



NÁNAY BENCE

# ELME ÉS EVOLÚCIÓ

NÁNAY BENCE

# ELME ÉS EVOLÚCIÓ

Az elmefilozófia és a kognitív tudomány evolúciós megközelítése

Kávé Kiadó, 2000

A kötet a Nemzeti Kulturális Alap támogatásával jelent meg.  
Sorozatszerkesztő  
KELEMEN JÁNOS

A fedélen Moholy Nagy László fotogramja látható.

© Nánay Bence, 2000

© Kávé Kiadó, 2000

## Tartalom

Előszó .....	7
I. Bevezetés .....	9
II. Az evolúció .....	15
1. Az evolúcióbiológia .....	16
2. Az evolúciós pszichológia .....	25
3. A neurális evolúció .....	34
Funkcionalizmus .....	49
1. A funkcionalizmus és előzményei .....	50
2. Mesterséges intelligencia és gépi funkcionalizmus .....	57
3. A funkcionalizmus bírálatai .....	68
IV. A redukció problémája .....	76
1. Redukálható-e az elme? .....	77
2. Nem-reduktív materializmus .....	83
3. Mentális okság .....	89
V. Intencionalitás .....	98
1. A reprezentáció tárgya .....	100
2. Téves reprezentáció .....	109
3. Intencionalitás és evolúciós változás .....	117
VI. Konnekcionizmus .....	126
1. A klasszikus szimbólumkezelő modell .....	127
2. A konnekcionista fordulat .....	132
3. Kettős modellek .....	141
VII. Magasabb szintű mentális reprezentációk .....	151
1. A metareprezentáció .....	152
2. Az elmeteória .....	156
3. A nyelv evolúciója .....	168
VIII. A kulturális evolúció .....	172
1. A kultúra és a gének .....	174
2. A mémek .....	177
3. Egy csoportközpontú modell .....	182
IX. Bibliográfia .....	193

## Előszó

Az embert mindig izgatta, hogy miként működik az elme: hogyan gondolkodunk, hogyan emlékezünk, miként működik a képzelet, miért van az, hogy néha nem jut eszünkbe egy olyan dolog, amit egyébként mindig tudni szoktunk. Az alábbi könyv két olyan diszciplína – az elmefilozófia és a kognitív tudomány – kérdésfeltevéseit és legfontosabb irányzatait mutatja be, amelyek éppen ezeket a kérdéseket feszegetik: azt próbálják megmagyarázni, miként működik az elme.

Az utóbbi években egyre jelentősebbé vált ezen problémák evolúciós megközelítése. Ahhoz, hogy megértsük, miként gondolkodunk, meg kell tudnunk magyarázni, hogyan alakult ki az elménk, és mennyiben más, mint például egy csimpánz elméje. Bár igen sok területen bizonyultak hasznosnak az evolúciós szemlélet elmefilozófiai, illetve kognitív tudományi alkalmazásai, ezek többnyire elszigetelt, egymástól független kísérletek maradtak. E könyvben megpróbálok egységes elméleti keretbe foglalni a klasszikus elmefilozófiai és kognitív tudományi kérdésfelvetéseket és azok evolúciós színezetű megoldásjavaslatait.

A könyv célja tehát kettős: egyszerre nyújt előképzettséget nem igénylő bevezetést az elmefilozófiába és a kognitív tudományba, ugyanakkor azonban új szempontok szerint tárgyalja a témaköröket, megpróbálja egy egységes evolúciós megközelítés egészébe ágyazni e tudományágak klasszikus problémáit.

Szeretném megköszönni mindazok segítségét, akik hozzájárultak ahhoz, hogy ez a könyv megjelenhetett. Mindenekelőtt Kelemen Jánosnak, e filozófiai könyvsorozat szerkesztőjének köszönöm, hogy felkért a kötet megírására. A legnagyobb köszönet Csányi Vilmost és Pléh Csabát illeti. Pléh Csaba ismertette meg az elme tanulmányozásának legérdekesebb tudományos megközelítéseivel és eredményeivel. Csányi Vilmostól pedig azt a szemléletet tanultam el, hogy az emberi elme vizsgálata nem elképzelhető a szociális szféra elemzése nélkül. Szathmáry Eörs a könyv evolúcióbíológiai háttérének tisztázásában segített. Kampis Györggyel, Forrai Gáborral, Huoranszki Ferencsel, John Searle-lel, Donald Davidsonnal, Howard M. Robinsonnal, Káldy Zsuzsával és Nemes Lászlóval folytatott konzultációim, illetve vitáim nagyon sokat segítettek a könyv megírásában. Köszönöm továbbá Csányi Vilmos, Forrai Gábor, Kampis György, Pléh Csaba, Kelemen János, Vasák Benedek és Blazsó Marianne a kéziratához fűzött részletes kommentárjait, amelyeket igyekeztem beépíteni a végleges szövegbe. Végetetül az, hogy e tudományágakkal és tudós művelőivel kapcsolatba kerülhettem, a Láthatatlan Kollégiumnak köszönhető, amelynek 1995-től 1999-ig tagja voltam.

Budapest, 1999. június 1.

*Nánay Bence*

# I. Bevezetés

Mind az elmefilozófia, mind a kognitív tudomány alapvető kérdése az, hogy miként működik az emberi elme. Néhány kézenfekvő választ már most érdemes kiszűrni. Az egyik ezek közül az a hagyományos nézet, amely szerint a gondolkodás és más mentális tevékenységek voltaképpen a lélek sajátjai, amely a testtől elkülönülten létezik, és irányítja mindennapi cselekedeteinket. Ezt az álláspontot szokás test-lélek dualizmusnak nevezni.<sup>1</sup>

A másik nézet szerint nem szükséges elméről beszélni, elég az aggyal foglalkozni: minden, ami a fejünkben történik, végső soron idegsejtek aktiválódásából áll össze, a megismerés és a gondolkodás megértéséhez tehát elégséges eszköz a neurológia, az agyban zajló folyamatok magasabb szintű megértésére semmi szükség. Ezt a nézetet szokás radikális redukcionista vagy eliminativista álláspontnak nevezni.<sup>2</sup> Ha tehát elfogadnánk, hogy az elme működésének vizsgálatához a neuronok aktivitásmintázatait kell tanulmányoznunk, akkor a „miként működik az elme?” kérdés megválaszolása teljes egészében a biológia és az idegtudomány feladata lenne.

Az elmefilozófia és a kognitív tudomány legtöbb képviselője középutat próbál keresni a két nézet között. Egyfelől elutasítják a test-lélek dualizmust, másfelől azonban a radikális redukcionizmust sem tartják megfelelőnek. Mentális folyamatainkat – az emlékezést, a gondolkodást – persze idegsejtek tüzelése idézi elő, de azt, hogy egy csokitortára gondolok, nem igazán érdemes olyan módon leírni, hogy most az A1632-es neuron aktiválódik, majd az A3412-es stb.

Ezen a ponton érdemes az elmefilozófia és a kognitív tudomány fővonalát szembeállítani egy harmadik nagy irányzattal, amely szintén az elme vizsgálatát tűzte ki céljául, ez pedig a behaviorizmus.

A behaviorizmus azt állította, hogy az emberi elméről semmi mást nem tudhatunk, csak azt, hogy adott ingerekre hogyan reagál. A behavioristák szerint az elme fekete doboz, nem lehet tudni, hogy milyen folyamatok zajlanak benne, a tudós kizárólag azt vizsgálhatja, hogy milyen összefüggés van a rendszert érő ingerek és a rendszer által adott válaszok között.<sup>3</sup>

A kognitív tudomány és a modern elmefilozófia legtöbb képviselője ezzel nem elégszik meg: szerintük ennél lényegesen többet tudhat meg a tudomány az elme működéséről. Hasznosabb, ha feltesszük, hogy az elme egy bizonyos modell szerint működik, majd e modell alapján megpróbáljuk előre jelezni, hogy milyen ingerekre hogyan válaszolna.<sup>4</sup> Ha a posztulált válaszok többé-kevésbé megegyeznek saját válaszainkkal, akkor az elmemo-  
dell többé-kevésbé használható.

Nincs tehát közvetlen rálátásunk arra, hogy milyen is az elme szerkezete, csak éppen jobban le tudjuk írni az elme működését, ha úgy tekintjük, *mintha* egy bizonyos elmemo-  
dell szerint működne.

Ezen a ponton kapcsolódik össze az elmefilozófia és a kognitív tudomány a mesterségesintelligencia-kutatással. Ha ugyanis feltételezzük, hogy az elme egy bizonyos modell szerint működik, akkor egy, az adott modell szerint funkcionáló számítógép – szerencsés esetben – hasonló ingerek hatására hasonló módon fog viselkedni, mint mi. Az elvont elmemo-  
dellek tehát empirikusan tesztelhetők a mesterségesintelligencia-kutatás segítségével. A számítógép-elme analógia végigvonul a modern elmefilozófia és a kognitív tudomány történetén, hiszen ez utóbbi tudományág nem titkolt szándéka a mesterséges értelem létrehozása.

Az elmefilozófia tehát – ellentétben a filozófia szóhoz tapadó sztereotípiákkal – igenis gyakorlati diszciplína. A filozófia azon nagyon kevés területeinek egyike, amelyeknek van, vagy lehet közvetlen gyakorlati hasznuk. A sok hibával dolgozó elmemo-  
működést próbálja leírni, és ennek mintájára próbál mesterséges értelmet létrehozni – vagy megmutatni, miért lehetetlen a mesterséges intelligencia létrehozása. Az elmefilozófiára – legalábbis annak egy részére – tehát nem illenek azok a jelzők, amelyekkel általában a filozófiát szokás jellemezni, nem valóság-tól elszakadt spekuláció, hanem empirikusan tesztelhető hipotézisek sorozata.

A kognitív tudomány viszont valódi tudomány. Igazi újdonsága az interdiszciplinaritás. Nem része ugyanis egyik klasszikus tudományágnak sem. A filozófia, a biológia, a mesterségesintelligencia-kutatás, a pszichológia, az etológia és a nyelvészet együttes vonzáskörében helyezkedik el. Valamennyi itt felsorolt tudományág terminológiáját és klasszikus kérdésfelvetéseit használja, de egyik diszciplína szerepe sem meghatározó jelentőségű benne. Egy több tudományból összegyűrt hibrid tehát, amelyet az tart össze, hogy mindegyik tudományból ki-szededegeti azokat a részterületeket, amelyek hasznosak lehetnek az elme tudományos modellezése szempontjából.<sup>5</sup>

Mind az elmefilozófia, mind a kognitív tudomány arra a kérdésre keresi tehát a választ, hogy miként működik az emberi elme, lényeges különbség van azonban módszerük és terminológiájuk között. Míg az elmefilozófiát a logikai elemzés és a klasszikus filozófiai kifejezések használata jellemzi, addig a kognitív tudomány módszerre inkább empirikus, összehasonlító, főleg pszichológiai, biológiai és nyelvészeti terminusokat használ.

<sup>1</sup> Cf. pl. Descartes 1994a, 1994b, 1996.

<sup>2</sup> Cf. pl. Churchland 1986.

<sup>3</sup> Cf. pl. Skinner 1953.

<sup>4</sup> Cf. pl. Haugeland 1996.

<sup>5</sup> Cf. Gardner 1985, Lycan 1990, Osherson et al. 1990, Pléh 1991, 1996, 1998b, Nánay 1997b, Wilson-Keil 1999.

Érdekes kapcsolódási pontokat keresni a két tudományterület között, hiszen sok megoldatlan kérdés éppen abból ered, hogy a kognitív tudományból hiányzik az elméleti megalapozás igénye, az elmefilozófia pedig sokszor nem veszi figyelembe a kognitív tudomány – főleg empirikus – eredményeit. A két diszciplína közötti átjárások és keresztthivatkozások segítségével talán lehetséges egy tágabb elméleti kontextust teremteni, amely egyaránt alkalmas a hagyományos elmefilozófiai problémák, valamint az újabb kognitív kérdésfeltevések kezelésére.

Egy ilyen közös elmefilozófiai-kognitív tudományi elméleti kontextusnak bizonyosan alapkategóriája a *reprezentáció* fogalma. Reprezentációnak nevezzük az elme építőelemeit, azokat a gondolatokat, ideákat, emlékeket, memórianyomokat, álmokat, amelyek benépesítik az elménket. A reprezentációk tehát az elme alapelemei, van azonban egy igen fontos sajátosságuk: mindig utalnak valamire: mindig *valamit* reprezentálnak. A borosüveg reprezentációja például a borosüveget mint fizikai tárgyat reprezentálja. Előfordul persze, hogy nincs olyan fizikai tárgy, amely a reprezentáció tárgya lenne, de a reprezentáció ekkor is vonatkozik valamire. Ha például elképzelek egy 1000 hektoliteres borosüveget (ami valójában nem létezik), akkor is van tárgya ennek a reprezentációnak, jóllehet, csak a képzeletemben létezik. A képzeletem azonban valóban létező hét deciliteres borosüvegek reprezentációja alapján alkotta meg e reprezentációt. Tovább bonyolítja a képet, hogy léteznek olyan elmeállapotok is, amelyek látszólag nem vonatkoznak semmilyen konkrét tárgyra, ilyen például a szorongás. A reprezentáció és a külvilág közötti viszony problémái az elmefilozófia központi kérdésfelvetését jelentik, erről szól e könyv V. fejezete.

Szintén klasszikus probléma, hogy minek alapján tudjuk azonosítani a reprezentációkat az agyban. Meg tudjuk-e mutatni, hogy hol van az én agyamban a francia forradalomról kialakított reprezentációm? Meg lehet-e mutatni azokat az idegsejteket, amelyeknek az és csak az a feladatuk, hogy a francia forradalom reprezentációt kódolják bennem? És ha igen, akkor van-e szükség egyáltalán reprezentációkról beszélni, nem elég-e, ha a neuronokat vizsgáljuk? Ezekkel a kérdésekkel, a reprezentációk redukciójának problémájával a III. és a IV. fejezet foglalkozik.

Az eddigi két problémacsoport klasszikusan elmefilozófiai kérdéseket feszegetett. A kognitív tudomány számára szintén igen fontos a reprezentáció fogalma, de egészen más szempontok szerint vizsgálja azt. Az egyik legfontosabb ilyen szempont az, hogy végül is mik is ezek a reprezentációk: szavak, képek, emléktöredékek, bonyolult mondatok vagy esetleg egyszerű numerikus vektorok. Minden reprezentációnk nyelvi jellegű, vagy a nyelv csak a jéghegy csúcsa, és szavaink egyszerűbb, nem-nyelvi reprezentációkból épülnek fel? E kérdéseket a VI. fejezet tárgyalja.

Szintén jellegzetesen kognitív tudományi probléma, hogy mi a különbség az emberi és az állati elme között. Másként fogalmazva, mi a különbség az emberi, illetve állati elmét felépítő reprezentációk között? Továbbá, lehet-e a reprezentációkról úgy beszélni, hogy a társadalmi közeget – a reprezentációk tanulását, átadását – kihagyjuk a képből, azt a kérdést tehát, hogy milyen módon viszonyul az én reprezentációm a másik ember reprezentációjához. Az előbbi problémát a VII. fejezet, ez utóbbit a VIII. fejezet vizsgálja.

Egy igen fontos és bonyolult elmefilozófiai problémakörrel nem foglalkozik a könyv, ez pedig a tudatosság kérdése. Ez a kérdéskör, amely ma hozzátétőlegesen ugyanolyan széles szakirodalommal rendelkezik, mint az összes többi elmefilozófiai probléma, egyszerűen nem fért bele e kötetbe. Emellett igaz az is, hogy a tudatosság kérdésköre a könyv fő gondolatmenetét alkotó evolúciós megközelítés szempontjából másodlagos jelentőségű.

A könyv első fele inkább az emberi és az állati elme közötti azonosságokat hangsúlyozza, míg a második fele a különbségekre helyezi a hangsúlyt, anélkül azonban, hogy áthághatatlan és csodás határvonalat tételezne az ember és az állat mentális működése között. Mindvégig alapfeltevésként szerepel, hogy az ember elméje az állati elméből világosan leírható evolúciós mechanizmusok segítségével alakult ki. Ezekről az evolúciós mechanizmusokról szól a következő fejezet.

## II. Az evolúció

Mind az elmefilozófiában, mind a kognitív tudományban egyre nagyobb teret kapnak az utóbbi tíz-húsz évben az evolúciós elmémmodellek. Mivel azonban ezek az evolúciós megközelítések nagyon különböző elméleti megalapozottságúak, és nagyon különböző céllal vonják be az evolúció elméletét az elme vizsgálatába, érdemes mindenekelőtt rendszerezni őket.

Az első és legelterjedtebb kutatási irány az emberi *elme biológiai evolúcióját* próbálja megmagyarázni. Azt a tényt hangsúlyozza ez az irányzat, hogy az ember az állatból evolválódott, és ez nemcsak az emberi testre, de az emberi elmére is igaz. Az emberi és az állati elme között tehát nincs éles a határvonal. Ha elfogadjuk a fenti állítást, akkor radikálisan átértelmeződnek az elmefilozófia és a kognitív tudomány kérdésfelvetései, hiszen a legbonyolultabb gondolati tevékenységeinknek is a viszonylag egyszerű főemlős-elméből kellett kialakulnia az evolúció mechanizmusának segítségével.

Az evolúciós megközelítés másik nagy kutatási iránya – amelyet itt az *elmén belüli evolúciónak* nevezek – abból indul ki, hogy az emberi gondolkodás ugyanúgy próbálkozások és hibakorrekciók sorozata, mint az evolúció. Kérdés tehát, hogy a gondolkodás folyamata leírható-e evolúciós terminusokkal.

A harmadik olyan pont, ahol az elme vizsgálatában hasznosak lehetnek az evolúciós terminusok, a *kulturális evolúció* világa. Divatok, politikai ideológiák és művészeti irányzatok ugyanúgy versengenek egymással, és ugyanúgy a túlélésért küzdenek, mint az egyes állatfajok a biológiai evolúció folyamatában.

Három olyan pont van tehát, ahol az elme működésének magyarázata és az evolúciós elmélet összekapcsolható: az emberi elme kialakulása, az elmén belüli evolúciós folyamatok és a kulturális evolúció jelensége.<sup>6</sup> Érdemes áttekinteni, hogy e három alkalmazási terület közül melyek azok, ahol valóban egy biológiailag elfogadható evolúciós elméletet alkalmazhatunk az elme vizsgálatára, és melyek azok, ahol az evolúció legjobb esetben is csak metaforaként szolgál. Ehhez azonban röviden tisztázni kell, pontosan mit is értünk evolúción.<sup>7</sup>

### 1. Az evolúcióbiológia

Az evolúció legfontosabb, bár nem kizárólagos mechanizmusa a természetes szelekció. A természetes szelekció egy nagyon egyszerű mechanizmussal képes elérni, hogy mindenféle előzetes terv nélkül igen bonyolult és az adott környezet feltételeinek megfelelő organizmus jöjjön létre. A természetes szelekció két lépésből áll, a véletlen variáció kialakulásából és az ezt követő szelektív kipusztulásból.

A véletlen variáció lépésének felel meg az, amikor például egy nyúl szül tíz kölyköt, amelyek igen eltérő tulajdonságokkal rendelkeznek: az egyik gyorsan fut, de nem lát élesen, a másoknak nagyon jó a szaglása etc. A szelektív kipusztulás lépése ennél kicsit brutálisabb. A tíz nyúlkölyök környezete általában nem túl barátságos: különböző ragadozók, kutyák, emberek próbálják elkapni őket. Általában a tizből jó esetben kettő éri meg az ivarérett kort, a többi elpusztul. Azok fognak elpusztulni, amelyek nem tudnak jól reagálni a környezet veszélyeire: nem tudnak elmenekülni a rókától, nem veszik észre a kutyát etc. Akik megbirkóznak ezekkel a veszélyekkel, azok túlélnek, és újabb nyulakat nemzenek. Ekképpen a következő nyúlgeneráció – ha a gyors futás és az éles látás öröklődik – valószínűleg jobb esélyekkel fog indulni a ragadozók ellen, hiszen annak a nyúlnak a lezármazottai, amely – ellentétben testvéreivel – éles látással rendelkezett és gyorsan futott. Néhány generáció múlva igen gyors és jól látó nyúlpopulációnk lesz, és ennek kialakulása világosan megmagyarázható pusztán a véletlen variáció és a szelektív kipusztulás segítségével.

Egy kicsit absztraktabban fogalmazva, a természetes szelekció első lépése mindig egy replikációs folyamat: egy nyúlból lesz tíz például. Ez a replikáció azonban – ahhoz, hogy működhessen a szelekció – nem lehet teljesen pontos. Ha tíz ugyanolyan kisnyúl születne, akkor a szelekció folyamata nem lenne képes a leggyorsabbakat kiválogatni, következésképpen semmilyen hatással nem lenne a következő generáció nyulainak gyorsaságára nézve. A replikáció tehát nem az eredeti organizmus lemásolása egy az egyben, mindig van eltérés az eredetitől.

A természetes szelekció második lépése a szelektív kipusztulás ebből az alapanyagból, a véletlen variáció által létrehozott sokféleségből dolgozik. A sokféle egyed közül az fog túlélni, aki legjobban alkalmazkodik a környezethez. Ezt mondja ki a híres „Survival of the fittest” elv: aki alkalmazkodik, az lesz a túlélő.<sup>8</sup> Különböző környezeti feltételek között más és más tulajdonság fog hasznosnak bizonyulni. Mivel nem tudható előre, hogy milyen környezeti feltételek fognak érvényesülni, minél nagyobb variabilitást biztosít a replikáció az utódoknak, annál valószínűbb, hogy lesz olyan utód, amely képes lesz túlélni az adott környezetben. A természetes szelekció

<sup>6</sup> Megjegyzendő, hogy ez a hármasság igen hasonlít az evolúció három játékerére, amelyet Karl Bühler különböztet meg (Bühler 1922): az ösztön, a szokás (dresszúra), illetve az értelem terére. Cf. Pléh 1995. Bühlernek emellett igen nagy szerepe volt az emberi és az állati elme közötti folytonosság és strukturális hasonlóság hangsúlyozásában, amelyben – tanítványának, Poppernek a közvetítésével – a mai evolúciós elmefilozófia előfutárának tekintendő. Cf. Pléh 1999.

<sup>7</sup> Darwin 1859.

<sup>8</sup> Cf. Walsh 1996.

innentől újra kezdődik: az az utód, amely a legjobban alkalmazkodott a környezethez, újabb utódokat hoz létre, megintcsak a véletlen variáció követelményei szerint.

Több ponton érdemes pontosítani az itt vázolt modellt. Egyrészt eddig úgy tekintettük, mintha egy organizmus – jelen esetben a nyúl – egy fix és változatlan környezethez alkalmazkodna, holott a környezet is megváltozhat, hiszen a nyúl környezete is az evolúciós folyamatok által formált élőlényekből áll. A rókák természetes szelekciójában például igen lényeges szelekciós tényező lehet az, hogy azok az egyedek, akiknek nem sikerül nyulat fogniuk, éhen halnak. Következésképpen az egyre gyorsabban futó és egyre élesebben látó nyulak mellett a természetes szelekció egyre gyorsabban futó és magát egyre hatékonyabban álcázó rókákat hoz létre.

Ezt a folyamatot szokás koevolúciós folyamatnak nevezni: az egyik állat alkalmazkodik a másik állathoz, a másik pedig az egyikhez.<sup>9</sup> Olyan ez, mint a fegyverek és a páncélok párhuzamos változása. Ha kitalálnak egy olyan fegyvert, amely a jelenleg használatos páncélok atüti, akkor előbb-utóbb megjelenik egy olyan páncél, amely ezt kivédi, majd megint egy hatékonyabb fegyver, és így tovább. A nagy különbség e két példa között persze az, hogy míg a páncélok és a fegyvereket gondolkodó emberi lények hozzák létre, addig a nyúl és a róka pusztán a természetes szelekció egyszerű, nem előre megfontolt és nem tudatos mechanizmusainak eredményeképpen változik meg.

Egy másik probléma, amelyet a fenti modell kissé elnagyoltan kezel, az öröklés problémája. Ez két okból is fontos. Egyrészt a természetes szelekció mechanizmusa csak akkor működik, ha a sikeres túlélő át tudja adni utódainak azokat a tulajdonságait, amelyeknek köszönhetően túlélte. A gyors nyúlak tehát gyors utódokat kell létrehozni. Ezt a követelményt teszi nyilvánvalóvá az úgynevezett génközpontú vagy más néven replikációközpontú evolúciómódel.

A génközpontú módel azt mondja ki, hogy igazából nem is a nyúl az, amely replikálódik, és amely küzd a túlélésért, hanem a nyúl sejtjeinek örökítőanyaga, tehát a gén.<sup>10</sup> A gén az egyetlen replikátor, célja, hogy mindenáron sokszorosítsa magát. A nyúl csak eszköz a génje kezében, azért próbál minél több életképes és sikeres utódot létrehozni, hogy génjét minél hatékonyabban terjessze szerte a világban.

Bár ez a módel nem minden evolúcióbiológust győzött meg,<sup>11</sup> az tagadhatatlan, hogy minden természetes szelekciós folyamat esetében a gén az, amely lehetővé teszi a viszonylag pontos replikációt, amely ugyanakkor a mutációk miatt elegendő variációs lehetőséget is biztosít. Szintén a gén teszi lehetővé, hogy a sikeresen túlélő organizmus utódai ugyanazzal a géntípussal, következésképpen durván ugyanazzal a tulajdonságokkal rendelkezzenek, mint a szülő.

Az öröklött tulajdonságok szerepe azonban egy másik okból is elemi fontosságú. A szerzett tulajdonságok ugyanis a darwini módelben nem öröklődnek. A véletlen variáció által létrehozott utódok már születésüktől fogva hordozzák azokat a tulajdonságokat, amelyek jóvoltából képesek túlélni, és amelyeket majd az utódaiknak is át fogják adni. Egy gyorsan futó nyúl tehát egyrészt már úgy született, hogy gyorsan fut, másrészt az utóda is gyorsan fog futni. Hiszen hiába sajátít el az állat valami, a túlélése szempontjából igen hasznos tulajdonságot, ha ezt az utódaiknak nem tudja átadni, akkor a következő generáció szempontjából ez az új tulajdonság értéktelen. Hiába talál ki például a nyúl egy fenyegető kiáltást, amelytől minden róka visszaretten, mivel az a kiáltás nem genetikailag kódolt, az utódok nem lesznek képesek ilyen csatakiáltásra, következésképpen a rókák gond nélkül el fogják kapni őket.

A szerzett tulajdonságok öröklődését lamarcki evolúciónak hívja a szakirodalom, és általános közmegegyezés alakult ki arról, hogy a növények és az állatok evolúciója nem lamarcki.<sup>12</sup> Az ember esetében persze a helyzet már távolról sem ilyen egyszerű. Az ember utóda ugyanis nemcsak azokkal a hasznos tulajdonságokkal rendelkezik, amelyeket a génjei által örökölt, hiszen például tanulás útján is képes elsajátítani olyan tulajdonságokat, amelyek a szülei túléléséhez hozzájárultak. Itt azonban már a kultúra és a kulturális evolúció igen bonyolult jelenségével van dolgunk, amelyről majd az utolsó fejezet fog szólni. Mindenesetre az ember – sőt, más, fejlettebb gerincesek – esetében már nem kizárólag a gének a hasznos tulajdonságok átadásának egyedüli eszközei, az evolúció itt épp emiatt már nagyon bonyolult és összetett folyamat lesz.

Az evolúcióbiológia tehát elveti a lamarcki evolúció elvét, a szerzett tulajdonságok öröklődését. A zsiráf nyaka nem azért ilyen hosszú, mert addig nyújtogatta, amíg a kis zsiráf nyaka egy kicsit hosszabb lett, az unokái még hosszabb. A valódi ok persze az, hogy a magas fákat a rövidnyakúak nem érték el, és ezek kipusztultak, a hosszúnyakúak viszont túléltek, és ők nemzették az utódokat, akiknek – mivel génjeiket a testvéreiknél kicsit hosszabb nyakú szüleiktől kapták – már néhány centivel hosszabb lett a nyakuk, mint az előző generációnak.

A radikális lamarckizmus tehát nem igazán meggyőző. Másfelől azonban az sem teljesen igaz, hogy egyáltalán nem számít, hogy mit csinál az állat hosszú élete során. Az állat ugyanis – és itt a már említett koevolúció jelenségéhez kell visszanyúlnunk – megváltoztathatja a környezetét, és ekkor a szelektív kipusztulás feltételei mások lesznek, mint az eredeti környezetben. Tegyük fel például, hogy egy kisnyúl túl vastag bundával született, és emiatt nagyon lassan fut. Ekkor nem tudna túlélni jelenlegi környezetében, ahol sok róka van. Még ha véletlenül

<sup>9</sup> Cf. Brandon–Antonovics 1996, Thompson 1994.

<sup>10</sup> Cf. Dawkins 1986. Cf. még Hull 1989.

<sup>11</sup> A génközpontú módel bírálatáról az utolsó fejezetben lesz részletesen szó.

<sup>12</sup> Lamarck 1809. Ellene cf. pl. Weismann 1893.



sikerülne is túlélnie, utódaira – akiknek szintén vastag bundájuk lesz – semmi jó nem vár: hosszú távon biztosan kipusztulnak. Ha viszont északra vándorol, ahol hidegebb van, ott mind neki, mind esetleges utódainak nem hátrányt, de előnyt jelent a vastag bunda: nem fagynak meg. Az állat tehát nem tudja megváltoztatni azt a génanyagot, amit majd utódaira fog hagyni, de kereshet olyan környezetet vagy alakíthatja olyanná aktuális környezetét, hogy az eleve adott tulajdonságok, amelyeket utódaira örökít, hasznosak legyenek utódainak az adott környezetben. Ezt a furcsa fából vaskarikát hívjuk Baldwin-effektusnak.<sup>13</sup>

A zsiráfnyak példájában nagy szerepet játszott a folyamatos változás elve. Ez is az evolúcióbiológia központi gondolatainak egyike. Egyik generációról a másikra ugyanis csak kis változások lehetségesek.<sup>14</sup> Ha egy zebrának például születne egy olyan utódja, amelynek ugyanolyan hosszú nyaka lenne, mint egy zsiráfnak, akkor az minden valószínűség szerint elpusztulna, holott a hosszú nyak hasznos tulajdonság lenne, mivel annak segítségével a magas fákról is tudna enni. A hosszú nyakú zebrának azonban a többi testrésze nem változott meg annak érdekében, hogy az állat képes is legyen kihasználni a hosszú nyak nyújtotta előnyöket, egy zebra testű, zsiráfnyakú állat mindig felborulna, helyett hogy a magas ágakról legelne. De még ha valahogy sikerülne is elvegetálnia, nem valószínű, hogy sok nőtényt el tudna csábítani, és ekképpen sok utódot tudna nemzeni. Ha viszont a zebra nagyságú és testfelépítésű állat nyaka csak egy fél centiméterrel lesz hosszabb egy generációs változás alatt, akkor egyrészt az állat többi testrésze is képes a nyakhoz idomulni, másrészt nem egyetlen egyed lesz hosszú nyakú, hanem az egész populáció, következésképpen az utódnemzés sem lesz probléma.

Láttuk tehát, hogy a természetes szelekció mechanizmusa az evolúció legfontosabb hajtóereje. Ehhez a kérdéshez kapcsolódik az evolúciós elmélet legnagyobb nyilvánosságot kapott vitája, az adaptáció-exaptáció vita. Az adaptacionisták azt állítják, hogy minden evolúciós folyamat természetes szelekció, következésképpen minden élőlény minden tulajdonsága a természetes szelekció eredménye, tehát minden tulajdonságunk adaptív.<sup>15</sup> Ezzel szemben vannak, akik szerint bár a természetes szelekció a legfontosabb, de nem ez az egyedüli hajtóerő az evolúcióban. Léteznek például olyan tulajdonságok, amelyek nem voltak hasznosak az adott organizmus számára, de szükségszerű velejárói voltak egy adaptív, a túlélés szempontjából hasznos tulajdonság megjelenésének. Tehát adaptív folyamatok melléktermékei voltak. Ezek a melléktermékek persze később más, immáron adaptív funkciót nyerhetnek, de eredetileg nem voltak adaptívak.

Steven Jay Gould vezette be az exaptáció fogalmát az evolúcióbiológiai diskurzusba, hogy rámutasson: abból, hogy egy tulajdonság most adaptív, egyáltalán nem lehet következtetni arra, hogy miért alakult ki.<sup>16</sup> Gould exaptációnak nevezi az olyan tulajdonságok megjelenését, amelyet nem adaptív folyamatok alakítottak ki. Egy tulajdonság tehát olyan okokból is kialakulhatott, melyek mai funkciójától igen távol állnak, és mai funkciójából nem következtethető ki. Következésképpen nem érdemes egy tulajdonság mai adaptív funkciója alapján történeteket gyártani e tulajdonság kialakulására nézve: a kialakulás nem található ki egyértelműen a mai adaptív funkcióból.

Érdemes megemlíteni azt a példát, amelyet a leggyakrabban szoktak felhozni az exaptáció jelenségének demonstrálására. A panda hüvelykujja eredetileg nem a többi ujjal szembe fordítható, fogásra alkalmas szerv volt, hanem egy egyszerű nem adaptív csontkinövés.<sup>17</sup> Később azonban ez a csontkinövés fokozatosan ujjá módosult. Már ezen a példán is jól látható, és az egész exaptáció vitára jellemző, hogy az exaptáció fogalma, amely pusztán nem-adaptív folyamatot jelent, igen könnyen összekeverhető az exaptáción alapuló funkcióváltozással, azzal a jelenséggel tehát, hogy az eredetileg nem adaptív szervek a későbbiekben hasznosnak bizonyulhatnak, és ekképpen mégis adaptívak lehetnek. Ez az utólag elnyert adaptív funkció azonban nem magyarázza a szerv megjelenését.

Az exaptacionisták tehát a túlzott adaptacionizmust próbálták ellensúlyozni. Az adaptációs sztorizgatást figurázzák ki.<sup>18</sup> abból, hogy egy szerv most adaptív egy élőlény számára, nem lehet arra következtetni, hogy milyen hatóerők alakították ki e szervet, hiszen lehetséges, hogy a most adaptív jegy kialakulása nem adaptív folyamat volt; az adaptivitás csak utólagos. Ha jobban megnézzük azonban, kiderül, hogy az adaptacionisták és az exaptacionisták álláspontja között nincs is olyan nagy különbség (talán épp ezért olyan éles a vitájuk). Az exaptacionisták is elismerik, hogy például a panda hüvelykujját igenis adaptív folyamatok alakították olyanra, amilyen. Másfelől az adaptacionisták is hajlanak arra, hogy elismerjék, léteznek nem adaptív folyamatok is, de ezek nem képesek komplex struktúrákat létrehozni, az exaptációs folyamatok előbb-utóbb átadják helyüket az adaptív folyamatoknak, amelyek tovább csiszolják a nem-adaptív folyamatok eredményét. Végző soron mind egy, hogy a panda csuklóján megjelent csontkinövést egy adaptív folyamat hozta-e létre, vagy egy egyszerű mutáció, hiszen miután – így vagy úgy – megjelent, a csontkinövésből hatodik ujjat úgyis a természetes szelekció csinált.

<sup>13</sup> A Baldwin-effektusról cf. Baldwin 1896.

<sup>14</sup> A gradualizmusról cf. pl. Gould–Eldredge 1993.

<sup>15</sup> Cf. pl. Dennett 1998a, Dawkins 1986, Barkow–Cosmides–Tooby 1992.

<sup>16</sup> Gould–Vrba 1982, Gould 1990, Gould–Lewontin 1979. Az exaptáció fogalmának persze jelentős előzményeri voltak már korábban is az evolúcióbiológiai irodalomban, például a preadaptáció fogalma. Minderről cf. Gould 1996, 1997.

<sup>17</sup> Cf. Gould 1990.

<sup>18</sup> Cf. különösen Gould–Lewontin 1979.

Érdeemes megemlíteni még egy evolúciós mechanizmust, amely a természetes szelekció mellett a legnagyobb jelentőségű, ez pedig az evolúció nagy átmeneteinek mechanizmusa, vagy más néven az újabb szerveződési szintek megjelenése.<sup>19</sup>

Az evolúció igazán radikális lépései nem magyarázhatóak pusztán a természetes szelekció segítségével. A többsejtűek megjelenése az egysejtűek világában például nem igazán kezelhető a fent vázolt sémában. A többsejtűek megjelenésével ugyanis az addig egymással a túlélésért versengő egysejtűek immár egy magasabb egység fenntartása érdekében kooperálni kezdenek. A sejtnek már nem az az érdeke, hogy a saját túléléséről gondoskadjon, hanem az, hogy a többsejtű organizmus túlélését biztosítsa, akár azon az áron, hogy saját magát feláldozza. A májsejt például nem a saját túléléséért küzd, hanem annak a szervezetnek a túléléséért, amelynek ő része. A többsejtűek evolúciója tehát egy olyan evolúciós folyamat, amely nem magyarázható a természetes szelekció mechanizmusával, hiszen itt az egymással versengő sejtek (egysejtűek) egyszer csak egy nagyobb egység részeként kezdenek el funkcionálni; heteronóm helyzetbe kerülnek ehhez a nagyobb egységhez képest. A szelekció szintje tehát megváltozik: kezdetben egysejtűek versengenek egymással, majd többsejtűek, és a kettő közötti átmenetet nem magyarázza a természetes szelekció, hiszen ott a szelekció szintje mindig rögzített.

Az evolúció során több ilyen nagy átmenetet lehet kimutatni, ezek közül csak a legfontosabb a prokarióta-eukarióta átmenet, az egysejtű-többsejtű átmenet és talán az egyén-csoport átmenet, ahol az egyén kerül függő helyzetbe a csoporthoz képest, és a csoport érdeke lesz az ő érdeke is, amelyet akár saját maga feláldozása árán is képviselni fog.<sup>20</sup> Az evolúció nagy átmenetei során tehát korábban külön-külön replikálódó elemek együtt fognak replikálódni, ez azonban néha a komponensek és a rendszer konfliktusához vezet. A rák például nem más, mint egy ilyen konfliktus, ahol is a többsejtű rendszerben az egyes komponensek (a rákos sejtek) nem hajlandóak alávetni magukat a nagyobb egységnek. A nagyobb egység, a többsejtű szervezet replikációjának elősegítése helyett ők maguk kezdenek el replikálódni, ami általában a nagyobb egység replikációjának megakadályozásával, azaz halálával jár együtt.

Az evolúció nagy átmeneteinek elmélete az elme vizsgálatában is fontos szerepet fog kapni, a neurális evolúció területén, de előbb az elme és az evolúció összekapcsolásának egy másik, talán egyszerűbb esetét érdemes megvizsgálni.

## 2. Az evolúciós pszichológia

Az evolúció elméletének legfontosabb elmefilozófiai és kognitív tudományi alkalmazása az emberi elme biológiai evolúciójának magyarázata. Kétségtelen, hogy az emberi faj kialakulására érvényes az evolúcióból modellje. Az emberi fajt is – mint minden más fajt – az evolúció csiszolta olyanná, amilyen. Ezen belül, ahogy például az emberi kéz a főmlősmancsból tisztán evolúciós folyamatok segítségével alakult ki, úgy az emberi elméről is jó okkal feltételezhetjük, hogy a főmlős-elméből evolválódott. Vannak persze nyilvánvaló különbségek egy gorilla és egy ember gondolkodása között, de a majom mancsa és az emberi kéz között is hasonlóan nagy különbségeket találunk, és e különbségeket is meg kell tudnunk magyarázni az evolúciós elmélet segítségével.

Az elme evolúciós megközelítésén belüli első nagy terület kiindulópontja tehát az, hogy az emberi elme működésének megértéséhez azt kell megvizsgálni, milyen evolúciós múlttal rendelkezik: miből alakult ki, milyen hatások nyomására. A jelenlegi elmeműködés magyarázata tehát előfeltételezi az ember és az emberi elme evolúciós előtörténetének leírását. Akkor értjük meg az elmét, ha tudjuk, hogyan és miért alakult olyanná, amilyen.

Ennek a megközelítésnek a legjelentősebb mai képviselője az evolúciós pszichológia irányzata, amelynek segítségével igen jól be lehet mutatni az egész megközelítési mód előnyeit és gyenge pontjait.<sup>21</sup> Fő állításuk, hogy jelenlegi mentális képességeink a környezethez való adaptáció következtében lettek olyanok, amilyenek jelenleg is. Az ember, akár a többi élőlény, a természetes szelekció műve: az az egyed tudott túlélni, aki jobban alkalmazkodott a környezetéhez. Az alkalmazkodást pedig egy olyan bonyolult élőlény esetében, mint amilyenek mi vagyunk, nem elsősorban az egyes szervek megváltozása biztosította, hanem elménk megváltozása. A környezethez való alkalmazkodás azt jelentette, hogy olyan szellemi képességeket fejlesztettünk ki, amely növelte túlélési esélyeinket.

A probléma csak ott van, hogy amikor ezek a túlélést jelentő szellemi képességek megjelentek, akkor egészen másmilyen volt a környezetünk – természetesen a társadalmi környezetünk is –, mint most. Az evolúciós folyamat lényegesen lassabb, mint a környezet változása, ezért a mostani emberi elme nem a mostani környezethez alkalmazkodott, hanem egy már rég nem létező környezethez.<sup>22</sup> Következésképpen sok olyan mentális ké-

<sup>19</sup> Maynard Smith–Szathmáry 1997. Cf. még Csányi 1978.

<sup>20</sup> Maynard Smith–Szathmáry 1997 nyolc ilyen átmenetet mutatott ki az evolúció során.

<sup>21</sup> Az evolúciós pszichológiáról: Barkow–Cosmides–Tooby 1992, Buss 1994, 1995, Wright 1994, Pinker 1997, Dennett 1998a, Cosmides–Tooby 1997, Stich 1998, Plotkin 1997, l. még Bereczkei 1998, Pléh 1998a áttekintését az evolúciós pszichológia irányzatáról.

<sup>22</sup> Ezt hívják az evolúciós pszichológusok az evolúciós adaptáció környezetének, amely nem azonos a jelenlegi környezettel. Cf. Cosmides–Tooby 1997. p. 14.

esség, amely a kőkorszaki környezetben a túlélést biztosította, a mai környezetben már egyáltalán nem hasznos, sőt, esetenként káros. A kőkorszaki környezethez adaptálódott elmével kell boldogulnunk ma is. Ha tehát meg akarjuk érteni az elmét, nem azt kell vizsgálni, hogy miként funkcionál ma, hanem hogy miként funkcionálhatott régen, hiszen akkor működött rendeltetésszerűen (a környezethez alkalmazkodva), ma már nem feltétlenül.

