

VI. Konnekciónizmus

A redukció és az intencionalitás problémája, illetve a funkcionalizmus kérdései ugyanúgy alkalmazhatók az állati elmére is, mint az emberi elmére. Az eddigi gondolatmenetek tehát nem feltétlenül és nem kizárólag az emberi elmére vonatkoztak.

Az emberi és az állati elmék között azonban olyan nyilvánvaló és szembevető különbségek is vannak, amelyekről az elmefilozófiának és a kognitív tudománynak számot kell adnia. Az emberek képesek olyan bonyolult absztrakt, nyelvi reprezentációk kezelésére, amelyre az állatok – úgy tűnik – nem képesek. Az evolúciós megközelítés ezen a ponton igen nehéz helyzetben van: egyfelől az ember és az állat közötti folytonos, graduális átmenetet kell hangsúlyoznia, mivel az emberi elme az állati elméből evolválódott, másfelől azonban az emberi gondolkodás néhány jellemzője, elsősorban a szabálykövetés, a nyelv és az absztrakció igazi válaszfalnak tűnik, s ez az emberi és állati elme közötti diszkontinuitás mellett szól.

A következő fejezetek kifejezetten a csak az emberre jellemző, humánspecifikus reprezentációkkal, illetve azok evolúciójával foglalkoznak, amely végül az igen bonyolult nyelvi, absztrakt reprezentációk és szabályok megjelenéséhez vezetett. E fejezet az átvezetést jelenti az eddigi, az emberi és állati elme között nem különbséget tevő, a kettő hasonlóságát hangsúlyozó rész és az ezután következő, a különbségekkel foglalkozó rész között. Az utóbbi tíz-tizenöt év legjelentősebb kognitív tudományi fordulata, a konnekciónizmus ugyanis az elme leírásának általános, mind az állati, mind az emberi elmére jellemző modelljét adja, de ugyanakkor rámutat arra a fontos különbségre is, hogy a szabálykövetés olyan képesség, amely csak az emberi elmére jellemző.

1. A klasszikus szimbólumkezelő modell

A kognitív tudományba a nyolcvanas években robbant be a konnekciónista megközelítés.¹⁷⁵ A konnekciónizmus néhány igen radikális állítása alapjaiban kérdőjelezte meg az elme leírásának uralkodó nézeteit, amelyek hozzávetőlegesen megfeleltethetők a gépi funkcionalizmus nézeteinek. Mielőtt a konnekciónizmus állításait elemeznénk, érdemes viszonylag részletesebben áttekinteni, mi is az a klasszikus elmemodell, amely ellen a konnekciónista irányzat küzd.

A klasszikus elmemodellekben a mentális reprezentáció diszkrét, egymástól független szimbólumokból áll. Ezek a szimbólumok megfelelnek természetes nyelvi fogalmainknak, és nem függenek a kontextustól, tehát zárt, oszthatatlan egységek. Az emberi elme működése így nem más, mint diszkrét szimbólumok manipulációja, amelyet explicit (vagy azzá tehető) szabályok vezérelnek. A szimbólumok közötti kapcsolat nem túl rugalmas; vagy fennáll egy meghatározott típusú kapcsolat – például kauzalitás –, vagy nem: közbülső eset nincs. A szabályokat pedig sorban, egymás után alkalmazza az emberi elme; egyszerre mindig csak egy műveletet végezhet. Ez a – több szempontból karteziánus ihletésű – modell sok tekintetben emlékeztet egy mai személyi számítógép működésére, hiszen mindkét rendszer kontextusfüggetlen, diszkrét szimbólumokból és az ezeket manipuláló explicit, szükségszerű szabályokból áll.¹⁷⁶

A klasszikus, szimbólumkezelő elmemodellek jellemzésére érdemes bevezetni a *szemantikai áttetszőség* fogalmát. Egy rendszer akkor szemantikailag áttetsző, ha az általa használt reprezentációk egyértelműen megfelelnek természetes nyelvi fogalmainknak – ilyen például a „labda” vagy a „macska”. A hangsúly itt az egyértelműsége van, hiszen minden reprezentáció körülírható természetes nyelvi fogalmainkkal, de csak a szemantikailag áttetsző rendszereknél van egyértelmű megfeleltetés a mentális reprezentáció és a szavak között. Az elme ilyen modellje azonban több ponton is támadható.

Mindenekelőtt az emberi elme mentális reprezentációi általában nem függetlenek a kontextustól: a labda szóhoz például más-más kontextusban egészen különböző mentális reprezentációk tartoznak, a kicsi és kemény golfabdától a nagy és puha strandlabdáig. Egy szónak tehát több reprezentáció is megfelelhet, attól függően, milyen kontextusban jelenik meg. A szimbólumok, amelyek egymástól és a kontextustól is független, zárt egységek, nyilvánvalóan nem képesek az elme ilyen rugalmas működésének magyarázatára.

A szimbólumkezelő rendszerek legnagyobb hiányossága ezek szerint az, hogy csak meghatározott mentális folyamatok leírására képesek: például a sakkjáték vagy egy geometriafeladat megoldásának leírására. De az intuitív képességeket igénylő feladatok ilyen modellálása már nem túl meggyőző (a művészi alkotásról nem is beszélve). A percepció pedig, amely folyamatos, gyors feldolgozást igényel, elképzelhetetlen a lassú, egyszerre csak egy művelet elvégzésére képes szimbólumkezelő rendszerek segítségével. Ez az ellenérv azért különösen érdekes, mert egy erre adott megoldáskísérlet – Jerry Fodor modularizmusa – az egyik legnagyobb befolyású

¹⁷⁵ Rumelhart–McClelland 1986a, Hinton 1984, Feldman–Ballard 1982, Clark 1996, a filozófiai kérdésekről cf. Ramsey–Stich–Rumelhart 1991, cf. még Macdonald–Macdonald 1995, Pléh 1992, 1997, Nánay 1996, Pinker–Mehler 1988. Meg kell említeni a perceptron-elméletet is mint a konnekciónizmus legfontosabb, hetvenes évekbeli előképét. Rosenblatt 1962. Ellene: Minsky–Papert 1969.

¹⁷⁶ Fodor 1975, 1983, Pylyshyn 1984, Newell 1980, 1990.

szimbólumfeldolgozó modell, amellyel sokan az egész klasszikus programot azonosítják. Érdekes tehát ezt részletesebben tárgyalni.

A klasszikus megközelítés és a gépi funkcionálisak szerint az észlelés reprezentációi mind kvázinyelvi szimbólumok. Ha egy kicsit elfordítom a fejemet, és így a retinára vetülő kép eltolódik vagy elfordul, akkor eközben szimbólummanipulációt hajtok végre. A szimbólummanipuláció azonban igen komoly és sok időt igénylő folyamat, és semmiképpen nem alkalmas arra, hogy a percepció gyorsaságát és rugalmasságát modellálja.

