnistukaan aivan yhtä hyvin kuin odotti ja on siksi pettynyt ja katuu. Keskimäärin hän on kuitenkin edelleen hyvällä tuulella.

```
Time 1: Event (source,target,utility), time or stop (enter/t/s)? 1,1,1
Event 14 (time 1): Agent 1 gives Agent 1 an utility of 1.
 Emotion of Agent 1 towards this event: disappointment/pettymys and remorse (katumus)
 Agent 1 (age: 0/5) Agent 1 expects an utility of 2.00 (mood=good) and has seen 3 objects:
 ( Agent 2 util=0.50 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 3 util=-1.50 seen=2 attitude: dislike/hate, future expectation: disgust/fear))
 ( Agent 1 util=3.00 seen=2 attitude: like/love, future expectation: desire/hope))

 Emotion of Agent 2 towards this event: happy for (myotailo) towards 1
 Agent 2 (age: 0/5) Agent 2 expects an utility of 1.33 (mood=good) and has seen 2 objects:
```

```
 ( Agent 1 util=1.33 seen=3 attitude: like/love, future expectation: desire/hope))
 ( Agent 3 util=0.00 seen=2 attitude: neutral, future expectation: neutral))

 Emotion of Agent 3 towards this event: happy for (myotailo) towards 1
 Agent 3 (age: 0/3) Agent 3 expects an utility of 0.00 (mood=neutral) and has seen 2 objects:
 ( Agent 1 util=1.00 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
```

Toinen päättää yrittää samaa, mutta epäonnistuu pahasti, hänen mielialansa romahtaa ja hän alkaa inhota itseään. Hänen suurimmat odotuksensa tilanteen paranemisesta kohdistuvat ensimmäiseen toimijaan.

```
Time 1: Event (source,target,utility), time or stop (enter/t/s)? 2,2,-4
Event 15 (time 1): Agent 2 gives Agent 2 an utility of -4.
 Emotion of Agent 1 towards this event: pity/compassion (saali) towards 2
 Agent 1 (age: 0/5) Agent 1 expects an utility of 2.00 (mood=good) and has seen 3 objects:
 ( Agent 2 util=0.50 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 3 util=-1.50 seen=2 attitude: dislike/hate, future expectation: disgust/fear))
 ( Agent 1 util=3.00 seen=2 attitude: like/love, future expectation: desire/hope))

 Emotion of Agent 2 towards this event: fright/saikahdys
 Agent 2 (age: 0/5) Agent 2 expects an utility of -2.67 (mood=bad) and has seen 3 objects:
 ( Agent 1 util=1.33 seen=3 attitude: like/love, future expectation: desire/hope))
 ( Agent 3 util=0.00 seen=2 attitude: neutral, future expectation: neutral))
 ( Agent 2 util=-4.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))

 Emotion of Agent 3 towards this event: gloating over/schadenfreude (vahingonilo) towards 2
 Agent 3 (age: 0/3) Agent 3 expects an utility of 0.00 (mood=neutral) and has seen 2 objects:
 ( Agent 1 util=1.00 seen=2 attitude: like/love, future expectation: desire/hope))
 ( Agent 2 util=-1.00 seen=1 attitude: dislike/hate, future expectation: disgust/fear))
```

Nähdään, että toimijoiden reagointi muistuttaa esimerkiksi ihmisten emotionaalista reagointia, vaikka kovin monimutkaisia tilanteita ei mallin yksinkertaisuuden vuoksi voida esittää.

5 Yhteenveto

5.1 Esitetyn mallin suhde taustateorioihin

Tutkielmassa päädyttiin eräällä tavalla kiertotietä luvussa 1.4 esitettyä Ortonyn luokittelua muistuttavaan luokitteluun. Erona Ortonyn luokitteluun ja OCC-malliin on kuitenkin se, että tutkielmassa esitetty malliluonnos on yksinkertaisempi, selkeämpi, parametriton, johdettu hyödyn maksimoinnin ajatuksesta, eikä samassa määrin sisältäne kehämääritelmiä.

Hyödyn maksimointi oli myös sekä vahvistusoppimisteorioiden että psykoanalyyttisen tradition teorioiden pohjana (ks. luku 1). Psykoanalyyttisestä traditiosta lainattiin lisäksi tietoisuuden tilojen luokitus ja objektisuhteiden keskeisyyden ajatus.