Jogos kérdés, hogy miért nem alkalmazkodott elménk a mai környezethez, miért alkalmazkodik még mindig a kőkorszaki környezethez. Az evolúciós biológia szerint az emberi evolúció az utóbbi néhány ezer évben lelassult, a kulturális változás viszont felgyorsult. Az evolúciós léptékkal egy pillanatnak megfelelő kétezer év például olyan radikális kulturális változásokat hozott, amelyet az igen lassú evolúciós folyamatok nem képesek követni. Ráadásul populációbiológiai tény, hogy minél nagyobb a népesség, annál lassabbak az evolúciós változások.<sup>23</sup> Egy hatmilliárd fős populációban gyakorlatilag nincsenek evolúciós hatások.<sup>24</sup> A kulturális változásokkal tehát nem tudott lépést tartani a biológiai evolúció,<sup>25</sup> ezért van az, hogy a mostani emberi elme egy rég nem létező környezethez alkalmazkodott, és nem az aktuális környezetünkhöz.

Az evolúciós pszichológia tehát azt hangsúlyozza, hogy az emberi elme – mivel a természetes szelekció terméke – csakis akkor érthető meg, ha nem önmagában véve szemléljük, hanem azzal a környezettel együtt, amelyben kialakult. Az elme működését elvileg meg lehet magyarázni, ha azt vizsgáljuk, milyen veszélyek elől kellett kitérnie, milyen kihívásokkal szembesült, egyszerűen, milyen környezeti feltételek között élt több százezer évvel ezelőtt. Némi gondot okoz, hogy legfeljebb elképzelésünk lehet arról, milyen is lehetett ez a környezet, így tehát hiába vezettük vissza az elme vizsgálatát az elmét formáló környezet vizsgálatára, ez utóbbiról semmilyen közvetlen információnk nincs, legfeljebb csak hipotéziseink lehetnek.

E hipotéziseket felhasználva azonban mégis sok mindent elárul az elméről az ember kőkorszaki életmódjának vizsgálata. Erre példa az evolúciós pszichológus Buss híres, ám sokat vitatott elmélete az ember azon ízléspreferenciájáról, amellyel az édes ételeket előnyben részesíti.<sup>26</sup> Amikor még a szavannán éltünk, és igen kevés kalóriához juthattunk, a cukor a maga viszonylag magas kalóriatartalmával igen nagy kincs lehetett. Következésképpen a cukor iránti preferencia evolúciósan hasznos volt, és ezért azoknak az embergyerekeknek, akik a szüleiiktől olyan géneket örökölték, amelyekben kódolva volt a cukor iránti preferencia, nagyobb volt az esélyük a túlélésre, vagyis az éhhalál elkerülésére. Ma azonban a csokoládé, a fagyi és a cukorbaj korában, amikor bőségesen sok kalóriát vehetünk magunkhoz, ez a genetikailag kódolt cukorimádat inkább káros, mint hasznos, hiszen rossz fogakhoz és elhízáshoz vezet.<sup>27</sup> Cukorban szegény környezetünk megváltozását tehát nem követte azon mentális diszpozíciónk megváltozása, amely arra indít minket, hogy minél több cukrot együnk, holott ez az egykor adaptív tulajdonság ma már egyáltalán nem adaptív.

Az evolúciós pszichológiát három ponton szokták támadni: túlzott innátizmusa, túlzott adaptacionizmusa és túlzott modularitása miatt. Mivel mindhárom vád – akár igaz, akár nem – egy-egy igen nagy horderejű elmefilozófiai, illetve kognitív tudományi vitában való elméleti állásfoglalást is jelent, érdemes nemcsak az evolúciós pszichológiát érő támadásokat, de az egyes vitás kérdéseket is részletesebben ismertetni.

Az evolúciós pszichológia egyik legfontosabb előfeltevése az innátizmus, tehát az öröklött, velülnszületett tulajdonságok jelentőségének hangsúlyozása. Az evolúciós pszichológia érvei nem működnek, ha mentális képességeink nagy része nem lenne velülnszületett, azaz genetikailag kódolt. Ezzel az evolúciós pszichológia erőteljesen elkötelezi magát az elméről való gondolkodás egyik legrégebbi, újra és újra felmerülő kérdésében, amely azt vitatja, hogy egyes szellemi képességeink velülnszületettek-e vagy tanultak.

A kérdés tehát az, hogy gondolatainkkal, reprezentációinkkal és szellemi képességeinkkel születésünk óta rendelkezünk-e, vagy tanulás útján sajátítjuk el őket. A két szélsőséges álláspont persze nem túl meggyőző. Egyrészt nem minden képességünk és reprezentációnk velülnszületett, más néven innát, hiszen ez esetben már az ókori görögöknek is rendelkezniük kellett volna például a helikopter reprezentációjával, ez pedig kevésbé valószínű.<sup>28</sup> Másrészt azt sem érdemes mondani, hogy minden tanult, ami az elménkben van, hiszen egy újszülött már sok információval rendelkezik a világról, tudja például, hogy miként kell szopni.

Születésünkkor tehát elménk nem tiszta lap, mint ezt John Locke és a XVIII. századi brit empiristák gondolták,<sup>29</sup> de nem is egy eleve adott elme, amelyben már ekkor kódolva van minden olyan tudás, amelyet egy felnőtt tud, mint ezt valamivel korábban Descartes és az őt követő racionalista tradíció gondolta.<sup>30</sup> A későbbi filozófusok és pszichológusok már megpróbálták elhatárolni mindkét szélsőségtől, de nagy különbségek voltak és vannak aközött, hogy az innátizmus vagy a tanulás pólusához állnak-e közelebb. Anélkül, hogy a probléma előtörténetét részletesen tárgyalnám, érdemes megemlíteni, hogy a behaviorizmus egyik alapállítása az ember szinte

<sup>23</sup> Cf. pl. Maynard Smith 1958.

<sup>24</sup> Még ha igaz is, hogy a hatmilliárd fős populáció erősen megosztott és nem homogén, így nem igaz, hogy mindegyik tagja minden másik tagjával kapcsolatba léphet.

<sup>25</sup> Lumsden–Wilson 1981 vitatja ezt az állítást, és amellett érvel, hogy az utóbbi néhány ezer évben is volt humán genetikai változás, amely a viszonylag friss kulturális hatásokat is beépítette. E témát részletesebben tárgyalja az utolsó fejezet.

<sup>26</sup> Buss 1994, 1995.

<sup>27</sup> Legalábbis a fejlett országokban.

<sup>28</sup> Ugyanakkor Jerry Fodor néhány írásában épp emellett érvel. Cf. pl. Fodor 1983.

<sup>29</sup> Locke 1979, cf. még Hume 1976, 1995.

<sup>30</sup> Descartes 1994a, 1994b, 1996.

korlátlan taníthatósága volt.<sup>31</sup> A hatvanas évek óta, nem utolsósorban a nyelvész Noam Chomsky hatására a tudományos közvélemény egyre nagyobb figyelmet szentelt az öröklött tulajdonságoknak.<sup>32</sup>

Fontos hangsúlyozni, hogy a velülnszületett, illetve tanult tulajdonságok kérdése nem belterjes, csak a tudósok számára lényeges probléma, hiszen igen komoly, mindenki számára jelentős kérdésekről szól. Azért vagyok okos, mert a szüleim is okosak voltak, vagy azért, mert neveltetésem során okos lettem? Ha azt akarom, hogy a gyerekeim intelligens legyenek, akkor intelligens anyát kell választanom neki, vagy jó iskolákba kell járatnom?

Az evolúciós pszichológia a tulajdonságok velülnszületettségét hangsúlyozza, hiszen ez előfeltétele annak, hogy az elme megértésében segítséget nyújtson az evolúciós magyarázat.<sup>33</sup> Az eddig egymást kizáró innátista-tanulásközpontú kettősséget azonban nem állítják élesen szembe. A velülnszületett és a tanult tulajdonságokról félrevezető úgy beszélni, mintha ezek csak egymás rovására létezhetnének. Nem igaz, hogy ha eggyel több velülnszületett tulajdonságunk van, akkor eggyel kevesebb tanult, és viszont. Az evolúciós pszichológia egyik legfontosabb állítása szerint minél több velülnszületett apparátussal rendelkezünk, annál hatékonyabban leszünk képesek a tanulásra. A velülnszületett és a tanult tulajdonságok tehát nem kizárják, hanem épp ellenkezőleg, erősítik egymást.<sup>34</sup>

Az a baj azonban, hogy az evolúciós pszichológusok nem mondják meg, hogy e kompromisszum konkrétan miként is valósul meg, és emiatt megoldásjavaslata önmagában üres marad. Az mindenesetre bizonyos, hogy az evolúciós pszichológia képes kikerülni az innátista elméletek legtöbb hibáját, a túlzott innátizmus vádját tehát jogosulatlanul bizonyult.

Érdekes módon az elme evolúciós megközelítésének egy másik nagy területe, az elmében zajló, más néven „neurális” evolúció vizsgálata ugyanerre az eredményre jut, de egy világosan leírt mechanizmust is ad ennek magyarázatára, amelyről a következő alfejezet fog szólni. Külön kiemelendő, hogy az evolúciós pszichológia és a neurális evolúció elméletei, amelyek kölcsönösen nem vesznek tudomást egymásról, ezen a ponton, az innát és tanult tulajdonságok viszonyának kérdésében nagyon hasonló megoldáshoz jutottak.

A második visszatérő ellenvetés az evolúciós pszichológiával szemben a túlzott adaptacionizmus vádját. Ez a vád valószínűleg jogos. Az adaptáció-exaptáció vita, amelyről az előző alfejezetben volt szó, az elme evolúciós megközelítése esetében is felmerül. A kérdés triviális formában úgy merül fel, hogy a sok képesség és tudásdarab közül, amely a fejünkben van, túlélésünk szempontjából mindegyik hasznos-e.<sup>35</sup>

Ez persze a probléma leegyszerűsítése, hiszen nyilvánvalóan még egyszer sem mentette meg az életemet az, hogy tudok másodfokú differenciálegyenleteket megoldani, sőt, primer túlélésem szempontjából még az sem volt segítségemre, hogy tudom, ki volt IV. Fülöp francia király harmadik fiúgyermeké. Az evolúciós pszichológia kicsit fordított ezen a kérdésfeltevésen, hiszen az érdeklődik, mi volt adaptív kőkorszaki őseink számára, nem az, hogy mi adaptív most. De annyiban igaza van a bírálóknak, hogy néhány evolúciós pszichológus, köztük az irányzat két alapítója, Leda Cosmides és John Tooby azt vallja, hogy ha valamilyen mentális reprezentációval rendelkezünk, akkor az szükségképpen adaptív volt valamikor számunkra; segített a túlélésben.<sup>36</sup> Ez az álláspont valóban vitatható, valószínűleg sok olyan tudás és képesség van a fejünkben, amely soha nem volt adaptív, hanem adaptív képességek mellékterméke, azaz exaptáció.

Soha nem segített a túlélésben az, hogy ismerem a gyökvonás műveletét, de a számolás már igenis sokszor segített, és a gyökvonás a számolás általános adaptív képességének nem-adaptív mellékterméke. Ráadásul sokszor számomra nem adaptív valamilyen szellemi képesség, de a szűkebb vagy tágabb környezetem túlélési esélyeit – s ezzel közvetve az enyémet is – növeli. Nekem személy szerint nem nőnek meg a túlélési esélyeim azzal, hogy tudom, miként működik például az elektromosság vagy a számítógép, de azáltal, hogy ez a találmány nagyban megváltoztatta tágabb környezetem életét, valószínűleg az egész társadalom számára adaptív az elektromosság, következésképpen közvetve az én számomra is az.

További bonyodalmakat okoz a kultúra világa. Primer túlélésemet nem segíti elő, hogy tudom, ki volt IV. Fülöp francia király harmadik fiúgyermeké, de ha megtanulom, akkor jobb jegyet kapok az iskolában, később jobb állást kapok, és esetleg nem kerülök utcára, és nem halok éhen.

Kevés adaptív képességet találunk tehát, ha tüzetesen megvizsgáljuk elménket, és valószínűleg kőkorszaki őseink is sok nem-adaptív tulajdonsággal rendelkeztek. Az evolúciós pszichológia alapötlete azonban nem kérdőjeleződik meg, ha elvetjük azt a tézist, hogy minden, ami a fejünkben van, adaptív. Ekkor persze a nem adaptív tulajdonságok megjelenésére is biológiailag korrekt elméletet kell adnunk.

Szerencsére az evolúciós pszichológia több részirányzatból áll, és ezeknek nem mindegyike fogadja el a totális adaptacionizmus előfeltételét. Az evolúciós pszichológia egyik legjelentősebb képviselője, Steven Pinker pél-

<sup>31</sup> Cf. Skinner 1953.

<sup>32</sup> Chomsky 1966. Cf. még Chomsky 1985, 1995. Az egész innátizmus problémát némileg más megvilágításban tárgyalja Elman et al. 1996.

<sup>33</sup> Pinker például az innátizmus modern kori atyjának, Noam Chomskynak a tanítványa, az ő örökségét viszi át az evolúciós megközelítésbe.

<sup>34</sup> Pinker 1997.

<sup>35</sup> Gould 1996, 1997, Dennett 1983.

<sup>36</sup> Cosmides–Tooby 1997. p. 14. Cf. még Stich 1998, Dennett 1998a

dául több mentális tulajdonságunkról is kimutatta, hogy exaptáció, és nem volt soha adaptív.<sup>37</sup> A túlzott adaptacionizmus ellenvetése tehát az evolúciós pszichológia egy részére áll, más részére viszont nem.

A harmadik ellenvetés az evolúciós pszichológiával szemben a túlzott modularizmus vádja. A modularizmus-interakcionizmus vita a percepcióelmélet régi vitája, innen került át az általánosabb kognitív folyamatok vizsgálatába.

A kognitív tudomány egyik legfontosabb területét alkotó percepcióelmélet középpontjában több évtizede az áthatolhatóság problémája áll: az a kérdés, hogy milyen viszonyban áll egymással az észlelés és a kogníció (a magasabb szintű elmeműködés, például memória, logika). Két markáns álláspont különíthető el. A modularisták szerint a percepció a magasabb szintű kognitív folyamatok számára áthatolhatatlan, a percepció folyamata nem használ semmilyen magasabb szintű tudást,<sup>38</sup> az interakcionistaik szerint viszont a kogníció tetszés szerint kontrollálhatja, irányíthatja és akár le is állíthatja a percepciót.<sup>39</sup>

A modularisták szerint a percepció folyamata önálló, automatikus és alapvetően buta: a retinát érő ingerből egy kimeneti szimbólumot hoz létre a perceptuális rendszer. Ez a folyamat mechanikus, abban az értelemben, hogy sem előzetes tudás, sem más perceptuális csatornák működése nem befolyásolja a rendszer működését. Ha valamilyen inger éri az érzékszervet, akkor automatikusan legyártódik a kimeneti szimbólum is, amelyet már a magasabb szintű elmeműködés fog használni.

Az interakcionistaik szerint a percepció és a kogníció több ponton is érintkezik: a perceptuális folyamatok használnak különböző rövid, közepes, illetve hosszú távú memóriát, és nincsenek elzárva a többi percepció csatornától: a percepciót befolyásolja egyrészt a magasabb szintű mentális tevékenység (sőt, a jókedv-rosszkedv is), másrészt a többi perceptuális folyamat is.

Általánosabban fogalmazva, a modularisták azt állítják, hogy az elmeműködés egymástól viszonylag független modulok működésének sorozata, tehát kisebb, elkülönült egységekre bontható. A látás modulja nem kommunikál a beszédértés moduljával, és mindkettő független a betűfelismerés moduljától. Az interakcionistaik viszont amellet érvelnek, hogy nincsenek ilyen merev és áthághatatlan határvonalak az egyes mentális feldolgozások között, az egyes területek kölcsönösen felhasználják egymás eredményeit.

A modularizmus-interakcionizmus vita – vagy ahogy másként nevezik, a területspecifikus-területáltalános vita – nem eldöntött.<sup>40</sup> Valószínűleg mind a törzsfejlődés, mind az egyedfejlődés során változik az elme felépítése, egy inkább interakcionista alaphelyzetből egyre elkülönültebb, modulárisabb struktúra alakul ki. Mindenesetre az igazság valahol a radikális modularizmus és a radikális interakcionizmus között van.

Steve Mithen például amellet érvel, hogy a modulok az egyedfejlődés során is változnak, de az emberré válás folyamatának során sem mindig voltak ugyanolyanok.<sup>41</sup> Mithen azt állítja, hogy a törzsfejlődés során többször is megváltozott az elme moduláris felépítése. A modularizmus-interakcionizmus vitát tehát nem lehet egy általános, absztrakt szinten eldönteni: az emberré válás folyamatának különböző szakaszaira a modularizmus más és más foka volt jellemző. Mithen szerint három ilyen szakasz különíthető el: a mai elme a teljes interakcionizmus kezdőpontjától az egymástól teljesen zárt modulokon keresztül a modulok közötti részleges átjárásig jutott el. Az emberré válás egyik legfontosabb lépése éppen a modulok közötti átjárás kialakulása volt. Mithen elméletében ezt a lépést a publikus nyelv és a tárgyak készítésének képessége tette lehetővé, lehet azonban amellet is érvelni, hogy a nyelv későbbi fejlemény, amelyet éppen az elme egyes területei közötti kommunikáció tett lehetővé. Ezekről a problémákról a VII. fejezetben lesz bővebben szó.

Néhány evolúciós pszichológus azonban, köztük a Cosmides–Tooby szerzőpáros, hajlik a radikális modularizmusra.<sup>42</sup> Szerintük egy mentális képesség egy modulnak felel meg. Ahhoz, hogy az egyes mentális képességek evolúcióját megmagyarázzák, izolálják őket egymástól. A beszédértés modulját ilyen és ilyen körületesi feltételekhez való alkalmazkodás alakította ki, a veszélyes ragadozók felismerésének modulját más körületesi tényező, következésképpen a két modul – mivel kialakulásuk feltételei különbözők – maguk is egymástól függetlenek lesznek, és nem is hozhatók kapcsolatba egymással.

Ezek a teoretikusok valószínűleg joggal kritizálhatók túlzott modularizmusuk miatt, de szerencsére ez a kritika nem érinti az evolúciós pszichológia egészét, hiszen ezen irányzat képviselői között nemcsak modularisták akadnak.<sup>43</sup> Akár a túlzott adaptacionizmus vádja, ez az ellenérv is csak az evolúciós pszichológia egy részére érvényes, más részére viszont nem.

Érdekes azonban felvetni egy negyedik, ritkábban hangsúlyozott problémát az evolúciós pszichológiával szemben. Mivel az evolúció az ember génkészletére van közvetlen hatással, a gének pedig az agy neurális struktúráját kódolják, ezért ha egy elmélet az elme evolúciós magyarázatára tesz kísérletet, akkor harmadik lépésként

<sup>37</sup> Pinker 1997. utolsó fejezet. Cf. még Plotkin 1997.

<sup>38</sup> A modularista álláspont: Fodor 1983, 1996a, Marr 1982, róla cf. Kovács 1991, Pléh 1986.

<sup>39</sup> Az interakcionista megközelítésről: pl. Gregory 1974, Rock 1980. A vitáról cf. Nánay 1997c.

<sup>40</sup> A modularizmus–interakcionizmus vita állóháborúja helyett többen egy rugalmasabb fogalompár, a területspecifikusság–területáltalánosság bevezetése mellett érvelnek. Cf. Elman et al. 1996.

<sup>41</sup> Mithen 1996.

<sup>42</sup> Cosmides–Tooby 1997. p. 31.

<sup>43</sup> Egy nem-modularista evolúciós pszichológia kidolgozására tesz kísérletet Pinker 1997, aki a modulok helyett – Chomsky nyomán – mentális szervek működéséről beszél.

az agy–elme kapcsolat kérdését is meg kell oldania<sup>44</sup>, azt tehát, hogy az agy evolúció által formált neurális struktúrája miként hoz létre mentális reprezentációkat, emóciókat vagy éppen ételpreferenciákat. Ezt a lépést az evolúciós pszichológia teljes egészében figyelmen kívül hagyja.<sup>45</sup>

Ezen a ponton válhat fontossá az elme evolúciós magyarázatának másik nagy területe, az elmén belüli evolúció kérdése, amely éppen arra a kérdésre keresi a választ, hogy egy individuum elméjén belül az egyszerű neurális kapcsolatokból miként alakulnak ki bonyolult reprezentációk, és ez a változás mennyiben tekinthető evolúciós jellegűnek.

A neurális evolúció elmélete tehát nem a törzsfajlódással foglalkozik – mint az evolúciós pszichológia –, hanem az egyedfejlődésre próbál evolúciós magyarázatot adni. Nem azt vizsgálja, hogy az előemberből miként evolválódott az ember, hanem azt, hogy az újszülött elméjéből milyen evolúciós mechanizmusok hozzák létre a felnőtt elméjét. Az evolúciós pszichológia az emberi elme filogenezisével foglalkozik, a neurális evolúció elmélete viszont az elme ontogenezisével. Az evolúciós pszichológiát az érdekli, mi történt a törzsfajlódásban, tehát az egyed születése előtt, a neurális evolúciót viszont az, mi történik a születés után.<sup>46</sup>

Különböző szférákban mozog tehát a két megközelítés, és ennek megfelelően kölcsönösen nem vesznek tudomást egymásról, holott az elme átfogó evolúciós megközelítésének mindkét problémát egységes elméleti keretbe kellene foglalnia.<sup>47</sup>

### 3. A neurális evolúció

William James, a múlt század végének igen nagy hatású pszichológusa és filozófusa mondta ki először, hogy az elme működése alapvetően evolúciós folyamat. Úgy érvelt, hogy mivel a környezettel összhangban lévő komplex rendszereket kizárólag evolúciós mechanizmusok hozhatnak létre, és mivel a gondolkodás kétségkívül a környezettel összhangban lévő komplex folyamat, ezért a gondolkodás folyamata evolúciós folyamat.<sup>48</sup>

Később sokan próbálták továbbvinni ezt a gondolatot,<sup>49</sup> igazán elterjedtté azonban Karl Popper, David Campbell és az evolúciós episztemológiának nevezett irányzat művei nyomán vált.<sup>50</sup> Popper szerint minden intelligens tevékenység az amőba mozgásától Einstein felfedezéseiig nem más, mint próbálkozások és hibakorrekciók egymásutánja.<sup>51</sup> Az amőba helyet változtat, megpróbál az egyik irányba állabakat növesztetni, majd, ha akadályokba ütközik, nem folytatja tovább ezt a mozgást, hanem más irányokkal próbálkozik, ahol nem állja útját semmilyen akadály, és arrafelé halad tovább. Hasonlóképpen, amikor matematikapéldát oldunk meg, akkor is több megoldási úttal próbálkozunk, és ha az egyikkel elakadunk, a másikkal folytatjuk. Popper szerint ez hasonló a természetes szelekció folyamatához: a próbálkozás felel meg a véletlen variációnak, a hibakorrekció pedig a szelektív pusztulásnak.

Popper és az evolúciós episztemológia inkább metaforaként használja az evolúció elméletét az elme működésének leírására. Az a kérdés, hogy az elmén belül zajló evolúció eszméje csak hasonlatként használható, vagy lehetséges olyan biológiailag korrekt elméletet is alkotni, amely az evolúciós folyamatok segítségével tudományos magyarázatot ad az elme működésére. A nyolcvanas évek közepén két kutató is kísérletet tett – egymástól függetlenül – a neurális evolúció tudományos elméletének kidolgozására, Jean-Pierre Changeux és a Nobel-díjas Gerald Edelman.

Changeux abból a jól ismert neurológiai tényből indul ki, hogy születésünkkor sokkal több neurális kapcsolat alakul ki az agyunkban, mint amennyire később szükségünk van.<sup>52</sup> Az, hogy ezek közül a kapcsolatok közül me-

<sup>44</sup> Érdemes ezen a ponton megemlíteni Poppert, aki szerint szintén kell hogy legyen kapcsolat az elme és az agy evolúciója között, de ez a kapcsolat épp fordított: az elme evolúciója határozza meg az agy evolúcióját, a viselkedés változik először, a biológiai szervek – így az agy is – csak követik e változást, cf. Popper 1972.

<sup>45</sup> Fodor hasonló ellenvetést tett. A konklúzió persze az evolúciós megközelítést keményen támadó Fodornál az evolúciós megközelítés teljes elvetése: „[...] köztudott, hogy a viselkedésünk evolúciója agyunk evolúciójának közvetítésével történik. Ha tehát az a kérdés, hogy az elmén adaptálódott-e vagy sem, nem azt kell vizsgálni, hogy mennyire komplex a viselkedésünk, hanem hogy milyen változásoknak kellett végbemenniük a főemlősök agyához képest, hogy agyunk képes legyen az emberi elme kognitív struktúráját produkálni. Erről viszont egyáltalán semmit nem tudunk, mivel nem tudjuk, hogy a mentális struktúránk milyen módon függ az agy szerkezetétől. [...] Agyunk igen hasonló a főemlősök agyához, míg az elménkről ez egyáltalán nem mondható el. Viszonylag kis változás az agy szerkezetében igán nagy viselkedésbeli eltéréseket eredményezhet főemlős őseinkhez képest. Ez esetben viszont vitathatóvá válik, hogy a kognitív komplexitást a viselkedési fenotípus graduális darwini szelekciója alakítaná. És itt nem segítenek az organikus evolúcióból vett analógiák sem, amelyek pedig hemzsegek az evolúciós pszichológiai szakirodalomban. Ha a zsiráf nyaka egy kicsit hosszabb lesz, az állat jobban el fogja érni a fa tetején lévő gyümölcsöt. Hihető tehát, hogy a szelekció fokozatosan egyre hosszabb nyakat alakít ki. De ha egy főemlős agya egy kicsit nagyobb (vagy barázdáltabb) lesz, ettől még viselkedése bármilyen irányban változhat.” Cf. Fodor 1998. p. 15.

<sup>46</sup> A törzsfajlódás és az egyedfejlődés, a filogenezis és az ontogenezis közötti kapcsolatról cf. Gould 1977.

<sup>47</sup> Az evolúciós pszichológia legfontosabb könyveiben csak egy-egy lábjegyzet, vagy még annyi sem található a neurális evolúció elméleteire.

<sup>48</sup> James 1880 különösen p. 456. Cf. még James 1890.

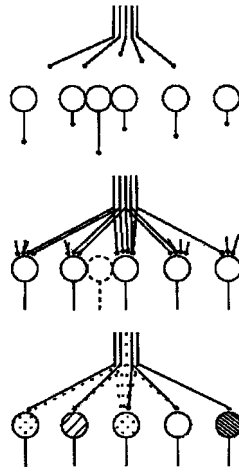
<sup>49</sup> A legfontosabb példa Mach 1927. Ernst Machról mint evolúciós gondolkodóról cf. Pléh 1998c.

<sup>50</sup> Popper 1972, Campbell 1974.

<sup>51</sup> Cf. Popper 1963, 1972, 1996, 1998, róla cf. Pléh 1995.

<sup>52</sup> Changeux 1985.

lyek maradnak meg és melyek halnak el, a környezeti hatások függvénye. A neurális kapcsolatok tehát olyanok, mint az ősvények az erdőben: ha használjuk őket, megmaradnak, ha nem, eltűnnek. Changeux felhasznál egy jól ismert érvet annak alátámasztására, hogy miért kell, hogy ez így legyen: az agyban több neurális kapcsolat van, mint amennyi génje van az embernek. Következésképpen az emberi agy összes neurális kapcsolata nem lehet genetikailag meghatározott. Éppen ezért az agyban a később szükségesnél lényegesen több kapcsolat alakul ki, azt pedig, hogy ezek közül melyek maradnak meg, a környezeti hatások fogják eldönteni. A neurális kapcsolatok között tehát a környezeti hatások szelektálnak.



A neurális darwinizmus elméletének illusztrációja. Az első kép az embrió idegsejtjei között éppen kialakuló kapcsolatokat, a második ábra a születés pillanatában már kialakult neurális kapcsolatokat mutatja. A harmadik képen pedig a felnőtt egyed idegsejtjei között fennálló összeköttetések láthatók, amelyek a születéskor meglévő kapcsolatok közül szelektálódtak a tapasztalati megerősítések által. Azok a kapcsolatok, amelyek nem kaptak semmiféle tapasztalati megerősítést, elhalnak (ezt szaggatott vonal jelöli). Az ábra Jean-Pierre Changeux: *L'homme neuronal*. (Paris: Fayard, 1983.) című könyvéből származik.

Ez a modell több szempontból is jól használható. Születésünkkor elegendő variabilitást biztosít az, hogy több neurális kapcsolatunk van, mint amennyi szükséges lenne, és emiatt sokkal könnyebben tudunk adaptálódni az új környezethez. Ha már születésünkkor fix lenne agyunk felépítése, akkor csak egy meghatározott környezetben lennénk képesek megmaradni, következésképpen nagyon szerencsés mutációk sorozatának kellett volna bekövetkeznie ahhoz, hogy ugyanaz az emberi faj az Északi-sarktól az Egyenlítőig mindenhol jelen lehessen. A neurális variabilitás tehát lehetővé teszi, hogy ne kelljen a szerencsés mutációra várni ahhoz, hogy egy adott környezetben is sikeresen túléljünk. Ezen a ponton érintkezik az evolúciós pszichológia által vizsgált probléma és a neurális evolúció: minél nagyobb a neurális variabilitás, minél több szerepe van tehát a környezeti hatásoknak a végleges neuronhálózat kialakításában, annál könnyebben tud egy élőlény a környezete változásaihoz alkalmazkodni.

Másrészt érdemes észrevenni, hogy a neurális evolúció modellje sikeresen egyesíti a velünk született és a tanult tulajdonságokat. Velünk született minden – vagy majdnem minden – neurális kapcsolatunk, de az, hogy ezek közül melyek maradnak meg, az a környezeti hatásoktól, tehát a tanulástól függ. Itt van tehát az a mechanizmus, amely képes biztosítani, hogy a velünk született és a tanult tulajdonságok nem zárják ki egymást, hanem éppen kölcsönösen erősítik egymás hatását. Minden tulajdonságunk velünk született a lehetőség szintjén, de minden tulajdonságunk tanult, mivel a környezet dönti el, mi fog ebből megvalósulni és mi nem.

Itt azonban nem szabad elfeledkezni arról, hogy Changeux az idegsejtek közötti kapcsolatok szelekciójáról beszél, és nem gondolatok, eszmék, vágyak szelekciójáról. Changeux elméletének önkényes kiterjesztése a bonyolultabb gondolatokra és reprezentációkra ugyanolyan idegen a tudományos megalapozottságtól, mint az evolúciós episztemológia metaforái.<sup>53</sup> Másfelől azonban amikor a gondolkodást akarjuk megérteni, nem sokat segít, ha tudjuk, hogy a neurális kapcsolatok között milyen szelekciós folyamatok zajlanak: ez nem magyarázza meg azt például, hogy miként oldok meg egy matematikapéldát. Valamilyen átkötésre volna tehát szükség a neurális szelekció biológiailag jól kezelhető, de mindennapi intuíciónk számára nem túl érdekes jelensége és a gondolkodás leírása között.

Erre tett kísérletet Gerald Edelman, aki Changeux-höz hasonlóan szintén egy szelekción alapuló elmémódellet állít fel, amelyet neurális darwinizmusnak hív.<sup>54</sup> Nála azonban a szelekció egységei nem az idegsejtek közötti kapcsolatok – mint Changeux-nél –, hanem nagyobb egységek: 100-1000 neuronból álló neuroncsoportok. Edelman szerint ezek a neuroncsoportok – amelyek bonyolultságban valahol az idegsejt és a mentális reprezen-

<sup>53</sup> Cf. Piatelli-Palmarini 1996.

<sup>54</sup> Edelman 1987, 1990.

táció között vannak – a szelekció egységei, ezek között szelektálnak a környezeti hatások. A neuroncsoport Edelman elmélete szerint tehát az összekötő kapocs az idegsejtek szintje és a gondolatok szintje között.

Az idegsejtek szintjéről tudjuk, miként működik a szelekció mechanizmusa, de ugyanezt szeretnénk tudni a gondolatok szintjéről is. Egy olyan közbülső szintre volna tehát szükség, amelyről még bizonyítható, hogy a szelekció szabályai szerint működik, de már nem áll olyan messze az intuitíve felfogható gondolkodástól. Edelman neuroncsoportjai viszont nem igazán képesek teljesíteni ezt a kettős kritériumot. Egyrészt a neuroncsoportok közötti szelekció létezéséről már távolról sincs olyan egyöntetűen elfogadott kísérleti bizonyíték, mint a neurális kapcsolatok közötti szelektív folyamatokról.<sup>55</sup> Másrészt a neuroncsoportok még nagyon távol vannak a gondolatoktól, még ha igaznak bizonyulna is, hogy valóban a környezeti hatások szelektálnak a neuroncsoportok között, ettől még semmivel nem jutottunk közelebb az emberi gondolkodás meggyőző és konkrét esetekre is alkalmazható magyarázatához.

Evolúcióbiológiai szempontból azonban további problémák is felmerülnek mind az edelmani, mind a changeux-i neurális evolúció elméletével szemben. A természetes szelekció két lépése közül ugyanis csak a második működik itt. A szelektív elhalás valóban jelen van, de a véletlen variáció kialakulásának lépése hiányzik. A születés után már eleve adva van az idegsejtek közötti kapcsolatok teljes variabilitása, ebből az alapanyagból fognak majd szelektálni a környezeti feltételek. A véletlen variáció kialakulásának lépése viszont nincs jelen.

Még súlyosabb probléma, hogy a neurális evolúció esetében nem beszélhetünk replikációról, legalábbis nem a szó eredeti evolúcióbiológiai értelmében. A sejt vagy az élőlény esetében a replikáció egyszerű: egy sejtből kettő lesz, vagy egy nyúlból kettő lesz. A neurális evolúció során azonban nem lesz több neurális kapcsolat, sőt, egyre kevesebb lesz, ahogyan a környezeti hatásoknak megfelelően egyre több szinapszis hal el. Ugyancsak nincs mutáció, hiszen ha nincs replikáció, akkor nem lehetséges hibás replikáció sem.

A replikáció fogalmát tehát újra kell definiálni ahhoz, hogy a neurális evolúció biológiailag tartható elmélet legyen. Erre tett kísérletet Paul Adams, aki úgy határozta meg ebben a kontextusban a replikáció fogalmát, mint a két idegsejt közötti kapcsolat megerősödését.<sup>56</sup> Adams abból indul ki, hogy a sejt esetében a replikáció a sejt minden komponenséből kettőt csinál, amikor egy gyerek születik, az anya egy tüdije, szíve, lépe helyett mind-ebből kettő lesz. A neurális kapcsolatra alkalmazva ez azt jelentené, hogy két sejt közötti kapcsolaterősség kétszer olyan nagy lesz. Ha viszont a szinaptikus átvitel erősödését tekintjük replikációnak, akkor érdemes ezt a fogalmat lazábban kezelni. Ha nem kétszeres lesz a szinaptikus átvitel, csak másfélszeres, akkor is lehet replikációról beszélni.

Sőt, mivel az emberi agy igen dinamikus, folyton változó és átrendeződő neuronhálózatot alkot, ahol egy kapcsolaterősség magától nem marad fenn, ebben a folyamatosan változó közegben már az is replikációnak tekinthető, ha egy kapcsolat nem hal el. Ahogy az erdei ösvények is eltűnnek, ha nem használják őket, úgy a neurális kapcsolatok sem maradnak fenn maguktól. Ha fennmaradnak, akkor beszélünk replikációról. Nem szükséges tehát sem az átviteli potenciál kétszeresére növekedését kimutatni, de még az egyszerű növekedését sem. Elég, ha nem csökken a két idegsejt közötti kapcsolat erőssége. A következőkben a neurális evolúció esetében a replikáció fogalmát úgy használjuk, mint egy neurális kapcsolat nem-elhalását.

Ez azonban megint újabb problémákat vet fel, hiszen ekkor a replikáció folyamata nem térbeli, hanem időbeli replikáció. Nem megduplázódik egy rendszer, hanem egyszerűen fennmarad. Kérdés, hogy az időbeli replikáció lehet-e evolúciós folyamatok alapja.<sup>57</sup> Fontos észrevenni, hogy a neurális evolúció nem az egyetlen olyan terület, ahol az evolúciós megközelítés nem térbeli, hanem időbeli replikációt előfeltételez. A társadalom szintén időben replikálódik: komponensei, az egyes személyek cserélődnek, a társadalmi berendezkedés mégis fix marad.

Mielőtt választ próbálnék adni ezekre a kérdésekre, érdemes áttekinteni, milyen problémák is merültek fel a neurális evolúcióval szemben. Három ilyen probléma volt: egyrészt nincs véletlen variáció, csak szelektív elhalás. Másrészt nem képes átkötést biztosítani az idegsejtek szintje és a bonyolultabb reprezentációk szintje között. Végül: az időbeli replikáció fogalmát feltételezi, amelyről nem világos, miként viszonyul az evolúcióbiológiában elfogadott replikációfogalomhoz. Ezt a három problémát kellene tehát megoldani.

Ezen a ponton vezetem be azt a modellt, amely a könyv gondolatmenetének vázát adja, ez pedig az elmén belüli, neurális evolúció és a nagy evolúciós átmenetek mechanizmusának összekötésén alapul. Amellett fogok érvelni a könyv első felében, hogy ha nem a természetes szelekció, hanem a nagy evolúciós átmenetek modelljét használjuk az elmén belüli evolúció leírásánál, akkor talán sikeresebben lesz használható az evolúciós megközelítés az elme leírására.

Ahogy a többsejtűek kialakulása esetében is előfordulhatott egy új szelekciós szint megjelenése, úgy a neurális evolúciót is lehetséges úgy tekinteni, mint nagy evolúciós átmenetek sorozatát. Az elméleti evolúció során tehát addig külön replikálódó egységek együtt kezdenek replikálódni. Az egyszerű szinapszisok tehát na-

<sup>55</sup> Edelman egyik legfontosabb kritikáját Francis Crick adta, aki az eredeti „neurális darwinizmus” cím helyett neurális edelmanizmusnak nevezi Edelman elméletét, rámutatva, hogy a darwinista evolúciómódelnek több ponton nem felel meg. Cf. Crick 1989.

<sup>56</sup> Adams 1998.

<sup>57</sup> Az időbeli replikáció az autopoietikus és az autogenetikus evolúciós elméletek egyik kulcsfogalma: ezek az elméletek egy olyan általános evolúciós keretelméletet próbálnak adni, amely egyaránt magában foglalja a klasszikus biológiai evolúciót és az időbeli replikáción alapuló evolúciós jellegű jelenségeket is. Cf. Maturana–Varela 1980, 1987, Csányi 1979, 1988, Kampis 1991.



gyobb egységekbe rendeződnek, és ezek az egységek versengenek a túlélésért, majd ezek a még nagyobb egységekbe szerveződnek. E modell szerint tehát a reprezentáció olyan, mint a többsejtű szervezet: kisebb elemek egysége, amelyek korábban egymással versengtek a túlélésért, de most együtt dolgoznak a nagyobb, őket magukba foglaló egység túléléséért.

Hogy világosabb legyen ez a mechanizmus, érdemes egy példát hozni. Tegyük fel, hogy egy béka csak legyekkel táplálkozik, és ezt oly módon teszi, hogy ha a retináján egy fekete pöttyöt észlel, akkor kinyújtja a nyelvét, és elkapja ezt a tárgyat, amely fekete pöttyként jelent meg a retináján. Ha a béka környezetében a légy az egyetlen rovar, akkor a béka igen hatékonyan tud ezzel a nagyon egyszerű módszerrel táplálékhoz jutni, hiszen, akárhányszor fekete pöttyöt lát, és kinyújtja a nyelvét, biztos lehet benne, hogy amit elkap, az légy lesz, és nem más. A „fekete pötty” reprezentációja kiváltja a nyelvkiváltás motoros akcióját, és végül a béka jóllakik.

Tegyük fel azonban, hogy a béka környezetében megjelennek a darazsak. A darázscsípés igen rossz hatással van a békára, nem elég, hogy nem lakik jól tőle, ha bekapja, de meg is csípi, ami miatt a béka legyöngül. Ha tehát a béka az előző stratégia szerint táplálkozna ebben a megváltozott környezetben, akkor minden valószínűség szerint elpusztulna, hiszen a darázs is fekete pöttyként jelenik meg a retináján, s ezért szegény állat ugyanúgy kapna be darazsakat is, mint legyeket.

Ekkor azonban egy különösen kreatív béka elméjében a „fekete pötty” reprezentáció helyett a „fekete pötty sárga csíkok nélkül” reprezentáció lesz aktív. Ez a béka tehát összekapcsolja a „fekete pötty” és a „sárga csíkok” reprezentációkat, és ezt a kapcsolt, bonyolultabb reprezentációt használja, amikor enni akar. Ha a „fekete pötty sárga csíkokkal” reprezentáció aktiválódik, akkor menekül, ha viszont a „fekete pötty sárga csíkok nélkül” reprezentáció, akkor boldogan bekapja azt a tárgyat, ami az orra előtt van, s ami jelen esetben mindig légy lesz, és soha nem darázs. Ez a kreatív béka óriási szelekciós előnnyel rendelkezik fajtársaival szemben, akik még mindig a primitív „fekete pötty” módszert használják, nem sok sikerrel.

Érdeemes megvizsgálni, hogy miként lehet leírni ennek az új reprezentációnak a megjelenését evolúciós nyelvezeten. Úgy gondolom, hogy itt az evolúció nagy átmeneteinek korábban bemutatott elméletét lehet leginkább használni. A „fekete pötty” és a „csíkos” reprezentációja, amely addig külön, egymástól függetlenül replikálódott, mostantól együtt replikálódik. Hasonlóan a sejtmaggal rendelkező eukarióta sejtek eredetéhez – amikor egy prokarióta (sejtmag nélküli) sejt bekapott egy másik prokarióta sejtet, és ahelyett, hogy szépen megemésztette volna, furcsa módon együtt kezdtek el replikálódni, és a bekapott sejt sejtmagként kezdett funkcionálni – itt is az történik, hogy két korábban külön életet élő reprezentáció egyszerre csak egy egységet alkot.<sup>58</sup>

Ha elfogadjuk, hogy ilyen evolúciós átmenetektől áll a mentális evolúció, akkor talán meg tudjuk válaszolni a neurális evolúcióval szemben felmerült három problémát. A véletlen variációt ekkor az újabb és újabb szelekciós szintek megjelenése biztosítja. Az idegsejtek közötti kapcsolatok szelekciója csak az első lépés, ezzel nem áll le az evolúciós folyamat. A neurális kapcsolatok nagyobb, néhány sejtől álló csoportokba rendeződnek, amelyek között újra beindul a szelekció mechanizmusa, majd ezek még nagyobb egységekbe, és így tovább. A variabilitást tehát nem a legalsó szint, az idegsejtek szintje szolgáltatja, hanem az újabb, egyre bonyolultabb egységek megjelenése, amelyek között szintén szelektálnak a környezeti hatások.

Szintén jobban kezelhető ezzel a modellel a bonyolultabb reprezentációk és az idegsejtek közötti összefüggés kérdése, hiszen az itt leírt példa épp azt mutatja, hogy az egyszerű reprezentációkból miként lesznek bonyolultabbak, majd még bonyolultabbak, pusztán a környezeti hatások nyomására. Amikor egy neurális kapcsolat egy nagyobb egység része lesz, többé már nem erősödhet vagy gyengülhet tetszés szerint a szinaptikus átvitele. A kapcsolaterősséget a nagyobb egység túlélése fogja meghatározni. Például a „borosüveg” reprezentációm felépítő neurális kapcsolatok nem változhatnak saját szakállukra, a „borosüveg” reprezentációm túlélését kell szolgálniuk, mint ahogy a májsejtem is az én fennmaradásomat szolgálja elsősorban, és csak másodsorban a saját túlélését.