Ezt a nehézséget próbálta meg leküzdeni Jerry Fodor elmélete.¹⁷⁷ Nála az elmének csak egy része – az úgynevezett „Központi Feldolgozó Egység” – explicit, kvázinyelvi szimbólumrendszer, és nem az egész elme. Maga a Központi Feldolgozó Egység pontosan az előbb leírtaknak megfelelően működik: egy egységes formális nyelv – Fodor ezt a Gondolat Nyelvének nevezi – kijelentéseinek összessége. A Központi Feldolgozó Egységet modulok kötik össze az érzékszervekkel, a modulok bemenete tehát az érzet – például a hőses képe a retinán –, kimenetük pedig a Gondolat Nyelvére lefordított egyedi kijelentés: „esik a hó”. Ez a feldolgozási folyamat következtetések sorozata, ugyanakkor teljes mértékben független a központi feldolgozó egységtől. A modulok tehát elkülönülnek mind egymástól, mind a formalizált szimbólumoktól: a percepció és a kogníció – bár mindkettő következtetésekkel dolgozik – két független szférát alkot. A kogníció információi nem hozzáférhetők a percepció számára, és a percepciót nem kontrollálhatja a kogníció. Fodor tehát a második fejezetben már bemutatott modularizmus-interakcionizmus vitában a modularista álláspontot képviseli.

Fodornak ezzel sikerült kiküszöbölnie a gépi funkcionális egyik legkevésbé meggyőző tézisét, amely azt állítja, hogy az egész elmében, a percepció legalsó szintjén is explicit kvázinyelvi szimbólumokat kell feltételezni. De ezért súlyos árat kell fizetnie. A modulok elméletének ugyanis több vitatható pontja is van. Egyrészt a Központ és periférián elhelyezkedő modulok képe erősen emlékeztet a test-lélek dualizmusra: az önmagukba zárt, buta, mechanikus modulok a testből érkező idegpályák meghosszabbításai, míg a központi feldolgozó egység a testtől teljes mértékben független szubsztanciát alkot, amely végső soron nem más, mint a lélek fogalmának új, bár leszűkített változata. Ez az elmemodell azonban más, ennél súlyosabb és kézzelfoghatóbb problémákat is felvet.

Fodor szerint a modulok működése következtetések sorozata. A szükségszerű következtetések – nevezhetjük őket explicit szabályoknak is – állandók, nem változhatnak. Semmilyen módon nem befolyásolhatja a következtetést tehát az, hogy korábban milyen ingereket érzékelt. Fodor szerint a retina képének egy eleme – például egy borosüveg – szükségszerűen, minden korábbi tapasztalat nélkül meghatározza a modul kimenetének megfelelő kvázinyelvi kijelentést – „Látok egy borosüveget”. Ezen a ponton viszont problematikussá válik a tanulás kérdése. Hogyan szereztem meg azt a tudást, hogy ez egy borosüveg? A korábbi ingerek a következtetés szempontjából irrelevánsak, tehát ha a percepció valóban következtetések sorozata, akkor Fodor számára egyetlen magyarázat lehetséges: a következtetésekből álló modulok (és így a borosüveg fogalma is) velünk születettek.

Nem nehéz észrevenni, hogy a modulok velünk születettsége szükségszerű következménye annak, hogy a modulok működését következtetések sorozataként fogjuk fel. A probléma igazi forrása az, hogy Fodor a percepciót következtetések láncolatának tekinti. Ezt a feltevést azonban Gilbert Ryle már a század közepén meggyőzően támadta.¹⁷⁸ Ryle elsődleges fontosságúnak tekinti a perceptuális tanulást, amely kizárja az állandó, változásra képtelen következtetés fogalmát. Szerinte a percepció nem alapulhat következtetéseken, hiszen az egyes ingerfeldolgozását minden esetben befolyásolja az, hogy korábban milyen ingereket észleltem.¹⁷⁹

Fodor modulelmélete egy másik oldalról is támadható. Karmiloff-Smith számos fejlődéslélektani kísérlet alapján amellet érvel, hogy az újszülött még egyáltalán nem rendelkezik modulokkal, azok csak később, fokozatosan fejlődnek ki.¹⁸⁰ Az elme modularista leírásának megfelelő, külső interakció számára zárt idegsejtcsoportok létezésére van neurobiológiai bizonyítékunk, Fodor tehát nem akkor tévedett, amikor modulok segítségével írta le az emberi elmét, hanem amikor a modulokat merev, változásra képtelen következtetések sorozatának tekintette. Nem modulokról kell tehát beszélni az elme vizsgálatokor, hanem a modulok kialakulásáról, a modularizáció folyamatáról.

E rövid kitérő, Fodor modularizmusának bemutatása után térjünk vissza a klasszikus szimbólummanipuláción alapuló elmemodellhez és annak bírálatához. Ezek közül kettőt szeretnék kiemelni. Az első egy evolúciós jellegű érv. Egy evolúciósan sikeres rendszer – márpedig ilyen az ember is – jelenlegi formájában sok mindent őriz fejlődésének korábbi stádiumaiból. Az evolúciós sikeresség ugyanis előfeltételezi, hogy a rendszernek kifejlődése minden pillanatában életképesnek kell lennie.¹⁸¹ Olyan ez, mintha egy autót úgy kellene felújítani, hogy a felújítás közben minden pillanatban gurulnia kell. Az ilyen rendszer tehát nem képes radikális változásokra, csak finom, fokozatos összecsiszolódásra. Ezért az emberi elme magasabb szintű tevékenységei – így a sakk és a

¹⁷⁷ Fodor 1983, 1996a. Róla cf. Pléh 1986.

¹⁷⁸ Ryle 1974.

¹⁷⁹ Hasonló eredményre jut Gibson is. Cf. Gibson 1966, 1979. Róla: Fodor–Pylyshyn 1981, Ullman 1980, Nánay 1997c.

¹⁸⁰ Karmiloff–Smith 1992, 1996.

¹⁸¹ Ezt az evolúciós sajátosságot a „barkácsolás” (bricolage, tinkering) jelenségének szokták nevezni. Cf. Jacob 1986, Gould 1990, 1996, Clark 1996.

geometriapélda megoldása is – nem egyik pillanatról a másikra jöttek létre, hanem fokozatosan fejlődtek ki az ember alacsonyabb szintű mentális folyamataiból: például a percepcióból és a mozgáskoordinációból. Az ember szimbólummanipuláló képességei tehát erősen magukon viselik egy alacsonyabb szintű mentális folyamat jegyeit.

A másik fontos érv a szimbólummanipuláló rendszerek ellen John Searle már említett kínai szoba argumentuma.¹⁸² Searle szerint, ha elfogadjuk, hogy az elme szimbólumkezelő gép, ezzel azt állítjuk, hogy bár használnunk bizonyos fogalmakat, nem ismerjük e fogalmak – saját szavaink – értelmét. Képzeljünk el egy kínaiul nem tudó embert, akinek kínai szimbólumokat kell párosítania, kizárólag a formájuk alapján. Ez az ember – bár a párosítást talán kifogástalanul elvégzi – egyáltalán nem érti a kínai szimbólumokat: nem ismeri jelentésüket. Searle szerint pontosan ezt csinálja a szimbólumkezelő gép is, következésképpen nem érti saját szimbólumait. Bár manipulálja őket, számára ezek semmiféle jelentéssel nem bírnak. Ha pedig az elmét szimbólumkezelő gépnek tekintjük, szükségszerűen azt állítjuk, hogy az emberi elme nem képes a saját maga által használt fogalmak és szavak jelentésének megértésére, márpedig ez ellentmond a mindennapi tapasztalatnak.¹⁸³

Sokan ezek miatt a problémák miatt az egész elme-gép analógiát elvetik, a nyolcvanas évek közepén azonban megjelent egy markáns irányzat, amely bár továbbra is gépnek tekintette az emberi elmét, de egy, a klasszikus szimbólummanipuláció elveitől radikálisan különböző gépnek. Ez az irányzat a konnekciónizmus.