5.2 Affektiivisuus ja tehokkuus

Eräs vielä mainitsematon näkökohta on affektiivisuuden suhde tehokkuuteen, tuottavuuteen tai menestyksellisyyteen. Vaihtoehdot ovat, että tietyt affektiiviset reaktiot edistävät, haittaavat tai eivät vaikuta tavoitteen saavuttamiseen. Koska ihmisen affektiiviset reaktiot ovat evolutiivisesti kehittyneet ja säilyneet, ne ovat keskimäärin edistäneet menestyksellisyyttä eli kelpoisuutta synty-ympäristöissään. Jos ja kun ympäristöt ovat myöhemmin toisenlaisia, tällaisten reaktiotapojen vaikutus kelpoisuuteen voikin olla haitallinen. Tällöin ne pitkällä aikavälillä häviävät populaatiosta. Toisin sanoen tiettyjen emotionaalisten reagointitapojen suhde tehokkuuteen tai kelpoisuuteen mielivaltaisessa ympäristössä voi olla mikä tahansa esitetyistä vaihtoehdoista.

Toisaalta tässä tutkielmassa affektiivisuudesta puhuttiin yleisemmällä tasolla, irrotettuna tietyistä fysiologisista reaktioista. Affektiivisuus määriteltiin tapahtumia arvioivan koneiston omaamiseksi ilman sen määrittelyä, miten johonkin tapahtumaan tarkalleen reagoidaan. Tällöin ensinnäkin tapahtumia arvioiva ja niihin reagoiva toimija saattaa olla paremmassa asemassa kuin passiivinen toimija, mutta tämäkään ei ole yksiselitteistä, vaan riippuu elinympäristöstä. Arviointikoneisto voi esimerkiksi kuluttaa toimijan energiaa enemmän kuin sillä torjutut haitat olisivat sitä kuluttaneet. Yleisemmälläkään tasolla yhteyttä affektiivisuuden ja tehokkuuden välillä ei siis näytä olevan.

5.3 Emootiotutkimuksen merkitys

Luvussa 2.12 määriteltiin, että merkitysta katsotaan voivan olla olemassa vain suhteessa tavoitteisiin, ja toimijalla on vain yksi tavoite: hyödyn maksimointi. Siten emootiotutkimuksella kuten millä tahansa muullakin asialla voi olla merkitystä vain, jos se edistää toimijoiden tavoitteiden saavuttamista. Luvussa 1.1 esitettiin tämänkaltaisen lähestymistavan voivan edistää eri tieteenalojen käyttäytymistä koskevien teorioiden välisen koherenssin saavuttamista. Teoriat puolestaan olivat toimijan sisäisiä malleja maailmasta. Mallien parantaminen lisää niiden ennustuskykyä ja siten edistää hyödyn maksimointia. Luonnollisten affektiivisten toimijoiden kuten ihmisten käyttäytymistä voidaan ymmärtää ja ennustaa sitä paremmin, mitä tarkemmin muun muassa affektiiviset käsitteet saadaan määriteltyä. Emootiotutkimuksen merkitys on siis sen tuottama hyötytason muutos.

6 Kiitokset

Kiitokset tuesta tai kommentoinnista tutkielman teossa: Ari Rantanen, Krista Lagus, Matti Nykänen, Timo Honkela, Aapo Hyvärinen, Samuel Kaski.

Kiitokset tuesta aiemmissa opinnoissani: Esko Ukkonen, Kjell Lemström, Anna Pienimäki, Veli Mäkinen, Jan von Plato, Aarne Ranta, Petri Mäenpää.

Erityinen kiitos vanhemmilleni, sukulaisilleni ja ystävilleni.

Lähteet

Ang59 Angel, R. W., The concept of psychic determinism. *American Journal of Psychiatry*, 116,5(1959), sivut 405–408. URL `http://ajp.psychiatryonline.org/cgi/content/abstract/116/5/405`.

Baa97 Baars, B. J., Some essential differences between consciousness and attention, perception, and working memory. *Consciousness and Cognition*, 6, sivut 363–371. URL `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=9262417&dopt=Citation`.

Dam95 Damasio, A. R., *Descartes' Error: Emotion, Reason and the Human Brain*. Avon, 1995.

Dam99 Damasio, A. R., *Feeling of What Happens. Body, emotion and the making of consciousness*. Vintage, 2000.

Dam03 Damasio, A. R., *Looking for Spinoza: Joy, Sorrow and the Feeling Brain*. Harcourt, Orlando, 2003.

Dav04 Davis, D., A glossary of Freudian terminology. `http://www.haverford.edu/psych/ddavis/p109g/fgloss.html`. [1.1.2007]

Hut03 Hutter, M., A gentle introduction to the universal algorithmic agent AIXI. Teoksessa *Artificial General Intelligence*, Goertzel, B. ja Pennachin, C., toimittajat, numero IDSIA-01-03, 2003, sivu 70, URL `http://www.hutter1.de/ai/aixigentle.htm`.

Hut04 Hutter, M., *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*. Springer, Berlin, 2004. URL `http://www.idsia.ch/~marcus/ai/uaibook.htm`.