A neurális evolúció új elmélete képes hidat teremteni az Edelman és a Changeux által kifejtett szelekciós elméletek között. Changeux szerint az idegsejtek között folyik a szelekció, Edelman szerint viszont a 100-1000 neuronból álló neurális csoportok között. Első látásra a két elmélet kizárja egymást, de ha elfogadjuk a nagy evolúciós átmenetek mechanizmusán alapuló új neurális evolúciós elméletet, akkor lehetővé válik, hogy mindkét szelekciós folyamatot egy elméleti keretben kezeljük. Az idegsejtek közötti kapcsolatok (amelynek kérdéséről Changeux elmélete szól) nagyobb egységekbe rendeződnek, melyek még nagyobbakba, egészen a 100-1000 neuront magukba foglaló egységekig (amelyekről Edelman beszél), és még tovább, még bonyolultabb reprezentációkig, mint amilyen például a „derékszögű háromszög” reprezentációja.

A mentális evolúció tehát újabb és újabb szelekciós szinteket hoz létre, az egyes szintek egymásba ágyazódnak, az alsó szint elemeiből épülnek fel a felsőbb szint egységei. Olyan ez, mint a katonai alakulatok rendszere. Tíz katona alkot egy szakaszt. A katonák autonómiája szinte teljesen megszűnt, a szakasz fennmaradását kell szolgálniuk, és nem saját érdekeiket. A szakaszok alkotják a század alakulatát. Az egyes szakaszok sem csinálhatják azt, amit akarnak, csakis a század érdekeit kell követniük, a század egysége és fennmaradása a céljuk, nem az egyes szakaszoké. Több század hasonló módon alkotja az ezred még nagyobb egységét, és így tovább.

<sup>58</sup> Az eukarióták evolúciójáról cf. Cavalier-Smith 1987.

Az idegsejtek ugyanígy alkotnak egyre nagyobb egységeket, amelyek mindig még nagyobb egységekbe rendeződnek. Az analógia persze csak az egyes szintek egymásba ágyazottságára vonatkozik, a hadseregben ugyanis sem szelekcióról, sem evolúcióról nem beszélhetünk, sőt, a katona valójában nem egy magasabb szintű egységtől kapja a parancsokat, amelynek ő is része, hanem egyszerűen a tizedestől.

Felmerül a probléma, hogy valóban új szelekciós szintek jelennek-e meg, vagy a tudós értelmezésének kérdése, hogy együtt kezeli – a jobb megértés kedvéért – a „fekete pötty” és a „csikos” reprezentációját. Valódi, objektíve elkülöníthető egységekbe rendeződnek az idegsejtek, vagy mindez csak a tudós szubjektív belevetítése? Ami miatt feltehető, hogy ezek a neuroncsoportok valódi egységeket képeznek, az az észleléssel és a motoros akciókkal való kapcsolat. A béka a nyelvkinyújtás motoros akcióját a „fekete pötty csikok nélkül” reprezentációjával kapcsolja össze, és nem mással, csak akkor ölti ki a nyelvét, ha ez a reprezentáció, illetve az ezt alkotó összes neurális kapcsolat aktív. Én a nagymamám arcát igen gyorsan felismerem, a „nagymama” reprezentáció tehát olyan stabil és összefüggő, valódi egységet alkotó reprezentáció, hogy a pillanat egy törtrésze alatt az egész reprezentáció, minden neurális kapcsolatával együtt aktiválódni tud.

Ezzel a példával megint visszajutottunk az időbeli replikáció kérdéséhez. Egy „nagymama” reprezentációból nem lesz több, nem osztódik, nem is lesz egyre erősebb (egy idő után), mégis lehet a replikációjáról beszélni, abban az értelemben, hogy ez a reprezentáció fennmarad egész életünkben, dacára annak, hogy a reprezentációt felépítő idegsejtek millió más fontos feladatot tudnának teljesíteni. Mégis ezt a reprezentációt szolgálják ki. A társadalom esetében is hasonló helyzetet találunk: a társadalmi berendezkedés korlátozza a társadalomban élő emberek autonómiáját, az emberek mégis betartják ezeket a korlátozó szabályokat, ahelyett hogy szabadon élnék világukat. A társadalom tehát aktívan replikálódik egyik pillanatról a másikra, azáltal hogy a szabad emberek mozgásterét csökkenti, és saját fennmaradásának szolgálatába állítja őket. Hasonlóképpen a reprezentáció is aktívan replikálódik, hiszen függő viszonyban tartja az idegsejteket.\*

Ha tehát elfogadjuk, hogy a gondolkodás evolúciós nyelvezeten jól leírható, ha nem a természetes szelekciót, hanem a nagy evolúciós átmenetek modelljét használjuk, akkor véleményem szerint biológiailag korrekt elméletet kapunk. Ez a modell lenne e könyv újdonsága, még sokszor elő fog kerülni a továbbiakban, amikor sorra véve az elmefilozófia és a kognitív tudomány klasszikus problémáit megvizsgálom, mennyiben jelent e modell új megközelítési módot e problémák kezelésében. Hogy ez a biológiával terhelt beszédmód jobban érthetővé váljon, az elméről alkotott, a mindennapi nyelvezethez közelebb álló leírások nyelvezetébe próbálom csomagolni.

Mindenekelőtt azonban lássuk, hogyan lehet és hogyan érdemes leírni az elme működését.

\* E bekezdés után a könyvbeli szövegből törölték a következőket:

A reprezentáció–társadalom analógia persze sántít, hiszen a mai társadalomban soha nincs az ember annyira függő helyzetben a társadalomtól, mint amilyenben a neuron van a mentális reprezentációtól, de mégis fontos közös pontok vannak, ha például olyan tárgyakkal vetjük össze a társadalom, illetve a mentális reprezentáció rendszerét, amelyekről intuitíve is beláthatjuk, hogy nem replikálódnak időben. Ilyen tárgyra volt példa a kilinc, amely szintén ugyanolyan marad az idő múlásával, mégsem szívesen mondjuk azt, hogy az ajtómon lévő kilinc időben replikálódik. A kilinc esetében éppen az a különbség, hogy nem áll önálló replikációra is képes komponensekből, sőt, komponensei – atomjai – rögzítettek, nem cserélődnek.

Egy lépéssel közelebb áll a mentális reprezentációhoz például a Duna, amelynek – bár komponensei cserélődnek – magasabb szintű szervezetsége azonos marad. Ugyancsak jó példa itt a kezünk, amelynek bőrszöveti cserélődnek és elhalnak, de a kéz alakja, működése és szerkezete azonos marad. E két példában tehát a komponensek változtak, de a struktúra megmaradt. A mentális reprezentáció és talán a társadalom is még egy lépéssel továbbmegy ennél. A mentális reprezentációt felépítő idegsejtek képesek lennének önálló replikációra (a fent leírt értelemben), ez a képesség azonban vissza van fogva, és alá van rendelve a reprezentáció replikációjának. Az „nagymama” reprezentációt felépítő idegsejt nem tehet mást, mint hogy a reprezentáció változásaihoz igazodik, nincs tehát önálló replikációja.

Hasonlóképpen egy katonai alakulat, például a díszzászlóalj is időben replikálódik. A katonák, akik ott szolgálnak, nem tehetnek bármit, például nem mehetnek el délelőttként úszni. A katonák tehát elsősorban a nagyobb egység, a díszzászlóalj egységét és fennmaradását biztosítják, mintsem saját replikációjukat és boldogságukat. A kilinc viszont nem rendelkezik önálló replikációra képes részekkel, ezért nem fejt ki semmilyen aktivitást annak érdekében, hogy fennmaradjon. Emiatt nem érdemes úgy leírni a kilincset, mint időben replikáló tárgy.

Fontos észrevenni, hogy az időbeli replikáció fogalma csak a nagy evolúciós átmenetek elméletén belül értelmezhető. Csak akkor van értelme a kilinc és például a reprezentáció közötti különbségről beszélni, ha elismerjük, hogy a reprezentáció olyan, mint a többszötti szervezet: kisebb elemek egysége, amelyek korábban egymással versengtek a túlélésért, de most együtt dolgoznak a nagyobb, őket magukba foglaló egység túléléséért.

### III. Funkcionalizmus

Az emberi elme tudományos leírásakor két nyilvánvaló megoldás adódik. Az egyik az agy idegsejtjeinek leírása. Ezzel több probléma is van, egyrészt ma még kevés tudással rendelkezünk erről a területről, és egy ideig nem is várható az agy pontos és részletes feltérképezése. Másrészt, még ha ismernénk is az agy teljes neuronhálózatát és annak aktivitásmintázatait, ez az idegsejtmintázat különböző egyedeknél más és más lesz. Az elme olyan, mint a kúpalakúra nyírt bukszus a parkokban. Két ugyanolyan alakú kúp belsejében az ágak soha nem ugyanúgy helyezkednek el, a külső alak mégis ugyanaz lesz.<sup>59</sup> Hasonlóképpen minden ember képes – kisebb felkészülés után – bebizonyítani Püthagorász tételét, de a neurális struktúra, amellyel mindezt megvalósítja, nagyon különböző lehet.

További problémát jelent, hogy ha ismernénk is az agy részletes térképét, és még ha ugyanaz a mentális tevékenység különböző emberek esetében ugyanazokkal az aktivitásmintázatokkal járna, még ekkor sem segítene mindez túl sokat annak megértésében, hogy mi is történik a fejekben, amikor például a nagymamámra gondolkok, vagy vörösborot akarok inni. Szükség van tehát egy olyan, az idegsejtek szintjénél magasabb leírási szintre, amely segít megmagyarázni mindennapi gondolkodásunkat.

Amikor az emberi gondolkodást akarjuk megmagyarázni, nem elég az idegsejtekről beszélni. Ez volna az egyik nyilvánvaló megoldás. A másik triviális válasz a behaviorizmus megoldása: az elméről nem tudunk semmi mást, csak hogy adott ingerekre miként reagál. Csak azt tudjuk tehát, hogy ha meglátok egy tigrist, akkor elszaladok. Az, hogy mindezt milyen elmeműködés segítségével teszem, nem megismerhető. Látható, hogy a neurológia és a behaviorizmus két szélsőséges modellt nyújt az elme leírására. A neurológia szerint az elme vizsgálata szempontjából lényeges egység az idegsejt, a behaviorizmus szerint pedig maga az emberi elme. Emiatt mindkét elmélet kénytelen elhanyagolni a sejt és az egész elme közötti köztes bonyolultságú egységeket, pedig mindennapi fogalmaink (gondolat, képzet, vágy, idea, ötlet) épp ezeknek felelnek meg. Az elme jobb leírásához tehát egy olyan közbülső szintre volt szükség, amely bonyolultságban valahol az idegsejtek szintje felett helyezkedik el, de ugyanakkor az egész elménél kisebb egységek kezelésére is alkalmas.

Az első ilyen irányzat, amely egységes elméletet adott e közbülső szint leírására, a funkcionalizmus volt.

#### 1. A funkcionalizmus és előzményei

A funkcionalizmus tehát a két végtől egyaránt megpróbál elhatárolódni.<sup>60</sup> Egyfelől többet akar mondani az elméről, mint a behaviorizmus, beszél tehát az elme építőköveiről, a mentális állapotokról. Másrészt viszont ezeket a mentális állapotokat (vagy reprezentációkat) nem az azokat megvalósító idegsejtek segítségével azonosítja, hanem funkcionális szerepük segítségével.<sup>61</sup> Amikor tehát egy mentális állapotról van szó, akkor nem azt vizsgáljuk, hogy mely idegsejtek tüzelése hozza létre azt, hogy a dugóhúzóra gondolok, hanem hogy milyen ingerek váltják ki ezt a gondolatot, milyen más reprezentációkat fog aktiválni, és végül milyen cselekvést vált ki. A „dugóhúzóra gondolni” mentális állapot esetében például ha szomjas vagyok, és látok egy bontatlan borosüveget, nem látok viszont dugóhúzót, akkor a dugóhúzóra fogok gondolni, és ez azt a cselekedetet fogja kiváltani, hogy felállok és megkeresem a dugóhúzót, majd kinyitom az üveget. Ez a funkcionális szerep fogja azonosítani a „dugóhúzóra gondolni” mentális állapotomat, és nem az azt megvalósító neurális struktúra.

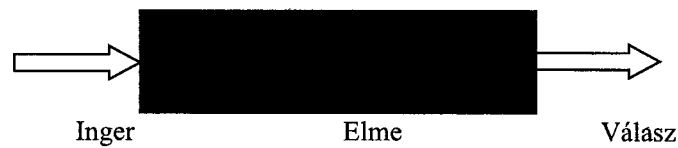
Talán ebből a példából is látszik, hogy a funkcionális leírás szempontjából az a fontos, hogy milyen ingerek válhatnak ki egy mentális állapotot, az milyen más mentális állapotokat vált ki, és végül milyen motoros válaszokhoz vezet. Mellékes, hogy az idegsejtek eközben éppen mit csinálnak. Ami lényeges, az a mentális állapot funkcionális szerepe: annak leírása, hogy őt milyen ingerek okozzák, és hogy ő milyen válaszokat okoz. Ha egy mentális állapot minden lehetséges funkcionális szerepét sorra vettük és leírtuk, akkor már mindent tudunk róla, anélkül hogy akár egy idegsejtet is megvizsgáltunk volna.

<sup>59</sup> A hasonlat persze Quine-tól származik, bár ő a különböző emberek nyelvi fogalmainak felépítésének különbözőségét illusztrálta ezzel. Cf. Quine 1960. p. 8.

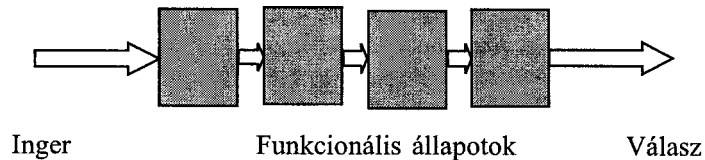
<sup>60</sup> Cf. pl. Putnam 1967, Hárman 1973, Block 1978, Fodor 1968, Smart 1971, Armstrong 1968, Lewis 1972, 1980, Shoemaker 1975.

<sup>61</sup> Cf. Block 1978, Lewis 1972, 1980.

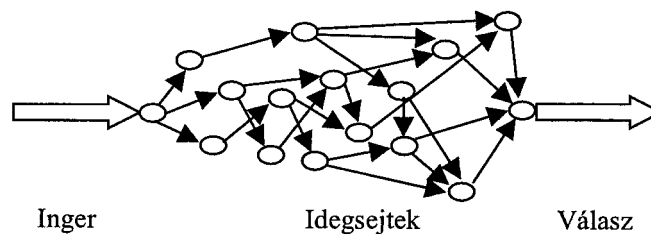
### Behaviorizmus:



### Funkcionalizmus:



### Redukcionizmus:



A funkcionalizmusnak tehát két alapvető állítása van, egyrészt az, hogy az elme egyes állapotairól lehet beszélni anélkül, hogy az azokat megvalósító neurális struktúráról beszéljünk, másrészt az, hogy helyett a funkcionális szerepekkel kell jellemezni ezeket a mentális állapotokat. Az első állítást úgy szokás emlegetni, mint az implementációfüggetlenség, vagyis a megvalósító közegetől való függetlenség téziséét. Ezt Hilary Putnam, a funkcionalizmus egyik alapító atyja úgy foglalta össze, hogy lehetnének ementáli sajtból is, ez nem változtatna semmit a funkcionális leírásról.<sup>62</sup>

A hetvenes években az elmefilozófia meghatározó irányzatává vált funkcionalizmus ezen a ponton megkísérel komoly elméleti előzményeket felmutatni. Arisztotelész az a filozófus, akiről megpróbálják megmutatni, hogy már ő is funkcionalista volt, hiszen a *De Anima* című könyvében a lélek egyes részeinek vizsgálatakor nem azt elemzi, hogy ezek a lélekrészek hogyan épülnek fel, hanem hogy mit csinálnak, tehát milyen funkcióval bírnak.<sup>63</sup> Anélkül hogy eldöntenénk azt a filológiai kérdést, hogy Arisztotelész funkcionalista volt-e, vagy sem, érdemes felhívni a figyelmet a modern funkcionalizmus és a régebbi, szellemi elődként felsorolt funkcionalisták közötti fontos különbségre.

A funkcionalizmus hetvenes-nyolcvanas évekbeli térhódítása nem választható el a mesterséges intelligencia kutatásának fejlődésétől. Ha ugyanis hiszünk abban, hogy egyszer mesterséges értelmet tudunk majd előállítani, akkor elengedhetetlen egy olyan elméleti leírás, amely az emberre is és a mesterségesen előállított értelmes rendszerekre is alkalmazható. Az implementációfüggetlenség tézise épp ezt a gyakorlati igényt elégíti ki. Amikor gondolkodási folyamatokról beszélünk, használjunk olyan nyelvezetet, amely nem kötelezi el magát azok mellett az intelligenciaformák mellett, amelyek véletlenül idegsejtekből épülnek fel, hiszen ezek csak egy alosztályát képezik a gondolkodó rendszereknek.

Nem az ementáli sajtból felépülő elme tehát az igazi kihívás a funkcionalizmus számára, hanem a szilikoncsipekből álló gondolkodó számítógép. A funkcionalizmus ennyiben a mesterségesintelligencia-kutatás szükségszerű elméleti hozadékának is tekinthető. A funkcionalizmus és a mesterségesintelligencia-kutatás közötti nem minden probléma nélküli kapcsolattal a következő alfejezet fog foglalkozni, érdemes azonban a funkcionalizmusnak ezt az elmefilozófiai irodalmon belül ritkán hangsúlyozott gyakorlati vonatkozását mindvégig szem előtt tartani.

Az egyik visszatérő filozófiai probléma például az elme leírásánál, hogy ne legyen se túl szűk, se túl tág a lényeknek az a köre, amelyre illik ez a leírás. A behaviorizmus kritériumai valószínűleg túl tágak: ha csak azt nézzük, hogy adott ingerekre miként reagálunk, akkor sok olyan rendszert is hozzánk hasonlóként fogunk leírni,

<sup>62</sup> Putnam 1967.

<sup>63</sup> Cf. Nussbaum–Putnam 1992. Ellene: Code–Moravcsik 1992, Code 1991, a vitáról jó képet ad a Nussbaum–Rorty 1992 kötet. Arisztotelész azonban a hetvenes évek funkcionalista hulláma előtt is a modern elmefilozófia felől olvasta újra például Gilbert Ryle. Cf. Ryle 1990. Putnamék voltaképpen Ryle olvasatát használják fel.

amelyek valójában nem gondolkoznak. Viszonylag könnyű lenne például olyan gépet kreálni, amely az öt érő ingerekre ugyanúgy reagál, mint a regényhős Oblomov, azaz sehogy, fekszik tovább a díványon, de ez a gép ettől még nem gondolkozna, pláne nem úgy, mint Oblomov. A behaviorizmus tehát túlzottan megengedő.

Másfelől, ha a megvalósító közeget tekintjük kritériumnak, akkor valószínűleg túl szűkre vontuk a gondolkodó lények körét, hiszen elképzelhető, hogy a marslakóknak, ha léteznek egyáltalán, nem idegsejtekből épül fel az elméjük. Ez a megközelítés tehát túl sovíniszta, kirekeszti a nem idegsejtalapú gondolkodó lényeket, a behaviorizmus viszont túl „liberális”, olyan rendszerekről is feltételezi, hogy gondolkoznak, amelyek valójában nem képesek erre.<sup>64</sup> A funkcionalizmus a túlzott liberalizmus és a túlzott sovínizmus között szeretné a középutat megvalósítani, amely nem rekesztene ki semmilyen gondolkodó marslakót vagy gépet, ugyanakkor nem is vonná túl tágra a gondolkodó lények kategóriáját. Ezt biztosítaná az implementációfüggetlenség kritériuma.

A funkcionalizmus második alapvető állítása, amely szerint a mentális állapotokat az absztrakt funkcionális szerepek segítségével lehet meghatározni, már közvetlenül következik az implementációfüggetlenség eszméjéből, hiszen ha a megvalósító közeg, tehát az idegsejtek szintje nélkül, attól függetlenül kell meghatározni a mentális állapotokat, akkor valamilyen más kritériumra van szükség, és ez a kritérium lesz a funkcionális szerep. Nem azt kell nézni, miből épül fel egy mentális állapot, hanem hogy mit csinál: mi okozza, és ő milyen más mentális állapotokat okoz.

Ezen a ponton azonban észre lehet venni a behaviorizmus örökségét a funkcionalista tanokon. A behaviorizmus szerint az elme egyetlen érvényes leírása az, hogy milyen ingerekre milyen választ ad. A funkcionalizmus az elmén belül azonosít ugyan bizonyos részegységeket, reprezentációkat, illetve mentális állapotokat, de ezeket pontosan ugyanúgy határozza meg, mint ahogy a behaviorizmus írja le az elmét: a mentális állapotok kiváltó ingereinek és motoros válaszainak segítségével. A behavioristák szerint az elme teljes egészében meghatározható, ha megmondjuk, milyen ingerekre milyen válasszal reagál. A funkcionalizmus továbbmegy ennél, és azt állítja, hogy az elméről ennél többet is lehet mondani: azt, hogy miből épül fel: milyen mentális állapotai vannak. Az egyes mentális állapotokat azonban behaviorista módon az ingerekkel és a válaszokkal definiálja. A funkcionalizmusban és a behaviorizmusban tehát közös, hogy kizárólag két magyarázati sémát ismer el az elmére vonatkozóan: az egyik a részekre bontás, a másik pedig az inger-válasz párokkal, azaz diszpozíciókkal való jellemzés. A különbség az, hogy más bonyolultsági szinten képzelik el e két magyarázati séma alkalmazását.

A funkcionalizmuson belül is vitatott azonban, hogy a mentális állapotok szintje, amely a funkcionális analízis tárgya, pontosan milyen leírási szintnek felel is meg.<sup>65</sup> A mentális állapotok funkcionális szerepe ugyanis más és más lesz, ha különbözőképpen írom le az ingereket és a válaszokat. Ha a „dugóhúzóra gondolni” funkcionális szerepét a kiváltó ingerrel és a motoros válasszal jellemzem, akkor azt is el kell döntenem, hogy milyen leírását adom az ingernek és a válasznak. Ha az ingert egyszerűen egy bontatlan borosüvegnek tekintem, a választ pedig annak, hogy megkeresem a dugóhúzót, akkor egy meghatározott nyelvezetet használtam a funkcionális elemzésben. Ha viszont úgy írom le az ingert, hogy egy szabálytalan henger alakú tárgy, amely piros színű folyadékot tartalmaz, a választ pedig úgy, hogy megfeszítem a combizomatot, majd a térdemet kinyújtom, akkor egészen más terminusokban beszéltem el ugyanazt a gondolkodási folyamatot. A második leírás empirikus, tapasztalaton alapuló, tudományos kifejezéseket használ, az első viszont eleve adott, a mindennapi szóhasználatnak inkább megfelelő terminusokat. Mindkét nyelvezet funkcionális leírást ad, el kell tehát dönteni, melyik az igazi.

Az ingerek tudományos, empirikus nyelven történő jellemzését használó funkcionalizmust pszicho-funkcionalizmusnak nevezzük.<sup>66</sup> Ez az irányzat értelemszerűen közelebb kerül a túlzott sovínizmus pólusához, hiszen az ingerek leírását a vizsgált rendszer konkrét működéséhez igazítja, és emiatt ha nem is idegsejtekről beszél, de konkrétan az emberi elme felépítéséhez kötődő terminológiát használ. A pszicho-funkcionalizmussal áll szemben a konceptuális funkcionalizmus, vagy más néven csak egyszerűen „funkcionalizmus”, amely az ingereket a mindennapi nyelv segítségével írja le.<sup>67</sup> A szoros értelemben vett funkcionalista leírás tehát a mindennapi szavainkat használja az egyes mentális állapotok és az elmét érő ingerek jellemzésére.

Itt a mindennapi nyelvnek különösen fontos szerepe van. A funkcionalizmus ugyanis úgy próbál beszélni az elméről, hogy nem távolodik el teljesen azoktól a terminusoktól, amelyeket egy nem szakmabeli használ, ha el akarja mondani, miről gondolkodik, mire vágyik, mit látott vagy mire emlékszik. Adva van tehát egy nyelvezet, amelyet sikeresen használunk az elméről beszélve, naivan és anélkül, hogy bármit is tudnánk a neuronok tüzeléséről. Ezt a nyelvezetet szokás népi pszichológiának nevezni: amikor másokról vagy saját magunkról beszélünk, ezt használjuk, olyan kifejezéseket, mint „akar”, „tud”, „gondol”, „vágyik”.<sup>68</sup>

A konceptuális funkcionalizmus megpróbálja úgy leírni az elmét, hogy annak építőkövei megfeleljenek a mindennapi nyelvezetünk terminusainak. Ebben a célkitűzésben persze észre lehet venni az analitikus filozófia egyik, Wittgensteinre és Ryle-ra visszanyúló nagy vezérvonalának hatását, amely szerint a nyelv, ezen belül is a beszélt, mindennapi nyelv olyan információkat hordoz rólunk és a világról, amelyet kár lenne nem hasznosíta-

<sup>64</sup> Cf. Lewis 1980, Block 1978.

<sup>65</sup> Cf. Block 1978.

<sup>66</sup> A pszicho-funkcionalizmus legfontosabb képviselői: Fodor 1968, Putnam 1967, Harman 1973, Block 1978.

<sup>67</sup> A konceptuális funkcionalizmus legfontosabb képviselői: Smart 1971, Armstrong 1968, Lewis 1972, Shoemaker 1975.

<sup>68</sup> Cf. Stich 1983.

ni.<sup>69</sup> A mindennapi nyelv elemzése tehát közelebb vezetheti a filozófust saját maga és a világ megismeréséhez. A funkcionalizmus végső soron ezt a wittgensteini leckét alkalmazza az elmé magyarázatára.

Ha sikeresen tudjuk értelmezni mások cselekedeteit azzal a leírással, hogy „mivel egy borosüveget látott, amelyet ki akart nyitni, és úgy gondolta, hogy a fiókban van a dugóhúzó, ezért odament a fiókhoz”, akkor az e leírásban használt kifejezéseket érdemes áttemelni az elmé filozófiai magyarázatába is. Ráadásul – két legyet ütve egy csapásra – ez még egy olyan általános elemzési szintet is lehetővé tesz, amely alkalmas a megvalósító közegetől független analízisre is.

A funkcionalizmus a hetvenes évek uralkodó elmefilozófiai irányzata volt, a nyolcvanas évektől kezdve azonban egyre több támadás érte különböző oldalról. Mielőtt azonban rátérnénk ezekre az ellenvetésekre, érdemes bemutatni a funkcionalizmus és a mesterségesintelligencia-kutatás kapcsolatát, annál is inkább, mert itt válik az absztrakt filozófiai elmélet a gyakorlatban is alkalmazhatóvá és – a filozófiában ez igen ritka jelenség – talán hasznossá.

## 2. Mesterséges intelligencia és gépi funkcionalizmus

A funkcionalizmus megjelenése nem elválasztható a mesterségesintelligencia-kutatás fellendülésétől. Akár az elmét, a legegyszerűbb gépeket is le lehet írni absztrakt funkcionális szerepek segítségével. Egy üdítőital-automata esetében például az ingerek – vagy más néven a bemenet – a pénzérmék bedobását jelenti, a kimenet, a motoros válasz pedig az üdítő és esetleg a visszajáró pénz kidobását.<sup>70</sup> Lehet azonban belső funkcionális állapotokat is hozzárendelni a gép működéséhez. Tegyük fel, hogy egy doboz kóla 100 forintba kerül, és a gép 50, illetve 100 forintos érméket fogad el. Ekkor a gépnek két belső állapota van, mindkettőt a funkcionális szerepe segítségével tudjuk azonosítani: az A állapotba akkor kerül a gép, ha bedobok egy ötvenest. A bemenet tehát az ötvenes, motoros válasz nincs, a gép A állapotba kerül. Ha viszont A állapotban még egy ötvenest bedobok, akkor egy doboz kóla lesz a kimenet, és a gép visszakerül B állapotba. B állapot az alapállapot, ha ekkor egy százast dobok be, akkor kimenet egy doboz kóla lesz, és a gép megintcsak visszakerül B állapotba.

A leírást lehetne tovább bonyolítani, de talán ebből is világos, hogy itt az egyes belső állapotokat ugyanazzal a módszerrel elemeztük, mint az elmé mentális állapotait, például „a dugóhúzóra gondolni” reprezentációt. Az absztrakt funkcionális vagy kauzális szerep számít, nem pedig az, hogy idegsejtek tüzelése vagy a Coca Cola Company ügyes szerkezete valósítja meg ezeket. A funkcionális leírás tehát gépekre és elmékre egyaránt jól passzol, ezért képezheti alapját a mesterséges értelem kutatásának is.

A mesterségesintelligencia-kutatást persze nem igazán az üdítőital-automata kérdése foglalkoztatja, sokkal inkább a számítógépek működése. A számítógép esetében pedig egy még szembevetőbb analógiát találunk a funkcionális leírásra, mégpedig a szoftver-hardver distinkciót.<sup>71</sup> Ugyanaz a szoftver különböző hardvereken is futhat: a szoftverszintű leírás, például az, hogy a szövegszerkesztőben begépelek egy szót, független attól, hogy milyen fizikai mechanizmusok, és-kapuk, illetve mikroprocesszorok valósítják meg. A szoftverszintű leírás tehát úgy viszonyul a hardverhez, mint a funkcionális leírás az idegsejtek tüzeléséhez: független tőle. Amikor arról beszélek, hogy villog a kurzor, akkor egyszerűbb azt számításba venni, hogy ezt a számítógépprogram melyik sora kódolja, mint pontosan kinyomozni, hogy a gép mely részén mely mikrokapcsolók változtattak állást. Ha tehát nem kell részleteiben ismerni a számítógép fizikai felépítését, hardverjét ahhoz, hogy számítógépprogramot írjunk, akkor, analóg módon, az agy konkrét neurális mechanizmusait sem kell ismernünk az elmé magyarázatához.

Az elmé és számítógépprogram közötti analógia hosszú ideig meghatározta, és részben ma is meghatározza az elméről való gondolkodást. Érdekes módon az elmé magyarázata és a gondolkodó gép létrehozásának terve kölcsönösen befolyásolja egymást. A gépeket úgy konstruálták, hogy hasonlítsanak az emberi elmére, az emberi elméről pedig számítógépes terminológiában beszéltek. E kétirányú hatás igen szövevényes kapcsolatot teremtett a két kutatási terület között, ráadásul ahogy a számítógépek egyre bonyolultabbá váltak, e kapcsolat folyamatosan átértékelődött.

E kétirányú hatásnak megfelelően a mesterségesintelligencia-kutatás filozófiai kérdései is két nagy csoportba sorolhatók. Egyrészt azt vizsgálják, hogy tudásunk az elméről mennyiben segít a mesterséges értelem létrehozásában. Ezen belül két szűkebb problémát lehet elkülöníteni: elvileg lehetséges-e az elméhez hasonló gépet létrehozni, illetve jelenlegi gépeink képesek-e erre. A mesterségesintelligencia-kutatás persze elsősorban nem az emberi elmé reprodukálására törekszik, hanem egyes specifikus feladatok, például a sakkozás képességének az emberi teljesítményt meghaladó színvonalúvá fejlesztését tűzte ki célul (mivel ennek van nagyobb piaca). Ez pedig többé-kevésbé sikerült is, elég, ha a Kaszparovot legyőző számítógépre gondolunk. Fontos azonban látni a különbséget e technikai bravúrok és az emberi elmé sokrétű gondolkodását reprodukáló gépek előállításának között.

<sup>69</sup> Ryle 1974, Wittgenstein 1992.

<sup>70</sup> Cf. Block-Fodor 1972, Nelson 1976.

<sup>71</sup> A szoftver-hardver analógiáról cf. Pylyshyn 1984, Haugeiand 1996.

Igaz, hogy a Deep Blue nevű számítógép talán jobban sakkozik, mint az ember (legalábbis, mint Kaszparov), de ugyanakkor a mesterségesintelligencia-kutatás a számunkra legtermészetesebb mindennapi feladatokat – például az arcfelismerést – sem tudja mind a mai napig igazán sikeresen megvalósítani. Az emberhez hasonló intelligencia létrehozása messze van tehát, de ez egyáltalán nem jelenti azt, hogy logikailag lehetetlen volna.

A mesterségesintelligencia-kutatáson belül kétféle álláspont különíthető el, a gyenge és az erős mesterséges intelligencia.<sup>72</sup> A gyenge mesterséges intelligencia álláspontja csak az elme vizsgálatának egy segédeszközét látja a számítógépben, anélkül hogy az elme és a számítógép közötti bármilyen lényegi hasonlóságról vagy azonoságról beszélne. Az erős intelligenciáé ezzel szemben azt állítja, hogy a számítógép működése valóban azonos lehet az elmeműködéssel. A számítógép tehát ugyanolyan értelemben gondolkodna, mint az ember. Hasonló in-gerekre ugyanolyan válaszreakciókat ad az ember és a gép, emellett ugyanolyan algoritmust is alkalmaz, más szóval ugyanolyan funkcionális állapotokat vesznek fel. Pusztán formális kritériumok alapján eldönthető tehát, hogy egy gép gondolkodik-e vagy sem.<sup>73</sup>

Eddig azt vizsgáltuk, hogy miként befolyásolta az elme kutatása a mesterségesintelligencia-kutatást. A fordított irányú hatás azonban legalább ilyen érdekes, mivel az elmefilozófiai és kognitív tudományi diskurzusra sokáig meghatározó befolyással volt a számítógép metaforája. A legkézenfekvőbb módon ezt az álláspontot a gépi funkcionalizmusnak nevezett irányzat képviselte.<sup>74</sup> Az emberi elmét számítógépprogramok tekintették: a számítógépprogramok alapegységeivel azonosították a gondolkodás építőköveit, és a gondolkodási folyamatokat az egyes programlépésekre próbálták visszavezetni. Vegyük észre, hogy itt éppen ellenkező előjelű jelenségről van szó, mint az elmefilozófia gyakorlati alkalmazásánál. Ott a gép egyre bonyolultabb, az emberi gondolkodáshoz egyre közelebb álló verzióinak megalkotásáról van szó, tehát egy egyszerű rendszer egyre komplexebbé tételéről. A gépi funkcionalizmus viszont egy komplex rendszert, az elmét próbálja egy egyszerűbb modellel leírni, amely nyilvánvaló leegyszerűsítésekhez vezetett.

De mit is jelent ez a leegyszerűsített elmemodell, amelyet a számítógépprogramok analógiája inspirált? Az alapgondolat az, hogy az emberi gondolkodás lépései olyan programsorokhoz hasonlatosak, mint „Ha  $A=1$ , akkor  $B=2$ ”. Az ehhez hasonló programsorok a hetvenes-nyolcvanas évek programnyelveire voltak jellemzők, a mai programozás már több ponton is bonyolultabb lehet ennél. Látható tehát, hogy az elmét egy meghatározott – és talán már túlhaladott – korszak számítógépprogramjainak nyelvezetével próbálta modellezni a gépi funkcionalizmus. Két aspektust kell kiemelni itt: az egyik a „ha... akkor” szerkezetek meghatározó szerepe, a másik pedig a szimbólumok használata. A gépi funkcionalizmus és őket követve sokáig az egész kognitív tudományi irodalom „ha... akkor” lépések, más néven következtetések sorozatának fogta fel az elmeműködést. A gépi funkcionalizmus szerint tehát a következtetések, az explicit szabályok az elme működésének irányítói.

A gépi funkcionalizmus és a számítógépprogram analógiájára felfogott elme elképzelésének másik sarkalatos pontja az, hogy a gondolkodást szimbólummanipulációnak tartják. A szimbólumok matematikai szimbólumok: a „ha  $A=0$ , akkor  $B=1$ ” programsor A, illetve B értékeinek felelnek meg: a szimbólum, akár csak a számítógép tárcseszei, tetszőleges értékeket felvehetnek, és ezeket változatlan formában elraktározzák.<sup>75</sup> Fontos észrevenni, hogy a szimbólumok teljesítik az implementációfüggetlenség kritériumát is. A gépi funkcionalizmus kiindulópontja tehát az, hogy az ember minden mentális reprezentációja formalizált, kvázinyelvi és explicit szimbólum. Minden, ami elménkben van, egy közös formális nyelvben van kódolva, akár a matematika szimbólumai. Mentális entitásaink tehát a formális nyelv egy-egy kijelentésének felelnek meg. A gondolkodás a gépi funkcionalisták szerint nem más, mint ezeknek a szimbólumoknak a manipulációja, betöltése, értékadása és tárolása.<sup>76</sup>

A gépi funkcionalizmus mindkét tézisének – mind a következtetések, mind a szimbólumok domináns szerepének – megkérdőjelezésére majd a konnekciónizmus irányzata tesz nagyszabású kísérletet, amelyről a VI. fejezetben lesz szó bővebben.<sup>77</sup> E ponton annyit kell csak megjegyezni, hogy a konnekciónista mesterségesintelligencia-kutatók nem a mesterséges intelligencia létrehozásának általános célkitűzését kérdőjelezték meg, hanem azt, hogy a hetvenes évek klasszikus számítógépprogramjai alapján bármi érvényeset is lehetne állítani akár az emberi elméről, akár az emberi elme hasonlatosságára létrehozott mesterséges gondolkodó gépekről. A konnekciónisták tehát belül maradtak a mesterséges intelligencia és az elme modellezése közötti vonzaskörön, de egy másik számítógépmodell alapján képzelték el ezt a kapcsolatot.

Most azonban – anélkül, hogy a klasszikus és a konnekciónista álláspont közötti vitákat ismertetnénk – érdemes azokról a leghíresebb bírálatokról szólni, amelyek nem a mesterséges intelligencia ilyen vagy olyan formáját kritizálják, hanem a mesterségesintelligencia-kutatás célkitűzését mint olyat. Vannak ugyanis, akik úgy gondolják, hogy nemcsak a jelenlegi számítógépek és a jelenlegi programok nem alkalmasak arra, hogy az emberéhez hasonló intelligenciát hozzanak létre, de ez a vállalkozás elméletileg is lehetetlen.

<sup>72</sup> A megkülönböztetés Searle-től ered. Cf. Searle 1996. p. 136.

<sup>73</sup> Ezt a kritériumot szokás Turing-tesztnak nevezni, cf. Turing 1964. Cf. még Hamad 1994.

<sup>74</sup> Fodor 1975, 1983, Pylyshyn 1984, Newell 1980, 1990.

<sup>75</sup> Csak a szimbólumok szintaxisa számít tehát, szemantikai tartalmuk irreleváns. Az elme szintaktikai, illetve szemantikai leírásának azonosságairól, illetve különbségeiről jó, bár kissé elfogult összefoglalást ad Fodor 1996b.

<sup>76</sup> Cf. főleg Fodor 1975, Pylyshyn 1984. Róluk Pléh 1998b.

<sup>77</sup> A legfontosabb konnekciónista szakirodalom: Rumelhart–McClelland 1986a, McClelland–Rumelhart 1986a, Smolensky 1996.

Három olyan elméleti ellenvetést érdemes megemlíteni, amely valódi kihívást jelenthet a mesterségesintelligencia-kutatás számára (a nagyszámú vallásos retorikát használó kritikát nem számolva). Az első Hubert Dreyfus nevéhez fűződik, aki beszédes, *Amit nem tudnak a számítógépek* című könyvében azt emeli ki, hogy az ember szociális lény, míg a számítógép nem az; az ember hihetetlen mennyiségű társadalmi szabályrendszert alkalmaz nap mint nap, öntudatlanul.<sup>78</sup> A vendéglőben például bal kézzel kell megfogni a villát, a borból először le kell tölteni a dugót magunknak, és csak ezután szabad a partnerünknek tölteni etc. Dreyfus szerint ha minden ilyen apró szociális szabályt és szokást explicit következtetésként kódolni kellene egy számítógépben, akkor olyan irdatlanul sok helyet foglalna el ez a sok szabály, hogy az már kezelhetetlenné tenné az alkalmazásukat. Dreyfus tehát a gépek szociális beágyazottságát hiányolja.

Egy másik átfogó kritikai ellenvetés viszont a perceptuális beágyazottságot kéri számon.<sup>79</sup> A számítógép sokkal lazább szálakkal kapcsolódik a környezetéhez, mint az emberi elme, amely folyamatosan ingerekkel szembe-sül. Ez az észrevétel nagymértékben támaszkodik a fenomenológia, különösen a francia fenomenológia elméletére, amely teljesen idegen mind az analitikus elmefilozófia, mind a kognitív tudomány filozófiai apparátusától. Merleau-Ponty, a francia fenomenológia egyik meghatározó alakja amellel érvel, hogy az elme és a környezet – amelybe beleértendő saját testünk is – nem választható el egymástól, az elme kialakulása mindig is kölcsönhatás: a környezethez való idomulás, a környezetbe való beágyazódás folyamata.<sup>80</sup> A minket körülvevő világ tehát ugyanúgy viszonyul hozzánk, mint a levél színe és visszája, egyik nincs a másik nélkül, s másik nincs az egyik nélkül. Zárójelben érdemes megjegyezni, hogy evolúciós terminológiában ez nem más, mint a koevolúció jelensége.

A szociális és a perceptuális beágyazottság mellett a harmadik nagy kritika John Searle kínai szoba argumentuma.<sup>81</sup> Searle az elme-számítógépprogram analógia lehetetlenségét egy gondolat kísérlet segítségével próbálja megmutatni. Képzeld el, hogy egy csak angolul beszélő ember ül egy zárt szobában, és kintről néha kínai írásjeleket tartalmazó papírokat kap. Egy angol nyelvű kis kézikönyve van, amely megmondja, hogy ha kap egy bizonyos rajzolatú írásjelet, akkor milyen vonalakat kell rajzolni a papírra, és azt kiadnia a szobából. Mindközben persze nem tud kínaiul, egy mukkot sem ért a bejövő és a kimenő írásjelekből. Searle azt állítja, hogy a mesterséges értelem hasonló ehhez a szituációhoz, a gép soha nem fogja *érteni*, mit is csinál. Mi viszont láthatóan értjük, amit csinálunk, következésképpen az emberhez hasonló mesterséges intelligencia létrehozása logikailag lehetetlen.<sup>82</sup>

Anélkül, hogy megválaszolnánk a gépi funkcionalizmus elleni három ellenvetést – a szociális beágyazottság és a perceptuális beágyazottság hiányát, illetve a kínai szoba argumentumot –, fontos kiemelni, hogy mind Dreyfus, mind Searle globális, minden jelenlegi mesterséges rendszer elleni kritikának szánta észrevételeit.<sup>83</sup> A konnekcionizmus megpróbálja megmutatni, hogy mindhárom kritika csak a klasszikus mesterségesintelligencia-kutatás modelljeire érvényes, de a két szerző ezt tagadja.<sup>84</sup> Mindhárom vitatott pontra visszatérünk még a konnekcionizmus tárgyalásakor. Az azonban bizonyosnak látszik, hogy a gépi funkcionalizmus modelljének érvényességét e három kritikai ellenvetés együtt meglehetősen kérdésessé teszi.