2. A konnekciónista fordulat

A konnekciónizmus az emberi elmét hálózatnak tekinti. A hálózat egyes csúcspontjai nem túl bonyolult egységek, a természetes nyelvi fogalmainknál lényegesen egyszerűbbek. A mentális reprezentáció tehát nem viszonylag kevés, de annál bonyolultabb, egymástól független szimbólumból áll, hanem éppen ellenkezőleg: a konnekciónizmus elmemodelljét sok, egyszerű műveletet végző egység közös, egymástól nem független működése határozza meg. Minden csúcspont sok másikkal van összekötve, változó intenzitású kapcsolatokon keresztül.

Ezek a változó intenzitású kapcsolatok helyettesítik a szabályokat, a konnekciónista modellben ugyanis nincsenek explicit (vagy explicitté tehető) szabályok.¹⁸⁴ Ez az egyik legfontosabb eltérés a klasszikus paradigmától. Egy szabály mindig szükségszerű: ha igaz, hogy „A-ból következik B”, és ha fennáll A, akkor szükségszerűen fennáll B is. Az elme konnekciónista modellje nem ismeri a szükségszerűség fogalmát. Ha a hálózatban A csúcspont aktív, és erős a kapcsolat A és B csúcspontok között, akkor nagy valószínűséggel B is aktiválódni fog, de ez egyáltalán nem szükségszerű.

A hálózat két csúcspontja közötti kapcsolaterősség ugyanis csak attól függ, hogy a két csúcspont milyen gyakran van egyszerre aktivált állapotban. Ha például a viharok és a villámlásnak megfelelő csúcspontokkal gyakran fordul elő, hogy egyszerre aktívak – tehát gyakran látunk egyszerre villámlást és vihart –, akkor a vihar-reprezentáció és a villámlás-reprezentáció közötti kapcsolaterősség nőni fog. Ez a kapcsolaterősség azonban csak elvileg érheti el a száz százalékot, gyakorlatilag soha. Az előbbi példánál maradva: bármilyen sokszor tapasztaljuk is egyszerre a villámlást és a vihart, ha Shakespeare *Vihar* című drámájáról beszélünk, nem feltétlenül fog aktiválódni fejünkben a villámlás-reprezentációja. A konnekciónista leírásban tehát két mentális reprezentáció között nem képzelhető el szükségszerű kapcsolat.

Nem nehéz felfedezni ezeken a gondolatokon David Hume több száz éves gondolatainak hatását: Az A és a B esemény közötti okozati összefüggés alapja nem a szükségszerűség, hanem a szokás, tehát az, hogy A és B esemény gyakran egyszerre történik meg.¹⁸⁵ Első közelítésben tehát azt lehet mondani, hogy a konnekciónizmus kétszáz év után visszahozta az ismeretelméleti gondolkodásba a szokás fogalmát a szabállyal szemben.¹⁸⁶ Látni fogjuk azonban, hogy a helyzet nem ennyire egyszerű.

Ha az elmében nincsenek szabályok, akkor nem szükséges feltételezni egy olyan központi egységet, amely a szabályok alkalmazását végezné – mint a szimbólumkezelő rendszerek esetében. Az elmének tehát nincs központi szabályalkalmazó egysége. Ebből viszont az következik, hogy mentális folyamataink nem feltétlenül sorban egymás után következnek be. Ha nincs centrális egység a tudatban, akkor elképzelhető, hogy párhuzamosan több mentális folyamat is zajlik a fejünkben. Ez lényegesen megnövelné az elme működésének sebességét. Ebben a kérdésben a konnekciónizmus lényegesen jobb leírását adja a valóságban hihetetlenül gyorsan működő

¹⁸² Searle 1996.

¹⁸³ Searle úgy gondolja, hogy érve nemcsak a szimbólumkezelő, de a konnekciónista rendszerek ellen is hatásos. Cf. Searle 1993.

¹⁸⁴ Röviden hivatkozni kell itt a szabály fogalma körüli legfontosabb filozófiai vitára, a szabály-externalizmus és szabály-internalizmus vitájára, amelynek fő kérdése az, hogy a szabályok a fejünkben vannak-e, vagy azon kívül. A szabály-internalizmus azt mondja ki, hogy a szabályok az elmében vannak, minden társadalmi szabály ebből ered. A szabály-externalizmus tagadja ezt, és képviselői azzal érvelnek, hogy a szabály nem lehet az elme belső ügye, csak a nyilvános, külső, társadalmi szabályrendszerekből eredeztethető. Cf. Kripke 1982, Wittgenstein 1992, Katz 1990, Schauer 1991.

¹⁸⁵ Hume szokásfogalmához, illetve a szükségszerű kapcsolatról l. Hume 1995 hetedik fejezet, pp. 58–76.

¹⁸⁶ Cf. Pléh 1992.

agynak, mint a szimbólumkezelő modell, amelynek egyik legsúlyosabb hiányossága éppen az, hogy lassú. Több szerző is a párhuzamosságot tekinti a konnekciónizmus igazán lényeges újításának.¹⁸⁷

Az elme konnekcionista modellje erősen emlékeztet az agy neuronhálózatára, innen származik a hálózat alapötlete is. Fontos azonban megjegyezni, hogy ez nem több, mint metafora. A konnekcionista hálózat csúcspontjai nem neuronok, hanem ennél nagyobb – de a nyelvi szimbólumoknál kisebb – egységek. Ezeket mikrojegyeknek vagy szubszimbólumoknak nevezzük.¹⁸⁸ Egy természetes nyelvi fogalmunknak tehát sok száz szubszimbólum összessége felel meg. Ha például meghalljuk azt a szót, hogy „macska”, akkor olyan szubszimbólumok fognak aktiválódni, mint „szőrös”, „nyávog”, „bajsz van”, „dorombol”, „kandúr”, de esetleg még olyanok is, mint „Tom és Jerry” vagy Spiró *Csirkefejének* porolóra felakasztott macskája. Ezekon kívül természetesen sok olyan szubszimbólum is aktiválódni fog, amelynek nem feleltethetők meg természetes nyelvi fogalmak: ilyenek például a képek és a funkcionális reprezentációk. Hogy ezek közül mely szubszimbólumok fognak valóban aktiválódni, illetve melyek erősen és melyek gyengébben, azt a kontextus és az illető személy korábbi – a macskákkal kapcsolatos – tapasztalatai határozzák meg.

Ez a példa elvezet a konnekciónizmus egyik legfontosabb állításához: a tudást a kapcsolatok hordozzák. Amit a macskákról tudok, kizárólag azon múlik, milyen erősségű kapcsolatok vezetnek a „macska” szótól az egyes szubszimbólumokig. Egy szó jelentése tehát nem magában a szóban van kódolva – ez volt a szimbólumkezelő megközelítés álláspontja. A jelentést kapcsolatok valósítják meg, a szó és más szavak, reprezentációk közötti kapcsolatok.