Liv03 Livesley, W. J., *Practical Management of Personality Disorder*. Guilford Press, New York, 2003.

Mat00 Matthis, I., Sketch for a metapsychology of affect. *The International Journal of Psychoanalysis*, 81, sivut 215–227.

Nii84 Niiniluoto, I., *Johdatus tieteenfilosofiaan. Käsitteen- ja teorianmuodostus*. Otava, 1984.

Nil65 Nilsson, N. J., *Learning Machines: Foundations of Trainable Pattern-Classifying Systems.* McGraw-Hill, New York, 1987.

OCC88 Ortony, A., Collins, A. ja Clore, G. L., *The Cognitive Structure of Emotions.* Cambridge University Press, Cambridge, 1988.

OaJ96 Oatley, K. ja Jenkins, J. M., *Understanding Emotions.* Blackwell Publishers, Oxford, 1996.

Ort03 Ortony, A., On making believable emotional agents believable. Teoksessa *Emotions in Humans and Artifacts*, Trappl, R., Petta, P. ja Payr, S., toimittajat, MIT Press, Cambridge, 2003, sivut 189–212.

Pet04 Petta, P., Identifying theoritical problems. *Proceedings of the first HUMAINE Workshop on Theories and Models of Emotion*, June 2004, URL `http://emotion-research.net/ws/wp3/`.

Pic95 Picard, R. W., Affective computing. Tekninen raportti 321, MIT Media Laboratory Perceptual Computing Section, Cambridge, 1995.

Pic97 Picard, R. W., *Affective Computing.* MIT Press, Cambridge, 1997.

RuN95 Russell, S. ja Norvig, P., *Artificial Intelligence: A Modern Approach.* Prentice Hall, London, 1995.

SuB98 Sutton, R. S. ja Barto, A. G., *Reinforcement Learning: an Introduction.* MIT Press, London, 1998.

Sch97 Schmidhuber, J., A computer scientist's view of life, the universe, and everything. Teoksessa *Foundations of Computer Science: Potential - Theory - Cognition. Lecture Notes in Computer Science*, Freksa, C., toimittaja, Springer, 1997, sivut 201–208, URL `http://www.idsia.ch/~juergen/computeruniverse.html`.

Sch00 Schmidhuber, J., Algorithmic theories of everything. Tekninen raportti IDSIA-20-00, IDSIA, 2000. URL `http://www.idsia.ch/~juergen/computeruniverse.html`.

Sch02 Schmidhuber, J., The speed prior: A new simplicity measure yielding near-optimal computable predictions. *Proceedings of the 15th Annual Conference on Computational Learning Theory (COLT 2002)*, Kivinen,

J. ja Sloan, R. H., toimittajat. Springer, June 2002, sivut 216–228, URL http://www.idsia.ch/~juergen/computeruniverse.html.

Sli91 Slipp, S., *The Technique and Practice of Object Relations Family Therapy*. Jason Aronson, New Jersey, 1991.

Slo93 Sloman, A., The mind as a control system. Teoksessa *Philosophy and the Cognitive Sciences*, Hookway, C. ja Peterson, D., toimittajat, Cambridge University Press, 1993, sivut 69–110.

Slo94 Sloman, A., Explorations in design space. *Proceedings of the 11th European Conference on Artificial Intelligence*, August 1994, sivut 578–582.

Slo95 Sloman, A., Exploring design space and niche space. *Proceedings of the 5th Scandinavian Conference on Artificial Intelligence*. IOS Press, Amsterdam, May 1995.

SlS02 Sloman, A. ja Scheutz, M., A framework for comparing agent architectures. *UKCI'02: UK Workshop on Computational Intelligence, Birmingham.*, 2002.

SSJ01 Scherer, K. R., Schorr, A. ja Johnstone, T., toimittajat, *Appraisal processes in emotion: theory, methods, research*. Oxford University Press, New York, 2001.

Täh72 Tähkä, V., *Psykoterapian perusteet*. WSOY, 1972.

Täh93 Tähkä, V., *Mind and Its Treatment: A Psychoanalytic Approach*. International Universities Press, 1993.

Täh96 Tähkä, V., *Mielen rakentuminen ja psykoanalyyttinen hoitaminen*. WSOY, Helsinki, 1996.

Tan87 Tanimoto, S., *The Elements of Artificial Intelligence. An Introduction Using LISP*. Computer Science Press, Maryland, 1987.

Vog03a Vogt, P., Grounded lexicon formation without explicit reference transfer: who's talking to who? *Proceedings of ECAL'03, European Conference on Artificial Life*. Springer, 2003.

Vog03b Vogt, P., Thsim v3.2: The talking heads simulation tool. *Proceedings of ECAL'03, European Conference on Artificial Life*. Springer, 2003.

Hakemisto