Végezetül érdemes megvizsgálni, hogy evolúciós szempontból hogyan néz ki a mesterségesintelligencia-kutatás és az elme modellezésének kapcsolata, és általánosabban az elme-számítógép analógia. Érdekes módon az evolúció kiindulópontja a mesterségesintelligencia-kutatás filozófiai kérdéseinek tárgyalásakor egymással radikálisan ellentétes érvekhez vezet. Egyfelől sokan használják annak bizonyítására, hogy a számítógép és ezzel minden számítógép alapú mesterséges értelem alapvetően különbözik az elmétől, hiszen az elmét az evolúció formálta, s ez nem mondható el a számítógépről. Az emberi elméhez hasonló mesterséges értelem létrehozása tehát nem lehetséges – szól az érv.

Ezzel szemben Daniel C. Dennett, az egyik legnépszerűbb kortárs filozófus egy kettős csavarral éppen az elme és a gép közötti hasonlóság alátámasztására használja az evolúció eszméjét.<sup>85</sup> Dennett bevallott szándéka, hogy megmutassa, nincsenek elméleti akadályai a mesterséges értelem létrehozásának, nincs tehát minőségi különbség az elme és a számítógép között. Ehhez azonban el kell tüntetnie egy elég kézenfekvő különbséget, mégpedig azt, hogy míg a számítógép működésébe mi csempészünk értelmet, addig mi magunktól, inherens módon vagyunk értelmesek. Minket nem programozott senki, hogy intelligensek legyünk, mi azok vagyunk magunktól is. A gép intelligenciája tőlünk származik, a miénk azonban eredendő.

Dennett igen frappánsan oldja fel ezt az ellentétet. A mi intelligenciánk sem eredendő, az is származtatott, mégpedig a minket kialakító evolúciós folyamatok terméke. Nem igaz tehát, hogy a számítógépet mi programoz-

<sup>78</sup> Dreyfus 1972.

<sup>79</sup> Cf. Varela–Thompson–Rosch 1992, Dreyfus 1972.

<sup>80</sup> Merleau-Ponty 1945.

<sup>81</sup> Searle 1996

<sup>82</sup> Searle érve tehát azt hangsúlyozza, hogy a funkcionalista rendszer nem képes kezelni a funkcionális állapotok tartalmát, valamire való irányulását, azaz intencionalitását. Erről a fogalomról részletesen az V. fejezetben lesz szó.

<sup>83</sup> Bár igaz, hogy Searle nem veti el egy, a mi neuronhálózatunkkal azonos szerkezetű, de mesterséges anyagból készült és mesterségesen létrehozott intelligencia lehetőségét.

<sup>84</sup> Dreyfus 1993, Dreyfus–Dreyfus 1986, 1988, Searle 1992.

<sup>85</sup> Dennett 1998a, 1998b.



tuk, de minket nem programozott senki. Mi is programozva vagyunk, mégpedig a szelektív folyamatok programoztak olyanra, hogy sikeresen túléljünk. Az emberi elmében ugyanúgy nincs semmi csodálatos, ahogy egy lemezjátszóban sincs. A különbség nem nagy: a lemezjátszót a Philips cég gyártotta, bennünket viszont a természetes szelekció.

### 3. A funkcionalizmus bírálatai

Az előzőekben a funkcionalizmus egyik vadhajtásának, a gépi funkcionalizmusnak a bírálatairól volt szó. Több ponton kritizálható azonban maga a funkcionalizmus általános eszméje is. Emlékeztetőül, a funkcionalizmus nem kötelezte el magát sem a „ha... akkor” típusú elmeműködés, sem a szimbólumok használata mellett, mindössze két dolgot állít. Egyrészt azt, hogy az elme állapotai implementációfüggetlenek, függetlenek tehát a megvalósító közegtől, az idegsejtektől. Másrészt az elme állapotai meghatározhatók az általuk betöltött funkcionális szerepek segítségével.

Három híres, sokat emlegetett kritikáról kell részletesen szólni. Ezek közül is a leghíresebb Putnam bírálata.<sup>86</sup> Putnam, aki a funkcionalizmus alapító atyái közé tartozott, később megtagadta saját korábbi elméletét, így az ő ellenvetése különösen sokat nyom a latban. Putnam érve szerint az elme állapotait nem kizárólag az elmén belüli történések határozzák meg, relevánsak lehetnek olyan tényezők is, amelyek az elmén kívül vannak.

Putnam sokat idézett és még többet vitatott iker-Föld gondolat kísérletében egy olyan bolygóról ír, amely pontos mása a mi Föld bolygónknak, azzal az egy kivétellel, hogy a víz kémiai felépítése nem H<sub>2</sub>O, hanem XYZ. Pontosan ugyanolyan tehát a két bolygó, és ezen belül az emberek is. Minden földi embernek, így nekem is létezik egy molekuláról molekulára azonos másolatom ezen a másik bolygón, amelyet Putnam iker-Földnek nevez. Az én agyamnak pontosan ugyanolyan a funkcionális felépítése, mint iker-földi másolatomé, de amikor arra gondol, hogy vizet akar inni, akkor mégis lesz valami különbség közöttünk, hiszen ő az XYZ vegyjelű szintelen, szagtalan, íztelen folyadékra gondol, én pedig a H<sub>2</sub>O vegyjelű szintelen, szagtalan, íztelen folyadékra.

Putnam ebből arra következtet, hogy mentális állapotaink azonosítási kritériumaiban szerepet játszanak olyan elmén kívüli, externális tényezők is, mint például a víz kémiai összetétele. Ezt az álláspontot nevezzük externalizmusnak, szemben az internalizmussal, amely azt állítja, hogy az elmén belüli tényezők önmagukban képesek meghatározni gondolatainkat, minden külső tényezőre való hivatkozás nélkül. Az externalizmus–internalizmus szembenállás az elmefilozófia egyik legélesebb vitájává vált.

Egy másik, szintén Putnamtól származó, kissé morbid gondolat kísérlet talán jobban megvilágítja, hogy mi is a vita tárgya. Képzeljünk el egy agyat, amelyet kioperáltak gazdája fejéből, és egy tartályban úszik már sok éve.<sup>87</sup> Egy ügyes professzor olyan ingerekkel bombázza az agy továbbra is működő receptorait, hogy az agy nem veszi észre, hogy kioperálták, úgy gondolja, hogy egy ép és egészséges testben éli világát, például épp egy virágos réten sétálgat. Putnam szerint nem ugyanaz a helyzet, amikor ez az agy virágos rétet lát, mint amikor én teszem ezt, hiszen én valóban a virágos réten járok, a tartályban úszó agy viszont nem.

E két gondolat kísérlet kétségkívül hatásos, érdemes azonban egy fontos distinkciót tenni. Két eset lehetséges: vagy meg tudom különböztetni a H<sub>2</sub>O és az XYZ vegyjelű folyadékokat, vagy nem. Ha igen, akkor a H<sub>2</sub>O és az XYZ különbsége valahol az elmémén belül is kódolva van, nem csak az elmén kívül. Ekkor tehát az externalizmus tézise nem áll. Ha viszont nem tudom megkülönböztetni e két folyadékot, akkor számomra mind egy, hogy mi a víz kémiai felépítése. Hogy világosabban megfogalmazzuk ezt a különbséget, érdemes átvenni Daniel Dennett „számunkra adott világ” fogalmát.<sup>88</sup> Dennett szerint két értelemben lehet a minket körülvevő világról beszélni, az én saját szemszögemből, illetve egy univerzális, mindent látó isteni nézőpontból. Az első szám egyes személyű szemszögből látott világot nevezi „számunkra adott világ”-nak. Ha nem tudom megkülönböztetni a H<sub>2</sub>O és a XYZ folyadékokat, akkor az iker-földi másolatom számára adott világ pontosan olyan, mint az én számomra adott világ. Jóllehet valóban vannak különbségek a minket fizikailag körülvevő világokban, de ez csak egy isteni nézőpont számára észrevehető.

A tartályban úszó agy esetében is mondhatjuk azt, hogy a számára adott világ pont olyan, mint bármelyik emberé, igaz, az őt körülvevő fizikai világ nem is hasonlít a miénkre. Kérdés, hogy az elme állapotainak meghatározásakor melyik világot kell figyelembe venni, a számunkra adott világot vagy az univerzális nézőpontból szemlélt környezetünket. Ha az első álláspontot fogadjuk el, akkor az internalizmushoz jutunk, ha a másodikat, akkor az externalizmushoz.

Anélkül, hogy állást foglalnék e vitában, feltétlenül fel kell hívni a figyelmet arra, hogy az externalizmus nemcsak a funkcionalizmus számára jelent vagy jelentene kihívást. Ha a mentális állapotokat külső tényezők is meghatározzák, akkor nemcsak a funkcionális állapotokra nem vezethetők vissza az elme állapotai, de az ideg-

<sup>86</sup> Cf. Putnam 1975a.

<sup>87</sup> Cf. Putnam 1981.

<sup>88</sup> Dennett „notional world” fogalmát fordítottam így, kissé szabadon. Cf. Dennett 1987

sejtekre sem, hiszen a fejünkön kívüli tényezők is beleszólnak abba, hogy mit is jelent a vízre gondolni. Erről a kérdésről azonban bőven lesz szó a redukció problémájáról szóló következő fejezetben.

A tartályban úszó agyak és molekuláris összetételekről szóló vita akadémikus fontoskodásnak tűnhet – talán az is. Annak tartja majdnem az egész kognitív tudományi irodalom is, nem igazán tud vele mit kezdeni. Egy olyan zárvány tehát, amelynek semmilyen gyakorlati alkalmazása nincs, ízig-vérig filozófiai kérdés. Van azonban egy ehhez szorosan kapcsolódó probléma, amely már nem ennyire a valóságtól elszakadt eseteket hoz fel, ez pedig a nyelvi és társadalmi externalizmus kérdése, amely a funkcionalizmus elleni második támadási pontot jelenti.

A nyelvi és társadalmi externalizmus is abból indul ki, hogy a mentális állapotainkat olyan tényezők is meghatározzák, amelyek nem a koponyánkban vannak, de ezen irányzat szerint ezek a külső tényezők nem képletek és tartályok, hanem a nyelv és a társadalmi szabályok.<sup>89</sup> A nyelvi externalizmus kiindulópontja az a tény, hogy szellemi kapacitásunk nem végtelen, ezért az életnek nem minden területén vagyunk maximálisan tájékozottak. Azokban a kérdésekben pedig, amelyekhez nem értünk, a szakértők véleményét tekintjük mérvadónak. Putnam nyelvi munkamegosztásnak nevezi ezt a jelenséget: sok ember van, aki nem tudja megkülönböztetni a nyárfát és a jegenyefát, de ha egy biológus megmondja neki, hogy a kertjében jegenyék vannak, akkor – bár még mindig nem tudja, mi is a különbség a két fa között – boldogan fogja újságolni ismerőseinek, hogy jegenyefa áll a háza előtt. Ez esetben tehát a „jegenye” reprezentációját nemcsak annak funkcionális szerepe határozza meg, de olyan, az elmén kívüli tényezők is, mint a szakértő véleménye és az a nyelvi közeg, amely a jegenye szó jelentését hősüinktől függetlenül meghatározza.

A nyelvi externalizmus valóban életszerűbb, mint a H<sub>2</sub>O-XYZ esete, hiszen nagyobb hatással van ránk és gondolkodásunkra a nyelvi közeg, amelyben felnőttünk, mint a víz vegyjele, de fontos észrevenni, hogy a putnami externalizmussal szembeni ellenérvek itt is működnek. A „számunkra adott világ” és a minket valóban körülvevő világ megkülönböztetése a nyelvi externalizmus esetében is sok nehézséget kiküszöböl, emellett ugyancsak igaz, hogy a nyelvi externalizmus is túllő a célon: a jegenye reprezentációja nemcsak a funkcionális szerepekre nem vezethető vissza, de az idegsejtek működésére sem. Ráadásul a nyelvi externalizmus csak a nyelvvvel rendelkező élőlények esetében tartható, így semmit nem mond a nem nyelvi állatok funkcionalista leírásáról.

Az externalista érvek ellen elképzelhető egy átfogóbb kritika is, mégpedig az evolúciós elmélet irányából. Ha én H<sub>2</sub>O-nak hiszem a vizet, iker-földi hasonmásom pedig XYZ-nak, annak ellenére, hogy az agyunk fizikai szerkezete molekuláról molekulára megegyezik, akkor nem adható evolúciós magyarázat a kétféle víz-reprezentáció (H<sub>2</sub>O-XYZ) közötti különbségre. A természetes szelekció a génekre hat, a gének az agy felépítését kódolják, az iker-földi hasonmásnak és nekem azonban ugyanolyan az agyunk. Az evolúció tehát ugyanolyanra formált bennünket, mégis különbözünk a víz megítélésében. Ha viszont a víz mint H<sub>2</sub>O és a víz mint XYZ reprezentáció közötti különbség nem evolúciós eredetű, akkor valamilyen más magyarázatot kell adni rá, és ezt az externalizmus elmulasztja megtenni. Ez az érv persze távolról sem jelenti azt, hogy az externalizmus nem tartható, ezen a ponton mindenesetre annyi bizonyos, hogy a funkcionalizmus első két kritikája nem bizonyult megsemmisítő erejűnek. A harmadik ellenvetés azonban már nehezebben kivédhető.

A harmadik támadási pont a funkcionalizmussal szemben az, hogy ez az irányzat nem tudja hova tenni a szubjektív érzetminőségeket, a kvália kérdését.<sup>90</sup> Intuitíve ez azt jelenti, hogy a számítógépnek nincsenek érzései, míg nekünk vannak. Érzetminőségnek nevezzük azt a szubjektív élményt, amely például a piros szín észlelését kíséri. A piros szín észlelését a funkcionalizmus természetesen absztrakt funkcionális szerepek segítségével határozza meg. Akkor kerülünk például ebbe az állapotba, ha a közlekedési lámpa legfelső színét nézzük, és ez olyan válaszreakciókat indít el, mint például hogy a fékre taposunk. Emellett azonban a piros színt kíséri egy szubjektív érzetminőség is. Nem tudhatom, hogy mások, akik a közlekedési lámpánál ugyanúgy viselkednek és ugyanolyan funkcionális állapotokba kerülnek, mint én, azok is ugyanazt az érzetminőséget érzik-e.

A színvakok és a szintévesztők például biztos, hogy nem ugyanazt érzik. Pirosnak látják azt a színt, amelyet mi zöldnek, és viszont. Ha azonban kellőképpen alkalmazkodnak a társadalmi elvárásokhoz, akkor pont úgy fognak viselkedni zöld színt látva, mint mi tesszük a piros szín észlelésekor. Az autóvezető szintévesztők a közlekedési lámpa legfelső színéről azt mondják, hogy piros, de a szubjektív érzetminőség, a kvália, amely ennek észlelését kíséri, a zöld szín szubjektív érzetminősége, az az érzetminőség, amelyet mi, nem színvakok a közlekedési lámpa legalsó színének észlelésekor szoktuk érezni. Megfordul tehát a két szín szubjektív érzetminősége, de a funkcionális szerep azonos marad. A szintévesztők is a pirosnál fognak fékezni, csak ők azt zöldnek látják. Az ő „zöld” mentális állapotuknak ugyanaz a funkcionális szerepe, mint az én „piros” mentális állapotomnak.

Bár nagymértékben gyengíti a funkcionalista érvelést a kvália-argumentum, de ebből sokan nem következtetnek a funkcionalizmus teljes elvetésére. Triviális megoldás például, hogy az érzetminőséget is fel kell venni

<sup>89</sup> Cf. Burge 1979, 1986, Putnam 1975b, Fodor 1994.

<sup>90</sup> A kvália, illetve a tudat témaköre igen kiterjedt irodalommal rendelkezik. A talán legfontosabb írárok: Nagel 1974, Baars 1988, Dennett 1991, Lycan 1990, McGinn 1991, Rosenthal 1986, Shoemaker 1975, Tye 1995, Flanagan 1992, Chalmers 1996, Searle 1992, Carruthers 1996, Churchland 1984, Dretske 1993, Jackson 1986, legjobb összefoglalások: Davies–Humphreys 1993, Block–Flanagan–Güzeldere 1997.

azon absztrakt funkcionális szerepek közé, amelyekkel a mentális állapot rendelkezik.<sup>91</sup> A piros észlelésének ugyanúgy kauzális következménye a piros érzete, mint ahogy az is, hogy rálépünk a fékre. A szintévesztők funkcionális állapota tehát csak részben azonos a miénkkel, hiszen nem váltja ki a „pirost érezni” mentális állapotot. A szubjektív érzetminőségek kérdése körül folyó vita tehát nem eldöntött, annyi azonban bizonyos, hogy jelentős mértékben gyengíti a funkcionalizmus érvelését.

Az evolúciós megközelítés és ezen belül az evolúciós pszichológia hajlik arra, hogy elfogadja az implementációfüggetlenség tézisé, hiszen amikor egy bizonyos mentális képesség eredetét vizsgálja, akkor el tud tekinteni attól, hogy milyen idegsejtek valósítják meg ezt a képességet. Amikor azt elemzik, miért volt adaptív a cukor iránti preferencia, akkor szükségképpen eltekintenek a hordozó közegtől, az idegsejtektől.

Az evolúciós pszichológiának tehát nincs ellenére az implementációfüggetlenség, de az egyes mentális állapotokat nem funkcionális szerepük segítségével azonosítja, hanem egykori, illetve mostani adaptív értékük segítségével.<sup>92</sup> A funkcionalizmus első állítását elfogadják tehát, de a másodikat nem. A második állítás elvetése is úgy jelenik meg azonban, hogy voltaképpen visszaadják a funkcionalizmusnak eredeti értelmét, amelyet a gépi funkcionalizmus elhomályosított. Visszahozzák tehát a funkció fogalmát a funkcionalista elemzésben, csak hogy ez a funkciófogalom nem absztrakt, hanem biológiai funkció lesz.<sup>93</sup> Az evolúciós pszichológia is egy funkcionalista elméleti keretben képzei el magát, belülről kritizálva azt.

Eddig tehát igazán megsemmisítő erejű elméleti ellenvetéssel nem találkoztunk a funkcionalizmus esetében, még az evolúciós megközelítés felől sem. Most azonban arra szeretnék rávilágítani, hogy a funkcionalizmus radikálisan kizár minden biológiailag tartható evolúciós megközelítést. Ha tehát komolyan vesszük az evolúció szerepét az elme vizsgálatában, akkor el kell vetnünk a funkcionalizmus legfontosabb állítását, az implementációfüggetlenséget. Vagy megfordítva, ha ragaszkodunk az implementációfüggetlenség téziséhez, akkor félre kell tenni minden evolúciós megfontolást.

Láttuk, hogy a mentális képességek elemzése az evolúciós előtörténet segítségével csak akkor elképzelhető, ha az agy és az elme közötti közvetítés megoldott. A környezeti szelektív hatások és az elme közötti összefüggés elemzésekor nem elég, ha csak azt vizsgáljuk, hogy a szelekciós hatások miként hatnak a génekészletre, a génekészlet és az agy szerkezete, illetve az agy és az elme közötti kapcsolatot is meg kell magyarázni. Ha tehát – mint láttuk – az elme evolúciós változását írjuk le, három lépésre van szükségünk. E három lépés: 1. a természetes szelekció hatása a genomra, 2. a génekészlet hatása az agy neurális felépítésére, 3. az agy neurális struktúrája és az elme közötti kapcsolat. Az implementációfüggetlenség tézisének elfogadása a harmadik lépés teljes kiiktatását jelenti.

Fontos észrevenni, hogy ez minden evolúciós megfontolásra igaz, nemcsak az elmén belüli evolúció elméletére. Az implementációfüggetlenség tézise e gondolatmenet szerint nemcsak a neurális evolúció elméletének, de minden elmére alkalmazott evolúciós gondolatmenetnek ellentmond. Ha elfogadjuk a neurális evolúció elméletét, akkor ez még nyilvánvalóbbá válik. Az agy–elme kapcsolatot jól képes kezelni a neurális evolúció elmélete: reprezentációink egyszerűbb reprezentációkból alakultak ki evolúciós mechanizmusok segítségével, az egyszerűbbek még egyszerűbbekből, és így tovább, egészen az idegsejtek szintjéig. Az idegsejtekből evolválódnak nagyobb egységek, neurális csoportok, majd ebből még nagyobbak, egészen az olyan bonyolult mentális állapotokig, mint „dugóhúzóra gondolni”. Ha viszont – provokatíven fogalmazva – a mentális állapotok az idegsejtekből evolválódtak, akkor nem lehetnek egyszerre függetlenek is tőle. Az implementációfüggetlenség és az elmén belüli evolúció kölcsönösen kizárják egymást.

Ez persze még önmagában nem jelenti azt, hogy a mentális evolúció a jó elmélet és az implementációfüggetlenség a rossz. Annyi azonban bizonyos, hogy ezt a két lovat nem lehet megülni egy fenékkal. Emellett még egy komoly érv szól az evolúciós modell mellett és a funkcionalizmussal szemben, ez pedig a tanulás szerepe. Ha a funkcionalizmus nem akarja azt állítani, hogy minden mentális állapotunk velünk született, akkor nem elég, hogy absztrakt funkcionális szerepeivel határozza meg ezeket a mentális állapotokat, kialakulásukat is meg kell tudnia magyarázni.

A tanulást – például a dugóhúzó reprezentációjának elsajátítását – igen nehéz kezelni a funkcionalista elmélet keretein belül, mégpedig azért, mert a tanulás, főleg a perceptuális tanulás az ingerek kategorizációján alapul. A kategorizáció viszont nem más, mint magasabb szintű mentális állapotok kialakulása alacsonyabb szintűekből, hiszen a kategória sem eleve adott. A kategorizáció – és így a tanulás – magyarázatához tehát nem egy, hanem legalább két szintet kell figyelembe venni az elemzésnek: a kategorizálandó ingerek szintjét és a kategória szintjét magát. Egy ilyen elemzés azonban nem elképzelhető a funkcionalista fogalmi kereten belül, amely csak egy elemzési szinttel foglalkozik, amelyet függetlennek tételez minden öt felépítő alacsonyabb szinttől.

<sup>91</sup> Cf. Lewis 1980

<sup>92</sup> Pinker ezt expliciten is bevallja, cf. Pinker 1997, különösen az első fejezetet.

<sup>93</sup> Fontos különbség, hogy a biológiai funkció többek szerint normatív fogalom, ezzel szemben a klasszikus elmefilozófiai funkcionalizmus funkciófogalma nem normatív. Cf. pl. Cummins 1975. A biológiai funkciót másképpen etiológiai funkciónak is nevezik. Cf. Wright 1973, Neander 1991, Cummins 1975.

Ezen a ponton érdemes Lycan teleológiai funkcionalizmusnak nevezett elméletét segítségül hívni.<sup>94</sup> Lycan szerint a funkcionalizmus nem tett mást, mint a behaviorizmus számára elemezhetetlen fekete dobozt jelentő elmélet komponenseire bontotta. Ezek a komponensek lettek a funkcionális állapotok, amelyek maguk egyenként pont olyan elemezhetetlen fekete dobozok voltak, mint a behaviorizmus számára az elme, csak kicsiben. Lycan azt kérdezi, hogy miért kellene itt megállni. A funkcionális állapotokat még tovább lehet bontani, még kisebb fekete dobozokra, azokat még annál is kisebbekre, és így tovább, egészen az idegsejtekig.

Lycan hosszú ideig tartó egy helyben toporgásból szakította ki a funkcionalizmust. Két kisebb probléma azonban felvethető kétségkívül frappáns megoldásával szemben. Egyrészt miközben a funkcionalizmus belső kritikáját próbálja adni, szép lassan kívül kerül a funkcionalista diskurzuson anélkül, hogy észrevenné. Az itt vázolt lépcsőzetes koncepció – ha komolyan vesszük – megkérdőjelezi az elme implementációfüggetlenségét, s így a funkcionalizmus elméleti kiindulópontját, hiszen Lycan egy funkcionális állapotot azoknak a komponenseknek a segítségével ír le, amelyek felépítik, tehát éppen az implementációt vonja be az elemzésbe. Lycan teleológiai funkcionalizmusa tehát nem funkcionalizmus, bár ő ezt nemigen akarná elismerni.

A második probléma ennél érdekesebb, és rávilágít a lycani elmélet és az elmén belüli evolúció elképzelésének különbözőségére is. Lycan felülről lefelé halad, az evolúciós megközelítés alulról felfele. Lycan egy adott funkcionális állapotot bont komponenseire, majd ezeket további komponenseire, ahogy az elemzés megkívánja. Ezzel szemben az evolúciós megközelítés fordított irányt követ. Az idegsejtek felől halad a bonyolultabb neurális csoportokon keresztül a funkcionális állapotokig. Emellett Lycannál azok a komponensek, amelyekre felbontja a funkcionális állapotokat, önkényesek, a filozófus dönti el, hogy elemzése szempontjából melyeket érdemes egyszerűbb egységekként tekinteni. Az evolúciós megközelítés viszont szigorú biológiai megkötésekkel dolgozik. Az idegsejtek nem tudják előre, milyen idegsejt-csoportokba fognak rendeződni, s e csoportok sem tudják, hogy milyen reprezentációt fognak kiadni. Az új és új szerveződési szintek tehát nem az elemző önkényes segédkonstrukciói, hanem az elme rendszerének belső működéséből következnek.

Mind Lycan, mind az evolúciós megközelítés egyetért azonban azzal, hogy a funkcionalizmus által elemzett egyetlen leírási szint helyett több szinten kell elemezni az elme működését, az idegsejtek szintjétől az egyszerűbb mentális állapotok szintjén keresztül a funkcionális állapotok szintjéig. A különbség az, hogy az evolúciós megközelítés e szintek egymáshoz való viszonyát és keletkezését is magyarázni próbálja.

Mindez persze nem jelenti az evolúciós megközelítés elsöprő győzelmét a funkcionalizmus felett. A következő fejezetben általánosabb elméleti keretben fogjuk vizsgálni a gondolatok, vágyak, eszmék világa és az idegsejtek közötti kapcsolatot, tehát az agy és az elme egymáshoz való viszonyát. Erről a viszonyról, amelynek elemzését a redukció problémájának szoktunk nevezni, a funkcionalizmus és az evolúciós megközelítés mást és mást állít.

<sup>94</sup> Lycan 1987

## IV. A redukció problémája

A funkcionalizmus problémaköre voltaképpen csak egy speciális esete a redukció problémájaként emlegetett kérdésnek. Ez a fejezet azt vizsgálja, hogy redukálható-e az elme az agyra, s ha igen, akkor mit jelent ez a redukció. A kérdés tehát az, hogy vissza lehet-e vezetni a mentális állapotainkat neurális, tehát fizikai állapotokra. Ha a dugóhúzóra gondolok, ezt a gondolatomat lehet-e redukálni az agyam idegsejtjeinek aktivitási mintázataira, vagy sem.

Nagy tétje van a redukció körül folyó vitáknak, ha ugyanis tagadjuk, hogy lehetséges lenne e redukció, akkor meg kell mutatni, hogy mi az a plusz a mentális állapotainkban, ami nem redukálható a fizikai világra, s ez könnyen egyfajta anyagtól elkülönült lélekfelfogás felé sodorhat el. Másfelől, ha elfogadjuk, hogy az elme redukálható az agyra, akkor az elme állapotainak vizsgálata ezzel feleslegessé válhat, hisz elég a neuronokról beszélni, a mentális állapotok nem is igazi állapotok, csak az idegsejtek tüzelése valódi, következésképpen ezekkel kell foglalkozni.

A helyzet persze nem ilyen egyszerű, több köztes út is elképzelhető az itt említett két véglet között. Ráadásul, mint a figyelmes olvasó észrevehette, mindeztidáig nem mondtuk meg, mit is jelent az, hogy redukció. A redukció fogalma többféle értelemben jelenik meg az elmefilozófiai irodalomban, nem mindegy, mely jelentését használjuk. Ez a fejezet úgy is tekinthető, mint az előző, funkcionalizmusról szóló rész tágabb, általánosabb kontextusba helyezése.

### 1. Redukálható-e az elme?

A redukció kifejezést alapvetően kétféle értelemben használjuk. Az első ezek közül a régebbi: egy elmélet redukciója egy másik elméletre. A kémia például redukálható a fizikára, ha minden kémiai törvényhez találhatunk egy, a fizika nyelvén megfogalmazott törvényt, amely azzal ekvivalens. Ehhez persze azt is le kell írni, hogy a törvényekben szereplő kémiai terminusok miként fejezhetők ki fizikai terminusok segítségével.

Egy elmélet tehát visszavezethető egy másik elméletre, ha egyrészt meg tudjuk adni azokat az „áthidaló törvényeket”, amelyek az egyik elmélet kifejezéseit definiálják a másik segítségével, másrészt ha a redukálandó elmélet törvényeit megadjuk a másik elmélet terminusaiban is.<sup>95</sup> A kémia tehát akkor redukálható a fizikára, ha minden szakkifejezést – például a vegyértéket – meghatározzuk fizikai terminusokkal, és az összes kémiai törvényhez – például Avogadro tételéhez – hozzárendelünk egy fizikai törvényt a fizika nyelvén és annak keretein belül.

A század első felében nagy redukciós hullám söpört végig szinte az összes tudományon. A végső alap, amelyre mindent visszavezettek, a fizika volt, a kémiát a fizikára, a biológiát a kémiára, a pszichológiát a biológiára, a társadalomtudományokat pedig a biológiára akarták redukálni.<sup>96</sup> Itt rögtön adódik az a kézenfekvő probléma, hogy miért lesz jobb, ha véghezvittük a redukciót, hogy tehát a gyakorlati alkalmazások terén hasznos-e a redukció, vagy sem. Érdemes-e például a francia forradalmat fizikai terminusokban, elektronpályákat és elemi részecskéket emlegetve elemezni?

Emellett a minket e pillanatban érdeklő lépés – a pszichológia biológiára való redukálása – esetében további gondokkal szembesülünk. Az elme működését leíró elméletet kellene az agy biológiai felépítését leíró elméletre visszavezetni. E két elmélet közül azonban egyik sem tekinthető olyan szép és komplett elméletnek, mint például a fizika vagy a kémia. A pszichológiának nem biztos hogy egyáltalán léteznek törvényei, de még ha léteznek is, akkor sem olyan világos, egyértelmű és kivétel nélküliek, mint például a fizika törvényei.<sup>97</sup> Az agy szerkezetének törvényeit pedig, bár lehet, hogy elvileg léteznek, jelen pillanatban még alig ismerjük; igen kevés információ van arról, hogy miként is működik az agy. Egy nem-elméletet kellene tehát egy még nem ismert elméletre visszavezetni, és ez nem könnyű.

A másik sokszor emlegetett probléma a pszichofizikai redukcióval (így nevezik az elme-agy redukciót) az, hogy egy mentális állapot sokféle fizikai, neurális állapotnak felel meg.<sup>98</sup> Amikor én a dugóhúzóra gondolok, azt egészen más neuronok egészen másképp fogják megvalósítani, mint amikor az amerikai elnök a dugóhúzóra gondol. Az áthidaló törvények, még ha léteznek is, nem olyan szép, egyszerű, egysoros összekötést biztosító definíciók lesznek, mint a legtöbb kémiai terminus fizikai meghatározása, hanem annál sokkal bonyolultabbak. Egy mentális állapothoz irtatlan sok, akár logikailag végtelen fizikai megvalósulást kell rendelniük, és kérdés, hogy egy ilyen áthidaló törvény mennyiben tekinthető egyáltalán törvénynek.

<sup>95</sup> Cf. Fodor 1974, Boyd 1980, Nagel 1961.

<sup>96</sup> Cf. pl. Neurath 1972.

<sup>97</sup> Millikan 1993. Cf még Rosenberg 1985.

<sup>98</sup> Boyd 1980, Fodor 1974, cf. Pereboom-Kornblith 1991.

A pszichológia redukciója a neurobiológia elméletére tehát ha nem is lehetetlen, de egy olyan bonyolult elméletet eredményezne, amely még sokkal áttekinthetlenebb, mint az eredeti elméletek bármelyike. Fontos hangsúlyozni, hogy természetesen vannak konkrét neurológiai eredmények, amelyek – különösen az elsődleges vizuális látókéregben – valóban azonosították egyes látási folyamatok neurológiai mechanizmusát, és minden valószínűség szerint egyre több területről lesznek neurológiai eredményeink. De még ha pontosan átlátnánk is az agy működésének minden részletét, akkor sem lenne sok fantázia bonyolult áthidaló törvények felállításában, amely mondjuk a feltékenység neurológiai terminusokban leírt megfelelőjét adná meg. A pszichológia redukciója tehát, az említett problémák miatt, elvesztette filozófiai érdekességét.

Éppen ezért, az utóbbi időben a redukció fogalmát egy egészen más értelemben kezdték használni. Nem az elméletek, hanem azok egyes építőköveinek visszavezethetősége került a vizsgálatok fókuszába. Nem az a kérdés tehát, hogy a pszichológia redukálható-e, hanem hogy az elme redukciója lehetséges-e. Ehhez azonban az áthidaló törvényekkel dolgozó előző kritériumok helyett újakat kellett találni.

A redukció új feltételrendszerét a szupervencia fogalma biztosítja.<sup>99</sup> A szupervencia függési reláció egy dolog két tulajdonsága között. Érdekes először egy nem elmére vonatkozó példán megvizsgálni, mit is jelent ez a ködös megfogalmazás. Sokat vizsgált kérdés például, hogy egy műalkotás, mondjuk egy szobor esztétikai értéke mennyire függ a szobor alakjától. Az az intuitív sejtésünk, hogy nagyon, hiszen mi mástól függene, mint a szobor alakjától. Ezt az erős kapcsolatot fejezi ki a szupervencia fogalma. A szobor esztétikai értéke szuperveniál a szobor alakján, ha két, ugyanolyan alakú szobor esetében az esztétikai érték is ugyanolyan lesz.

Ez persze triviálisnak tűnhet, de nem az. Marcel Duchamp „Fontaine” című installációja például nem volt más, mint egy piszoár, ugyanolyan alakú, mint az egyetem első emeleti vécéjének egyik piszoárja, a két tárgy esztétikai értéke mégis nagyon más lesz (egyébként a pénzügyi értéke is). Duchamp installációja esetében tehát az esztétikai érték nem szuperveniál a tárgy alakján. A szupervencia fogalma tehát két tulajdonság közötti függést fejez ki. Intuitíve érezzük, hogy valamilyen erős kapcsolat van a szobrok alakja és esztétikai értéke között, de hogy ez a kapcsolat milyen erős, annak egzakt jellemzésére jó a szupervencia.

Az elme vizsgálatában értelemszerűen a mentális tulajdonságok és a fizikai tulajdonságok közötti szupervencia lesz érdekes.<sup>100</sup> Az elme szuperveniál az agyon, ha két, neurálisan azonos agy pontosan ugyanolyan elmét fog létrehozni. Precízebben: egy  $F$  mentális tulajdonság szuperveniál egy  $G$  neurális tulajdonságon, ha minden olyan  $x$  tárgy, amely rendelkezik  $G$  neurális tulajdonsággal, szükségképpen rendelkezik  $F$  mentális tulajdonsággal is. A redukció új definíciója pedig az lesz, hogy az elme akkor és csak akkor redukálható az agyra, ha szuperveniál az agyon.

Ez a meghatározás elég a redukcióhoz: ha teljesül ez a kritérium, akkor igen szoros kapcsolat van a mentális és a neurális tulajdonságok között. Egy neurális tulajdonság csak egy mentális tulajdonságot kódolhat. A szupervencia bevezetése tehát ügyesen kiküszöbölte a redukció kérdésének egyik legproblematiszabb pontját, azt, hogy egy mentális állapotnak több neurális állapot is megfelelő lehet: az én dugóhúzó-gondolatom, az amerikai elnöké etc. A szupervencia irodalma nem foglalkozik azzal, hogy ha egy bizonyos mentális állapotban vagyunk, akkor milyen az idegsejtek aktivitásmintázata, amelyek mindezt megvalósítják, csak az ellenkező irányt vizsgálja. Ha egy bizonyos neurális állapotban vagyunk, akkor muszáj egy meghatározott mentális állapotban is lennünk. Jóllehet ezt a mentális állapotot más neurális struktúra is megvalósíthatja.

A korábbi, áthidaló törvényeken alapuló redukciófogalomhoz képest tehát megengedőbbek lettek a kritériumok. A szupervencia fogalmán alapuló redukció voltaképpen alulról felfelé irányuló redukció. Ez paradox állításnak tűnhet, hiszen a visszavezetés mindig egy bonyolultabb szintet vagy entitást próbál visszavezetni egy egyszerűbbre. Itt azonban a felső és az alsó szint, a redukálandó és az, amire redukálok, a fizikai és a mentális egészen más kapcsolatban van. Egyedül az a fontos, hogy ha adott egy a fizikai tulajdonság, akkor szükségképpen adott legyen a mentális tulajdonság is. Csak a fizikai felől a mentális felé irányuló kapcsolattal foglalkozik tehát a szupervencia kérdésköre, a másik iránnyal nem, ezzel kikerüli azt a feladatot, hogy a mentális tulajdonsághoz hozzá kelljen rendelni minden lehetséges neurális megvalósulását. Az áthidaló törvények mindkét irányú hozzárendelést biztosítanak, a szupervencia csak a neurális–mentális hozzárendelést, ami lényegesen kisebb feladat, de látni fogjuk: még ez sem biztos, hogy sikerül.

Ezen a ponton érdemes tisztázni egy terminológiai kétértelműséget. Az az állítás, hogy „ $X$  redukálható”, nem azonos azzal, hogy „ $X$  redukálható  $Y$ -ra”. Eddig csak a második kifejezéssel foglalkoztunk:  $X$  redukálható  $Y$ -ra, ha  $X$  szuperveniál  $Y$ -on. Meg kell azonban határozni azt is, hogy mikor lesz valami redukálható. Az elme például nyilvánvalóan redukálható az egész fizikai világra (ha csak nem akarunk dualisták lenni, és nem anyagi magyarázóelvéket is behozni), biztos azonban, hogy nem lesz redukálható kizárólag a retina idegsejtjeire.

Nyilvánvaló megoldásnak tűnik, hogy akkor mondjuk, hogy az elme redukálható, ha redukálható az agyra, sokan ezt a redukciómeghatározást fogadják el. Itt azonban érdemes egy gyakorlati megjegyzést tenni. Ha az elmeről kiderül, hogy nem redukálható az agyra, de redukálható az agyra és még egy, az agyon kívüli tárgy fizikai tulajdonságára, akkor praktikus szempontból ezen egyetlen tárgy egyetlen fizikai tulajdonságának leírása már semmiség lesz ahhoz képest, hogy az agy teljes neurális struktúráját feltérképeztük. Ha tehát például kiderítjük,

<sup>99</sup> A szupervencia fogalma az analitikus etikából került át az elmefilozófiába. Cf. Hare 1953.

<sup>100</sup> Cf. Davidson 1980.

hogy az elme működése az agy működésére és a pillanatnyi balatoni viharjelzésre redukálható, akkor a viharjelzés megállapítása gyerekjáték lesz ahhoz képest, hogy az agy pillanatnyi állapotát meghatározzuk.

Egy lazább definíciót ajánlok tehát: az elme akkor redukálható, ha redukálható fizikai tulajdonságok valamilyen véges halmazára. Az agyra való redukció természetesen kielégíti ezt a definíciót, de például az említett kévéssé valószínű viharjelzésre és agyra együttesen történő redukció is. Az agy önkényes kiragadása véges halmazok közül már csak azért is vitatható, mert egyrészt elképzelhető, hogy az agy fizikai tulajdonságain kívül más tárgyak fizikai tulajdonságai is hatnak az elmére, másrészt egyáltalán nem biztos, hogy az agy minden fizikai tulajdonsága releváns lesz.  $X$  tehát akkor és csak akkor redukálható, ha létezik egy olyan véges tulajdonsághalmaz, amelyen szuperveniál.

A szupervenciának több fajtáját is meg lehet különböztetni. Ezek közül a két legfontosabb a globális és az erős szupervencia (emellett még legalább tíz szupervenciatípust definiáltak).<sup>101</sup> Az erős szupervencia azonos a fenti definícióval: ha egy agy olyan neurális állapotban van, mint az enyém, akkor olyan mentális állapotban is kell lennie. Fontos kiemelni, hogy az erős szupervencia esetében csak az agy számít, az agy állapota teljes egészében meghatározza az elme állapotát.<sup>102</sup>

A globális szupervencia ellenben azt mondja ki, hogy ha két lehetséges világ megkülönböztethetetlen fizikailag, akkor megkülönböztethetetlenek a két világban élő emberek gondolatai is.<sup>103</sup> Ha erről a meghatározásról lebontjuk a lehetséges világokra való hivatkozás fontoskodó mázát, akkor azt kapjuk, hogy a globális szupervencia esetében az egész fizikai világ határozza meg az elme állapotát. Akkor szuperveniálnak neurális tulajdonságaimon a mentális tulajdonságaim, ha két fizikailag ugyanúgy felépülő világban ugyanolyan lesz az elmém állapota.

Duchamp „Fontaine” című provokatív műalkotásának esztétikai értéke például nem szuperveniál erősen a tárgy fizikai tulajdonságán, hiszen egy pontosan ugyanolyan fizikai tulajdonságokkal rendelkező tárgynak (az egyetem első emeleti piszoárjának) igencsak különböző az esztétikai értéke. Ugyanez a műalkotás azonban globálisan szuperveniál a fizikai tulajdonságokon, hiszen a „Fontaine” és a kortárs képzőművészek és művészetkritikusok agya, valamint a képzőművészeti folyóiratok együttesen már meghatározzák a műtárgy esztétikai értékét.

Az erős és a globális szupervencia között tehát az a különbség, hogy míg a globális esetében az egész fizikai világon szuperveniálnak mentális tulajdonságaim, addig az erős szupervencia esetében csak az én agyam fizikai tulajdonságain. A globális szupervencia tágabb fogalom, és igen nehezen tagadható, hacsak nem akarjuk azt is tagadni, hogy minden anyagi szubsztanciából épül fel. A globális szupervencia tagadása tehát a dualizmus elfogadását jelenti, azt a tézist, hogy léteznek nem anyagi valóságok, lelkek, szellemek. Ha ki akarjuk kerülni a dualizmust, muszáj elfogadni a globális szupervenciát. Az erős szupervencia viszont a redukcionizmust jelenti, következetesen igen lényeges kérdés a kétféle szupervencia egymáshoz való viszonyának tisztázása.

Jaegwon Kim mellett érvel, hogy a két szupervencia-fogalom azonosnak tekinthető.<sup>104</sup> Ha ez igaz lenne, akkor ez azt jelentené, hogy a dualizmus és a redukcionizmus között nincs középút, csak úgy tagadhatjuk a dualizmust, ha ezzel elfogadjuk a redukcionizmust, és viszont, ami nem tűnik nagyon kecsegtető alternatívának.