Ehhez a kérdéshez kapcsolódik a prototípus fogalma. A konnekcionista hálózat igazi erőssége az, hogy egyedi esetekből – az emberi agyhoz hasonlóan – képes rugalmasan általánosítani. Képes tehát prototipikus reprezentációt alkotni kizárólag egyedi ingerek alapján. Ha ugyanis a reprezentáció szubszimbólumok közötti kapcsolatok összessége, akkor az egyedi inger egy meghatározott erősségű kapcsolatot jelent két szubszimbólum között. Sok egyedi inger után pedig a két szubszimbólum között kialakul az egyedi kapcsolaterősségek súlyozott matematikai átlaga, és ez az átlag nem más, mint az egyedi ingerek prototípusa – erre a két szubszimbólumra nézve.

Ugyanez a gondolatmenet kiterjeszhető természetes nyelvi fogalmainkra is, amelyek több száz szubszimbólum közötti kapcsolatnak felelnek meg. Ha az egységek közötti kapcsolaterősségek mindegyike a korábbi egyedi ingerek kapcsolaterősségeinek átlaga, akkor a több száz szubszimbólum közötti kapcsolatok összessége nem más, mint a vizsgált természetes nyelvi fogalom prototipikus reprezentációja. Ezért képes a konnekcionista hálózat – hasonlóan az emberi agyhoz – olyan prototipikus reprezentáció létrehozására is, amely esetleg minden kapott egyedi inger mintázatától különbözik.¹⁸⁹

A hálózat pillanatnyi állapota – a korábbi egyedi ingerek általánosítása – prototipikus reprezentáció, azonban ez a prototípus is egyedi entitás: ugyanazon szubszimbólumok közötti kapcsolatokkal írható le, mint az egyedi ingerek. A konnekciónizmus tehát nem fogadja el az egyedi és az általános klasszikus szembeállítását, mert az általánost is egyediként írja le. Amikor tehát az egyedi inger feldolgozódik, nem absztrakt, általános kategóriák alá rendelődik, hanem egy másik egyedi reprezentációval hasonlítódik össze. Mivel mind az általános kategória, mind az egyedi tapasztalat ugyanazon szubszimbólumok közötti meghatározott erősségű kapcsolatok összessége, nincs értelme két különböző reprezentáció – egyedi és általános – ellentétéről beszélni, hiszen ez a két entitás csak néhány kapcsolat erősségének fokozatában különbözik egymástól.

Legvilágosabban ez a konnekcionista percepcióelméletben mutatkozik meg. Itt a klasszikus álláspont szerint – kis leegyszerűsítéssel – a kétdimenziós ingereket a tárgy háromdimenziós, nézőpontfüggetlen mentális reprezentációjával hasonlítom össze. A konnekcionista percepcióelmélet szerint azonban a tárgyaknak nincs is háromdimenziós mentális reprezentációjuk.¹⁹⁰ A percepció ezek szerint a kétdimenziós ingernek és a tárgy kétdimenziós reprezentációinak az összehasonlítása.

A konnekciónizmus állításai mögött észre lehet venni egy alapvető előfeltevést, amelyen mindegyik lényeges tétel alapul: az elme analóg rendszer. Egy rendszert akkor nevezünk analógnak, ha két állapota között nincs minőségi ugrás, csak fokozatos átmenet. Ebben különbözik az analóg rendszer a digitálistól, ahol a rendszer két állapota között csak minőségi ugrás képzelhető el, a folyamatos átmenet itt nem lehetséges. A konnekcionisták előfeltevése – az elme analóg rendszer – azt jelenti tehát, hogy az emberi agy működésének alapegysége nem a diszkrét, lehatárolt állapotok közötti ugrásszerű váltás, hanem a folytonos átmenet. Ezt a konnekciónizmus úgy valósítja meg, hogy a mentális reprezentációt nem diszkrét szubszimbólumokra vezeti vissza, hanem a szubszimbólumok közötti, folytonosan változó kapcsolaterősségekre.

Szorosan összefügg ezzel a hálózat egy kevésbé megnyugtató tulajdonsága is: a konnekciónizmus nem ismeri az ellentmondás fogalmát. Ha a hálózat két egymásnak ellentmondó adatot kap, a nekik megfelelő két különböző

¹⁸⁷ Innen származik az irányzat másik elterjedt elnevezése, PDP, amely a Parallel Distributed Processing (párhuzamosan megosztott feldolgozás) rövidítése. Ez a címe Rumelhart és McClelland 1986-os híres könyvének is.

¹⁸⁸ Clark mikrojegyeknek, Smolensky szubszimbólumnak nevezi ezeket az egységeket. Clark 1996, Smolensky 1996. Ma az utóbbi elnevezés az elterjedtebb.

¹⁸⁹ A prototipikus általánosításról lásd Rumelhart–McClelland 1986a.

¹⁹⁰ Cf. Edelman–Weinshall 1989, Poggio–Edelman 1990.

kapcsolaterősség súlyozott matematikai átlaga lesz az eredmény. Az elme konnekcionista modellje tehát nem képes az ellentmondás felismerésére, éppen azért, mert analógnak tekinti az agy működését.¹⁹¹

Az elme konnekcionista leírása meggyőző eredményeket mutatott a mesterséges intelligencia kutatásának területén. A konnekcionista hálózatok sok feladatot hasonlóan végeznek, mint az ember. A hálózat például képes az angol igék – rendhagyó és szabályos – múlt idejének megtanulására, minden előzetes tudás, beépített explicit szabályok nélkül, kizárólag a tapasztalat alapján. Sőt a hálózat eközben nagyjából ugyanolyan típusú hibákat követ el, mint egy átlagos angol gyerek.¹⁹²

A hálózatok és az emberi elme hasonlóan reagál a rendszer kisebb-nagyobb sérüléseire, rendellenességeire is. Ha egy szimbólumkezelő rendszer sérül – például törlődik belőle egy explicit szabály –, akkor az egész rendszer működése megáll. A konnekcionista hálózat viszont képes tovább működni a rendszer kisebb sérülése esetén is, akárcsak az emberi agy. A nagyobb károsodások is hasonló változásokat okoznak, például az amnézia betegsége az emberi agyban ugyanazokat a tüneteket mutatja, mint egy részben lerombolt konnekcionista hálózat.¹⁹³

A konnekcionizmus legradikálisabb, valóban új és igazán jelentős gondolatai a hálózat tanulásával foglalkoznak.¹⁹⁴ A konnekcionista hálózatban a tudást a kapcsolatok hordozzák – ezt láttuk. Csakhogy ezek ugyanazok a kapcsolatok, amelyeken az új ingerek feldolgozódnak. Az újabb tapasztalat ezeken az életrősségeken változtat. A következő tapasztalat a már megváltozott hálózaton keresztül megy végbe. Az egyes tapasztalatok tehát befolyásolják az újabb tapasztalatok feldolgozását. Ha a tüzről egyszer azt érzékelttem, hogy éget, akkor ez a hálózat megfelelő helyén egy (vagy több) életrősséget megváltoztat. Következésképpen a legközelebbi tűzpercepcióm már különbözni fog az előzőtől, hiszen a hálózat egy pontján kódolva van az információ, hogy a tűz égethet is. Leegyszerűsítve: ha egyszer megégetem a kezem, utána a tűz és a fájdalom csúcspontja közötti kapcsolat erősebb lesz.

Nagyon hasonlít ez Gilbert Ryle gondolataihoz. Ryle – akárcsak a konnekcionizmus – a tanulás szerepének hangsúlyozásával akarja elhatárolni az elme működését egyfelől a szabály, másfelől a szokás fogalmától.¹⁹⁵ Hiba tehát a konnekcionizmust a Hume-féle asszociáción alapuló ismeretelmélet egyszerű újjáélesztésének tekinteni. A konnekcionizmus valódi, nem pusztán felszíni egyezéseket mutató előzménye Gilbert Ryle – bár ő nem beszél sem hálózatokról, sem asszociációról.¹⁹⁶ A konnekcionizmus Ryle elvont elméletét tette megfoghatóvá és modellezhetővé, azáltal, hogy az elmét hálózatnak tekinti.¹⁹⁷

Ezen a ponton lesz egyértelmű a különbség a korábbi – például Hume-féle – asszociáción alapuló rendszerek és a konnekcionizmus között. Hume megkülönbözteti az idea és az impresszió fogalmát.¹⁹⁸ impresszióknak nevezi a minket érő ingerek összességét, ideának pedig a mentális reprezentációt.¹⁹⁹ Hume szerint az impresszió határozza meg az ideát.²⁰⁰ Kant és Schelling óta pedig rengeteg különböző kísérlet történt arra, hogy az ideát és az impressziót úgy fogják fel, mint amelyek egymást határozzák meg: az idea az impressziót, az impresszió az ideát.

A konnekcionista hálózat meglehetősen éles fordulatot hozott ebben a régi filozófiai problémában. Ha ugyanis a tudást (a mentális reprezentációt, az ideát) a hálózat életrősségei hordozzák, az ingert (a tapasztalatot, az impressziót) pedig a hálózat csúcspontjainak aktiválódása, akkor az idea és az impresszió nem két különböző dolog, hanem egy érem két oldala: ugyanaz a hálózat, különböző aspektusokból értelmezve. Az idea az élek erőssége, az impresszió a csúcspontok aktivitása. Itt tehát nem egy egyedi esetet kell általános törvény hatálya alá hozni, hiszen az idea is csupa egyedi esetből áll: egyedi esetek invarianciájából, prototípusából. Ezért képes minden egyes impresszió megváltoztatni az ideát, amely a következő impressziót így már kicsit másképpen fogja feldolgozni.

A konnekcionista hálózatokkal azonban egy ponton komoly probléma van. A magasabb szintű mentális tevékenységeket – például a sakkot vagy a tudományos bizonyítást – nem lehet kizárólag a hálózatok segítségével leírni. Igaz, hogy hétköznapi kifejezéseink túlnyomó többsége szubsimbólumokból áll, de például a tudományos nyelvhasználat szavainak általában egy és csak egy meghatározott jelentés felel meg. Tudományos fogalmaink

¹⁹¹ Ezt a hiányosságot emeli ki a konnekcionizmus leghíresebb kritikája: Fodor–Pylyshyn 1988

¹⁹² Rumelhart–McClelland 1986b, 1987, vannak, akik megkérdőjelezik ezt az eredményt, cf. Pinker–Prince 1988, 1994. A vitáról jó összefoglalást ad Lima–Corrigan–Iverson 1994.

¹⁹³ McClelland–Rumelhart 1986b.

¹⁹⁴ A tanulás szerepének felértékelődése miatt a konnekcionizmus gyakran tagadja az innát, velünk született mentális tulajdonságok létezését. Hogy néhány emblemikus megoldást lássunk a konnekcionizmus és az innatizmus közötti sokrétű kapcsolatra, Rumelharték klasszikus könyve radikálisan antiinnatista, cf. Rumelhart–McClelland 1986a. Egy kifinomultabb, de szintén innatizmussal szembeálló elméletet vázol Elman et al. 1996., Marcus 1998 viszont kísérletet tesz egy olyan konnekcionista modell kidolgozására, amelynek jelentős innatista implikációi vannak.

¹⁹⁵ „A pusztán szokásból végzett gyakorlat lényege az, hogy az egyik ténykedés másolata az előzőknek. Az értelmes gyakorlat lényege pedig az, hogy az egyes ténykedéseket módosítják a megelőzők. A cselekvő személy ugyanis még tanul.” – írja Ryle (Altrichter Ferenc fordítása), cf. Ryle 1974. pp. 57–58.

¹⁹⁶ A konnekcionizmus különböző előzményeiről cf. Nánay 1996, Nánay 1997a, Pléh 1992.

¹⁹⁷ Cf. Nánay 1997b.

¹⁹⁸ A magyar fordításban az impresszió fogalmát – kissé félrevezető módon – benyomásnak is fordítják.

¹⁹⁹ Ez persze a kognitív terminológia felőli némileg anakronisztikus utólagos értelmezés.

²⁰⁰ Hume 1976, első könyv, cf. még Hume 1995, második fejezet.

reprezentációja tehát szemantikailag áttetsző, ellentétben a konnekcionista hálózattal. Ezen kívül – mint láttuk – a konnekcionizmus nem képes az ellentmondás és az explicit szabályok kezelésére sem.

Ezen a ponton érdemes bevezetni egy újabb ryle-i distinkciót. Ryle megkülönböztetése szerint a „tudni, hogyan” tudásfajtája éppen abban különbözik a „tudni, mit”-től, hogy nem explicit tudás.²⁰¹ Amikor gondolkodom, amikor biciklizni tanulok, és amikor bekötöm a cipőfüzőmet, ezt nem explicit mentális reprezentációk, szimbólumok alapján teszem. Ryle gondolatmenetét kissé leegyszerűsítve a konnekcionista hálózat tudása Ryle „tudni, hogyan” fogalmának felel meg, a „tudni, mit” pedig a klasszikus szimbólum-feldolgozó rendszer tudására jellemző.²⁰² Ahogyan Ryle szerint az emberi tudás csak együttesen magyarázható a „tudni, mit” és a „tudni, hogyan” fogalmával, az elme leírása sem lehetséges csak a konnekcionizmus vagy csak a klasszikus paradigma segítségével.

Nem sokkal a konnekcionizmus megjelenése után tehát kiderült, hogy önmagában ez a felfogás nem képes az elme komplex működésének magyarázatára. A megoldás elég logikusnak látszott: a klasszikus szimbólumkezelő modell és a konnekcionizmust össze kell olvasztani. Meg kell próbálni egyesíteni a két ellentétes megközelítés előnyeit. Az ilyen kettős modellek a kilencvenes évek elejétől mindmáig az elme kognitív tudományi leírásának szinte kizárólagos modelljei.

3. Kettős modellek

A konnekcionizmus és a szimbólummanipuláción alapuló elmemodellek összebékítésére annyi és annyféle megoldási javaslat született, hogy elengedhetetlennek tűnik ezeknek a kettős modelleknek valamilyen rendszerezése. Első közelítésben meg kell különböztetni a hibrid, illetve az integrált modelleket.²⁰³

Hibrid modelleknek nevezzük azokat a leírásokat, amelyek szerint az elme egyik része konnekcionista módon működik, egy másik viszont szimbólumkezelő rendszer.²⁰⁴ A sakkozás például szimbólumkezelés, a percepció viszont konnekcionista folyamat. A hibrid modell tehát a konnekcionista és szimbólumkezelő rendszer közötti különbséget kizárólag az elme különböző területei, funkciói közötti különbségre vezeti vissza. A kettős modellek másik, a hibrid modelltől különböző típusa az integrált modell. Az integrált modell egy rendszer két-szintű leírása. Az elme működését tehát voltaképpen kétféle módon kell leírni, mint szimbólumkezelést és mint a konnekcionista hálózat változását. Egy jelenség van tehát, csak a leírási szintek különböznek.