## 2. Nem-reduktív materializmus

Nem mindenki ismeri azonban el, hogy csak e két út lehetséges: a dualizmus és a redukcionizmus. Azokat az elméleteket, amelyek egy köztes álláspont megalapozására tesznek kísérletet, nem-reduktív materializmusnak hívjuk, ahol a materializmus szónak semmi köze a dialmathoz, egyszerűen azt a tézist foglalja össze, hogy nem létezik nem anyagi szubsztancia, nem létezik a testtől független lélek. Ha tüzetesebben megvizsgáljuk, a funkcionális materializmus mindenképpen nem-reduktív materialista elmélet, hiszen egyrészt azt állítja, hogy az elme teljes egészében anyagi természetű, másfelől viszont az elme állapotait függetlennek (tehát nem redukálhatónak) tételezi az idegsejtektől, tehát a megvalósító közegtől.

Dualizmus	Nem-reduktív materializmus	Redukcionizmus
Nincs szupervencia	Globális, de nem erős szupervencia	Erős szupervencia

<sup>101</sup> Chalmers 1996, Kim 1984a, 1987, 1998 jó áttekintést ad ezek egy-egy részéről.

<sup>102</sup> Formálisan:  $F$  erősen szuperveniál  $G$ -n, ha minden  $x$ -re és minden  $F$  tulajdonságra igaz az, hogy ha  $x$  rendelkezik  $F$  tulajdonsággal, akkor rendelkezik a  $G$  tulajdonsággal is, akkor következetesen, minden lehetséges világban, bármely  $y$ , amely  $F$  tulajdonságú, egyben  $G$  tulajdonságú is lesz. Cf. Kim 1984a. p. 165.

<sup>103</sup> Formálisan:  $F$  globálisan szuperveniál  $G$ -n, ha két olyan lehetséges világ, amely  $G$  tulajdonságok tekintetében megkülönböztethetetlen,  $F$  tulajdonságok tekintetében is megkülönböztethetetlen lesz. Cf. Kim 1984a. p. 168.

<sup>104</sup> Kim 1984a.

Lássuk tehát, milyen érvekkel lehet a nem-reduktív materializmust a dualizmus Szköllája és a redukcionizmus Kharübdiszé között átkormányozni, tehát mindkét oldalról megvédeni.<sup>105</sup> Kétféle megoldás kínálkozik a funkcionalisták és minden más nem reduktív álláspont számára, az externalizmus és az instrumentalizmus.

Az instrumentalisták szerint az elme igenis redukálható fizikai állapotokra, de ettől még nem kell eldobni a mentális leírásokat, mert azok hasznosak lehetnek.<sup>106</sup> Az instrumentalizmus szerint az elmének nem valódi állapota az, hogy például a madridi Pradóra gondolok, vagy hogy palacsintát akarok enni.<sup>107</sup> Ezek a mi elmeről alkotott leírásunk részei. Természetesen elménk működése nem más, mint neuronok aktiválódása, de ezt mégis érdemes úgy leírni, mint mentális reprezentációk sorozatát, ha ekkor jobban képesek vagyunk megjósolni, miként fog az illető tárgy (esetünkben az ember) működni. A termosztátnak vagy a lemezejátszónak is lehet mentális állapotokat, vágynak, hiteket, gondolatokat tulajdonítani, csak nem biztos, hogy érdemes, nem biztos, hogy ezzel jobb predikciókat tudunk tenni a termosztát vagy a lemezejátszó működéséről. Nem ontológiai, hanem episztemológiai különbség van tehát köztem és a termosztát között: nem a valódi működésben van a nagy különbség, hanem abban, hogy miként érdemes leírni ezt a működést.

Látható, hogy az instrumentalizmus jól egyensúlyoz az eliminatív redukcionizmus és a dualizmus között. Nem redukcionista, hiszen – pragmatikus okok miatt – továbbra is a pszichológiai terminusokat használja, de nem is dualista, hiszen elismeri, hogy végső soron minden, ami az elmében történik, csak és kizárólag az idegsejtek műve. A mentális állapotok csak leírások, elméleti konstrukciók. Hasonló szerepet töltenek be az elme leírásában, mint például a súlypont a mechanikában: nincs valóságos létezése, mégis érdemes úgy tenni, *mintha* lenne, mert ez egyszerűbbé teszi a tárgyak mozgásának leírását. Akár a súlypont, a mentális reprezentáció is pusztán *hasznos elméleti konstrukció*, semmi több.<sup>108</sup>

Az instrumentalizmus meglehetősen cinikus nézet, hiszen ha komolyan vesszük, nem más, mint vérbeli redukcionizmus, megfejelve némi tiszteletkörökkel a mentális terminusok gyakorlati alkalmazhatóságáról. A két elkerülendő véglet közül tehát a redukcionizmushoz került közelebb. Sőt, sok, magát instrumentalistának valló szerző, például Dennett másként redukcionista néven nevezi magát, de egyúttal hangsúlyozza, hogy redukcionizmusa nem jelent eliminativizmust, tehát a mentális terminusok teljes kiküszöbölését az elme leírásából.<sup>109</sup>

Sokaknak azonban túl redukcionista tünik az instrumentalisták álláspontja. Az instrumentalizmus ugyanis tagadja a redukálhatatlan mentális tulajdonságok létezését. A dualizmus tagadásából azonban ez még nem következik. Létezhet egy olyan materialista elmemodell, amelyben mégis szerepelnek redukálhatatlan mentális állapotok. Ez a funkcionalisták célkitűzése is, akiknek szintén túl radikális az instrumentalizmus, és nem redukálható, de ugyanakkor nem is dualista mentális állapotokkal akarják modellezni az elmét.

Ezen a ponton érdemes visszatérni Jaegwon Kim téziséhez, amely szerint az erős és a globális szupervencia ekvivalens egymással. Ezt a tézist implicit módon az instrumentalisták is elfogadják, hiszen az erős szupervencia tagadását, ami csak a redukálhatatlan mentális tulajdonságok létezését jelentené, azonosnak veszik a globális szupervencia tagadásával, tehát a dualizmussal, és a dualizmus elutasítása miatt elvetik a vele ekvivalensnek hitt nézetet is, amely szerint léteznek redukálhatatlan mentális tulajdonságok. A kád vízzel tehát a gyereket is kiöntik.

Kim azonban téved, ezt formális logikai nyelven is ki lehet mutatni.<sup>110</sup> A globális és az erős szupervencia közötti különbség, amelyet Kim figyelmen kívül hagy, éppen az, hogy *mire* redukáljuk a mentális állapotokat. Az erős szupervencia esetében szigorúan az agy fizikai állapotaira, a globális esetében azonban az egész fizikai világra. Ha tehát az elme nemcsak az agytól, hanem egy, az agyon kívüli tárgy fizikai tulajdonságától is függ, akkor az elme nem redukálható az agyra, mivel nem szuperveniál erősen. A globális szupervencia viszont teljesül, mivel az elme az egész fizikai világra nagyon is redukálható. A materializmus is megmarad, és a redukcionizmus sem teljesül: ideális kompromisszum.

Le is csaptak az externalisták erre a lehetőségre.<sup>111</sup> A már emlegetett H<sub>2</sub>O–XYZ példa például itt is működik. A „vízre gondolok” mentális állapot például nem szuperveniál erősen az agyamon, mivel létezik egy másik agy – iker-földi hasonmásom agya –, amely pontosan ugyanolyan neurális állapotokkal rendelkezik, mint én, a „vízre gondolok” mentális állapota mégis más lesz, hiszen ő az XYZ-re gondol, én viszont a H<sub>2</sub>O-ra. A globális szupervencia viszont teljesül, hiszen az egész fizikai világ már meghatározza a mentális állapotomat, tehát azt, hogy milyen atomi összetételű vízre is gondolok. Voltaképpen itt a mentális állapot nem redukálható az agyra,

<sup>105</sup> Cf. Nánay 1998a, 1998b.

<sup>106</sup> Cf. pl. Dennett 1996b, 1996c, 1998b, Dennett instrumentalizmusáról pl. Dahlbom 1993, Pléh 1998c.

<sup>107</sup> Az instrumentalizmus címke itt egyértelműen a korai Dennett instrumentalizmusára vonatkozik, figyelmen kívül hagyva az instrumentalizmus tudományfilozófiai irányzatán belüli megosztottságokat. További leegyszerűsítést jelent, hogy Dennett az utóbbi néhány évben elég messzire jutott eredeti instrumentalizmusától. Erről azonban részletesebben lesz szó az utolsó fejezetben.

<sup>108</sup> Dennett súlypont példája Fregétől ered, de szintén a súlypont analógiáját használja az Én leírására Ernst Mach is. Frege 1953, Mach 1927, cf. még a kérdéskörrel Pléh 1998c.

<sup>109</sup> Cf. Dennett 1998a, különösen p. 423.

<sup>110</sup> Nánay 1998a

<sup>111</sup> Petrie 1987, Paull-Sider 1992, Hellman 1985.



csak az agyra plusz a víz kémiai felépítésére. Ez a különbség azonban épp elég ahhoz, hogy a nem-reduktív materializmus lehetséges legyen.

A funkcionalizmus tehát mégis képes volt megvédeni a nem-reduktív materializmus álláspontját. Vagy mégsem? Ha visszaemlékszünk, a funkcionalizmus egyik legjelentősebb bírálatát épp az externalizmus adta. A funkcionalizmus tehát csapdahelyzetbe került. Ha el akarja kerülni mind a dualizmust, mind a redukcionizmust, akkor harmadik útként az externalizmust kell választania, ami szintén ellentétes a nézeteivel.

Az externalizmust illetően azonban megfogalmazható még egy átfogóbb elméleti ellentetés is. Ha ugyanis komolyan vesszük, hogy az iker-Földön élő hasonmásom agya pontosan, molekuláról molekulára azonos az én agyammal, és mégis más a mentális állapotunk, akkor egy olyan állításhoz jutunk, hogy egy külső dolog, például a víz kémiai képlete úgy képes hatással lenni az elmére, hogy az agyra nincs hatással. Ez pedig igen erős állítás, és közel viszi az externalizmust a dualizmushoz, amely az agytól elkülönült, nem anyagi természetű gondolkodó lélekről beszél.

Egy további megjegyzés is adódik azonban. Az externalista megoldás ugyanis nem felel meg az elme redukálhatósági kritériumának. Igaz ugyan, hogy kimutatja, az elme nem redukálható az agyra, de azt nem bizonyítja, hogy fizikai tulajdonságok semmilyen véges halmazára nem redukálható. Esetünkben például biztos, hogy redukálható az agy állapotára plusz a víz kémiai képletére, amely még mindig véges halmazt alkot, ráadásul, ha már az agy állapotát meghatároztuk, ehhez képest semmiféle víz képletét is megmondani, a redukció tehát az  $H_2O$ -XYZ különbség hozzávételétől nem lesz lényegesen nehezebb.

Ahhoz, hogy valóban az elme redukálhatatlansága melletti érvet szolgáltasson, az externalista modellnek meg kellene tudnia mutatni, hogy nem véges az agyon kívüli olyan fizikai tulajdonságok száma, amelyek hatnak az elmére. Ha ezt nem teszi, akkor nem az elme redukálhatatlanságát bizonyította, csak azt a lokális érdekességű tételt, hogy speciel az agyra nem redukálható. Ha azonban az externalizmus megpróbálja bizonyítani, hogy az elmére ható tényezők halmaza nem véges, akkor igencsak megkérdőjelezhető lesz az az állítása, hogy ez a sok-sok olyan tulajdonság, amely az elmére mind hatással van, az agyra viszont semmilyen hatással nem bír. Mi tehát ezek szerint mindebből semmit nem érzékelnénk, mint ahogy a  $H_2O$ -XYZ különbségből sem. A nem véges számú, csak az elmére ható, az agy receptorait rejtélyes módon kikerülő hatás miatt azonban megintcsak a dualizmus felé fog eltolódnia az externalista elmélet.

Az instrumentalizmus tehát a redukcionizmus felé tolódik el, míg az externalizmus inkább a dualizmus felé. A valódi középutat egyik elmélet sem volt képes megvalósítani.

Úgy tűnik tehát, hogy ha el akarjuk kerülni a dualizmust, akkor el kell fogadni a redukcionizmust, vagy legalábbis annak valamelyik alesetét. Láttuk azonban, hogy a redukcionizmus egyre gyengébb tételeket jelent. Az áthidaló törvényektől az erős szupervenciától az egy véges halmazra való redukcióig a redukció fogalma egyre gyengébb és gyengébb kritériumokat alkalmaz. Ami maradt a redukcionizmusból ezen a ponton, az nem több, mint hogy az elme állapota minden időpillanatban redukálható az agy állapotára. Ennek a tézisnek pedig semmilyen gyakorlati relevanciája nincs, hiszen ha a jelen pillanatban mentális állapotomat idegsejtjeim állapotára redukáltam, akkor ebből még semmit nem fogok megtudni arról, hogy a következő pillanatban milyen mentális, illetve neurális állapotba fogok kerülni.

A mentális állapotok redukálhatósága továbbá egyáltalán nem jelent determinizmust, és nem jelenti azt sem, hogy elménk mechanikusan működne. A redukciót gyakran szembeállítják az új mentális entitások megjelenésének lehetőségével.<sup>112</sup> Ha azonban elfogadjuk a fenti gondolatmenetet, akkor a redukció egyáltalán nem zárja ki az új reprezentációk megjelenését.

Bár igaz, hogy az elme minden pillanatban redukálható, de ez a redukció nem teszi lehetővé, hogy a leírjuk a neurális állapotok ismétlődéseit, invarianciáit, márpedig az elme leírása éppen az ilyen invarianciák leírásánál kezdődik. A mentális állapot tehát minden időpillanatban azonos a neurális állapottal, de az egyes mentális állapotok *típusai* nem hozhatók átfedésbe a neurális állapotok *típusaival*, márpedig az elme leírásához a mentális jelenségek típusairól kell beszélni. A mentális állapotok tehát redukálhatók, de a mentális állapotok típusai nem. Ha tehát elfogadjuk a mentális állapotok redukálhatóságát, akkor nem muszáj eldobnunk és neurális leírásra cserélnünk az elme leírását.

### 3. Mentális okság

A redukció problémájához szorosan kapcsolódik a mentális okság kérdése.<sup>113</sup> Mi az, ami azt okozza, hogy kinyújtom a kezem a dugóhúzóért? Egy idegsejt tüzelése vagy egy mentális reprezentáció? Mi a cselekedeteink kiváltója, a gondolat vagy neurális állapotunk? Mivel a mentális reprezentáció mindig azonos egy neurális struktúrával, ezért a kérdés konkrétan az, hogy a mentális reprezentáció ok-okozati hatása redukálható-e az ideg-

<sup>112</sup> Az egyik legfontosabb redukcióval foglalkozó szöveggyűjtemény már címében is erre a szembeállításra utal: *Emergence or Reduction*, tehát új tulajdonságok megjelenése vagy redukció. Beckermann–Flohr–Kim 1992.

<sup>113</sup> Legjobb összefoglalás: Heil–Mele 1993. Cf. még Pereboom–Kornblith 1991, Marras 1993.

sejtek ok-okozati hatására. Ha szomjas vagyok, és ezért megkeresem a dugóhúzó, akkor a „szomjas vagyok” mentális állapot okozza azt, hogy felállok, vagy az e mentális állapotot megvalósító idegsejtek?

Eddig a redukcióproblémának csak egyik aspektusát láttuk: magának az elmének a redukcióját. Ez azonban a redukció problémájának csak egyik esete. Ettől független kérdés, hogy a mentális állapotok kauzális hatása redukálható-e az idegsejtek kauzális hatására. Most ezt, a redukció előző alestétől különböző kérdést kell megvizsgálnunk: az elme ok-okozati hatásának redukcióját. Lehetséges ugyanis, hogy a mentális reprezentációk – mint az előző alfejezetben láttuk – redukálhatók, de a mentális egységek mégis ok-okozati hatással bírnak. Eddig azt vizsgáltuk, milyenek az elme építőkövei, most az lesz a kérdés, mit csinálnak. A redukcionista válaszra ebben a kérdésben az, hogy cselekedeteink kiváltó okai nem a mentális reprezentációk, hanem az idegsejtek aktívációs állapota. A mentális reprezentációk tehát – szép kifejezéssel élve – kauzálisan impotensek, más néven epifenomenálisak: nem kiváltó okai semminek, minden az idegsejtek műve.<sup>114</sup>

Ezzel szemben egy nem-redukcionista elméletnek a mentális okság kérdésében azt kell megmutatnia, hogy mentális reprezentációink és mentális állapotaink kauzálisan nem impotensek.<sup>115</sup> Nem lehet tehát száműzni őket az elme leírásából. A vita körül forog, hogy az elmeműködés leírásánál az idegsejtek vagy a reprezentációk működtetik a rendszert. A nem-reduktív nézet védelmezői igen nehéz helyzetben vannak, mivel azt kell megmutatniuk, hogy egy reprezentáció oksági hatása nem redukálható a reprezentációt felépítő idegsejtek oksági erejére, ez pedig azt a kérdést veti fel, hogy ha nem az idegsejtek tüzelése adja a reprezentáció hatóerejét, akkor ugyan mi. Itt tehát a nem-reduktív nézet megint közel kerül ahhoz, hogy kénytelen legyen valamilyen nem anyagi szubsztanciát (például lelket) bevonni a magyarázatába, ez viszont nagymértékben veszélyezteti a magyarázat tudományosságát.

A frontvonalak tehát, úgy tűnik, befagytak: a redukcionista azt állítja, hogy csak az idegsejtek rendelkeznek oksági hatással, minden más hatás lehetetlen. Aki mást mond, dualista. A nem-redukcionista pedig azt állítja, hogy a mentális reprezentációknak igenis van saját oksági hatóereje, de miközben ezt a nézetet próbálják védeni a redukcionizmussal szemben, gyakran valóban a dualizmus felé mozdulnak el. A köztes megoldást képviselő nem-reduktív materializmusnak most sem maradt sok hely.

Ebben az állóháborús helyzetben Donald Davidson egy olyan premisszára mutatott rá, amelyet mindkét szembenálló fél feltesz érvelése részeként.<sup>116</sup> Ha megkérdőjelezzük e premisszát, akkor egyrészt alaptalanná válik a szembenállás, másrészt lehetővé válik egy nem-reduktív materialista álláspont.

Davidson szerint mindkét tábor oksági felfogása abból indul ki, hogy egy dolog *bizonyos tulajdonságainak köszönhetően* képes valamit okozni. Egy *x* tárgy tehát azért képes ok-okozati hatást létrehozni, mert rendelkezik *F* tulajdonsággal. A koponyánkban levő pépes anyag azért képes kiváltani azt, hogy kinyissam a borosüveget, mert ilyen és ilyen tulajdonságokkal rendelkezik: mert adott mentális állapotban van, vagy mert adott neurális állapotban van.

Davidson radikálisan tagadja a kauzalitásnak ezt a felfogását. Szerinte nem tulajdonságok rendelkeznek oksági erővel, hanem maguk az események. A tulajdonságok a mi leírásaink az eseményekről. A kauzalitás viszont az események dolga, és az események így is, úgy is kiváltanak oksági hatást, függetlenül attól, hogy én milyen tulajdonságok segítségével írom le őket. Egy *x* esemény tehát nem azért hoz létre ok-okozati hatást, mert részesezik egy tulajdonságban, hanem úgymond alanyi jogon, függetlenül attól, hogy milyen tulajdonsággal ruházom fel én, mint külső szemlélő. A koponyánkban rejlő kocsonyás anyag pedig adott esetben kiváltja azt, hogy kinyitom a borosüveget, függetlenül attól, hogy mentális vagy neurális tulajdonságokkal rendelkezik-e, hiszen a tulajdonságok nem vesznek részt a oksági hatásban, voltaképpen nem többek, mint a mi utólagos hozzátételeink. Visszatérő, mindkét tábor által használt példa ezen álláspont elemzésekor a hangtompító pisztoly analógiája.<sup>117</sup> Ahogy a lövés szempontjából is mellékes, hogy a pisztoly, amiből kilőtték a golyót, hangtompító volt-e, úgy a mentális okság szempontjából is mellékes, hogy a kiváltó gondolatot reprezentációként vagy idegsejtmintázatként írom-e le.

A redukcionista mellett érvelnek, hogy a mentális okság jelenségének kiváltója az elme mint idegsejtek tömege. A nem-redukcionista ezzel szemben azt állítja, hogy ez a kiváltó erő az elme mint reprezentációk csoportja. Mindkét álláspont elfogadja tehát, hogy a mentális okság jelenségének kiváltója az elme mint valami, a vitatott kérdés csak annyi, hogy mint micsoda. Davidson ezzel szemben mellett érvel, hogy az elme nem *mint valami* okozza cselekedeteinket, hanem magától, függetlenül attól, hogy miként írom le, idegsejtek tömegének vagy reprezentációk csoportjának.

Davidson megoldásának van némi pragmatikus bája: az események úgy vannak a világban, ahogy vannak, mi pedig megpróbáljuk őket leírni valahogy, több-kevesebb sikerrel, mindig azt a leírást választva, amely éppen a legjobban használható. Az egyetlen igazi és valódi dolog a tárgyak közötti ok-okozati viszony. Hogy ezt milyen tulajdonságok kísérik, hangtompítók avagy mentális reprezentációk, az másodlagos kérdés.

<sup>114</sup> Cf. pl. Kim 1984b, 1990, Bickle 1998.

<sup>115</sup> Cf. pl. Fodor 1974, Cummins 1983, de ide tartozik a funkcionalizmus összes képviselője.

<sup>116</sup> Davidson 1993. Cf. még LePore–Loewer 1987.

<sup>117</sup> A hangtompító analógiájáról cf. Sosa 1984, 1993, Davidson 1993. p. 17.

Ezt a megoldásjavaslatot azonban többen bírálják, furcsa módon mind a redukcionista, mind a nem-redukcionista oldalról.<sup>118</sup> Davidson ugyanis egy olyan metafizikai képben gondolkodik, amelyben a tulajdonságok nem reálisak, csak az események léteznek valójában, a tulajdonságok a mi leírásaink az eseményekről. Ezzel a metafizikai képpel azonban nem mindenki ért egyet. Anélkül, hogy részletesen elemeznénk, hogy milyen érvek és ellenérvek lehetségesek a tulajdonságrealizmus kérdésében, és a Davidson-féle álláspont mennyire tartható, egy másik, a tulajdonságok nem reális volta mellett nem elkötelezett megoldásjavaslatot szeretnék proponálni. Talán nem meglepő módon az evolúciós elmélet fogja adni e megoldásjavaslat keretét.

John Mackie volt az első, aki a mentális okság problémájában evolúciós jellegű érveket hozott fel.<sup>119</sup> Mackie a nem-reduktív materializmus védelmében azzal érvelt, hogy a mentális reprezentációk nem lehetnek kauzálisan impotensek, mivel az evolúciós mechanizmusok kiválasztották őket. Mackie érvelése szerint mivel léteznek mentális reprezentációk, ezért ezek adaptívak – vagy adaptívak voltak –, és csak akkor lehetnek adaptívak, ha valós oksági hatást fejtenek ki. Következésképpen, ha elfogadjuk, hogy az elme az evolúció terméke, akkor a mentális reprezentációk nem lehetnek kauzálisan impotensek.

Mackie érvelésének azonban van egy gyenge pontja. Az érv első lépése azt mondja ki, hogy ha egy mentális képesség vagy reprezentáció jelen van az elmében, akkor ebből az következik, hogy az adaptív vagy legalábbis valaha az volt. Ez az álláspont azonban igencsak vitatható, hiszen – mint ezt a II. fejezetben részletesen láttuk – egy mentális képesség nem feltétlenül kell, hogy adaptív legyen. Lehet, hogy más adaptív folyamatok mellékterméke, amely önmagában nem adaptív. Elképzelhető tehát, hogy a mentális reprezentációk sem adaptívak, hanem más folyamatok exaptációi, nem-adaptív melléktermékei.

A redukcionista pedig, kihasználva ezt a kikaput, érvelhet úgy, hogy az idegsejtek aktivitási mintázata az, amely valóban adaptív, az evolúciós mechanizmusok ezt választották ki, a mentális reprezentációk jelenléte pedig csak ezeknek az adaptív folyamatoknak a mellékterméke. Mackie érvelése tehát ezen a ponton csődöt mond: hiába igaz, hogy ami adaptív, az kauzálisan aktív, ha a mentális reprezentációról nem tudjuk belátni, hogy adaptív. Mackie maga is megemlíti ezt a lehetséges ellenérvet, de nem ad kielégítő választ rá.<sup>120</sup>

Ha viszont elfogadjuk a neurális evolúció második fejezetben kifejtett elmén belüli modelljét, és feltesszük, hogy az elme egyre komplexebb mentális reprezentációkat hoz létre, amelyek magukban foglalják azokat az egyszerűbb mentális entitásokat, amelyekből evolválódtak, akkor egészen más megvilágításba kerül a mentális okság problémája. Ha igaz, hogy a mentális reprezentációk az idegsejtekből evolválódtak a környezeti ingerek hatására, akkor nem lehetnek kauzálisan impotensek, hiszen nagy szerepet játszanak és játszottak az organizmus túlélésében.

Hogy megint a kreatív béka példáját vegyük, a béka, amelynek a környezetében csak legyek élnek, igen hatékonyan tud táplálkozni annak a kissé primitív, de hatékony módszernek a segítségével, hogy ha lát egy fekete pöttyöt, akkor azonnal kinyújtja a nyelvét, és elkapja azt a dolgot, ami a szeme előtt van. Mivel a környezetében minden, ami fekete pöttyként jelenik meg a retináján, légy lesz, ezért a béka problémamentesen tud táplálkozni e módszer segítségével.

Ha azonban megjelennek a béka környezetében a darazsak, akik nemhogy nem ehetők a béka számára, de csípésük jelentősen rontja a túlélési esélyeket, akkor már rögtön nem kifizetődő az előbbi stratégiával táplálkozni, hiszen ugyanolyan eséllyel kapna be legyet a béka, mint darazsat. Ha ekkor egy kreatív béka elméjében összekapcsolódik a „sárga csíkok” és a „fekete pötty” reprezentációja, és csak azokat a fekete pöttyöket kapja be, amelyek nem csíkosak, akkor életben marad. Mivel a többi béka kvázi halálra van ítélve az elavult táplálkozási módszer miatt, ezért mind a béka, a „fekete pötty sárga csíkok nélkül” reprezentáció nagy valószínűséggel meg fog maradni, és el fog terjedni.

Azt láttuk, hogy itt egy új mentális reprezentáció megjelenése kizárólag a környezeti hatásoknak köszönhető. Mielőtt a darazsak feltűntek volna szegény béka körül, sem a béka, de még a legokosabb biológus sem gondolta volna, hogy nemsokára a „fekete pötty sárga csíkok nélkül” mentális reprezentáció fog megjelenni a béka elméjében. A darazsak megjelenése viszont egy új, addig nem létezett mentális állapotot hívott életre. Ennek a mentális állapotnak a megjelenését magyarázza a neurális evolúció új elmélete.

A „fekete pötty sárga csíkok nélkül” reprezentáció megjelenését a környezeti hatások tették szükségessé: e reprezentáció nélkül a béka nem igazán lenne életképes a darazsak inváziója miatt. E reprezentáció oksági hatóereje nem redukálható az őt felépítő kisebb egységekre vagy idegsejtekére, hiszen *éppen ezekből alakult ki* a nagy evolúciós átmenetek segítségével. Nem redukálható a „fekete pötty sárga csíkok nélkül” reprezentációjának oksági hatása a „fekete pötty” és a „sárga csíkok” külön vett oksági hatására, hiszen éppen azáltal fog túlélni a kreatív béka, mert együtt kezeli, egy egységbe foglalja e két reprezentációt.

Mackie érvelésének második lépését tehát érdemes elfogadni: ha egy mentális reprezentáció az evolúció terméke, akkor nem lehet kauzálisan impotens, hiszen ekkor e reprezentáció által kiváltott ok-okozati hatás is hozzájárult az élőlény túléléséhez. Mackie érvének első lépését azonban módosítani kell. Hogy belássa, egy reprezentáció az evolúció terméke, Mackie felteszi, hogy a természetes szelekció alakította ki. Az élőlény ősei

<sup>118</sup> Cf. Kim 1993, McLaughlin 1993.

<sup>119</sup> Mackie 1985.

<sup>120</sup> Mackie 1985 p. 142.

számára hasznos volt ez a reprezentáció, és ezért rögzült az organizmus génjeiben. Ez azonban nem bizonyítható minden reprezentáció esetében. A gyökvonás jelének reprezentációja például nem valószínű, hogy bármikor is jelentősen hozzájárult volna az emberi faj túléléséhez. Mackie nem veszi figyelembe az elmén belüli evolúció jelenségét.

Ha a gyökjel reprezentációja nem is adaptív az emberi faj számára, minden olyan ember fejében, ahol megtalálható ez a képesség, az azt létrehozó folyamatok evolúciós jellegűek voltak. Az, hogy a gyökjel reprezentációja megmaradt és megmarad az elmében, az e képességért felelős idegsejtek visszafogásán alapul. Ezek az idegsejtek sok más okos dolgot művelhetnének, de mégis a gyökvonás jelének reprezentációját fogják továbbra is megvalósítani.

A különbség Mackie elmélete és az itt vázolt modell között az, hogy míg Mackie a természetes szelekció segítségével azt akarja igazolni, hogy minden reprezentációnk adaptív a faj számára, addig a neurális evolúció új elmélete olyan mechanizmust biztosít, amely képes megmagyarázni, hogy minden reprezentáció evolúciós eredetű, de nem feltétlenül adaptív. A gyökjel reprezentációja nem adaptív az emberi faj számára, de – paradox módon fogalmazva – adaptív magának a gyökjel reprezentációja számára.

Itt azt használtuk ki, hogy a nagy átmenetek evolúciója, az új szerveződési szintek megjelenése mindig adaptív az új szerveződési szint számára, különben nem jelenne meg. A többsejtűek evolúciója az egysejtűekből ugyanilyen elven történt. Az újonnan létrejött organizmus túlélésének és replikációjának elősegítése az elsődleges célja az őt felépítő sejteknek, és nem a saját túlélésük. Hasonlóképpen a gyökjel-reprezentációt felépítő idegsejtek sem azt csinálnak, amit akarnak. Elsődleges rendeltetésük a gyökjel reprezentációjának túlélésének biztosítása. Úgy táncolnak, ahogy a reprezentáció fütyül. Ezen idegsejtek működését nem lehet megmagyarázni tehát anélkül, hogy a reprezentáció oksági hatását ne vennénk számításba.

Ez azonban egy új, az előzőnél is erősebb érvet szolgáltat a mentális okság redukciója ellen. A redukció elve feltételezi, hogy az az entitás, amire redukálunk, független attól, amit redukálunk. A mentális okság redukciójának esetében azonban, ha elfogadjuk a neurális evolúció új elméletét, ez a feltétel nem áll. Az idegsejtek kauzális hatása ugyanis nem határozható meg a mentális reprezentáció kauzális hatásától függetlenül, sőt, épp a mentális reprezentáció, az idegsejtek szintű szervezettsége szabja meg, hogy milyen oksági hatásokat fog kifejteni az egyes idegsejt. Itt tehát az az eset áll fenn, hogy  $X$  entitást  $Y$  entításra akarnánk redukálni, de ugyanakkor  $Y$  entitás  $X$  entitástól függ, és ez eleve kizárja a redukció lehetőségét.

Az evolúciós fejezetben a reprezentáció és az azt megvalósító idegsejtek kapcsolatát a díszzászlóalj és a katona kapcsolatához hasonlítottam. Ahogy a katona sem mehet el strandolni, mert azt kell tennie, amit a nagyobb egység – melynek része – tesz, úgy a reprezentációt felépítő idegsejt működése is alá van vetve a reprezentáció működésének. A katona is feláldozza saját érdekeit a zászlóaljért, az idegsejt is a reprezentációért. Folytatva az analógiát, ahogy a közlegény díszlépése sem magyarázható magában, csak kizárólag a zászlóalj mozgásából származtatva, úgy az idegsejt működése is csak a reprezentáció működésének és oksági hatásainak feltételezésével adható meg.

Ha tehát helyes a neurális evolúció új elmélete, akkor a mentális reprezentációk nem lehetnek kauzálisan impotensek, sőt, ellenkezőleg, az idegsejtek oksági hatása is csak a reprezentációk oksági hatásával magyarázható.

## V. Intencionalitás

Mindeddig az elmét felépítő egységnek, a mentális reprezentációknak a belső felépítéséről volt szó: redukálhatók-e, milyen viszonyban állnak az őket megvalósító idegsejtekkel, miként lehet jellemezni oksági hatásukat. A jelen fejezet arról szól, hogy mik is ezek a reprezentációk: mi a különbség aközött, amikor a kutyára, illetve amikor a macskára gondolok. Eddig azt elemeztük, miből áll a reprezentáció, most itt az idő, hogy megvizsgáljuk: mit is reprezentál.

Az intencionalitás kifejezés éppen a reprezentációknak ezt a sajátosságát ragadja meg.<sup>121</sup> A reprezentáció mindig valamire vonatkozik, valamit jelent, valamit reprezentál. Intencionalitással rendelkezik tehát, mutat, utal valamire. Az intenció szó itt használt jelentését meg kell különböztetni a mindennapi értelmétől. Az intencionalitás a filozófiai diskurzusban nem szándékosságot jelent, hanem valamire való irányulást.<sup>122</sup> Az intencionalitás tehát az összekötő kapocs a reprezentáció és a reprezentált dolog között.

Az intencionalitás fogalmát Franz Brentano vezette be a filozófiai irodalomba.<sup>123</sup> Brentano azt próbálta megmagyarázni, hogy mi a különbség a gondolatok világa és a fizikai tárgyak világa között. Úgy érvel, hogy míg a gondolatok mindig valamire vonatkoznak, irányulnak, addig a tárgyak nem irányulnak semmire. Amikor a dugóhúzóra gondolok, akkor e gondolatom a dugóhúzóra mint tárgyra vonatkozik. A dugóhúzó maga viszont nem vonatkozik semmire. A gondolatok tehát rendelkeznek intencionalitással, a tárgyak viszont nem: ez a nagy különbség.

Brentano megoldásjavaslata persze több ponton is támadható: léteznek az elme világának is olyan elemei, amelyek nem vonatkoznak semmire, a szorongás fogalma például éppen abban különbözik a félelemtől, hogy nincs meghatározott tárgya. Másfelől a tárgyak között is találunk olyanokat, amelyek rendelkeznek intencionalitással, egy papírra felírt „dugóhúzó” szó például a dugóhúzóra utal, akárcsak dugóhúzó-reprezentációm. Itt azonban hatásos ellenérv, hogy a papírra felírt szó csak akkor és annyiban rendelkezik intencionalitással, ha elolvassa egy olyan lény, aki értelmezni tudja, és képes felismerni, hogy mire utal e szó. Az írás intencionalitása tehát nem eredendő intencionalitás, a szöveget író vagy elolvasó ember intencionalitásából származik, azon élösködik.

Függetlenül attól, hogy Brentano elméletének vannak hiányosságai, nem kerülhető ki az általa felvetett probléma: a reprezentációk kitüntetett helyzetben vannak minden más dologhoz képest: rendelkeznek intencionalitással. Ha viszont ez igaz – márpedig úgy tűnik, legalábbis legtöbb reprezentációnkra igaz –, akkor valamilyen meggyőző magyarázatot kell szolgáltatni e különbségre.

Brentano szerint a gondolatok, a mentális tartalmak világa alapjaiban különbözik a többi tárgy világától. Az elme csak szemantikai terminusokkal, csak az egyes mentális reprezentációk tartalmára, tárgyára, jelentésére való hivatkozással írható le. Ezzel szemben a többi tárgy leírásához nincs szükségünk ilyen kifejezésekre.

Ahhoz tehát, hogy az elmét ne valami csodálatos, a tudomány számára felfoghatatlan dolognak tekintsük, meg kell tudnunk magyarázni azt a jelenséget, hogy a mentális reprezentációknak intencionalitásuk van. Az intencionalitást, a mentális tartalmak világát tehát naturalizálnunk kell, azaz meg kell tudnunk magyarázni egy olyan nyelven is, amelyben nem használunk intencionális, azaz a mentális reprezentációk tartalmára vonatkozó, más néven szemantikai terminusokat.<sup>124</sup> A mentális reprezentáció és annak tárgya közötti kapcsolat nem-intencionális, nem-szemantikai terminusokban történő magyarázatát nevezzük az intencionalitás naturalizálásának.

### 1. A reprezentáció tárgya

A kérdés tehát az, hogy milyen kapcsolat van a reprezentáció és a reprezentált dolog, a reprezentáció tárgya között, és miként lehet megmagyarázni ezt a kapcsolatot, illetve kialakulását. Mielőtt azonban ezzel a kérdéssel foglalkoznánk, érdemes áttekinteni az intencionalitás fogalma körüli legfontosabb nézetkülönbségeket.

Az egyik legfontosabb, az angolszász nyelvfilozófián belül nagy múlttal rendelkező vita az intencionalitással kapcsolatban azt feszegette, hogy mely dolgok rendelkeznek eredendő intencionalitással, és melyek intencionalitása származtatott. Konkrétabban: a nyelv intencionalitásán alapul a gondolatok intencionalitása, vagy fordítva: a gondolatok intencionalitása teszi lehetővé a nyelv intencionalitását?<sup>125</sup> Anélkül, hogy részletebben elemeznénk a gondolkodás és a nyelv kapcsolatát, amelyről a következő fejezetben lesz szó, itt minden-

<sup>121</sup> Az intencionalitás kérdésköréhez cf. Searle 1983. Chisholm 1967, Pléh 1998c.

<sup>122</sup> Az intencionalitás fogalmának hétköznapi és filozófiai jelentése közötti kapcsolathoz cf. Dennett 1998b. pp. 7–8.

<sup>123</sup> Brentano 1983.

<sup>124</sup> Cf. Quine 1998.

<sup>125</sup> E vitáról lásd Sellars 1963, aki a nyelv elsőbbsége mellett érvel, Chisholm 1955, 1967, aki a nyelv származtatott intencionalitását hangsúlyozza. Köztes álláspontot javasol Davidson 1984, aki szerint sem a nyelv, sem a gondolatok intencionalitása nem vezethető vissza a másikra.

esetre annyit érdemes leszögezni, hogy ha nem akarjuk kirekeszteni az állati elmét a kutatás köréből, akkor semmiképpen nem lehet az a kiindulópontunk, hogy a nyelv az, amely eredendő intencionalitással rendelkezik, minden más, így a gondolatok, képzetek, vágyak is csak ebből az elsődleges intencionalitásból erednek. Ha vannak a rókának mentális reprezentációi – például a nyúlról –, akkor ezek a reprezentációk vonatkoznak valamire – esetünkben a nyúlra, amelyet meg lehet enni –, következésképpen rendelkeznek intencionalitással, noha a róka nem beszél semmilyen nyelvet.

Hozzá kell azonban tenni, hogy ez az érv, bármennyire hatásos is, nem teszi zárójelbe a gondolkodás és a nyelv intencionalitásának kérdését.<sup>126</sup> A gondolkodás intencionalitását ugyanis általában éppen a gondolkodást kifejező szavaink nyelvtani szerkezete alapján szoktuk elemezni. „Arra gondolok, hogy...”, „tudom, hogy...”, „arra vágyom, hogy...”, „azt látom, hogy...”, mindegyik olyan kifejezés, amely gondolkozási folyamatot jelöl, rendelkezik egy tárgyi mellékmondatral, amelyet a reprezentáció tárgya tölt be. A gondolkodás intencionalitását tehát nehéz anélkül elemezni, hogy egyúttal a nyelvvel ne foglalkoznánk.

További problémát jelent, hogy vannak olyan reprezentációink, amelyek bár irányulnak valamilyen tárgyra, de ez a tárgy nem létezik: ilyen például az, amikor a boszorkányra vagy az üveghegyre gondolok. Az üveghegy-reprezentációmnak (vagy a 17-es számról alkotott reprezentációmnak) van intencionalitása, de amire irányul, az nem egy létező tárgy. Az üveghegy problémájának részletesebb elemzéséhez Frege distinkciójához érdemes visszanyúlni.

Frege híres 1892-es cikkében megkülönbözteti a jel jelentését [Sinn] és referenciáját [Bedeutung].<sup>127</sup> Jelentősen leegyszerűsítve Frege gondolatmenetét azt lehet mondani, hogy a referencia az a fizikai tárgy, amelyre a jel utal, a jelentés viszont magában foglalja a meghatározás módját is. Frege jelentésfogalmát anakronisztikus módon a mai mentális reprezentáció fogalommal lehetne azonosítani, feltéve, hogy a különböző emberek mentális reprezentációjában van valami közös, hiszen Frege a szubjektív, személyenként változó képzet fogalmát világosan elhatárolja a jelentés fogalmától. A mentális reprezentáció és annak tárgya közötti kapcsolat tehát a fregei jelentés és referencia közötti kapcsolatnak felel meg.<sup>128</sup>

Frege fogalmaival tehát úgy lehet megfogalmazni a problémát, hogy bizonyos szavaknak – így a boszorkánynak vagy az üveghegynek – bár létezik jelentése, de nincs referenciája. Frege azonban további zavaró elemeket fedezett fel a jelentés és a referencia közötti kapcsolatban. Gyakran előfordul ugyanis, hogy egy referenciának több jelentés felel meg. Másként fogalmazva több reprezentáció ugyanarra a fizikai tárgyra utal.

Oidipusz király elméjében például Iokaszté reprezentációja és az anyareprezentáció két külön reprezentáció volt, annak ellenére, hogy a valóságban ugyanarra a személyre vonatkozott. A gondot az okozta, hogy szegény Oidipusz nem volt tudatában, hogy a két reprezentáció ugyanarra a nőre vonatkozik. Nem tudta tehát, hogy a két eltérő jelentéshez egy és ugyanaz a referencia tartozik. Az intencionalitást vizsgálva tehát két furcsa jelenségre is magyarázatot kell találnunk. Egyrészt léteznek olyan reprezentációk, amelyek bár rendelkeznek intencionalitással, de tárgyuk valójában nem létezik, mint ahogy üveghegy sem létezik. Másrészt lehetséges, hogy két különböző reprezentáció ugyanarra a tárgyra irányuljon.

Az intencionalitás problémája elleni legradikálisabb, voltaképpen övön aluli támadást Daniel Dennett elmélete jelentette.<sup>129</sup> Dennett szerint az intencionalitás nem valódi tulajdonság, intencionalitást mi, értelmező emberek tulajdonítunk bizonyos dolgoknak, hogy könnyebben le tudjuk írni működésüket. Az elmében semmi nem rendelkezik eredendő intencionalitással. Ha viszont úgy értelmezzük a szomszédunk cselekedetét, hogy azt gondolja, hogy kivágtuk a cseresznyefáját, és ezért ő is ki akarja vágni az én diófámat, akkor esetleg hatékonyabban tudok védekezni a szomszéd ezen akciója ellen.