Hibrid modellre a legjobb példa Hernád István, azaz Stevan Harnad elmélete.²⁰⁵ Harnad kiindulópontja a már bemutatott Searle-féle kínai szoba argumentum. Harnad továbbgondolja Searle érvét: ha valaki meg akarja tanulni kínaiul, és csak egy kínai–kínai szótár állna a rendelkezésére, akkor az egyik kínai szimbólum jelentését csak egy másikkal tudná megadni, és így tovább a végtelenségig. De ebből a gondolatmenetből Harnadnál nem az következik, hogy nincsenek szimbólumaink, hanem csak az, hogy szimbólumaink önmagukban, a világtól függetlenül nem képesek jelentést hordozni. A szimbólumok és külvilág – a percepció – között folyamatos kapcsolat van. Az elmemodelleknek tehát meg kell magyarázniuk a szimbólumok lehorgonyzását a külvilághoz való kapcsolatát, s ezt Harnad szerint egy konnekcionista alrendszer valósítja meg az emberi elmében.

Világos, hogy Harnad elmélete is kettős, mégpedig ezen belül hibrid modell. A magasabb szintű tevékenység – például a nyelv, a sakk – pusztán szimbólummanipuláció, amelynek nincs semmi köze a konnekcionista hálózatokhoz. A percepció viszont, amely a szimbólumok jelentését, illetve változását biztosítja, tiszta konnekcionista hálózat: független mindenfajta szimbólummanipulációtól. A Harnad-féle hibrid modell tehát két részre osztja az elmét: tisztán konnekcionista és tisztán szimbólumkezelő részre. Ezek élénk kapcsolatban vannak ugyan egymással, de az egyik soha nem írható le a másik fogalmaival. Visszatérő hiányosságuk az ehhez hasonló hibrid rendszereknek – és ez alól Harnad sem kivétel –, hogy nem igazán tudják modellezni a konnekcionista és a szimbólumkezelő alrendszerek közötti belső kommunikációt. Ha – mint Harnad is mondja – a percepció kimenetei konnekcionista típusú, kontextusfüggő, szubszimbolikus reprezentációk, a magasabb szintű mentális folyamatok bemenetei pedig kontextustól független szimbólumok, akkor a kétféle reprezentáció egymásba alakulását is meg kell magyarázni.

Harnad valójában a fodori modularista modellt terjeszti ki olyan módon, hogy nála az egyes modulok szerkezetét konnekcionista hálózatok alkotják. Ezzel kiküszöböli a Fodor-féle modularizmus számos ellentmondását (elsősorban a reprezentációk jelentésének eredetét és az innatizmust illetően), de az alapvető modularista feltevéseket – így például a kontextusfüggetlenséget – nem vitatja.²⁰⁶ Harnad rendszere ugyanis rendkívül leegyszerűsíti-

²⁰¹ Ryle 1974. Második fejezet.

²⁰² A tudni, mit és tudni, hogyan közötti különbségtevés persze sok esetben egyáltalán nem egyszerű feladat. A nyelvtani tudásról például nem könnyű eldönteni, hogy melyik kategóriába tartozik. Hasonló problémák merülnek fel a perceptuális tudással is.

²⁰³ Ez a felosztás Smolenskytól származik. Cf. Smolensky–Legendre–Miyata 1992. Van persze olyan konnekcionista modell is, amely kimarad ebből a klasszifikációból, pl. Shastri–Ajjanagadde 1993.

²⁰⁴ Hibrid modellek például: Hender 1989a, 1989b, Clark 1991, 1994, Harnad 1994.

²⁰⁵ Harnad 1994.

²⁰⁶ Harnad elmélete csak egyik esete a konnekcionizmus és a modularizmus igen változatos viszonyának. Az első konnekcionista modellek radikálisan interakcionisták voltak, olyannyira, hogy az elme hálózatában minden csúcspontot minden másik csúcsponttal összeköthető-

tett nyelvfogalmat használ. Egy szó jelentésének rögzítéséhez szerinte elégséges különböző képek invariáns jegyeiből összeálló újabb – immáron egyetlen – kép. A „ló” szó jelentését például a lovak különböző nézőpontú, színű és méretű képeinek invariáns jegyeiből összeálló kép egyértelműen meghatározza. Mindez a ló esetében még viszonylag hihető. A „hosszú” szó jelentését viszont egészen más képpel kellene meghatározni akkor, ha hosszú asztalról, illetve hosszú rúdról beszélünk (a hosszú fejezetekről nem is beszélve), és itt úgy tűnik, nem segít további invariáns jegyek keresése sem.

A „hosszú” szó jelentésének példája persze nem tőlem származik. Wittgenstein a *Filozófiai vizsgálódásokban* többek közt ezen példa segítségével kritizálta saját korábbi nyelvelméletét.²⁰⁷ Harnad modellje tehát ugyanúgy nem képes kezelni a kontextus fogalmát, mint a *Logikai-filozófiai értekezés*, hiszen mind a fiatal Wittgenstein, mind Harnad olyan túlegyszerűsített nyelvfogalmat használ, amelyben nincs helye sem a szavak használatának, sem kontextusuknak.

A konnekcionista kettős modellek hibrid formája tehát több ponton is támadható. Az integrált modellek azonban ki tudják kerülni a hibridek ellen felhozott ellenérvek egy részét. Az integrált modellek szerint a konnekcionista és a szimbólumkezelő működés különbsége nem az elme más és más területeinek feleltethető meg, hanem az elme különböző leírási szintjeinek.²⁰⁸ A szimbólummanipuláció nem valós jelenség, csak a konnekcionista hálózat néha úgy működik, hogy ez a működés jól közelíthető a szimbólummanipuláció leírásával. Hasonlóképpen a konnekcionista működés is csak jó közelítése az idegsejtek aktivitásmintázat-változásainak.

A különböző szintek egymáshoz való viszonya igen hasonlít a számítástechnikában jól ismert emuláció jelenségéhez. Ahogy egy Pentium is képes – megfelelő segédprogram segítségével – lefuttatni egy Commodore 64-es programot, képes tehát emulálni a Commodore 64 működését, úgy a konnekcionista rendszer is úgy tesz, mintha szimbólummanipuláción alapuló elmeműködés lenne, tehát csak emulálja a szimbólummanipulációt. Hasonlóképpen, az idegrendszer is csak emulálja a konnekcionista működést.

Az integrált modell végső soron nem más, mint az instrumentalizmus feltámasztása a konnekcionista fogalmi kereten belül. Akár Dennett instrumentalizmusa szerint, az integrált modellek szerint is minden, ami a fejünkben történik, idegsejtek tüzelése, de könnyebben tudjuk kezelni az elmét, ha úgy írjuk le, *mintha* magasabb szintű egységekből, mentális reprezentációkból épülne fel.²⁰⁹ Sem az instrumentalizmus, sem az integrált modellek szerint nincs a reprezentációknak valódi létezésük, ezek pusztán hasznos elméleti konstrukciók, amelyek a megértést segítik elő.