Dennett – mint a mesterséges intelligencia lehetséges voltának egyik lelelkesebb harcosa – nem engedheti meg, hogy valamiféle Brentano által sugallt alapvető és nem figyelmen kívül hagyható különbség legyen az elme és az összes többi tárgy között. Brentano nyomán ugyanis úgy érvelhetnénk – és érvelnek is – a mesterséges intelligencia ellenzői, hogy míg az elme intencionalitása igazi és eredeti, addig a számítógép mindig csak olyan feladatot képes elvégezni, amelyre az ember megtanítja.<sup>130</sup> Ha léteznek is reprezentációk egy számítógépprogramban, amelyek például egy szövegszerkesztő esetében a cikk karaktereinek számát reprezentálják, és még ha rendelkezik is emiatt egy program valamifajta intencionalitással, ez mindig csak azért lesz jelen a gépben, mert a programozó beletette, a számítógép intencionalitása tehát – ha van ilyen – mindig az ember eredendő intencionalitásán élőszködik.<sup>131</sup>

<sup>126</sup> Cf. Stalnaker 1999.

<sup>127</sup> Frege 1980, cf. még Russell 1956, 1971. A problémáról cf. Nánay 1997a.

<sup>128</sup> Frege persze mint radikális antipszichológista, ezzel az értelmezéssel valószínűleg nem értene egyet, a mai kognitív tudományi és elmefilozófiai viták áttekintésénél mégis hasznos, ha Frege distinkcióját közelítjük a mai fogalmakhoz.

<sup>129</sup> Dennett 1996a, 1996b, 1998b. Dennett intencionalitáselméletéről cf. Dahlbom 1993, Kampis 1999, Pléh 1998c.

<sup>130</sup> Cf. Searle 1992, 1996.

<sup>131</sup> Cf. pl. Stalnaker 1984, 1999.

Sokan próbálják visszautasítani ezt az érvet mondván, hogy a számítógép is rendelkezik eredendő, a programozó szándékától független intencionalitással.<sup>132</sup> Dennett az ellenkező, első pillantásra nehezebb utat választja. Nem a gépről mutatja meg, hogy van eredendő intencionalitása, hanem meglepő módon az emberről akarja belátni, hogy nem rendelkezik eredendő intencionalitással. Dennett szerint az elmében sincsenek olyan reprezentációk, amelyek úgymond maguktól, eredendő módon bírnának bármilyen irányultsággal. Az emberi elme ugyanúgy nélkülöz mindenféle csodás, nem fizikai elemet, mint a gép. Csak éppen ahogy a gép reprezentációit a programozó hozta létre, úgy az ember reprezentációit a vak evolúciós folyamatok.<sup>133</sup> Mint ezt a funkcionalizmusról szóló fejezetben már írtuk, a gép és az elme közötti különbség Dennett számára pusztán annyi, hogy míg a gépet az IBM cég gyártotta, addig az elmét a természetes szelekció.

Dennett azonban továbbmegy ennél az intencionalitás kérdésében. Azt is állítja, hogy a termosztát és az emberi elme között nincs különbség a tekintetben, hogy egyik sem rendelkezik eredendő intencionalitással. A különbség mindössze annyi, hogy míg az emberek akcióit jól ki tudom számítani, és meg tudom jóslani, ha intencionalitást tulajdonítok nekik, ha azzal számolok, hogy mire gondol, mit akar, addig a termosztát esetében ezzel nem érek el túl jó eredményt. Ha tudom, hogy a barátnóm tudja, hogy megcsalom, ha tehát intencionalitást tulajdonítok neki, akkor könnyebben ki tudom számítani a viselkedését hazatérve, mint ha nem tulajdonítanék intencionalitást neki. A sakkprogramnak is sokszor hasznos intencionalitást tulajdonítani: azért lépett a bástyával, mert azt gondolta, hogy leveszem a futóját, és akkor ő sakkot adhat. A joghurtos pohárnak viszont hiába tulajdonítok intencionalitást, ettől nem fogom tudni jobban megjósolni, miként fog viselkedni.

Az intencionalitás tehát értelmezés kérdése: senkiben és semmiben nincs meg magától, de néha hasznos lehet, ha bizonyos dolgokat vagy embereket úgy értelmezünk, *mintha* rendelkeznének intencionalitással, mert ekkor hatékonyabban meg tudom jóslani viselkedésüket. Az embernek, folytatja Dennett, evolúciósan rögzült tulajdonsága, hogy intencionalitást tulajdonít minden olyan mozgó dolognak, amely viselkedésének megjóslásakor ez hasznos lehet. Ez az intencionalitástulajdonítási képesség adaptív volt, hiszen a másik egyed manipulálását tette lehetővé, és ezzel nagyobb túlélési esélyt biztosított. Később, miután mások viselkedését úgy elemeztük, mintha intencionalitással bírnának, magunkat is ekképpen kezdjük tekinteni, s ezért érezzük azt, hogy saját gondolataink is intencionalitással bírnak.<sup>134</sup> Ez azonban illúzió, sem mi, sem a többi ember nem rendelkezik több intencionalitással, mint a termosztát, az intencionalitás csak mint mások értelmezésekor másoknak tulajdonított hasznos elméleti konstrukció létezik.

Dennett megoldásjavaslata meglehetősen radikális, de kétségkívül képes az intencionalitás jelenségét naturalista terminusokban elemezni, s ennyiben választ jelent Brentano kérdésére. Dennett kiindulópontja azonban az, hogy intencionalitás nem létezik, csak hasznos elméleti konstrukció mások viselkedésének elemzésekor. Ha ezt nem akarjuk elfogadni, és a reprezentációk intencionalitását mint valódi tulajdonságot akarjuk elemezni, akkor meg kell magyaráznunk a reprezentáció és a reprezentáció tárgya közötti kapcsolatot.

Három nagy típusba szokás sorolni a reprezentáció és a reprezentált dolog közötti kapcsolat magyarázatának megoldási javaslatait.<sup>135</sup> E háromféle reláció hozzávetőlegesen megfeleltethető Charles Sanders Peirce háromféle jeltípusának, az ikonikus, az indexikus, illetve a szimbolikus jelnek.<sup>136</sup> Az ikonikus jel esetében a jel és a jelölt dolog hasonlít egymásra, mint ahogy egy portré hasonlít arra a személyre, akit ábrázol. Az indexikus jel esetében a jel és a jelölt dolog közötti kapcsolatot az biztosítja, hogy a jelölt dolog okozza a jelet, így például a távoli füst azért utal egy égő házra, mert a tűz (a jelölt dolog) okozza a füstöt (a jelet). A szimbolikus jel esetében ez a kapcsolat a közmegegyezésen alapul, mint a matematikai szimbólumok vagy a nyelv esetében.

A mentális reprezentáció és a reprezentált dolog közötti kapcsolatra adott magyarázatok közül az első tehát a reprezentáció és annak tárgya közötti hasonlóságot tekinti az intencionalitás magyarázó elvének. A kutyáról alkotott reprezentációnk azért épp a kutyára vonatkozik, mert hasonlít hozzá.<sup>137</sup> Látható, hogy ez az elmélet a képeket tekinti a reprezentáció alapeseteinek. Itt két probléma adódik: az első az, hogy a képként felfogott reprezentáció modellje nem képes kezelni a Frege által felvetett kínos eseteket.

Egyrészt ekkor nem lehetnének reprezentációink nemlétező tárgyról, mint például az üveghegyről, hiszen nincs, amihez hasonlítson a reprezentáció. Másrészt ha két reprezentáció ugyanarra a tárgyra vonatkozik, akkor mindkettő hasonlít a tárgyra, következésképpen egymásra is. Ez viszont azt jelenti, hogy Oidipusz nem lehetné azt, hogy Iokaszté és az anyja két különböző személy, hiszen Iokaszté-representációja és anyarepresentációja hasonlítana egymásra. A másik probléma az, hogy az absztrakt fogalmak intencionalitását is meglehetősen nehéz

<sup>132</sup> Talán lehetségesek olyan (konnekcionista alapelveken működő) gépek, amelyek maguktól képesek új reprezentációkat kialakítani, tehát maguktól, a programozó szándékától és tevékenységétől függetlenül rendelkeznek intencionalitással. Cf. Fahlman–Lebiere 1990, Hinton et al. 1995. Ezek lehetőségét következetesen tagadja pl. Marcus 1998.

<sup>133</sup> Dennett 1998a, 1998b.

<sup>134</sup> Ernst Mach csaknem száz évvel Dennett előtt hasonlóan érvelt. Cf. Mach 1927, Mach és Dennett felfogásának kapcsolatáról cf. Pléh 1998c.

<sup>135</sup> Cf. Millikan 1993. pp. 3–12.

<sup>136</sup> Peirce 1975.

<sup>137</sup> A hasonlóság mint reprezentációs kritérium nézethez cf. Wittgenstein 1989, Shephard–Metzler 1971, a kérdés kognitív tudományi aspektusainak jó összefoglalása Block 1983.

kezelné a hasonlóságon alapuló modell segítségével, hiszen mire is hasonlítana például a felhajtóerő vagy a differenciálegyenlet reprezentációja?

A másik nagy megoldásjavaslat-típust a kauzális elméletek jelentik.<sup>138</sup> A kutyáról alkotott reprezentációm azért a kutyára vonatkozik, mert egy kutya okozta, hozta létre, egy kutyát meglátva keletkezett e reprezentáció.<sup>139</sup> Ez a magyarázat a legelterjedtebb az intencionalitás elmefilozófiai irodalmában, a fregei kérdések azonban itt is megoldatlanok maradnak. Miként magyarázza ez az elmélet a boszorkány reprezentáció intencionalitását, ha a boszorkány referenciája – mivel nem létezik – nem is volt képes ok-okozati hatással bírni ezen reprezentáció kialakulására. Vagy szegény Oidipusz Iokaszté-representációja miért nem egyezett meg az ő anyarepresentációjával, holott mindkettőt ugyanaz a fizikai entitás, Iokaszté kauzális hatása idézte elő?

A harmadik magyarázat a reprezentáció és annak tárgya közötti kapcsolatra a szó és a jelentés közötti kapcsolatból indul ki. A reprezentáció ezek szerint egyfajta „gondolkodás nyelve”, *lingua mentis* lenne, amelynek alapelemei – bár nem azonosak például a magyar nyelv szavaival –, mégis nyelvi jellegűek, más szóval kvázinyelvi.<sup>140</sup> A reprezentációk tehát hasonlóak a szavakhoz, az intencionalitásuk is hasonló a szó és a jelentés közötti kapcsolathoz.<sup>141</sup> Ez a megoldás mindenképpen képes kezelni például a nem létező referenciájú reprezentációk esetét. Igaz, hogy az üveghegy reprezentációja nem vonatkozik semmilyen valós tárgyra, de az üveg és a hegy reprezentációja külön-külön igen. Az üveghegy reprezentáció nem más, mint e két kisebb reprezentáció összetétele, kompozíciója. A kvázinyelviként felfogott reprezentációk esetében a kompozicionalitást könnyebb magyarázni, mint a kauzális elméleteknél.

A másik fregei problémát is jól tudja kezelni a kvázinyelvi reprezentációelmélet. Két reprezentáció – akár két szó – vonatkozhat ugyanarra a tárgyra annak ellenére, hogy ők maguk különbözőek. A kvázinyelvi reprezentációkról, illetve az általuk felvetett problémákról a következő fejezetben lesz szó. Előljáróban elég annyi, hogy bár az előző két felfogással ellentétben Frege problémáiban használható ez a megoldási javaslat, komoly gondok merülnek fel azt illetően, hogy miként alakult ki a reprezentáció és a reprezentált dolog közötti kapcsolat. Mivel a kvázinyelvi jellegű reprezentációk és tárgyak közötti kapcsolat önkényes (akárcsak a természetes nyelvek esetében), nehéz megmagyarázni, hogy miért éppen egy adott kvázinyelvi kifejezés fog a dugóhúzóra utalni, és nem egy másik. Ha el akarjuk kerülni, hogy minden reprezentációnkat velünk születettnek tételezzük, akkor meg kell magyarázni, hogy mi határozta meg a reprezentációk intencionalitását.

A három megoldásjavaslattal szemben felhozott ellenvetések persze inkább csak kijelölik a lehetséges támadási pontokat, sok érvet és ellenérvet lehetne még említeni. E problémák miatt mindenesetre sokan olyan megoldást keresnek az intencionalitás problémájára, amely ötvözi e három elmélet előnyeit, de ugyanakkor elkerüli hátrányait.<sup>142</sup> Ahelyett, hogy e kompromisszumos javaslatok bonyolult rendszereit sorra vennénk, inkább egy negyedik, az előző három – a hasonlóságon, a kauzális kapcsolaton, illetve a kvázinyelvi felépítésen alapuló – megoldási úttól különböző intencionalitáselméletéről kell szólni. Ez – talán nem meglepő módon – az evolúciót emeli be a reprezentációk intencionalitásának vizsgálatába.

Az alapelv világos és kézenfekvő: azért van kapcsolat a reprezentáció és a reprezentáció tárgya között, mert ha nem lett volna ilyen kapcsolat, akkor kihaltunk volna. Ha a tigris reprezentációja a gazellára utalt volna, akkor valószínűleg nem sok esélyünk lett volna védekezni a tigrisek támadása ellen. Azért van tehát intencionalitásuk reprezentációinknak, mert ez adaptív volt, segített a túlélésben.

Ez a magyarázat tehát adaptacionizmust előfeltételez. Nem minden reprezentációnk esetében bizonyítható azonban, hogy intencionalitása elősegíti vagy elősegítette valaha is túlélésünket. Az például, hogy cseresznyefa-representációnk a cseresznyefára utal, és nem másra, nem valószínű, hogy bárki túlélésében is szerepet játszott volna.<sup>143</sup> További gondot jelentenek a Frege által felvetett kérdések: a nemlétező dolgokra utaló reprezentációkról sem könnyű belátni, hogy adaptívak lettek volna, a két különböző, de ugyanarra a tárgyra vonatkozó reprezentáció pedig egyenesen negatív hatással is lehet a túlélési esélyekre, mint ezt Oidipusz király esete ékesen bizonyítja.

Az intencionalitás evolúciós felfogása persze ezen a ponton védekezhetne úgy, hogy nem minden reprezentáció-representált kapcsolat adaptív, hanem maga az a képességünk, hogy a reprezentáció és tárgya általában megfelel egymásnak. Természetesen sok, önmagában nem adaptív mellékterméke, exaptációja is van ennek a folyamatnak – ilyen lenne az üveghegy, a Iokaszté és a cseresznyefa intencionalitása –, de ettől még adaptív maga az a tény, hogy reprezentációink intencionalitással rendelkeznek, hogy tehát valamire utalnak. Oidipuszt valószínűleg kevésbé vigasztalna ez a konklúzió, mindenesetre az intencionalitás jelenségének négy különböző megoldásjavaslata közül mindegyik támadhatónak bizonyult, még az egyébként ígéretes evolúciós megközelítés is.

<sup>138</sup> Kauzális elméletekhez cf. Dretske 1981, Fodor 1987, 1990, Davidson 1991, illetve Locke 1979, Descartes 1994a mint két nagy előd.

<sup>139</sup> A kauzalitáselméletekről jó összefoglalást ad Forrai 1998.

<sup>140</sup> Cf. Sellars 1963, Pylyshyn 1984, Newell 1990. A formális nyelvek kérdésében a szellemi előd itt Leibniz és Raimundus Lullus. cf. Leibniz 1986, Lull 1993.

<sup>141</sup> E harmadik megoldási javaslat gyakran összekapcsolódik a funkcionalista ihletésű fogalmi szerep szemantikával. A kettő kapcsolatáról cf. Harman 1982, Forrai 1998. pp. 103–107.

<sup>142</sup> Pl. Block 1986, Fodor 1994.

<sup>143</sup> A gyilkos galóca reprezentációja esetében persze az adaptacionistáknak könnyebb dolguk van.



Éppen ezért érdemes az intencionalitás problémáját és az arra adott evolúciós választ egy másik oldalról, a téves reprezentációk felől is megvizsgálni. A téves reprezentáció kérdésköre nem egy kis érdeklődésre számot tartó részproblémája az intencionalitás naturalizációjának, hanem a probléma egészének újrafogalmazása.<sup>144</sup> Ráadásul a négyféle reprezentációelmélettel szemben eddig felhozott, némileg felületes ellenérvekkel ellentétben a téves reprezentáció problémája valódi kihívást jelent mind a négy elmélet számára. A következőkben ezt elemezve próbálunk megoldást találni az intencionalitás naturalizálására, és egyúttal részletesebben megvizsgálni a probléma evolúciós megközelítését is.

## 2. Téves reprezentáció

A téves reprezentáció kérdésköre egy igen egyszerű példán alapul. A „légy” mentális reprezentációnak a külvilágban a legyek felelnek meg: a „légy” reprezentációnak tehát a légy a tárgya.<sup>145</sup> A vizsgált organizmus azonban tévedhet, és légynek tekinthet egy olyan fekete zümmögő dolgot is, ami nem légy, hanem mondjuk darázs.

Több esetben is előfordulhat, hogy a béka azt gondolja, hogy legyet lát: ha legyet lát, ha darazsat lát, ha egy repülő kis kavicsot lát etc. A „légy” reprezentációt tehát több dolog is kiválthatja a külvilágban. De ezen dolgok közül kizárólag a legyek azok, amelyeket a „légy” reprezentáció valóban reprezentál, amelyekkel tehát valódi kapcsolat, valódi korrespondencia áll fent, a többi dolog (darázs, kavics) esetében téves reprezentációról kell beszélnünk. A légy reprezentációjának intencionalitását tehát a helyes reprezentációk határozzák meg. A téves reprezentációk és a helyes reprezentációk elkülönítése nélkül tehát nem elképzelhető az intencionalitás magyarázata.

Hogy egy másik, végre nem a rovarokhoz kapcsolódó példát említsünk, Csehov *Cseresznyés kertjének* első felvonásában Ljubov Andrejevna rég halott édesanyjának hiszi a kert egyik virágbaborult fáját. E hiedelmének valódi tárgya a fehér fa, ő azonban a hajnali fényekben anyjának hiszi ezt a fehér árnyat. Tévesen reprezentál tehát, hiszen reprezentációjának valódi tárgya a fehér fa. Az igazi kérdés persze az, hogy mi dönti el: egy reprezentáció valódi vagy téves reprezentáció-e, hiszen szegény béka vagy Ljubov Andrejevna maga nem tudja eldönteni (ezért is reprezentál tévesen). Egy naturalista intencionalitáselméletnek tehát független kritériumot kell adnia arra, hogy mi különbözteti meg a téves reprezentációt a valóditól.

Mielőtt rátérnénk a téves reprezentáció problémájának evolúciós jellegű megoldásaira, előbb a két leghíresebb, nem-evolúciós megoldásjavaslatot kell áttekintnünk. Jerry Fodor szerint a téves és a helyes reprezentáció közötti különbség kritériumát az aszimmetrikus függés adja.<sup>146</sup> A „légy” reprezentáció–darázs reláció (okszági kapcsolat) aszimmetrikusan függ a „légy” reprezentáció–valódi légy relációtól (okszági kapcsolattól). Precízebben: létezik olyan lehetséges világ, ahol van „légy”–légy reláció, de nincs „légy”–darázs reláció, nem létezik viszont olyan lehetséges világ, ahol nincs „légy”–légy reláció, de van „légy”–darázs reláció. Egyszerűbben fogalmazva: nem nézhetek egy darazsat légynek, ha (korábban) nem néztem valódi legyeket légynek.

Van ezzel a megoldásjavaslattal egy súlyos probléma: az aszimmetrikus függés elmélete előfeltételezi, hogy a fizikai tárgyak (például a legyek), amelyekkel való relációtól aszimmetrikusan függ a többi reláció, jól meghatározhatók az organizmus nézőpontjától függetlenül. Fodor ezt be is vallja, amikor amellet érvel, hogy ha történetesen mindig pónik okozták a „ló” reprezentációt, az akkor is a ló reprezentációja lesz, mert a pónik nem mint pónik, hanem mint lovak okozták a „ló” reprezentációt.<sup>147</sup> Ez az álláspont önmagában kauzalitásfelfogásáért is támadható. Davidson érvei, amelyeket a mentális okság tárgyalásánál láttunk például mind relevánsak itt.<sup>148</sup> Másrészt viszont Fodor elméletéből az is következik, hogy léteznie kell egy „lóság” tulajdonságnak, amely minden megfigyelőtől függetlenül létezik, és a „ló” reprezentációt csak azáltal okozhatja valami, ha részesedik a „lóság” eme tulajdonságában. Ez viszont egy olyan metafizikai világgépet feltételez, ahol a tulajdonságok tőlünk függetlenül léteznek, mi pedig csak ezeket az eleve meghatározott tulajdonságokat tesszük magunkévá. Mindez logikailag koherens, viszont – ha Fodor valóban naturalizálni akarja a mentális tartalmak intencionalitását – teljesen elfogadhatatlan, hiszen ezek a tőlünk függetlenül, eleve meghatározott tulajdonságok szemantikát csempésznek vissza oda, ahonnan Fodor azt kiküszöbölni szeretné. A mentális tartalmak szemantikáját visszavezette e tulajdonságokra, de mivel ezek a tulajdonságok sem nélkülözik a szemantikát, Fodor a mentális tartalmak naturalizálásában semmit nem haladt előre.

A téves reprezentáció problémájának legelterjedtebb megoldáskísérlete a *normális feltételekre* való hivatkozás.<sup>149</sup> A „légy” reprezentációt normális körülmények között csak a legyek okozhatják. Más tárgyak csak abnormális körülmények között képesek erre, a darazsat például csak akkor nézem légynek, ha nem normálisak a kö-

<sup>144</sup> Fred Dretske *Misrepresentation* című cikke volt e kutatási irány első írása, Dretske 1986.

<sup>145</sup> A következőkben a bevett jelölésnek megfelelően a légy reprezentációját az idézőjelbe tett „légy” szóval fogom jelölni, bár ez a jelhasználat erősen a kvázinyelvi elmélet melletti elkötelezettséget tükröz.

<sup>146</sup> Fodor 1990, 4. fejezet, főleg pp. 181–182. Cf. még: Fodor 1987, Fodor 1994.

<sup>147</sup> Fodor 1990, pp. 191–192.

<sup>148</sup> Davidson 1993, főleg p. 13.

<sup>149</sup> Ld. például: Dretske 1986, Dretske 1995.

rülmények, ha például túl sötét van. A „légy” mentális tartalom ezek szerint visszavezethető arra a nem intencionális tartalomra, hogy „normális körülmények között legyek okozzák”. Az intencionalitás tartalmakat tehát naturalizáltak.

Sajnos a helyzet nem ennyire rózsás. A normális körülményeket ugyanis nem lehet a mentális tartalomtól függetlenül meghatározni. A légy felismerésének normális feltételei mások lesznek, mint a szentjánosbogár felismerésének normális körülményei, amelyet viszont éppen sötétben tudok felismerni. Ezeket a normális feltételeket az határozza meg, hogy miről is van szó, tehát a mentális tartalom, amelyet viszont éppen e normális feltételekkel akartunk meghatározni. Körkörösséghez jutottunk tehát: a normális feltételek határozzák meg a mentális tartalmakat, a mentális tartalmak pedig a normális feltételeket.

Szintén ezt a hibát követi el a normális körülményekre való visszavezetés egy látszólag különböző megoldási javaslata is.<sup>150</sup> Eszerint a normális körülményeket meg tudjuk határozni a mentális tartalmaktól függetlenül is, ha azokat a körülményeket nevezzük normálisnak, amelyek esetében érzékszerveink nem tökéletlenül működnek. A legyet azért néztem darázsna, mert túl sötét volt, és ezért szemem nem működhetett tökéletesen. Ha ekkor érzékszerveim működését kifogástalanná lehetne tenni, rögtön észrevenném, hogy tévedtem. Amikor ellenben légynek nézem a legyet, hiába javitanék érzékszerveim felfogóképességén, ugyanazt gondolnám.<sup>151</sup>

Ez az érvelés mégis körkörösséget csempész a normális körülmények meghatározásába. Milyen normális körülményeket kellene felállítani a mozgókép percepciója esetében? A mozgókép mozgásban látása éppen érzékszerveink fogyatékoságát használja ki: másodpercenként csak huszonnégy képkockát vagyunk képesek megkülönböztetni. A gondolatmenetből az következik, hogy a mozgóképnek csak téves reprezentációja van, hiszen érzékszerveink tökéletlenségeinek kiküszöbölése esetén mindig rá fogunk jönni, hogy nem is mozgást látunk. Lehetne persze azzal védekezni, hogy az érzékszervek tökéletesítése nem feltétlenül az abszolút értelemben való tökéletességet jelenti, ez esetben viszont a tökéletesség minden egyes mentális tartalom esetében más és más kritériumokkal rendelkezik, tehát megint csak körkörösen határoztuk meg a mentális tartalmakat és a normális körülményeket.

A téves reprezentáció problémájának egy, az eddigiektől radikálisan eltérő megoldását jelenti az evolúcióra való hivatkozás. A „légy” mentális tartalom és a légy mint fizikai dolog között azért van kapcsolat, mert a béka evolúciós története során is volt ilyen kapcsolat, valamint a béka azért volt képes túlélni, mert fennállt ez a reláció a „légy” reprezentáció és a legyek között. Egy mentális tartalmat e felfogás szerint az határozza meg, hogy korábban milyen relációban állt a környezettel. E tekintetben nem fontos, hogy most mire vonatkozik a „légy” reprezentáció (hogy például éppen egy darazsat nézek légynek), a reprezentáció tartalmát kizárólag a múltbeli légyfelismerések határozzák meg.

Ruth Millikannek, az intencionalitás evolúciós felfogása nagy úttörőjének terminológiájával megfogalmazva:<sup>152</sup> a „légy” mentális tartalomnak az a *valódi funkciója*, hogy relációban álljon a legyekkel.<sup>153</sup> Millikan szerint a mentális reprezentációk ugyanúgy funkcióval rendelkeznek, mint például a szív. Ráadásul e funkciót ugyanúgy az evolúciós előtörténet határozza meg, mint a szív esetében. Ha kihagy a szívem, akkor is szívnek kell tekinteni, hiszen nem az a fontos, hogy most miként funkcionál, hanem hogy korábban, evolúciós előtörténetem során miként funkcionált. Hasonlóképpen a légyre vonatkozó reprezentáció attól légyre vonatkozó reprezentáció, hogy korábban légyre vonatkozó reprezentáció volt az organizmus evolúciós őseinél, függetlenül attól, hogy most mire használható, és mire nem. Fontos megjegyezni, hogy ez az előtörténet ontogenetikus előtörténet is lehet, nemcsak a filogenetikus, tehát az egyedfejlődést is jelentheti, nemcsak a törzsfjlődést.

Ez az elmélet a téves reprezentáció kérdését igen jól képes kezelni: a valódi reprezentációt éppen az különbözteti meg a téves reprezentációtól, hogy korábban is jelen volt az organizmusban, sőt, az éppen általa tudott túlélni, hogy rendelkezett ezzel a reprezentációval. Látható, hogy itt nincs szó körkörösségről, Millikannek sikerül a mentális tartalmak intencionalitását nem intencionális jelenségekre (a vak evolúcióra) visszavezetni.

Millikan számára egy reprezentáció attól reprezentáció, hogy korábban reprezentáció volt az organizmus evolúciós őseinél. A normális körülményekre való visszavezetés elmélete, mint láttuk, ezzel ellentétes álláspontot fogalmaz meg: a reprezentáció tartalmát az határozza meg, hogy mire használható most. Ezek szerint tehát a *jelenben* betöltött funkció határozza meg a reprezentáció tartalmát, míg az evolúciós megközelítés szerint a *múltban* betöltött funkció. A normális feltételek elmélete a jelenre vonatkozik, Millikan pedig a múltra vezeti vissza a reprezentációk funkcióját. Millikan tehát nem szembesül azzal a – mint láttuk – igen nehéz feladattal, hogy a nem téves reprezentáció jelenbeli normális körülményeit a mentális tartalomtól függetlenül határozza meg, hiszen az ő számára a múltbeli mentális tartalmak fixálják a jelenbeli mentális tartalmakat, körkörösség nélkül.

Millikan elméletének egyik leghíresebb kritikája az az empirikus megfigyelés, hogy sokszor evolúciósan hasznosabb, ha tévesen reprezentálunk, mint ha jól.<sup>154</sup> Az a katona, aki azt hiszi magáról, hogy nem sebesülhet

<sup>150</sup> Whyte 1991.

<sup>151</sup> Ibid. p. 72.

<sup>152</sup> Millikan 1984, Millikan 1993.

<sup>153</sup> Millikan 1984, pp. 27–28.

<sup>154</sup> Cf. Papineau 1993. Ch. 3.

meg, bátrabban fog harcolni, és esetleg akkor is győzelemre vihetné csapatát, ha ez nem sikerülne neki, ha félne a sebesüléstől. Előfordulhat tehát, hogy a téves reprezentáció szelekciós előnyt élvez a helyes reprezentációval szemben, ami viszont azt jelenti, hogy borul Millikan elmélete, amely szerint éppen az különbözteti meg a nem téves reprezentációt, hogy az az evolúció során biztosította az organizmus túlélését. Ez az ellenvetés azonban nem jelenti az evolúciós szemantika használhatatlanságát. Úgy tűnik, a sikerszemantika téziseinek segítségével kivédhető ez az ellenvetés.

A sikerszemantika szerint egy vélekedés, egy elképzelés akkor igaz, ha hozzásegíti az organizmust ahhoz, hogy beteljesítse egy vágyát, ha tehát sikeres.<sup>155</sup> Ha csokoládét akarok enni, akkor az a vélekedésem, hogy csokoládé van a hűtőszekrényben, pontosan akkor lesz igaz, ha ez a gondolat hozzásegít ahhoz, hogy valóban csokoládét egyek.<sup>156</sup> A sikerszemantika tehát nem evolúciós, de teleologikus színezetű megoldást ajánl a téves reprezentáció problémájára: egy reprezentáció akkor nem téves, ha sikeres, ha tehát hozzásegít egy vágyam beteljesítéséhez.<sup>157</sup> Könnyen látható viszont, hogy ez a magyarázat előfeltételezi a vágy beteljesítésének fogalmát.

A sikerszemantika szerint egy vágy akkor teljesül be, ha megszűnik létezni.<sup>158</sup> Látható, hogy az érvelés itt kerül mindenféle evolúciós terminust, implicit módon azonban mégis evolúciós színezetű gondolatmenetet előfeltételez. Azt állítja ugyanis hogy nem elégséges feltétel egy vágy beteljesítéséhez az hogy megszűnjön létezni, hiszen ekkor a már említett „csokoládét akarok enni” vágyat beteljesítené az is, ha valaki jól gyomorszájon rúgna, márpedig a sikerszemantika ezt az esetet szeretné kizárni a vágy beteljesítésének esetei közül. Ezért azt a további feltételt támasztja, hogy egy vágy akkor teljesül be, ha megszűnik létezni, mégpedig egy olyan mechanizmus által, amelyet megismételnék egy későbbi hasonló szituációban is. A gyomorszájon rúgást nyilván nem használnám legközelebb arra, hogy megszűnjön a csokoládéra irányuló vágyam, míg a csokoládéevést annál inkább.<sup>159</sup> A sikerszemantika szerint tehát akkor van kapcsolat a vágy és beteljesítése között, ha ez a kapcsolat a jövőben is fennállna ha a körülmények adottak lennének. Hogy világosabb legyen ez a kétségtelenül homályos megfogalmazás, a normális feltételekre való hivatkozás, az evolúciós előtörténet és a sikerszemantika közötti kapcsolatot érdemes úgy felfogni, hogy míg az evolúciós előtörténet-modell szerint a reprezentáció és a reprezentált dolog közötti *múltbeli* kapcsolat a meghatározó, addig a normális feltételek a *jelen* kapcsolatot, a sikerszemantika pedig a *jövőbeli* kapcsolatot tekinti relevánsnak.

A sikerszemantika tehát két lépésben naturalizálja a mentális tartalmakat. Első lépésben a vélekedések mentális tartalmát visszavezeti a vágyak mentális tartalmára. A második lépés az, hogy a vágyak mentális tartalmát naturalizálja. David Papineau – aki Millikan mellett az evolúciós intencionalitáselmélet másik nagy képviselője – elfogadja az első lépést, de elutasítja a másodikat.<sup>160</sup> Az első lépés segítségével képes kivédeni a „bátor harcos” ellenérvet. Ha egy harcos azt gondolja, hogy nem sebesülhet meg, ez a hit számára evolúciósan hasznos, ámbar intuitíve nem igaz. Egy téves reprezentáció is lehet tehát evolúciósan hasznos, tehát reprodukálódhat. Ha viszont elfogadjuk, hogy annyiban lehet egy reprezentációt helyes reprezentációnak tekinteni, amennyiben beteljesíti valamely vágyunkat, akkor a „nem sebesülhetek meg” vélekedés – bár szemléletesen nem igaz, mégis – helyes reprezentáció lesz, hiszen beteljesíti a harcosnak azt a vágyát, hogy megnyerje a csatát. Papineau tehát a sikerszemantika első lépését átveszi, a másodikra viszont nincs szüksége, mivel azt állítja, hogy az evolúciós előtörténet határozza meg a vágy és beteljesülése közötti korrespondenciát, a vágyak intencionalitásának visszavezetését tehát az evolúciós előtörténet biztosítja. Ha csokoládét akarok enni, akkor azt, hogy e vágyamat mi tudja beteljesíteni, a korábbi csokoládévágyak betöltése határozza meg. Papineau tehát a sikerszemantika és a tiszta evolúciós szemantika kombinálására tesz kísérletet.

### 3. Intencionalitás és evolúciós változás

A téves reprezentáció problémáját tehát, úgy tűnik, eddig az evolúciós szemantika volt képes legjobban kezelni. Van azonban egy súlyos probléma a klasszikus evolúciós szemantikával. Ha a mentális tartalmakat (akár a vélekedéseket, mint Millikannél, akár a vágyakat, mint Papineau-nál) az evolúciós előtörténet határozza meg, akkor az új mentális tartalmakat – amelyeknek értelemszerűen nincs evolúciós előtörténetük – nem határozza meg semmi, következésképpen ilyen mentális tartalmak nem léteznek. Ezt Millikan világosan meg is fogalmazza híres „replikáns” gondolkísérletében.<sup>161</sup>

<sup>155</sup> Whyte 1990, Whyte 1991.

<sup>156</sup> Whyte 1990, pp. 149–150.

<sup>157</sup> Fontos megjegyezni, hogy Millikan is különbséget tesz vágyak és vélekedések között, de – ellentétben Whyte-tal – naturalizálásukat hasonló módon képzeleli el.

<sup>158</sup> Whyte, 1991, p. 65.

<sup>159</sup> Ibid. p. 68.

<sup>160</sup> Papineau 1993, Papineau 1996, Cf. még: Papineau 1987.

<sup>161</sup> E gondolkísérlet első megfogalmazása (a mocsárember esete) Davidsonnál található. Davidson 1987.

Tegyük fel, hogy mesterségesen létrehozunk egy olyan élőlényt, amely atomról atomra azonos velem. Millikan szerint egy ilyen organizmusnak nincsenek mentális tartalmai, mint ahogy nincs mája, szíve és más biológiai szerve sem, hiszen nincs evolúciós előtörténete, ami meghatározná a szervek és mentális tartalmak „valódi funkcióját”.<sup>162</sup> Világos a tétel radikalitása: csak akkor rendelkezem mentális tartalmakkal, ha evolúciós előtörténetemben is léteztek már ezek a mentális tartalmak.

Nem kell azonban ilyen elrugaszkodott gondolat kísérletet segítségül hívni ahhoz, hogy megkérdőjelezzük az evolúciós történetre való hivatkozás egyetemes magyarázó erejét. Millikan elmélete az evolúció stabil állapotait jól magyarázza, nem képes viszont az evolúciós változások kezelésére – igaz, erre nem is nagyon törekszik. Millikan elsősorban az evolúció szélcsendes periódusait vizsgálja, nem elsődleges célja az evolúciós változások magyarázata. Így voltaképpen a tanulással sem tud mit kezdeni: az új reprezentációknak nincs tartalmuk, hiszen nem határozza meg őket azt az evolúciós előtörténet (ekkor persze az ontogenetikus, az egyedfejlődés során végbemenő evolúcióról van szó). Millikan persze tudatában van ennek a problémának, ezért különbséget tesz a direkt és a származtatott valódi funkciók között.<sup>163</sup> A direkt valódi funkciót kizárólag az evolúciós előtörténet határozza meg. Ezzel szemben származtatott valódi funkcióról akkor beszélünk, ha az előtörténet mellett a pillanatnyi környezet is befolyásolja a mentális tartalmat.

Tegyük fel például, hogy egy kaméleon még soha nem volt kék színű, viszont rendelkezik azzal az evolúciós eredetű képességgel, hogy mindig olyan színű lesz, amilyen a környezete, hogy a ragadozók ne vegyék észre. Ha ekkor a kaméleon kék színű környezetbe kerül, és kék lesz, akkor a bőr pigmentjeinek bekékkülése nem bír direkt valódi funkcióval, hiszen korábban sem e kaméleon, sem az ősei nem színeződtek még kékre. A pigmentek kék színűre változása viszont rendelkezik származtatott valódi funkcióval, amely két dologból származtatódik: egyrészt a pillanatnyi (kék) környezetből, másrészt a kaméleon azon direkt valódi funkciójából, hogy olyan színűvé kell változni, mint a környezet, mert ekkor a ragadozók nem vesznek észre.<sup>164</sup> A pillanatnyi környezet és az evolúciós előtörténet tehát együtt határozzák meg az új mentális tartalmak intencionalitását. Így tehát a kecske is jól lakik, a káposzta is megmarad: az evolúciós előtörténet is szerephez jut, ugyanakkor újdonság is megjelenhet a pillanatnyi környezet szerepe miatt. A struktúra evolúciósan adott, csak az változik a pillanatnyi környezetnek megfelelően, hogy éppen milyen módon töltődik be ez a struktúra. Millikan szerint a származtatott valódi funkciók elmélete képes kezelni az új reprezentációk és egyáltalán az újdonság evolúciós szerepét, de látnunk kell, hogy ez egy rendkívül konzervatív evolúciófelfogáshoz vezet. Egy ilyen modell szerint ugyanis minden változás egy merev és változatlan struktúra komponenseinek cserélődése, anélkül azonban, hogy maga az evolúciósan meghatározott struktúra megváltozna.

Sem a direkt, sem a származtatott valódi funkciók elmélete nem képes tehát kezelni a valódi, nem triviális evolúciós változásokat. Ez az első nagy probléma az evolúciós előtörténetre hivatkozó modellel. A másik igen csak vitatható előfeltevése ennek az elméletnek az, hogy minden mentális tartalmat adaptívnek tekint. A második fejezetben már kitértünk az adaptáció–exaptáció vitára, amely lényegi ellenvetést jelent az evolúciós megközelítés és főként Millikan számára.<sup>165</sup> Az úgynevezett „friss evolúciós előtörténet” megközelítés megpróbálja beépíteni az exaptáció jelenségét a millikani elméletbe. E szerint nem elégséges kritérium, hogy valamikor régen hasznos volt egy tulajdonság, annak az organizmus közvetlen előtörténetében is szerepet kellett játszania.<sup>166</sup> Ha a panda csuklóján exaptív módon megjelenő csontkinövésekből a közelmúlt adaptív folyamatai esiszoltak hatodik ujját, akkor mégis lehet funkciót tulajdonítani neki, annak ellenére, hogy keletkezése exaptáció volt, hiszen nem az a fontos, hogy mi történt régen, hanem hogy milyen evolúciós mechanizmusok hatnak a szervezetre „mostanában”.<sup>167</sup> Sajnos ez a megoldási javaslat nem képes kezelni a vakbél esetét, amely, bár „mostanában” nem adaptív, mégis megmaradt, egyszerűen amiatt, hogy nem is különösebben káros a szervezet számára. Szintén nem használható a „friss evolúciós előtörténet” modell az olyan kulturális fogalmak (cseresznyefa-reprezentáció, a Iokaszté-reprezentáció, üveghegy-reprezentáció) evolúciós eredetének magyarázatára, amelyek soha nem voltak adaptívak az egyed számára. Az adaptáció tehát továbbra is probléma marad az intencionalitás evolúciós magyarázata számára.<sup>168</sup>

A következőkben a téves reprezentáció problémakör egy olyan új, evolúciós elméletének lehetőségét körvonalazom, amely talán képes elkerülni a túlzott adaptacionizmus vádját.

Mindenekelőtt a téves reprezentáció kérdéskörében eddig használt leegyszerűsített reprezentáció-fogalom helyett a következőkben a reprezentációt úgy fogom meghatározni mint az inger és a válaszreakció összekötését.

<sup>162</sup> Millikan 1984, p. 92. Cf. Millikan 1993 1. fejezet.

<sup>163</sup> Millikan 1984, pp. 39–44.

<sup>164</sup> Ibid. p. 41.

<sup>165</sup> Millikan 1984-es könyvében az exaptáció problémájának első – és gyakorlatilag utolsó – említése a 136. oldalon történik. A származtatott valódi funkciók elméletét lehet úgy interpretálni, mint az exaptáció fogalmának bevezetésére irányuló kísérletet, de még ezen igen jóindulatú interpretációt elfogadva is igencsak dominál a könyvben az adaptacionista megközelítés.

<sup>166</sup> Cf. Griffith 1993, Godfrey–Smith 1994.

<sup>167</sup> Ez a „mostanában” persze több százezer év is lehet. A „friss evolúciós előtörténet” modell egyik hibája, hogy nem határozza meg világosan, mit is jelent ez a „mostanában”.

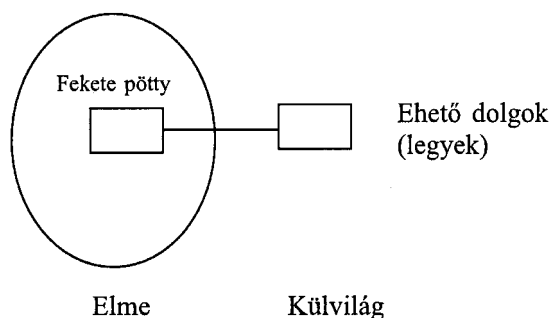
<sup>168</sup> Ezt persze többen megkérdőjelezzik, a funkciók evolúciós magyarázata és az adaptáció–exaptáció probléma kapcsolata a biológia filozófiájának egyik legvitatottabb kérdése. Cf. Walsh 1996.

Akkor lehet a légy reprezentációról beszélni, ha a légy valamilyen akciót jelent a számomra. Meglátom a legyet, és kiöltöm a nyelvemet. A reprezentáció tehát az akció lehetőségének felismerése az ingerben: az ingert mint az akció lehetőségét percipiálom. Fontos hangsúlyozni, hogy az akció nem feltétlenül jelent valóban végrehajtott akciót.<sup>169</sup> Legtöbbször csak az akció lehetőségét percipiálom, a voltaképpeni akció elmarad. Akkor beszélünk reprezentációról, ha egy (vizuális) ingerben a motoros akció lehetőségét ismerem fel: ha a fekete pöttyöt mint az evés akciójának lehetőségét percipiálom.<sup>170</sup>

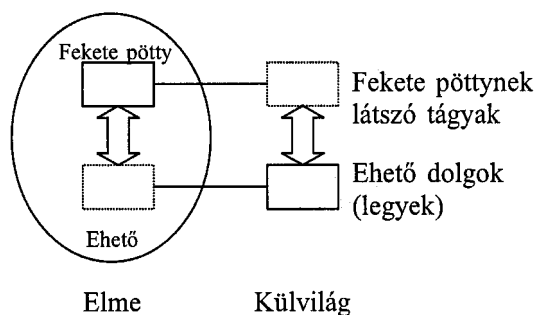
Bár az elmefilozófia tradícióitól távol áll ez a meghatározás,<sup>171</sup> a kognitív tudomány több képviselője is hangsúlyozta, hogy a reprezentációk tárgyalását és az ingereket nem szabad elválasztani az általuk kiváltott válaszreakciók kérdésétől.<sup>172</sup> A percepció egoista folyamat: mindig azt látom meg egy tárgyról, hogy mire tudom használni, milyen akciót kínál számomra. A szék reprezentációja például nem pusztán kontúrok és felületek összessége, hanem annak az akciónak a lehetőségét is magában foglalja, hogy le tudok rá ülni. Egy tárgy persze különböző pillanatokban és lelkiállapotokban más és más akció lehetőségét hordozhatja, egy újság által kiváltott ingerekhez például más akciókat kapcsolok akkor, ha egy légy röpköd a szobámban, amelyet le akarok csapni, mint akkor, ha a választási eredményekre vagyok kíváncsi.<sup>173</sup> Az akció az ember esetében nem kizárólag végrehajtott akció, de nem is kell feltétlenül motoros jellegűnek lennie: egy matematikapélda megoldása közben a + jelben az összeadás akciójának lehetőségét ismerem például fel.<sup>174</sup>

Ha azonban elfogadjuk, hogy a reprezentáció nem más, mint az akció lehetőségének felismerése az ingerben, akkor a téves reprezentáció problémája is leegyszerűsödik. Ha az ingert kiváltó tárgy azonos azzal a tárggyal, amelyre az akció irányul, akkor helyes reprezentációról beszélünk. Ha nem azonos, akkor kapjuk a téves reprezentációt.

A téves reprezentáció megértéséhez tehát nem elég, ha csak a reprezentáció és a reprezentált dolog közötti kapcsolatot vizsgáljuk. Nem használható tehát az a kétpólusú modell, amelyben a téves reprezentáció klasszikus magyarázatai gondolkodnak. Ehelyett egy négypólusú modellt javaslok, amelyben a reprezentáció és annak tárgya közötti direkt kapcsolat helyett lényegesen összetettebb mechanizmus áll.



*A téves reprezentáció kétpólusú modellje*



*A téves reprezentáció négypólusú modellje*

<sup>169</sup> Ez csak a nagyon alacsonyrendű élőlényekre, illetve például a gerinceleji reflexre lenne igaz.

<sup>170</sup> Bár látszólag e magyarázati séma igen hasonló a diszpozicionális elemzések logikájához (cf. Ryle 1974), óriási különbség van e két modell között, hiszen a diszpozicionális elemzések épp a reprezentációk helyett vezetnek be a viselkedés diszpozicionális leírását, míg az itt javasolt modell a reprezentációkat magukat írja le az inger és a kiváltott akció közötti kapcsolat segítségével.

<sup>171</sup> Bár Leibniz sok tekintetben szellemi elődnek tekinthető. Cf. Leibniz 1986.

<sup>172</sup> Cf. Edelman 1987, Bunge 1980, és elsősorban Gibson 1966, 1979, róla: Nánay 1997c, ellene: Fodor–Pylshyn 1981, Ullman 1980.

<sup>173</sup> A példa David Marrtól származik. Marr 1982.

<sup>174</sup> Az emberi reprezentáció bonyolultabb eseteit, az absztrakt, nem akciószerű reprezentációk kialakulását az utolsó előtti fejezet fogja tárgyalni.

Négy entitást kell tehát bevonni a magyarázatba, az eddigi kettő helyett, kettőt az elmén kívül: az ingert kiváltó tárgyat és azt a tárgyat, amelyre az akció irányul, valamint kettőt az elmében: a válaszreakciót kiváltó mentális entitást és az ingert. Akkor beszélünk reprezentációról, ha a két elmén belüli entitás megegyezik, ha tehát az elme azonosnak tekinti őket; másként fogalmazva: ha az ingerben felismeri a válaszreakció lehetőségét, ha összeköti az ingert a válasszal. Ez a reprezentáció pedig akkor lesz helyes, ha a két fizikai tárgy – az ingert kiváltó tárgy, illetve az, amelyre a reakció vonatkozna – megegyezik. Egy tárgy által kiváltott ingerben az elme felismeri egy akció lehetőségét, és ha az a tárgy, amelyre ez az akció irányul (vagy irányulna) azonos az ingert kiváltó tággyal, akkor beszélünk helyes reprezentációról. Minden más esetben a reprezentáció téves.

Eddig a téves reprezentáció problémájára kerestünk választ. Miként használható azonban ez az új megközelítés az intencionalitás naturalizálásának kérdésében?

Ehhez először is lássuk, hogyan alkalmazható a fenti modell a már jól ismert kreatív béka példájára. Kezdetben, amikor csak legyek repkednek a béka körül, a béka annak a mechanizmusnak a segítségével táplálkozik, hogy minden olyan tárgyat bekap, amely fekete pöttyként jelenik meg a retináján. Ez a módszer egyszerű, de hatékony, mivel a fekete pötty helyesen reprezentálja a legyeket. De mit is jelent ez? Ha egy tárgy kiváltja a fekete pötty ingerét, és a béka ebben az ingerben felismeri a táplálkozás akciójának lehetőségét, akkor ez az akció az ingert kiváltó tárgyra fog vonatkozni.

Sajnos azonban a darazsak megjelenésével ez az egyszerű módszer nem lesz túl hatékony, hiszen az ingert kiváltó tárgy nem mindig lesz azonos azzal, amelyre a béka táplálkozási akciója irányul. Az akció mindig legyekre irányul (mert ezt tudja megemészteni a béka), az inger kiváltója viszont sok esetben darázs lesz. Ebben a környezetben tehát téves a béka reprezentációja: az ingert kiváltó tárgy sok esetben nem azonos azzal a tárggyal, amelyre a válasz irányul. Valaminek változnia kell, hiszen különben kipusztul a békapopuláció. A táplálkozás akciója mindig a légyre fog irányulni, hiszen ez a béka csak legyeket képes megemészteni. Az ingernek kell tehát megváltoznia, ha a béka el akarja kerülni a téves reprezentációt, márpedig el kell kerülnie, különben kipusztul.

Ebben a szorongatott helyzetben jelenik meg a kreatív béka elméjében a „fekete pötty sárga csíkok nélkül” mentális entitás. Ezt az ingert kiváltó tárgy már minden esetben azonos lesz azzal a tárggyal, amelyre a táplálkozás akciója irányul, tehát a legyekkel. Újra helyesen reprezentál tehát a béka, de eközben a reprezentációja megváltozott: most már nem a „fekete pötty” ingerében ismeri fel az evés akciójának lehetőségét, hanem a „fekete pötty sárga csíkok nélkül” ingerben. Egy bonyolultabb reprezentáció jelent meg.

Ez a modell tehát – Millikan elméletével ellentétben – képes kezelni az evolúciós változásokat, a radikálisan új mentális reprezentációk megjelenését is. Ráadásul az adaptivitás kérdése sem merül fel, hiszen ha az új reprezentáció nem jelent volna meg, akkor az állat kipusztult volna.

A környezeti változás és a téves reprezentáció elkerülése új, az előzőnél bonyolultabb reprezentáció megjelenését tette lehetségessé és szükségessé (a „fekete pötty sárga csíkok nélkül” kialakulását a „fekete pötty”-ből). A „fekete pötty” igen egyszerű mentális entitás: intencionalitása könnyen naturalizálható, visszavezethető a retina csap- és pálcikasejtjeinek működésére. Ha viszont elfogadjuk a fenti modellt, amely szerint a reprezentációk egyre bonyolultabbá válhatnak, mivel el kell kerülniük azt, hogy tévesek legyenek (ekkor ugyanis eltűnnek), akkor igaz az, hogy minden magasabb rendű, komplex reprezentáció az ilyen egyszerű, naturalizálható reprezentációkból fejlődött ki az itt leírt, jól meghatározható és semmi nem-naturalisztikus elemet nem tartalmazó evolúciós mechanizmus által.

Ezt a mechanizmust persze a neurális evolúció új elmélete biztosítja. A mentális reprezentációk intencionalitása ugyanis visszavezethető azon egyszerűbb reprezentációk intencionalitására, amelyből evolválódott, ez a még egyszerűbbekre, egészen a neurálisan azonosítható olyan egyszerű reprezentációkig, mint a „fekete pötty”. Mivel a visszavezetés mindegyik lépése kizárólag evolúciós mechanizmusokat feltételez, az utolsó lépés pedig az idegtudomány feladata, a legbonyolultabb reprezentációk intencionalitását is meg lehet magyarázni naturalisztikus terminusokban. Az intencionalitást ezek szerint az evolúciós szemantikák korábban bemutatott problémáit elkerülve sikerült naturalizálni.

## 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.<sup>175</sup> 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.<sup>176</sup>

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 egyértelműségeen 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 strandabdáig. 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

<sup>175</sup> 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.

<sup>176</sup> 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.<sup>177</sup> 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.<sup>178</sup> 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.<sup>179</sup>

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.<sup>180</sup> 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.<sup>181</sup> 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

<sup>177</sup> Fodor 1983, 1996a. Róla cf. Pléh 1986.

<sup>178</sup> Ryle 1974.

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

<sup>180</sup> Karmiloff–Smith 1992, 1996.

<sup>181</sup> 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.<sup>182</sup> 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.<sup>183</sup>

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.<sup>184</sup> 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.<sup>185</sup> Első közelítésben tehát azt lehet mondani, hogy a konnekciónizmus kétszáz év után visszahozta az ismeretelméleti gondolkodásba a szokás fogalmát a szabállyal szemben.<sup>186</sup> 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

<sup>182</sup> Searle 1996.

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

<sup>184</sup> 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.

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

<sup>186</sup> 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.<sup>187</sup>

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.<sup>188</sup> 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.<sup>189</sup>

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.<sup>190</sup> 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

<sup>187</sup> 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.

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

<sup>189</sup> A prototipikus általánosításról lásd Rumelhart–McClelland 1986a.

<sup>190</sup> 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.<sup>191</sup>

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.<sup>192</sup>

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.<sup>193</sup>

A konnekcionizmus legradikálisabb, valóban új és igazán jelentős gondolatai a hálózat tanulásával foglalkoznak.<sup>194</sup> 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.<sup>195</sup> 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.<sup>196</sup> A konnekcionizmus Ryle elvont elméletét tette megfoghatóvá és modellezhetővé, azáltal, hogy az elmét hálózatnak tekinti.<sup>197</sup>

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.<sup>198</sup> impresszióknak nevezi a minket érő ingerek összességét, ideának pedig a mentális reprezentációt.<sup>199</sup> Hume szerint az impresszió határozza meg az ideát.<sup>200</sup> 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

<sup>191</sup> Ezt a hiányosságot emeli ki a konnekcionizmus leghíresebb kritikája: Fodor–Pylyshyn 1988

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

<sup>193</sup> McClelland–Rumelhart 1986b.

<sup>194</sup> 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.

<sup>195</sup> „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.

<sup>196</sup> A konnekcionizmus különböző előzményeiről cf. Nánay 1996, Nánay 1997a, Pléh 1992.

<sup>197</sup> Cf. Nánay 1997b.

<sup>198</sup> A magyar fordításban az impresszió fogalmát – kissé félrevezető módon – benyomásnak is fordítják.

<sup>199</sup> Ez persze a kognitív terminológia felőli némileg anakronisztikus utólagos értelmezés.

<sup>200</sup> 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.<sup>201</sup> 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.<sup>202</sup> 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 elmamodellek összebékítésére annyi és annyiféle megoldási javaslat született, hogy elengedhetetlennek tűnik ezeknek a kettős modelleknek valamilyen rendszerezése. Első közelítésben meg kell különböztetni a hibrid, illetve az integrált modelleket.<sup>203</sup>

Hibrid modelleknek nevezzük azokat a leírásokat, amelyek szerint az elme egyik része konnekcionista módon működik, egy másik viszont szimbólumkezelő rendszer.<sup>204</sup> 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.<sup>205</sup> 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 elmamodelleknek 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.<sup>206</sup> Harnad rendszere ugyanis rendkívül leegyszerűsíti-

<sup>201</sup> Ryle 1974. Második fejezet.

<sup>202</sup> 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.

<sup>203</sup> 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.

<sup>204</sup> Hibrid modellek például: Hendlér 1989a, 1989b, Clark 1991, 1994, Harnad 1994.

<sup>205</sup> Harnad 1994.

<sup>206</sup> 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.<sup>207</sup> 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.<sup>208</sup> 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 elméműködés 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.<sup>209</sup> 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 elméműködéssel 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 elméműködés 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.

<sup>207</sup> Wittgenstein 1992. p. 135. § 251.

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

<sup>209</sup> 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.<sup>210</sup>

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.<sup>211</sup> 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.<sup>212</sup>

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.

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

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

<sup>212</sup> 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 implementációfüggelenség elve, s így a funkcionizmus sem.

Hogy emlékezetes katonai példánknál maradjunk, a klasszikus szimbólumkezelő modell valami olyasminnek 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 zászlóaljaknak. 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.<sup>213</sup>

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.

<sup>213</sup> 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.

## VII. Magasabb szintű mentális reprezentációk

A könyv eddigi fejezeteiben a reprezentáció fogalmát leszűkített értelemben használtam: úgy kezeltem, mint ami közvetlenül kapcsolódik az őt kiváltó ingerekhez. Az emberi gondolkodás egyik legfontosabb sajátossága azonban éppen az, hogy képes akkor is felidézni egy gondolatot, képzetet, reprezentációt, ha az e reprezentációt kiváltó inger nincs jelen. Az állat nem jelenlévő dolgokról (vagy a jelenlévő ingerekhez semmilyen módon nem kapcsolódó dolgokról) valószínűleg nem tud gondolkodni, az ember viszont igen. Az állati reprezentáció tehát függ az ingerektől, míg az emberi reprezentáció függetlenedhet tőlük. Ez az ingerfüggetlen reprezentáció teszi lehetővé, hogy például a múltra emlékezzünk. Az állat számára a múlt mindig csak a jelennel kapcsolatban, mint a jelen szempontjából releváns múlt jelenik meg.<sup>214</sup> Az ember azonban képes a múltra a jelentől függetlenül, a jelen ingerektől elszakadva is emlékezni. Az a kérdés, hogy az ingerektől való függetlenedés miként és milyen evolúciós mechanizmussal jött létre, illetve milyen kapcsolatban áll a nyelvi reprezentáció megjelenésével.

A nyelv evolúciójának és a többi magas szintű mentális képesség evolúciójának egymáshoz való viszonya az elme vizsgálatának egyik legfontosabb problémája: a nyelv az alapja minden más, az emberre jellemző bonyolult mentális képességnek, vagy fordítva, a nyelv csak a jéghegy csúcsa, sok már meglévő bonyolult elmeműködés tette lehetővé egyáltalán a nyelv megjelenését. E vita igen lényeges kérdésre világít rá, ha ugyanis elfogadjuk, hogy a nyelv megjelenése volt az oka minden más az emberre jellemző mentális változásnak, akkor maga a nyelv éles határvonalat fog jelenteni az ember és a többi főemlős között, ami evolúciós szempontból vitatható gondolat. Ha viszont a nyelvet egy késői, sok más elmebeli változást előfeltételező jelenségnek tartjuk, akkor egyrészt jobban tudjuk magyarázni a nyelv megjelenését is, másrészt pedig az ember és a többi állat közötti szakadék sem lesz olyan mély.<sup>215</sup> E fejezet az utóbbi álláspont mellett próbál érvelni.

### 1. A metareprezentáció

Ezen a ponton érdemes bevezetni egy, az utóbbi évek elmefilozófiai és kognitív tudományi diskurzusában nagy karriert befutott kifejezést, a *metareprezentáció* fogalmát. A metareprezentáció – igen tág definíció szerint – a reprezentáció reprezentációja.<sup>216</sup> A magasabb szintű emberi reprezentációk megjelenését vizsgáló teoretikusok nagy része nem a bonyolult (nyelvi, ingerfüggetlen stb.) reprezentációk evolúciós, filogenetikus kialakulásának vizsgálatára helyezik a hangsúlyt, hanem – Jean Piaget hasonló kérdésfelvetését követve – a fejlődéslélektani kérdésekre, arra tehát, hogy a gyermek fejlődése során mikor és hogyan jelennek meg a magasabb szintű reprezentációk.<sup>217</sup> A metareprezentáció kifejezés is ebben a fogalmi keretben jelent meg.

Alan Leslie azt a még Piaget-től eredő kérdést vizsgálja, hogy a gyerek reprezentációi hogyan válnak függetlenné az aktuális ingerektől. Ezen belül a mintha-játék jelenségét elemzi, amely kétéves kor körül jelenik meg. Akkor beszélünk mintha-játékról, ha a gyerek például egy banánt a füléhez emel, és azt mondja, hogy „halló”, ha tehát egy tárgyról azt tetteteli, hogy egy másik tárgy. Igen fontos, hogy a gyerek pontosan tisztában van azzal, hogy ez a két tárgy nem azonos, ő mégis azonosnak tekinti (a játék kedvéért). Leslie amellett érvel, hogy a mintha-játék jelensége nem magyarázható kizárólag ingerfüggő reprezentációk segítségével (ezt ő elsődleges reprezentációnak nevezi). A mintha-játék ugyanis nem más, mint két elsődleges reprezentáció párhuzamos koordinálása. Ez viszont nem lehetséges, ha csak elsődleges reprezentációkkal dolgozik a gyerek elméje, hiszen ekkor az elsődleges reprezentációk – a banánreprezentáció és a telefonkagyló-reprezentáció – a játék közben megváltoznának, egymáshoz idomulnának.<sup>218</sup> Szükség van tehát bonyolultabb, ingerfüggetlen reprezentációkra is ahhoz, hogy a mintha-játékot kielégítően írjuk le.

Leslie ezt az ingerfüggetlen reprezentációt azonosítja a metareprezentáció fogalmával. Szerinte a gyerek fejlődése során kétéves kor körül megjelenik a metareprezentáció képessége. Ekkorra a gyerek képes reprezentációkat gyártani nemcsak a világ dolgairól, de saját reprezentációiról is. Képes tehát reprezentációit manipulálni. A metareprezentáció Leslie (és Zenon Pylyshyn) szerint szintaktikai szervezettségű; minden metareprezentáció egy egységes „Gondolkodás Nyelvére”, *lingua mentis*re kódolva tárolódik. Ez úgy képzelendő el, hogy minden elsődleges reprezentációról készül egy leírás a „Gondolkodás Nyelvére” fordítva, és ezek lesznek a reprezentáció reprezentációi. Könnyen látható, hogy ez valóban elegendő ahhoz, hogy ingerfüggetlen legyen a reprezentáció, hiszen a nyelvi leírások már nem függenek azoktól az ingerektől, amelyek invarianciaosztályát leírják.

<sup>214</sup> Cf. Donald 1991.

<sup>215</sup> Donald szerint például a nyelv megjelenése előtt már több százezer évig majdnem minden mai mentális képességgel rendelkezett a Homo Erectus, csak éppen a kommunikációt nyelv helyett mimetikus eszközökkel, utánzással valósította meg. Cf. Donald 1991.

<sup>216</sup> A metareprezentáció kifejezést Zenon Pylyshyn vezette be (Pylyshyn 1978), majd Alan Leslie híres cikke nyomán vált egyre elterjedtebbé (Leslie 1987).

<sup>217</sup> Cf. Piaget 1962.

<sup>218</sup> Leslie ezt a jelenséget szemléletesen reprezentacionális abúzusnak nevezi. Cf. Leslie 1987. p. 415.



Leslie tehát azonosítja az ingerfüggetlen reprezentáció és a metareprezentáció fogalmát. Ez az azonosítás azonban önkényes. A metareprezentáció szükségképpen ingerfüggetlen, az azonban egyáltalán nem biztos, hogy nem létezhet olyan ingerfüggetlen reprezentáció, amely nem rendelkezik a metareprezentáció igen bonyolult, szintaktikai, nyelvi szervezetségével. Elképzelhető, hogy nem kétféle, hanem háromféle reprezentációs formát kell megkülönböztetni. Nem elég az elsődleges reprezentáció és a metareprezentáció közötti dichotómiáról beszélni, hanem egy harmadik, köztes kategóriát is érdemes bevezetni.

Éppen ezt tette Josef Perner, aki szerint a mintha-játékhoz nem szükséges metareprezentáció, elég az úgynevezett „másodlagos reprezentáció” is.<sup>219</sup> Perner másodlagos reprezentáció fogalma azt jelenti, hogy a gyerek képes párhuzamosan két modellt szem előtt tartani, képes különböző szituációk (múlt-jelen, lehetséges-valóságos) párhuzamos kezelésére. Perner a metareprezentációt ennél lényegesen bonyolultabb és a fejlődés során jóval később megjelenő reprezentációs módnak tekinti, amelyet ő is a „reprezentáció reprezentációja” terminussal definiál. Leslie elsődleges reprezentáció és metareprezentáció fogalma közé tehát beiktat egy közbülső szintet – a másodlagos reprezentációt –, és azt állítja, hogy a mintha-játékhoz nem szükséges metareprezentáció, csak másodlagos reprezentáció. Ahhoz, hogy megmagyarázzuk azt a jelenséget, hogy a banánnal úgy játszik a gyerek, mintha telefonkagyló lenne, nem szükséges feltenni, hogy a gyerek képes metareprezentációra. Elég a másodlagos reprezentációról beszélni, hiszen ez csak két különböző modellt, a valóságos és a lehetséges szituáció párhuzamos manipulálását igényli. A másodlagos reprezentáció a gyerek kétéves kora táján alakul ki, míg a metareprezentáció csak négyéves korára. Ezt támasztják alá azok a kísérletek is, amelyek szerint egy egészséges gyerek kétéves korában már képes a mintha-játékre, míg egészen négyéves koráig nem mutat olyan bonyolultabb elméműködésre utaló jeleket, amelyek a metareprezentáció meglétére engednek következtetni (ilyen például a hamis vélekedés teszt, ahol a gyerekeknek azzal kell számolni, hogy mit tud a másik). Leslie és Perner elmélete közötti különbségeket az alábbi táblázat foglalja össze.

Reprezentációs mód	Ingerek invarienciaosztálya	Ingerfüggetlen reprezentáció	Reprezentációk reprezentációja
A gyerek életkora	2 éves korig	2–4 éves korig	4 éves kor után
Leslie terminológiája	Elsődleges reprezentáció	M e t a r e p r e z e n t á c i ó	
Perner terminológiája	Elsődleges reprezentáció	Másodlagos reprezentáció	Metareprezentáció

Leslie és Perner elmélete a reprezentációs módokról

Perner elmélete tehát differenciáltabb és a fejlődéslélektan empirikus eredményeivel inkább összhangban álló képet ad, mint Leslie-é. Ráadásul evolúciós szempontból támadható az a Leslie elméletében implicit módon benne rejlő állítás, amely szerint a metareprezentációs képesség varázsütésszerű megjelenése jelentené a választóvonalat az állati és az emberi elme között, hiszen ekkor a metareprezentáció megjelenését egy lépésben kellene megmagyarázni. Ez azonban biológiailag nem túl plauzibilis magyarázatokhoz vezet, nehéz ugyanis azt az álláspontot védeni, hogy a metareprezentáció képessége az emberi genom egy véletlen mutációjának eredménye. Ehelyett sokkal meggyőzőbb egy többlépcsős modellt felállítani, amely kisebb, de biológiailag meggyőzőbb változások sorozatán keresztül jut el a metareprezentáció képességének kialakulásához.

Perner és Leslie – mint fejlődépszichológusok – nem foglalkoznak azzal a kérdéssel, hogy a különböző reprezentációs módok miképpen evolválódtak, de a Perner által adott osztályozás mindenképpen alkalmasabb arra, hogy egy evolúciós reprezentációelmélet kiindulópontja legyen. Ehhez azonban ezen a ponton be kell vezetnünk egy, a magasabb szintű reprezentációk magyarázata szempontjából kulcsfontosságú fogalmat, az elméleti fogalmát.

## 2. Az elméleti

Az utóbbi tíz év elmefilozófiai és kognitív tudományi vitáiban az egyik leglényegesebb terminus az elméleti (theory of mind) fogalma.<sup>220</sup> Az elméleti az a képesség, hogy másoknak vágyakat, illetve vélekedéseket, gondolatokat tulajdonítsunk. Az ember ugyanis képes arra, hogy döntéseiben ne csak azt vegye figyelembe, hogy

<sup>219</sup> Perner 1991.

<sup>220</sup> A „theory of mind” angol kifejezés szó szerinti fordítása elme-elmélet lenne, e szó azonban a magyarban nem igazán használható, ezért a magyar nyelvű szakirodalom két fordítással próbálkozik. Pléh Csaba könyvében a tudatelmélet (Pléh 1998b), Csányi Vilmos könyvében az elméleti szerepel mint a „theory of mind” fordítása (Csányi 1999). Én a magyarul rosszabbul hangzó elméleti kifejezést veszem át, egyrészt mivel az elmefilozófiai diskurzus számára a tudat, illetve az elme élesen elkülönítendő fogalom, másrészt egy később teendő diszinkció (elméleti-percepcióelméleti) csak e terminológiát felhasználva értelmes.

ő mit tud, hanem azt is, hogy egy másik individuum mit tud. Másképpen viselkedem, ha tudom, hogy a barátnőm tudja, hogy megcsalom, mint akkor, ha tudom, hogy nem tudja. Az elméletória jelensége teszi lehetővé például a becsapást vagy a hazugságot is.

Az elméletória képességének megjelenése a gyerek körülbelül négyéves korára tehető. A fejlődépszichológusok általában az úgynevezett Sally-Ann teszt segítségével szokták kimutatni, hogy egy gyerek rendelkezik-e ezzel a tulajdonsággal (az autisták például nem vagy sokkal kisebb mértékben rendelkeznek vele). A Sally-Ann tesztben a gyerek előtt van két doboz és két ember, Sally és Ann. Mind Sally, mind Ann látja, hogy a jobboldali dobozban van elrejtve valami (például egy darab csokoládé). Ezután Sally kimegy a szobából, Ann pedig a csokit az egyik dobozból átteszi a másikba, majd ismét bejön Sally, és megkérdezik a gyereket, hogy Sally melyik dobozhoz fog menni a csokiért. Négyéves koruk után a gyerekek probléma nélkül rámutatnak arra a dobozra, amelyben eredetileg volt a csoki, hiszen Sally nem tud arról, hogy Ann közben áttette azt a másik dobozba. Négy év alatt azonban azt várják, hogy Sally ahhoz a dobozhoz menjen, ahol a csoki valóban van, mivel nem tudják kezelni, hogy Sally mást tud, mint ők (például nem tudja, hol a csoki).

Az elméletória jelenségének elemzése kulcsfontosságú lett a kognitív tudományban, de az elmefilozófiában is.<sup>221</sup> A kérdés filozófiai aspektusai közül itt számunkra legfontosabb az a probléma, hogy az elméletória megjelenése milyen kapcsolatban áll a nyelv megjelenésével.<sup>222</sup> Logikailag négy lehetőség különíthető el:

1. Függetlenek.<sup>223</sup>
2. A nyelv előfeltétele az elméletóriának.<sup>224</sup>
3. Az elméletória előfeltétele a nyelvnek.<sup>225</sup>
4. A nyelv és az elméletória kölcsönösen hívták életre egymást, koevolválódtak.<sup>226</sup>

Anélkül hogy megpróbálnám eldönteni ezt a kérdést, érdemes kitérni a primatológia, illetve a kognitív etológia néhány eredményére, amelyek arra vonatkoznak, hogy rendelkeznek-e a főemlősök elméletóriával. Ha rendelkeznek, akkor a fenti 2. és a 4. pont erősen támadhatóvá válik, hiszen a főemlősök nem rendelkeznek nyelvvel, és mégis rendelkeznek elméletóriával, ennyiben ellenpéldát jelentenek mind a koevolúciót, mind a nyelv prioritását hangsúlyozó elméletekkel szemben.

Igen régi vita, hogy vajon az emberek kívül más állat is rendelkezik-e az elméletória képességével, vagy sem. Premack és Woodruff<sup>227</sup> (a „theory of mind” terminus megalkotói) mellett érveltek, hogy a csimpánz például képes, később Cheney és Seyfarth azt próbálták kimutatni, hogy bár a cercopitheciformok nem képesek erre, de a csimpánzok igen.<sup>228</sup> Tomasello és Call, illetve Heyes viszont amellel érvel, hogy nincs okunk elméletóriát feltételezni a főemlősöknél.<sup>229</sup> Érdemes leírni néhány olyan megfigyelést, amelyre ezek az érvek támaszkodnak.

Az elméletória legjellemzőbb példái a primátakutatásban a becsapás jelenségei.<sup>230</sup> Az egyik leghíresebb megfigyelés szerint két verekedő cercopitheciform közül az egyik igen rosszul állt, de a már majdnem megvert cercopitheciform ekkor egy leopárdvészjelkiáltást hallatott (holott a közelben sehol nem volt leopárd), amire a másik (az erősebb) cercopitheciform felmenekült a fára, abbahagyva a verekedést. Ez leginkább úgy magyarázható, hogy a gyengébb cercopitheciform számolt azzal, hogy a másik állat tudja, hogy a vészjel azt jelenti, hogy leopárd van a közelben, ezért fel fog menekülni a fára, abbahagyva a verekedést. Páviánoknál figyelték meg azt, hogy egy fiatalabb pávián a domináns hím háremének egyik tagjával kopulált (amit, ha meglát a domináns hím, akkor megveri a fiatal hímét), de – hogy elkerülje a verést – mindezt egy szikla mögött tette, mégpedig úgy, hogy ő maga látszott, de a nőtényt, akivel kopulált, eltakarta a szikla a domináns hím szeme elől. A páviáncsoportokban ha a domináns hím meglátja, hogy egy fiatalabb hímnek merev a pénisze, akkor megveri az illető hímét, hiszen csak neki van joga a csoporthoz tartozó nőtényeket megtermékenyíteni. Többször is megfigyelték, hogy az élelmesebb fiatal páviánok a domináns hím jelenlétében eltakarják merev péniszüket, ami szintén az elméletóriát támasztaná alá, hiszen a másik egyed gondolatait is szem előtt tartják viselkedésükben. Hasonló, bár kicsit még bonyolultabb eset volt, amelyben egy olyan ételosztó helyet állítottak fel a kutatók, amelyből a domináns hím csimpánz nem kaphatott enni, de mindenki más igen. Egy idő után a fiatal csimpánzok megkapták a banánt, de a domináns hím rögtön elvette tőlük. Ezek után csak akkor mentek az ételosztó helyre a fiatal csimpánzok, ha a domináns hím nem volt a közelben. Ez már önmagában is bizonyíték lehetne az elméletória létezésére mellett, de a domináns hím még tovább ment, elbújt, amikor látta, hogy egy fiatal hím jön, és – mivel nem látja őt – elveszi a banánt, majd természetesen előbújt, és elvette tőle.

<sup>221</sup> Cf. Carruthers–Smith 1996, Davies–Stone 1995.

<sup>222</sup> Egy másik sokat tárgyalt, filozófiai is érdekes probléma az elméletória fogalmával kapcsolatban a szimulációelmélet–elméletelmélet vita, amely azt a kérdést vizsgálja, hogy milyen kapcsolat van az elméletória és a saját gondolat között, tehát például a másoknak tulajdonított borosüveg-reprezentáció és saját borosüveg-reprezentáció között. Cf. Stone–Davies 1995, Davies–Stone 1995, Carruthers–Smith 1996, 11–139.

<sup>223</sup> Cf. pl. Millikan 1984

<sup>224</sup> Cf. pl. Bickerton 1990, Dennett 1991

<sup>225</sup> Cf. pl. Sperber–Origgi 1998

<sup>226</sup> Cf. pl. Pinker 1999

<sup>227</sup> Premack–Woodruff, 1978.

<sup>228</sup> Cheney–Seyfarth 1990. Cheney–Seyfarth 1995.

<sup>229</sup> Tomasello–Call 1997. Heyes 1994.

<sup>230</sup> Cf. Byrne–Withen 1988, Withen–Byrne 1997.

A leghíresebb kísérlet azonban kétségkívül az, amelyben rövid videófelveteleket vetítettek a csimpánzoknak, amelyeken egy ember valamilyen problémával küszködött, például vacogva ült egy, a konnektorba nem bedugott radiátor mellett, vagy ágaskodott egy számára túl magasan levő szőgre akasztott kalap után. Az egyes klipek vetítése után tíz állóképet mutattak az állatnak, amelyek közül probléma nélkül kiválasztotta azt, amely a klipen látható probléma megoldását jelenti, egy bedugott konnektor képét vagy egy létra képét.

Szintén kísérleti szituációban figyelték meg, hogy egy csimpánz előtt két doboz volt, ő tudta, melyikben van étel, melyikben nem. Két ápoló jött be, megkérdezték, melyikben van az étel, majd az egyik megosztotta azt a csimpánzzal, a másik pedig nem. Ezután a csimpánz a barátságos ápolónak továbbra is megmutatta, melyik dobozban van az étel, az önző ápolót azonban mindig megpróbálta félrevezetni, az üres dobozt mutatta neki, holott tudta, hogy abban nincs étel.

Ha ezek a kísérletek, illetve megfigyelések valóban alátámasztják, hogy a primáták rendelkeznek elmeteóriával, akkor az elmeteóriának triviálisan nem előfeltétele a nyelv. Kérdés azonban, hogy valóban alátámasztják-e. Két oldalról szokták támadni ezt a tézist.<sup>231</sup> Egyrészt azt emelik ki, hogy a legerősebb evidenciaként szolgáló megfigyelések anekdotikusak, nem reprodukálható eredmények. Másrészt azt állítják, hogy az itt leírt és az ezekhez hasonló jelenségek magyarázhatók egyszerű kondicionálásos tanulással is: a fiatal himet mindig megverte a domináns hím, amikor a szeme láttára kopulált az egyik nőténnyel, következőképpen legközelebb megpróbálja elkerülni ezt a konfliktust, és úgy kopulál, hogy a domináns hím ne lássa.

Anélkül hogy eldöntenénk ezt az empirikus problémát, filozófiai szempontból kiemelendő, hogy az utóbbi években egyre több olyan elmélet születik, amely nem azt az eldöntendő kérdést vizsgálja, hogy rendelkeznek-e a primáták elmeteóriával, hanem azt, hogy az elmeteória milyen fokozatával rendelkeznek. Evolúciós szempontból sem igazán tartható az az álláspont, amely szerint a főemlősök elméjéből teljesen hiányzik ez a képesség, majd az emberré válás során hirtelen teljes egészében jelen lesz. Ezzel szemben érdemes különböző fokozatokat elkülöníteni az elmeteórián belül. Erre tett kísérletet Andrew Withen, aki megkülönbözteti a viselkedés-attribúciót, valamint az elme-attribúciót, és e két jelenség között graduális átmenetet tételez fel.<sup>232</sup> Az első esetben az állat azt veszi figyelembe, hogy a másik individuum miként szokott viselkedni, a második esetben azonban azt is, hogy mi van a fejében.

A primátakutatás területén kívül, de Withen distinkciójához hasonló koncepciót vázol a fejlődépszichológus Gergely György is. Gergely – akárcsak Leslie és Perner – a magasabb szintű reprezentációk egyedfejlődését, ontogenetikus kialakulását vizsgálja, és nem a filogenetikus kialakulást, mint a primátakutatók. Gergely kutatásainak középpontjában az elmeteória kialakulása áll. Láttuk korábban, hogy a gyerek körülbelül négyéves korára sajátítja el ezt a képességet. Gergely egy híres kísérlete azonban megmutatta, hogy már az egyéves gyerek is bizonyos elvárásokkal viszonyul a környezetében feltűnő cselekvő figurákhoz. A kísérlet azt mutatta ki, hogy az egyéves gyerek racionális viselkedést vár el a másik individuumtól (amelyet e kísérletben a számítógép képernyőjén megjelenő kör testesít meg).<sup>233</sup> Anélkül hogy a kísérletet részletesen ismertetnénk, a racionális viselkedés itt azt jelenti, hogy a viselkedés célorientált, a cselekvő a legegyszerűbb utat választja célja eléréséhez.

Gergely azonban amellett érvel, hogy a racionális, célorientált viselkedés tulajdonítása még nem elmeteória, hiszen nem feltétlenül szükséges, hogy a gyerek mentális tartalmakat tulajdonítson valakinek. Sokkal kézenfekvőbb egy „teleológiai magyarázatot” adni: a gyerek nem azt veszi tekintetbe, hogy a cselekvő mit tud vagy mit akar, hanem azt, hogy miként viselkedik: a viselkedés az elért (vagy elérendő) végállapot felől – a cél felől – értelmeződik. Gergely állítása az, hogy ez az egyéves korra kialakult teleológiai hozzáállás bár még nem elmeteória, de az elmeteória kialakulásának közvetlen előzménye és előfeltétele.<sup>234</sup> A teleológiai magyarázatban a másik individuum céljait veszi figyelembe a gyerek, de ezek a célok még nem mentális állapotok, hanem cselekvésének végállapotai. Vegyük észre, hogy Gergely „teleológiai magyarázat” fogalma a Perner féle másodlagos reprezentáció képességével ekvivalens: a gyerek egyszerre tartja szem előtt a jelenlegi és a jövőbeli állapotot (célállapotot). Gergely kísérleteinek egy értelmezése szerint tehát a másodlagos reprezentáció képessége a fejlődéslélektanban az elmeteória előfeltétele.

Érdemes még egy elmélet bemutatásának erejéig visszatérni a primatológia területéhez, tehát a gyerekek helyett ismét a majmokhoz. Andrew Withen az imént bemutatott distinkciójánál még használhatóbb Juan-Carlos Gómez elmélete, aki szintén az elmeteória viselkedési előzményeit keresi. Gómez megkülönbözteti a látható mentális jelenségek tulajdonítását, illetve a rejtett mentális jelenségek tulajdonítását (ez utóbbi lenne ekvivalens az elmeteóriával).<sup>235</sup> A látható mentális állapotok (például hova néz a másik állat) sokkal inkább befolyásolják az állat döntését, mint a semmilyen külső jegy által nem kísért rejtett mentális állapotok. Gómez szerint a látható mentális állapotok tulajdonítása jól bizonyítható a primáták esetében, míg a rejtett mentális állapotok tulajdonítása kevésbé valószínű.

<sup>231</sup> Heyes 1997 jó összefoglalást ad erről.

<sup>232</sup> Withen 1996.

<sup>233</sup> Gergely et al. 1995.

<sup>234</sup> Gergely–Csibra 1998.

<sup>235</sup> Gómez 1996.

Gómez fogalmi kettősségét érdemes továbbfejleszteni és kiterjeszteni. Nevezzük továbbra is elmeteóriának azt a képességet, hogy vélekedéseket, gondolatokat tulajdonítunk egy másik egyednek, tehát azzal kalkulálunk, hogy a másik mit tud. Ez felelne meg Gómez „rejtett mentális állapotok tulajdonítása” terminusának. Az elmeteóriától érdemes megkülönböztetni és bevezetni a „percepcióteória” fogalmát, amely azt a képességet jelenti, hogy percepciót tulajdonítunk egy másik egyednek, tehát azzal kalkulálunk, hogy a másik mit lát. Belehelyezkedünk a másik nézőpontjába. A percepcióteória fogalma Gómez „látható mentális állapot tulajdonítása” terminusával hozható átfedésbe. Viselkedésemet az aktuális ingerek mellett az is befolyásolja, hogy egy másik egyed adott ingerek hatására miként viselkedik. Igen fontos, hogy nem az van hatással az én viselkedésemre, hogy a másik mit tud vagy akar, hanem az, hogy mit lát.

Látható, hogy a percepcióteória sokkal gyengébb kritérium, mint az elmeteória. Továbbá a fent idézett példák mindegyikében voltaképpen az állat azt vette figyelembe, hogy a másik egyed mit lát (vagy mit nem), nem pedig azt, hogy mit tud. A leopárdvészjelet leadó cercóf csak azzal kalkulált, hogy a másik egyed mit fog hallani, és erre hogyan fog reagálni. A kopuláló fiatal hím pedig csak az érdeklődött, hogy látja-e a domináns hím, hogy mi van a fejében, az nem releváns számára. Tehát a fenti példákból csak a percepcióteória meglétére következtethetünk, az elmeteóriára nem.

Érdemes ezen a ponton röviden kitérni az elmeteória és a humán csoportszerkezet közötti kapcsolatra. Az elmeteória nyilvánvalóan csak olyan állatok esetében adaptív, amelyek csoportban élnek, ahol tehát igen komoly szelektív előnyt jelent, ha képes vagyok előrelátni a csoporttársam viselkedését. Az elmeteória és az állati csoport mérete közötti összefüggést mutatta ki Robin Dunbar is, aki szerint egyenes arányosság van a csoportméret és a taktikai becsapások esetének gyakorisága között a különböző primáta fajok esetében.<sup>236</sup> Valószínűleg az emberi elme evolúciója esetében sem lehet elvonatkoztatni attól, hogy milyen csoportokat alkottak őseink.<sup>237</sup> Az emberi elme komplexitásának titka talán éppen az, hogy az ember minden primátánál szorosabb csoportkötelékben élt, és emiatt rendkívül fejlett elmeteóriára volt szüksége, amely aztán a nyelv kialakulásához vezetett.

Ha viszont igaz, hogy az emberré válás igen szoros csoportkötelékekben ment végbe, akkor az emberi elme megváltozását nem egy önmagában álló folyamatnak kell elképzelni, hanem egy szoros csoport keretein belül végbemenő változásnak. Az egyéni túlélés mellett a csoport túlélése is fontos volt. Az egyedek egymással való kapcsolata, az, hogy mennyire tudják megjósolni egymás viselkedését fontosabb volt, mint az egyéni siker. Az elmeteória tehát az egyik legfontosabb adaptív tulajdonság a csoport szempontjából.

Következésképpen valószínű, hogy az elmeteória nem az elme inherens, a szociális kapcsolatoktól független változásának eredménye. Ha a szelekció szempontjából elsősorban a csoport integritása és a csoporton belüli kohézió volt releváns, akkor a szelekciós nyomás azoknak a mentális változásoknak kedvezett, amelyek (mint az elmeteória) a másik csoporttag viselkedését is figyelembe tudták venni. A magasabb szintű elméműködés kialakulásában tehát nem elhanyagolhatók a csoport hatékony működését biztosító mentális képességek (így az elmeteória) kialakulása. Hiba volna tehát egy individualista elmémódelből kiindulva az elmeteóriát származtatott tulajdonságnak tekinteni. Sokkal valószínűbb, hogy az elmeteória megjelenése tette lehetővé az elme azon inherens változásait, amelyek például a metareprezentáció vagy a nyelv kialakulásához vezettek.