Az integrált modell annyiban nyújt újat a klasszikus instrumentalizmushoz képest, hogy nemcsak a szimbólumokat tekinti hasznos elméleti konstrukciónak, hanem közbülső szintként a szubszimbólumokat is. Mindazonáltal az integrált modell ugyanolyan cinikus és végső soron üres elmélet, mint az instrumentalizmus: nem azzal foglalkozik, hogy miként épül fel az elme, szerinte ez a kérdés nem vizsgálható, hanem azzal, hogy az elme milyen leírása jár a legnagyobb gyakorlati haszonnal, mely elmemodellek segítségével lehet a legjobban megjósolni az elme működését. Az integrált modell a klasszikus instrumentalizmus már bevett két elemzési szintje – a neurális és a mentális szint – mellett feltételez, és praktikus szempontból hasznosnak ítél egy harmadik, köztes elemzési szintet, és ez lesz a konnekcionista hálózat szintje. A konnekcionista hálózat azonban ugyanúgy csak elméleti konstrukció, mint a súlypont vagy a kvázinyelvi szimbólum.

Az integrált modell – akár az instrumentalizmus – nem igazán cáfolható és nem is támadható saját keretein belül, ennyiben igen becsületes elmélet: nem ígér túl sokat, de azt teljesíti is. Az elmét mindkét szinten le kell írni, konnekcionista hálózatként is és szimbólummanipulációként is, s ezután azt a leírást kell használni, amely jobban közelíti az adott jelenséget. Ez a sakkjáték esetében valószínűleg a szimbólummanipuláció lesz, míg a biciklizés esetében a konnekcionista hálózat, de tudnunk kell: az, hogy milyen leírási szintet választunk, a mi pragmatikus céljainktól függ, hiszen végső soron minden csak az idegsejteken múlik.

Érdekes evolúciós szempontból megvizsgálni a kétféle kettős modellt, annál is inkább, mert a konnekcionista elmeműködéssel szemben felhozott érvek mind kizárólag az emberi elmére vonatkoztak. Az állati elmék leírása igen jól elképzelhető a konnekcionista kereteken belül. Vannak azonban olyan képességek, amelyek hiányoznak a konnekcionista leírásból, de jelen vannak az emberi elmében – ilyenek például az explicit szimbólumok, a szabályok, az ellentmondás. Az az intuíció tehát, hogy az emberi elme, amely a konnekcionista mechanizmusok mellett explicit szimbólumok manipulációját is végzi, az állati elme alapvetően konnekcionista működéséből alakult ki. Ez esetben viszont evolúciós mechanizmust kellene adnunk a szimbólummanipuláción alapuló elmeműködés kialakulására.

Erre az integrált modellek nyilvánvalóan nem képesek, hiszen szerintük az állati és az emberi elme között nincs inherens különbség, a különbség csak annyi, hogy milyen modellek segítségével tudjuk megjósolni működésüket. Az állati elmére több esetben alkalmazható a konnekcionista modell, mint az emberi elmére, ez minden,

nek tekintettek. Cf. Rumelhart–McClelland 1986a, Smolensky 1988. Később megjelentek olyan konnekcionista elméletek is, amelyek éppen a modularizmus valamilyen formáját próbálták beépíteni a konnekcionizmusba. Cf. Elman et al. 1996.

²⁰⁷ Wittgenstein 1992. p. 135. § 251.

²⁰⁸ Integrált modellre elsősorban Smolensky elmélete a példa: Smolensky 1996, Smolensky–Legendre–Miyata 1992.

²⁰⁹ Érdekes módon Dennett a konnekcionizmust mégsem integrált modellként, hanem hibridként építi be elméletébe, cf. Dennett 1993.

amit az emberi és állati elme közötti különbségről az integrált modellek mondani tudnak. Hangsúlyozni kell, hogy persze ez önmagában koherens és jól védhető álláspont, de ha tovább akarunk lépni, és többet akarunk tudni az elméről az integrált modellek instrumentalizmusánál, akkor más megoldást kell keresnünk.

A hibrid modellek esetében, ha elfogadjuk, hogy az evolúció formálta az elmét, igen kevésbé meggyőző elméletet kapunk. A hibrid modellekben az emberi elme két részre osztható, az egyik konnekcionista hálózat, a másik szimbólumkezelő rendszer. Ha igaz, hogy a szimbólummanipuláció képességét az evolúció alakította ki az embernél, akkor azt kapjuk, hogy míg az evolúciós folyamatok az elme egyik felét érintetlenül hagyták, addig a másik felét olyan radikálisan megváltoztatták, hogy ott semmilyen emléke sem maradt az egykori konnekcionista működésnek. Ez az elképzelés pedig nemcsak intuitíve kevésbé valószínű, de számos evolúciós alapelvnek ellentmond, legvilágosabban a folytonos változás elvének, amely azt mondja ki, hogy a szervezetnek úgy kell megváltoznia, hogy közben minden időpillanatban életképes legyen.

Sem az integrált, sem a hibrid modell nem igazán meggyőző tehát, különösen nem az evolúciós szempontból. A következőkben egy harmadik lehetséges kettős modellt szeretnék körvonalazni, amely evolúciós alapokon nyugszik. Ennek vázolásához azonban fel kell hívni a figyelmet az eddig elemzett konnekcionista, szimbólumkezelő és kettős rendszerek egy közös problémájára, amely mindegyik – egyébként nagyon különböző – elmélettel szemben felvethető. Ez pedig a szubszimbólum-lehorgonyzás problémája.

Harnad látszólag megoldotta annak kérdését, hogy a percepciótól független szimbólumokat a percepcióhoz kapcsolja, szerinte a kontextusfüggetlen szimbólumokat a konnekcionista hálózatok töltik fel jelentéssel, azok horgonyozzák le a perceptuális tapasztalatban. Van azonban ennek az érvelésnek egy komoly hibája. Harnad ugyanis a szimbólumok jelentését visszavezeti a szubszimbólumok jelentésére, nem mondja meg azonban, hogy a hálózat csúcspontjai, a szubszimbólumok miként töltődnek fel jelentéssel. Nem ad választ tehát arra, hogy ha a szimbólumok a szubszimbólumokban vannak lehorgonyozva, akkor a szubszimbólumok hova horgonyozódnak le.

Ront a helyzeten, hogy a konnekcionista irodalom meglehetősen kétértelműen kezeli a szubszimbólum fogalmát. Néha egészen egyszerű, az idegsejteknél nem sokkal bonyolultabb egységeknek, funkcionális vagy geometriai vektoroknak tekinti őket, néha viszont a szimbólum bonyolult szemantikai szerkezetével rendelkező, de azért kicsit egyszerűbb szimbólumoknak, mint például a „macska” esetében a „szőrös” vagy a „dorombol” szubszimbóluma. Szinte minden konnekcionista teoretikus keveri a kétféle koncepciót.²¹⁰

Az egyszerű funkcionális vagy geometriai egységeknek megvan az az előnyük, hogy könnyen megoldható a lehorgonyozásuk. Komoly problémát jelent viszont, hogy ezekből az alacsony szintű reprezentációkból hogyan tud összeállni egy komplex szemantikával rendelkező szimbólum. A szubszimbólumok e felfogása tehát túl alacsony szintű. A másik koncepció („szőrös”, „dorombol”) viszont túl magas szintű: a lehorgonyozásuk problémája továbbra is megoldatlan.

Könnyen belátható, hogy e problémák oka a konnekcionizmus alapvetően funkcionalista jellege. A konnekcionizmus ugyanis nem igazán tudott kilépni a funkcionalizmus kereteiből. A konnekcionizmus funkcionalista elkötelezettségét több teoretikus ki is mondja.²¹¹ Bár radikálisan tagadja a funkcionalizmus egyik oldalának, a gépi funkcionalizmusnak az előfeltételeit, de mégis megmarad a funkcionalizmus keretein belül, elsősorban amiatt, mert továbbra is ragaszkodik az implementációfüggetlenség elvéhez.²¹²

Míg az egész konnekcionista fordulat a szimbólumként felfogott mentális reprezentációk és az idegsejtek szintje közötti átkötésként is értelmezhető, ez a közvetítés a két szint között nem igazán lett sikeres. A köztes szint, a konnekcionizmus szintje ugyanis ugyanúgy implementációfüggetlen, mint ahogy a szimbólumok szintje volt, s ennyiben nem biztosít igazi összekötő kapcsolatot a szimbólumok és az idegsejtek között. A szubszimbólumok e funkcionalista konnekcionizmus szerint homogén, egy meghatározott komplexitási szintnek megfelelő egységek, valahol a szimbólumok és az idegsejtek bonyolultsága között. Ekkor azonban önkényes lesz, hogy az idegsejtek és a szimbólumok közötti bonyolultsági spektrum mely pontját jelöljük ki mint a szubszimbólumok szintjét. Bárhova helyezük is, vagy áthidalhatatlan lesz a távolság a szubszimbólumok és a szimbólumok között, vagy túl nagy ugrás lesz az idegsejtektől a szubszimbólumokig.

Olyan ez, mintha az Empire State Building tetejére úgy akarnánk feljutni, hogy valahova félútra egy vízszintes platót teszünk. Vagy túl alacsonyra helyezük a platót, ekkor fel tudunk rá mászni, de onnan még messze van az épület teteje. Vagy pedig túl magasra, s ekkor nem tudunk feljutni rá. A probléma itt persze az, hogy önkényes, miért csak egy ilyen platót teszünk a föld és az épület teteje közé, azaz miért csak egy szubszimbolikus szintet tételezünk fel.

A következőkben egy olyan modellt vázolok, amely az Empire State Building-es példa esetében inkább a lépcsőjárás régi és jól bevált módszerére emlékeztet. A kiindulópont megint a neurális evolúció új elmélete. Nem megoldás tehát, ha az idegsejt és a szimbólum között csak egyféle, önkényesen kiragadott bonyolultságú egységet feltételezünk, és ezt nevezzük szubszimbólumnak. Sokkal jobban képes kezelni a szubszimbólum-lehorgonyzás problémáját egy olyan modell, amelyben a szubszimbólum az egyszerű idegsejt és a bonyolult szimbólum közötti spektrumon elhelyezkedő, bármilyen köztes komplexitású egységet jelenti.

²¹⁰ Ezt részletesen ki lehet mutatni Rumelhart, Clark és Smolensky elméletében is, cf. Nánay 1996.

²¹¹ Az irányzat egyik alapművében például találunk egy *Mikrofunkcionalizmus* című fejezetet. Cf. Clark 1996 p. 57.

²¹² A konnekcionizmus funkcionalista elkötelezettségéhez cf. elsősorban Ramsey–Stich–Rumelhart 1991.

Ha az idegsejtek kisebb csoportokba rendeződve egységeket alkotnak, majd ezek még újabb, bonyolultabb egységeket, majd még bonyolultabbakat adnak ki, egészen a szimbólumokig, akkor a szimbólumlehorgonyzás problémája nem probléma többé, mindegyik egység az egy szinttel lejjebb levőtől nyeri jelentését, egészen le az idegsejtekig. Mivel a felső szint egységei mindig az alsó szint egységeiből állnak, azokból evolválódtak, azok működését fogják vissza, a közvetítés a két szint között nem csak látszólagos, mint a funkcionista konnekciónizmus esetében, hanem kézzelfogható, és világos evolúciós mechanizmussal magyarázható. Ha azonban valóban azt feltételezzük, hogy a magasabb szint az alacsonyabból nyeri a jelentését, hogy oda legyen lehorgonyozva, akkor a felső szint nem lehet független az alsótól, nem tartható tehát ebben a keretben az implementációfüggetlenség elve, s így a funkcionizmus sem.

Hogy emlékezetes katonai példánknál maradjunk, a klasszikus szimbólumkezelő modell valami olyasminak felelne meg, ha a katona és a NATO hadserege között nem lenne semmilyen közbülső, közvetítő egység, minden katona közvetlenül a NATO-parancsnokságtól kapná a parancsokat. A konnekciónizmus egy közbülső szintet iktatott közbe, mondjuk a zászlóalj egységét, az egyes zászlóaljak a parancsnokságnak vannak alárendelve, míg a katonák közvetlenül a zászlóaljnak. Ezzel szemben a nem-funkcionista, evolúciós ihletésű konnekciónizmus egy olyan hadseregnek felelne meg, ahol a katonák a szakasz fennhatósága alá tartoznak, a szakasz a század alá, az az ezred alá etc. Itt – ellentétben az előző két példával – az egyes egységek között valós kommunikáció és információáramlás van.²¹³

A katonák példája persze egy ponton sántít: az ezred nem néhány századból evolválódott, hanem a sorozás durva és az evolúció mechanizmusaira cseppet sem hasonlító eszközeinek eredménye. Az egyszerű szubszimólumok viszont az idegsejtekből evolválódtak, a neurális evolúció mechanizmusával, a bonyolultak az egyszerűekből alakultak ki, majd – utolsó lépésként – a bonyolult szubszimólumokból kialakultak a szimbólumok. Ez az utolsó lépés azért különösen fontos, mert egyrészt ez a lépés választja el az emberi és az állati elmét, másrészt pedig ez az evolúciós lépés nemcsak az elme egységeit változtatja meg, de a közöttük levő kapcsolat módját is, az alacsonyabb szintű, hálózatszerű, rugalmasan működő kapcsolatok helyett itt kialakulnak a következtetések és az explicit szabályok.

Ez az utolsó lépés minőségi váltást hoz, hiszen itt nemcsak egyszerűen még bonyolultabb reprezentációk jöttek létre, hanem olyan reprezentációk, amelyek már egy szabálykövető rendszer elemei lehetnek. A kérdés tehát az, hogy miképpen alakultak ki e szimbólumok és az őket manipuláló szabályok. Erről fog szólni a következő fejezet.

²¹³ Itt a szemléletesség kedvéért megint elvonatkoztattunk attól a tényről, hogy a katona valójában nem az őt magába foglaló magasabb szintű egységtől kapja a parancsokat (ez felelne meg a reprezentációk egymásba ágyazódásának), hanem a feljebbvalójától.