Mind Leslie, mind Perner származtatott tulajdonságnak tekintik az elmeteóriát. Leslie szerint az elmeteóriának előfeltétele a metareprezentáció, Perner – hasonlóan az ingerfüggetlen reprezentációk kérdéséhez – itt is egy differenciáltabb koncepciót képvisel, mint Leslie. Perner szerint a másodlagos reprezentáció önmagában nem elég az elmeteóriához, de a másodlagos reprezentáció másik individuumra alkalmazása teremti meg az elmeteória még kialakulatlan formájának, a „mentalisztikus viselkedéstulajdonításnak” a lehetőségét. Az elmeteóriát pedig csak a metareprezentáció másik individuumra alkalmazása teszi lehetővé. Látható, hogy Perner is – Leslie-hez hasonlóan – az elméműködésben szolipszista módon végbement változásból, egy új reprezentációs mód megjelenéséből eredezteti az elmeteória kialakulását. Ha azonban elfogadjuk, hogy a humán csoport integritása elsődleges fontosságú tényező volt az emberi elme kialakulásában, akkor fordított utat kell járnunk.

A következőkben megpróbálok egy olyan modellt körvonalazni, amelyben a csoportszintű változások megelőzik az individuális változásokat, és amely képes a másodlagos reprezentáció–metareprezentáció distinkció, illetve a percepcióteória–elmeteória distinkció egy elméleti keretbe foglalására. Egy ötszintű modellt szeretnék javasolni, amelyben az elméműködés újabb, egyre magasabb szintjeinek megjelenésére biológiailag meggyőző magyarázatot lehet adni.

Az ötlépcsős modell legelső szintje az elsődleges reprezentáció, amely ingerek invarianciaosztályával azonos. Innen kiindulva az első lépés a percepcióteória, a másik egyed nézőpontjába való behelyezkedés megjelenése, amely a csoport szempontjából adaptív folyamat, hiszen a csoport szempontjából előnyös, ha az egyes egyed a másikat érő ingereket is figyelembe veszi.

A második lépés, a másodlagos reprezentáció megjelenése exaptációs folyamat. Miután megtanultam, hogy a másik egyedet érő ingereket figyelembe vegyem, azaz hogy behelyezkedjek a másik egyed nézőpontjába, képes leszek párhuzamosan különböző modelleket manipulálni, hiszen a percepcióteória is a saját nézőpontom és a másik egyed (akinek a nézőpontjába behelyezkedem) nézőpontjának szétválasztását előfeltételezi. Ha egy cso-

<sup>236</sup> Dunbar 1996.

<sup>237</sup> A humán csoportról és annak az ember evolúciójában betöltött szerepéről részletesebben lesz szó a következő fejezetben.

porttársamat megtámadja egy leopárd, akkor nem én ugrok hátra, nem hiszem azt, hogy engem fenyeget veszély, de mégis világosan érzékelem, hogy a csoport egy másik tagja veszélyben van.

Ennek a lépésnek tehát nincs közvetlen adaptív haszna. A percepcióteória megjelenésének mellékterméke. Másként fogalmazva a percepcióteóriát most nem egy másik egyedre, hanem saját magamra alkalmazom: viselkedésemet az is befolyásolja, hogy az aktuálistól különböző ingerek hatására miként viselkednék. Itt a hangsúly a kontrafaktuáliston van: miként *viselkednék*. Ezen a szinten jelenik meg a mintha-játék, a múlt „lefűződése”, függetlenedése a jelentől, és a reprezentáció függetlenedése az aktuális ingerektől.

A harmadik lépés ismét adaptív: az elmeteória megjelenése. Itt viselkedésemet az eddigieken túl az is befolyásolni fogja, hogy a másik egyednek milyen reprezentációi vannak. Nemcsak azt veszem figyelembe, hogy most milyen ingerek érik a másikat, hanem azt is, hogy az aktuálistól különböző ingerek hatására miként viselkedne. Ez a lépés egyértelműen adaptív a csoport szempontjából, hiszen jobban meg tudom jósolni a másik egyed viselkedését, jobban össze tudom hangolni saját viselkedésemet az övével, ha nemcsak a másikat érő ingereket, hanem reprezentációit is figyelembe veszem. Ez a lépés nem elképzelhető a másodlagos reprezentáció képessége nélkül, hiszen ahhoz, hogy figyelembe vegyem, hogy a másik mit tud (esetleg nem tudja, amit én már tudok etc.), párhuzamosan kell kezelnem két modellt, azt, amit én tudok a világról, és azt, amit a másik (és e kettő csak ritkán esik egybe). Megjegyzendő, hogy ennek a lépésnek a fejlődéslélektani vetületét írja le Gergely György elmélete, amelyben a mentális oksági magyarázatok teleológiai gyökeréről (tehát az elmeteória másodlagos reprezentációs előfeltételéről) beszél.

Az negyedik és utolsó lépés, amely az ötödik szintre vezet, a metareprezentáció megjelenése. Ez megint exaptív folyamat, nem más, mint az elmeteória alkalmazása saját magamra. Most nem a másik egyednek tulajdonítok vélekedéseket, gondolatokat, hanem saját magamnak. Hangsúlyozni kell, hogy ez nem más, mint az elmeteória mellékterméke, hiszen ha képes vagyok másoknak vélekedéseket, gondolatokat tulajdonítani, azaz mások reprezentációiról gondolkodni (mások reprezentációit reprezentálni), akkor képes leszek saját reprezentációimról is gondolkodni, azokat manipulálni, képes leszek tehát reprezentációim reprezentációjára, a metareprezentációra. Összefoglalva:

#### *1. Elsődleges reprezentáció*

Aktuális ingerek invarianciaosztálya.

Kialakulása adaptív az egyed számára.

#### *2. Percepcióteória (Theory of Perception, másik egyed nézőpontjába belehelyezkedés)*

Viselkedésemet az aktuális ingerek mellett az is befolyásolja, hogy egy másik organizmus adott ingerek hatására miként viselkedik. Fontos, hogy nem az van hatással az én viselkedésemre, hogy a másik mit tud vagy akar, hanem az, hogy mit lát.

Kialakulása adaptív folyamat a csoport számára: a csoport szempontjából előnyös, ha az egyed a csoport más tagjait érő ingereket is figyelembe veszi.

#### *3. Másodlagos reprezentáció (kontrafaktuális elmeműködés)*

Viselkedésemet az aktuális ingerek mellett az is befolyásolja, hogy az aktuálistól különböző ingerek hatására miként viselkednék. Itt jelenik meg a „mintha”-játék, valamint a különböző szituációk (múlt-jelen, lehetséges-valóságos) párhuzamos kezelése. Ez a reprezentációs mód lefűződik a közvetlen perceptuális ingerekről.

Kialakulása exaptív folyamat: a percepcióteória alkalmazása saját magamra.

#### *4. Elmeteória (Theory of Mind)*

Reprezentáció tulajdonítása más egyedeknek, viselkedésemet az eddigiek mellett az is befolyásolja, hogy a másik egyed az aktuálistól különböző ingerek hatására miként viselkedne, tehát hogy a másik egyednek milyen reprezentációi vannak.

Kialakulása adaptív folyamat a csoport számára: a csoport szempontjából előnyös, ha az egyed nemcsak a csoport más tagjait érő ingereket veszi figyelembe, hanem a csoport más tagjainak kontrafaktuális elmeműködését is.

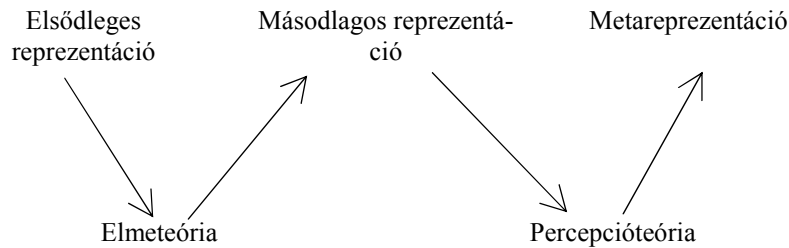
#### *5. Metareprezentáció*

A reprezentáció reprezentációja, címkézhető, manipulálható, tetszés szerint kombinálható szimbolikus reprezentáció, a nyelvi reprezentáció közvetlen előfeltétele.

Kialakulása exaptív folyamat: az elmeteória saját magamra alkalmazása.

Az egyes lépések tehát cikkcakkban mozognak. Egy új, a csoport szempontjából hasznos tulajdonság megjelenése után azt magamra alkalmazom, majd az ezáltal bonyolultabbá váló reprezentációs módot megint a csoportra, majd megint magamra, és így tovább.

### A magasabb szintű reprezentációk kialakulásának ötszintes modellje



Fontos észrevenni, hogy a harmadik és az ötödik lépés, tehát a metareprezentáció, illetve a másodlagos reprezentáció megjelenése exaptív folyamat: egy olyan képességet, amely eddig mások értelmezésére szolgált, magamra alkalmazok. Ez a lépés viszont azt jelenti, hogy viszonylagos átjárás van az elme egyes képességei, moduljai között.

Ezen a ponton az itt körvonalazott modell párhuzamos eredményeket mutat Steve Mithen már bemutatott elméletével.<sup>238</sup> Akárcsak Mithennél, itt is kulcsszerepe van az elme egyes területei közötti kommunikációnak, hiszen az elmeteória magamra alkalmazása és így a metareprezentáció kialakulása sem volna lehetséges, ha az egyik mentális képesség nem hathatna a másikra. E modell szerint azonban – és ez lényeges különbség Mithenhez képest – a nyelv megjelenése nem mindezen változások oka, hanem egy késői következménye, amelynek magyarázata az elme vizsgálatának egyik legbonyolultabb kérdése.

## 3. A nyelv evolúciója

A metareprezentáció megjelenését tehát lépésről lépésre magyaráztuk, egy lépés azonban még hátravan, a nyelvi reprezentáció megjelenése.

1866-ban a Société Linguistique de Paris betiltotta az emberi nyelv eredetét vizsgáló tanulmányok közreadását, mivel a kérdést reménytelenül bonyolultnak tartotta, amelynek vizsgálata feleslegesen elfecsérelt idő. Azóta mégis igen sok és sokféle elméletet adtak a nyelv evolúciójának problémájára. Anélkül hogy részletesen tárgyalnánk a nyelv evolúciójának elméleteit, érdemes röviden áttekinteni néhány jellegzetes megoldási javaslatot, amelyek az utóbbi években a nyelvevolúció vizsgálatának néhány lehetséges irányát képviselik.<sup>239</sup>

Négy elméletet fogok itt érinteni, Robin Dunbar, Richard Dawkins, Derek Bickerton és Steven Pinker elméletét. Mindenekelőtt érdemes megkülönböztetni a többtől az első két teoretikus elméletét, amelyek a szintaxis, a nyelvtan evolúciójának magyarázatára nem tesznek kísérletet. Ezen belül Dunbar a csoportból indul ki, míg Dawkins az individuumból.

Robin Dunbar híres nyelvi kurkászás-elmélete szerint a nyelv eredetét voltaképpen az magyarázza, hogy az emberi csoport mérete megnőtt (ami viszont annak következménye, hogy a szavannán több ragadozó volt, mint az őserdőben, így biztonságosabb volt nagyobb csoportokban élni).<sup>240</sup> A primátacsoportok fennmaradása szempontjából kulcsfontosságú kérdés a csoport integritása és az agresszió visszaszorítása, amelyet a legtöbb főemlős kurkászással ér el. A megnövekedett csoportméret miatt azonban a humán csoport tagjai nem tudtak elegendő időt fordítani a kurkászásra (mert ekkor nem lett volna idejük ételmezt szerezni), túl sok volt a kurkászandó csoporttárs, így az agressziót valamilyen más módszerrel kellett visszaszorítani. A nyelv azért volt adaptív az emberi csoport számára, mivel segítségével kevesebb idő alatt több emberrel lehet tartani a szociális kontaktust. Kurkászni csak egy embert lehet, beszélgetni azonban többel is lehetséges, így sokkal hatékonyabban lehetett ápolni a csoporton belüli kapcsolatokat.

Richard Dawkins szerint ezzel szemben nem a csoport érdeke hívta életre a nyelv megjelenését, hanem az egyén érdeke.<sup>241</sup> A nyelv nem annak az egyednek a számára adaptív, aki hallgatja a beszédet, hanem annak, aki beszél, hiszen ő manipulálni tudja a hallgatót. Dunbar elméletével szemben itt egy igen önző emberkép van középpontban: a nyelv azért volt adaptív, mert lehetővé tette, hogy hatékonyabban lehessen becsapni a másik egyedet. Sem Dunbar, sem Dawkins koncepciója nem ad azonban választ arra, hogy miért volt szükség egy bonyolult nyelvtannal rendelkező nyelv evolúciójára, hiszen sem a nyelvi kurkászáshoz nem szükséges szintaktikailag

<sup>238</sup> Mithen 1996.

<sup>239</sup> Cf. pl. Deacon 1997, Lieberman 1984, Lieberman 1998, jó összefoglalás a Hurford–Studdert–Kennedy–Knight 1998. szöveggyűjtemény.

<sup>240</sup> Dunbar 1996.

<sup>241</sup> Dawkins–Krebs 1978.

jólformált nyelv, sem a másik becsapásához; a nyelvtanilag nem formált odavetett szavak is elegendőek lettek volna.

Derek Bickerton komolyan veszi a szintaxis evolúciójának kérdését, szerinte a nyelvtani struktúra nélküli protonyelv kialakulását (amelyre egyébként a csimpánzok is megtaníthatók) a közös feladatok koordinálásának szükségessége világosan magyarázza.<sup>242</sup> Ahhoz, hogy egy mamutot el tudjon fogni egy csoport, a részfeladatokat meg kellett beszélni, ehhez azonban valamilyen ősnyelv volt szükséges: „Te kergeted, én kiugrom a bokorból.” Ennek az evolúciója a kisebbik probléma. A nyelvtan megjelenése az igazán nagy lépés, ez különbözteti meg az embert az állattól, ez a lépés pedig egy véletlen mutációval magyarázható. Bickerton tehát látja a nyelvtanilag jólformált nyelv evolúciójának problémáját, de a mutációra való hivatkozással nem igazán válaszolja meg, inkább a szőnyeg alá söpri a kérdést.

Pinker ennél plauzibilisebb koncepciót vázol: szerinte a nyelvtan azért adaptív, mert hatékonyabb udvarlást tesz lehetővé.<sup>243</sup> Pinker azt állítja, hogy azok a hímek, akik csak annyit mondtak a nősténynek, hogy „gyere!” szelektív hátrányba kerültek azokkal szemben, akik valahogy úgy fogalmaztak, hogy „arra gondoltam, hogy esetleg nem bánná-e, ha megnéznénk ezt a szép majomkenyérfát”.

Fontos rámutatni, hogy mindegyik itt bemutatott elmélet egy alapvető változásra próbálja visszavezetni a nyelv kialakulását. Egy apró változás (akár a csoportszerkezetben, akár a genomban) a nyelv megjelenéséhez vezetett. Ezzel szemben biológiailag plauzibilisebb modellt jelentene, ha sikerülne több apró, egymást követő fokozatos, graduális változások sorából felépíteni a nyelv megjelenéséhez vezető evolúciós folyamatot. Minden olyan elmélet ugyanis, amely egy princípiummal magyarázza a nyelv megjelenését, implicite egy, az embert az állattól mereven elválasztó kategorizációt foglal magában. Ha viszont komolyan vesszük, hogy az emberi és az állati elme között folytonosság van, akkor meg kell mutatni azokat a kisebb lépéseket, amelyek segítségével az állati elméből fokról fokra kialakult az emberi elme.

Az itt vázolt ötlépcsős modell erre tesz kísérletet. Fontos azonban megjegyezni, hogy a metareprezentáció önmagában még nem feltétlenül nyelvi reprezentáció, de mindenképpen szabálykövető mentális feldolgozást lehetővé tevő, manipulálható, címkézhető reprezentációs rendszert képez. A címkézhetőség biztosítja, hogy a nyelv szavait és e metareprezentációkat könnyen össze lehet kapcsolni. Emellett a metareprezentáció a szükséges behelyettesíthető alapelemeket, a manipulálható szimbólumokat is biztosítja a szabálykövetés mentális folyamatához. Ez még nem jelent ugyan nyelvi reprezentációt, de talán tartalmazza a nyelvi reprezentáció és a szabálykövetés (illetve a szintaxis) megjelenésének minden előfeltételét. A reprezentáció reprezentációjaként felfogott metareprezentációtól a nyelv kialakulásáig eljutni mindenesetre ugyanolyan viszonylag kis lépés, mint az eddigiek voltak.

Az előző fejezetekben kifejtett, mind az emberre, mind az állatra jellemző individualista, a szociális szférát nem hangsúlyozó modellekhez képest az e fejezetben vázolt, csak az emberre jellemző evolúciós folyamat igen erőteljesen összefonódik az ember szociális környezetével. A reprezentációk e fejezetben kifejtett egyre komplexebbé válásának vizsgálata nem elképzelhető egy individualista fogalmi kereten belül. A csak az emberre jellemző metareprezentációs, illetve nyelvi reprezentációs képességek és szabályok éppen amiatt alakultak ki, mert az ember társas lény.

<sup>242</sup> Bickerton 1981, Bickerton, 1990.

<sup>243</sup> Pinker 1999, Pinker–Bloom 1990.

## VIII. A kulturális evolúció

Az emberi elme evolúciós megközelítésének talán legkényesebb pontja a kultúra kérdése. Egyrészt a kultúra – bármit jelentsen is pontosan – úgy tűnik csak az emberre jellemző jelenség, és merev választóvonalat képez az ember és az állat között. Másrészt az a tény, hogy az ember igen bonyolult kultúrával rendelkezik, megnehezíti az evolúcióbíológiai fogalmak elmére való közvetlen alkalmazását. Nehéz amellet érvelni például, hogy evolúciós szempontból bármilyen szerepet is játszottak volna Rilke szonettjei vagy Beethoven késői vonósnygyesei. A kultúra szférája tehát minden evolúciós elmélet legkényesebb területei közé tartozik. Ráadásul a kultúra fogalmának definiálásában távolról sincs konszenzus, szinte mindenki mást és mást ért e fogalmon.

Tovább bonyolítja a helyzetet, hogy a kultúra működésére többen is evolúciós ihletésű elméletet dolgoztak ki, amely szerint a kulturális egységek szelektálódnak. Az evolúció és az elme vizsgálatának ez a harmadik nagy területe, az elmen belüli evolúció és az elme törzsfjldésének kérdése mellett. Ráadásul a kultúra vizsgálata nem választható el az elme vizsgálatától már csak azért sem, mert a kulturális változások gyakran visszahatnak az elmére, a külső információtárolás, tehát az írás feltalálása például az emlékezési technikákat igencsak átalakította.<sup>244</sup> E fejezetben a kultúra és az evolúciós elméletek viszonyáról, s ezen belül is a kulturális evolúció problémájáról lesz szó.

Mindenekelőtt le kell szögezni, hogy hamis az a szembeállítás, mely szerint a kultúra kizárólag az emberre jellemző jelenség, amely az állatok világában teljesen ismeretlen. Az utóbbi évek egyik legizgalmasabb kutatási területe az etológiában éppen a nem emberi kultúrák létezésének kérdése.<sup>245</sup> A főemlősök társas viselkedésében ugyanis sok olyan elemet találunk, amely az emberi kultúrára is jellemző. Nem lehet tehát megkerülni azt a kérdést, hogy miként viszonyul az emberi kultúra a majmok, ezen belül is főleg a csimpánzok társas viselkedéséhez.

Ha hűek akarunk lenni ahhoz az alapelvhez, hogy az evolúciót folytonos, graduális átmenetek, és nem nagy ugrások sorozatának tekintjük, akkor nem tartható egy olyan álláspont, amely szerint kultúrához hasonló jelenség még a legfejlettebb emberszabásúak esetében sem létezik, az embernél viszont hirtelen, varázsütésre megjelenik. Másfelől nyilvánvaló különbségek léteznek a csimpánz társas viselkedése és az ember kultúrája között.

Érdeemes tehát a kultúra fogalmát úgy meghatározni, hogy valamilyen előképe és korai formája a főemlősök-nél is azonosítható legyen. A csimpánzok is használnak tárgyakat, híres megfigyelés például, hogy a pálmadió feltöréséhez – ami igen nehéz feladat – a csimpánzok támasztékot használnak. A diót egy lapos köre teszik, és egy másik kövel próbálnak úgy ráütöni, hogy a kemény héj eltörjön. A talaj azonban nem mindig vízszintes, ezért a lapos követ gyakran alá kell támasztani egy harmadik kövel, hogy a pálmadió le ne guruljon róla. Ez az eszközhasználat – amely lényegesen bonyolultabb eset, mint például a sokat emlegetett természetfalászathoz használt kis bot – nem öröklött, hanem tanult: generációról generációra adódik át, a kölykök az anyjuktól tanulják.<sup>246</sup>

Egy másik híres, a csimpánzkultúrák létezése mellett gyakran felhozott példa a majmok öngyógyítási viselkedése. A csimpánzok viszonylag sok olyan növényt ismernek, amelyet egészségesen nem fogyasztanak, csak akkor esznek belőlük, ha valamilyen betegséggel küszködnek.<sup>247</sup> Azért különösen figyelemreméltó ez a jelenség, mert több gyógynövényt is ismernek, és mindegyikről tudják, milyen bajra jó, nem tévesztik össze őket. A kultúra szempontjából azért megvilágító erejű a gyógynövények példája, mert a patika intézménye és a csimpánzok viselkedése közötti kapcsolat hozzávetőlegesen jellemzi, milyen értelemben lehet a csimpánzok kultúráját az emberi kultúra előképének feltételezni. Voltaképpen az emberi kultúra legbonyolultabb teljesítményei is az igen egyszerű állati kultúrák továbbfejlesztett és továbbfinomult változatai.

### 1. A kultúra és a gének

Az eddigi fejezetekben az emberi elme önmagában vett működéséről volt szó. A kultúra szférája azonban az egyes elmék közötti kapcsolatokról, hatásokról, illetve ezek intézményesült formáiból áll. Kérdés, hogy lehet-e ezekre is evolúciós modellt adni. Láttuk, hogy az egyéni emberi elme igen jól vizsgálható az evolúció elméletének segítségével, ebből azonban egyáltalán nem következik, hogy az elmék közötti kapcsolatokat – tehát a kulturális jelenségeket – is az evolúció segítségével kellene tanulmányozni.

A kultúra nem evolúciós mechanizmusokon alapuló modellje mellett érvel Dan Sperber, aki szerint a kultúra és az elme vizsgálata úgy viszonyul egymáshoz, mint a járványtan az orvostudományhoz.<sup>248</sup> A járványtan, amely a betegségek és a kórokozók továbbadását kutatja, egészen más elméleti konstrukciókat használ, mint az orvostudomány, amely egy szervezeten belül vizsgálja a betegségeket és a kórokozókat. Hasonlóképpen, a kultúra ta-

<sup>244</sup> Donald 1991. A kérdés ma nagyon divatos filozófiai aspektusairól cf. még Havelock 1963, Nyiri 1994.

<sup>245</sup> McGrew 1994, 1998. Boesch 1993, Wrangham et al. 1994.

<sup>246</sup> Boesch–Boesch 1990.

<sup>247</sup> Huffman–Wrangham 1994.

<sup>248</sup> Sperber 1985, 1994.



nulmányozása, amely a mentális reprezentációk továbbadásával foglalkozik, nem használ szükségképpen az elme vizsgálatának elméleti keretét alkotó evolúciós modelleket.

Sperber megoldását olyan, az elme átfogó evolúciós megértésére törekvő tudósok is átveszik, mint például Pinker.<sup>249</sup> Csakhogy ennek az egyébként igen sikeres megközelítésnek van egy támadható pontja: megmarad egy olyan szolipszista módszertani modell keretében, amelyben a magába zárt elme működése független attól az összetett társadalmi kontextustól, amelyben működik. A kultúra Sperberék szerint teljes mértékben függetlenedni tudott az evolúciós folyamatoktól és a genetikai kötöttségektől.

A szociobiológia irányzatának képviselői ezzel ellentétes utat járnak: szerintük a kultúra teljes mértékben függ az evolúciós folyamatoktól és a genetikai kötöttségektől.<sup>250</sup> Ezt az álláspontot mondja ki a szociobiológia egyik legtöbbet idézett metaforája, amely szerint a gének pórázon tartják a kultúrát. A póráz ugyan elég hosszú, de a kultúra nem tud elszakadni a genetikailag meghatározott alapoktól.<sup>251</sup> A szociobiológia ennek megfelelően a kulturális jelenségek mindegyikét úgy próbálja magyarázni, hogy az valamilyen szempontból hasznos az egyén túlélése, az egyén génjei szempontjából. Márpedig komoly nehézségekbe ütközik, amikor ezt olyan kulturális konstrukciókról kell megmutatnia, amilyenek például a Loire-völgyi kastélyok.

Tehát ez az irányzat is egy individuumból indul ki, ahol az egyének a saját túlélésükről és minél hatékonyabb szaporodásukról gondoskodnak, s ehhez ideiglenesen – és ha éppen hasznos számukra – igénybe veszik egy másik egyedet, illetve a kultúra által nyújtott segítséget is, de csak ha a haszon nagyobb, mint a ráfordítás. Távolról emlékeztet ez a klasszikus szerződéselméletekre, amelyek szerint szabad individuumból azért alkottak társadalmat, mert belátták, hogy csoportba tömörülve jobbak az esélyeik a túlélésre, mintha egyedül vándorolnának a szavannán.<sup>252</sup>

A két ellentétes felfogás mindegyike – a kultúra teljes függését kimondó szociobiológia és a teljes függetlenedés mellett érvelő járványtani modell – az individuumból indul ki. A két szélsőséges álláspont között érdemes közbülső megoldást keresni, amely szerint a kultúra, bár természetesen függ a genetikai alapoktól, ugyanakkor el is tud szakadni attól. A velünk született és tanult tulajdonságok kérdésében egy olyan álláspont tűnt leginkább tarthatónak, amely szerint mentális tulajdonságaink nem mind velünk születettek, azaz nem függenek teljesen a génjeinktől, másfelől azonban nem is mind tanultak, tehát nem függetlenek teljes mértékben a génkészlettől.<sup>253</sup> Ha ez igaz, akkor a kultúra kérdésében is egy köztes álláspontot célszerű elfogadni, amely szerint a kultúrát részben meghatározzák genetikai adottságaink, de csak részben. Igen nagy szabadságfokunk van a tekintetben, hogy a genetikailag rögzített kereteken belül milyen kultúrát alakítunk ki.

A gének és a kultúra függetlensége mellett érvelő járványtani modell a kulturális relativizmusban talál szövetségesre, amely a kultúrák radikális különbségét hangsúlyozza.<sup>254</sup> A szociobiológia ezzel szemben a közös jegyeket és az univerzálisan jellemző vonásokat keresi az egyes kultúrákban.<sup>255</sup> minden kultúra strukturálisan egyforma, az esetleges különbségek figyelmen kívül hagyhatók.<sup>256</sup>

Egy köztes álláspont is elképzelhető azonban, amely szerint minden kultúra rendelkezik egy genetikailag meghatározott alappal, de erről az alapról kiindulva és a genetikai megkötöttségeken belül egymástól igen távol álló kulturális rendszereket képes létrehozni. Ráadásul ez a közbülső álláspont talán képes elkerülni a mindkét szembenálló nézetre jellemző individualista gondolatú modellt.

## 2. A mémek

Richard Dawkins mémelmélete a leghíresebb kulturális modell, amely evolúciós alapul, és emellett rugalmasabban képes kezelni a kultúra szféráját.<sup>257</sup> Dawkins szerint a kultúra működése is evolúciós folyamat, de itt nem a gének küzdenek a fennmaradásért, hanem egy új evolúciós egység, amelyet ő szellemes módon mémnek nevez. A mém nem más, mint egy mentális reprezentáció, amelynek – akárcsak a génnek – az a célja, hogy minél több példányban terjedjen el. A mémek esetében ez azt jelenti, hogy egy mém annál sikeresebb, minél több másik elmébe juttatja el magát. A mémek tehát elméinkért küzdenek. Mivel elménk kapacitása véges, ezért csak viszonylag kevés mém befogadására képes. Ezért a korlátozott kapacitásért szállnak versenybe a mémek, a sikerte-

<sup>249</sup> Pinker 1997, pp. 208–210.

<sup>250</sup> A szociobiológia legfontosabb alpművei: Wilson 1975, 1978, Lumsden–Wilson 1981. L. még Bereczkei 1998. Ellene például: Kitchner 1982, Lewontin–Rose–Kamin 1984.

<sup>251</sup> Wilson 1978, p. 167.

<sup>252</sup> Cf. főleg Hobbes 1970, Rousseau 1963, Locke 1989, Hume 1976.

<sup>253</sup> Ez volt a neurális evolúció megoldási javaslata.

<sup>254</sup> Cf. Montesquieu 1981, a kulturális relativizmus kérdéskörének pszichológiai és filozófiai vonatkozásaihoz cf. pl. Shweder 1991. Berlin 1996.

<sup>255</sup> Ide tartozik a humánológia tudományának nagy része is, az univerzalizmus–relativizmus problémaköréről cf. pl. Eibl-Eibesfeldt 1989.

<sup>256</sup> Meg kell jegyezni, hogy ez az univerzalizmus csak a szociobiológia egy részére igaz, hiszen például Lumsden–Wilson 1981 amellett érvel, hogy mivel az emberi génkészlet hozzávetőleg ezer évente jelentősen megváltozik, s ezért a különböző kultúrák különböző génkészlet pórázán vannak. A gén–kultúra függés tehát itt is jelen van, de a kultúra világát meghatározó gének kultúráról kultúrára változnak.

<sup>257</sup> Dawkins 1989.

lenek nem tudnak bekerülni elménkbe, s ezért eltűnnek, a sikeresek viszont jó sok ember elméjébe befészkelik magukat. A mémek tehát gondolatdarabok, mém lehet a liberalizmus eszméje, az Eroica szimfónia első tételének fő témája vagy a fogmosás.

Dawkins elmélete radikális következményekkel jár. A génközpontú evolúcióbíológiai modell szerint – amelyet többek között szintén Dawkins vezetett be – az állatoknak nincs saját akaratuk, voltaképpen azt teszik, amit önző génjeik diktálnak nekik túlélésük és szaporodásuk biztosítása érdekében.<sup>258</sup> A szociobiológia szerint ez a modell változtatás nélkül illik az emberre is. Dawkins szerint azonban az ember esetében valami radikálisan új történik az evolúciós folyamatban, megjelenik egy új replikátor, a mém, és ettől kezdve az emberi elmét a mémek és a gének túlélésért való együttes küzdelme határozza meg. Az ember nem kizárólag a génjeinek engedelmeskedik, ez nagy különbség az állatokhoz képest, de azért persze a gének is jelentős mértékben meghatározzák a viselkedését.

Dawkins tehát a kulturális jelenségeknek épp azt a köztes modelljét képviseli, amely mellett az előzőekben érveltünk: a kultúra szféráját nem határozzák meg teljesen a genetikai faktorok, de nem is teljesen független tőlük. Ez a függetlenedés azonban Dawkins esetében leginkább a mémektől való függést jelenti. Az állatok kizárólag a génjeiktől függtnek, az emberek azonban két tűz között vannak: a géneknek és a mémeknek is engedelmeskedniük kell. Nincs tehát emberi akarat, az állathoz képest a különbség pusztán annyi, hogy kettős meghatározottsággal rendelkezünk: úgy viselkedünk, ahogy a sejtjeinkben levő géneknek és a fejünkben levő mémeknek tetszik.

Ezek szerint a mémek az elme leigázói, nem mi határozzuk meg, mit gondolunk, a mémek irányítanak bennünket. Az elme a mémek egymással való küzdelmének csatatere, a sikeres mémek benn maradnak az elmében, a kevésbé sikeresek kiesnek. A mémelmélet tehát igen radikális: tagadja, hogy lenne valamilyen központi Én, amely a lélek fogalmához hasonlatosan irányítja cselekedeteinket. Csak a mémek küzdelme határozza meg, mit teszünk és gondolunk.<sup>259</sup>

Mint láttuk, a mémek elmélete kielégíti a kulturális elmélettel szembeni követelményeket: a kultúra világa nem teljesen független a genetikai alapoktól, a mémek viszont egy radikálisan új elemet hoznak a kulturális folyamatokba. Ráadásul a mémek terjedésének valamilyen előképét a főemlősöknél is lehet azonosítani. A lapos kő alátámasztása például a pálmadió feltörésének esetében sikeres mémnek bizonyult a csimpánzok körében, hiszen sok csimpánzelmébe befészkelte magát. Az állati és az emberi elme közötti folytonos átmenet is beépíthető tehát a mémelméletbe.

Felmerül azonban néhány nyilvánvaló ellenvetés is a mémelmélettel szemben. Az egyik leghíresebb az öngyilkosság értelmezésével kapcsolatos látszólagos ellentmondás. Az öngyilkosság mémje ugyanis látványosan sikeres, különösen ahhoz képest, hogy terjedésének közvetlen következménye a hordozó halála, a terjedés tehát a terjedés megszűnéséhez vezet.<sup>260</sup> Ez a jelenség felfogható azonban úgy is, hogy az öngyilkosság mémje akkor is létezhet a fejben, ha nem leszek öngyilkos, hiszen akkor is tudom, mi az öngyilkosság. Más dolog tehát az öngyilkosság mémjével rendelkezni, és más öngyilkosnak lenni. Az öngyilkosság mém hordozói közül csak kevesen lesznek valóban öngyilkosok.

Megfogalmazható azonban néhány valóban komoly kritika is a mémelmélettel szemben, amely megkérdőjelezi az egész megközelítés létjogosultságát. Mindenekelőtt Dawkins igencsak leegyszerűsíti a mémek átadásának kérdését, hiszen szerinte a mémek egyenként támadnak. Dawkins tehát úgy gondolja, hogy a mémek között nincs kapcsolat, függetlenek egymástól. Lehet azonban amellett érvelni, hogy az egyes mémek nem külön-külön, egymástól függetlenül, hanem meghatározott rendszerben adódnak át. Az Eroica első tételének fő témája mém nem igazán képes átadódni anélkül, hogy az első tétel melléktémájának mémje is át ne adódna. A mémek tehát – Dawkins elképzelésével szemben – rendszert alkotnak, amelyben egyes mémek közelebb állnak egymáshoz, mint mások. Akárcsak a gének esetében, ahol például az egy kromoszómán található gének gyakrabban járnak együtt, mint a más-más kromoszómán találhatóak, itt is fontos a replikáció egységei közötti kapcsolatrendszer. Ritkán szembesülünk egyes mémekkel, általában mémcsoportokkal, mémrendszerekkel találkozunk.

Ennél lényegesen komolyabb ellenvetést jelent az implementációfüggetlenség kérdése. A Dawkins-féle mémelmélet egyik nem rögtön látható előfeltevése ugyanis éppen az, hogy az elme implementációfüggetlen, független tehát az öt megvalósító neurális közegtől. A mémek ugyanis *per definitionem* implementációfüggetlen entitások, hiszen egy és ugyanaz a mém a te agyadban egészen más neurális struktúrákkal lesz azonos, mint az enyémben. A mém fogalma tehát független attól a neurális szerkezettől, amely megvalósítja. Érdekes, hogy az elme evolúciós megközelítésének egyik legfontosabb képviselője, Daniel Dennett is elfogadja a mémelméletet mint a kulturális evolúció átfogó elméletét. Dennett, aki az instrumentalizmus harcos védelmezőjeként nagy erőbedobással küzdött a funkcionalizmus és a mentális állapotok implementációfüggetlensége ellen, ezen a ponton mégis kénytelen feladni instrumentalizmusát, és a mémelmélet miatt igen közel kerül az általa olyannyira gyűlölt funkcionalizmushoz.<sup>261</sup>

<sup>258</sup> A mémelmélet a génközpontú evolúciós modell kiterjesztéseként indult. Cf. Dawkins 1986, 1989.

<sup>259</sup> Dennett például többek között emiatt az antikarteziánus hozadéka miatt fordul a mémelmélethez. Cf. Dennett 1991, 1998a.

<sup>260</sup> E kérdéskörrel lásd: Dawkins 1989. pp. 159–161., Dennett 1998a. pp. 390–391.

<sup>261</sup> Cf. Nánay 1999.

A mémek elmélete tehát ugyanúgy a megvalósító közegtől való függetlenség téziséen alapul, mint a funkcionalizmus. Következésképpen minden olyan evolúciós érv, amely a funkcionalizmus ellen irányult annak implementációfüggelensége miatt, ebben az esetben is felvethető. A dolog pikantériáját az adja, hogy egy evolúciós ihletésű elméletet bírálunk itt az evolúció felől. Az elmén belüli evolúció gondolata és az implementációfüggelenség nem összeegyeztethető gondolatok, mint ezt a III. fejezetben részletesen láttuk: ha igaz, hogy a mentális reprezentációk – a mémek –egyszerűbb idegsejtsortokból evolválódtak, akkor nem lehetnek függetlenek tőlük. Ha tehát elfogadjuk a neurális evolúció új elméletét, akkor az implementációfüggelenség téziséet s ezzel együtt a mémelméletet is el kell vetnünk.

Egy harmadik ellenvetés is felvethető azonban a mémelmélettel szemben. Dawkins elmélete ugyanis – akár csak a szociobiológia és a járványtani modell – egy individuumközpontú modellen alapul. Magányos, a társadalmi közegtől független egyedek azok, akik mémeket továbbítanak és befogadnak. Magában álló, társas szférától független egyedek alkotják a kiindulópontot, a mémáramlás ezek között az atomikus egyedek közötti másodlagos folyamat. Ugyanezt a kritikai ellenvetést támasztja alá az a tény is, hogy bár a mémelmélet evolúciós ihletésű modell, nem ad választ arra a kérdésre, hogy mi tette lehetővé a mémek mint új replikációs egységek megjelenését. Nem ad magyarázatot tehát a mémek kialakulására.

Miért és főleg kinek a számára volt hasznos a mémek megjelenése? Az egyén számára nem lehetett hasznos, hiszen éppen az egyént igázzák le a mémek. Ráadásul, ha az egyén számára lenne adaptív folyamat a memetikuss evolúció, akkor végső soron a mémelmélet visszajutna a szociobiológia kiindulópontjához, amelytől éppen el akart határolódní, hiszen ekkor azt állítaná, hogy a mém végső soron egy igen rafinált eszköz az egyén túlélésének biztosítására. Az egyén számára tehát nem lehet hasznos. Másfelől azonban, ha azt állítjuk, hogy maguknak a mémeknek volt hasznos a mémek megjelenése, akkor körben forgó magyarázatot adunk, hiszen az igazi kérdés éppen az, hogy miként történt a mémek felbukkanása. Ha nem akarunk valamiféle varázsütésszerű, csodás megjelenést elfogadni, akkor másfelé kell keresnünk a mémek kialakulásának eredetét.

Továbbá a mémek elmélete előfeltételez valamilyen társadalmi formációt és kommunikációt, hiszen ha ez nem létezne, akkor a mém nem tudna eljutni egyik elméből a másikba. Másfelől viszont Dawkins minden egyénen kívüli, egyének közötti jelenséget – tehát minden kulturális jelenséget – a mémelmélettel magyaráz. Megint csak logikai körben forgáshoz jutottunk: a mémek előfeltételezik a kommunikációt és a csoportot, a csoport viszont előfeltételezi a mémeket. Dawkins védekezhetne persze azzal, hogy e két sajátosság, a mémek, illetve a társadalmi szervezethez és a kommunikáció koevolválódtak, kölcsönösen hívta életre egymást. Ez jó megoldás lenne, hiszen a mémek megjelenését fokozatosan képzei el, és így az a csimpánzkultúrákra is kiterjeszhető lenne: a csimpánzkultúrák esetében még a mémek terjedésének csak egy kezdeti, nem kialakult formája volt jelen, mint ahogy a csimpánzok esetében az egyedek közötti kommunikációnak és a csoportnak is csak egy kezdeti stádiumáról lehet beszélni.

Ha azonban ezt elfogadná a mémelmélet, akkor el kellene fogadnia azt is, hogy a csoport kialakulása párhuzamosan zajlott a mémek kialakulásával. Ez viszont az individualista elméleti keret tagadását jelentené, márpedig ez ellenkezne Dawkins explicit célkitűzésével.

### 3. Egy csoportközpontú modell

Érdemesebb a másik irányból elindulni, és a társadalmat nem egyedülálló individuumok utólagos egyesülésének felfogni, hanem az embert már eleve társas lényként tekinteni. Az eddig bemutatott három kulturális modell mindegyike az egyénből indult ki, az egyének közötti kapcsolatrendszer, a kultúrát pedig utólagos hozzátételnek tekintette. A következőkben a kulturális evolúció egy olyan felfogása mellett fogok érvelni, amelyben az egyén logikailag nem előzi meg a társadalmi kapcsolatrendszer, amelyben tehát az ember már mindig is mint egy csoport része jelenik meg. Az, hogy az ember társas lény, nem utólagos hozzátétel az alapvetően individualista emberi természethez, hanem annak szükségszerű velejárója.

Egy ilyen elmélet azonban igen komoly biológiai megalapozást igényel. Az evolúciós biológia egyik legvitatottabb kérdése, hogy létezik-e csoportselekcio. Lehet-e tehát a csoportról mint az evolúciós folyamatok önálló egységéről beszélni? Lehet-e a csoport és az egyén közötti viszonyt a szervezet és sejt közötti viszony analógiájára felfogni? Sokáig a biológusok teljes mértékben elvetették ezt a gondolatot.<sup>262</sup> Később Boyd és Richerson megmutatta, hogy a csoportselekcio logikailag lehetséges, de arra nézve semmilyen állítást nem tettek, hogy az emberi csoport esetében beszélhetünk-e erről a jelenségről.<sup>263</sup> Az utóbbi években két filozófus, Elliott Sober és David Sloan Wilson igen meggyőzően érvelt amellet, hogy az emberi csoport esetében is létezhet csoportselekcio.<sup>264</sup> Sober és Wilson a különböző szekták felépítését vizsgálva jutottak arra a konklúzióra, hogy bizonyos esetekben a humán csoport képes úgy működni, mintha organizmus lenne, az egyedek pedig éppen olyan

<sup>262</sup> A legnagyobb hatású munka a csoportselekcio gondolatának ellenében George C. Williams könyve volt (Williams 1966).

<sup>263</sup> Boyd–Richerson 1985.

<sup>264</sup> Wilson–Sober 1994, Sober–Wilson 1998.

heteronóm, függő helyzetben vannak a csoporthoz képest, mint a sejtek a szervezethez képest. Az organizmus-társadalom hasonlat persze nem újdonság, Sober és Wilson pusztán annyit tett, hogy e gondolat biológiailag korrekt megalapozását adta.<sup>265</sup> Ha az egyes csoportok közötti variabilitás elég nagy, a csoporton belüli variabilitás pedig elég kicsi, akkor a csoportok közötti szelekció „elhomályosítja” az egyedek közötti szelekciót.<sup>266</sup>

A csoport mint szelekciós egység gondolata ugyanúgy az időbeli replikáció második fejezetben már tárgyalt gondolatát előfeltételezi, mint az elmén belüli evolúció. A csoport replikációját nem úgy határozzuk meg, hogy egy csoportból kettő lesz, hanem úgy, hogy képes fennmaradni annak ellenére, hogy az őt felépítő komponensek folyton cserélődnek. Az elemek változnak, a struktúra azonban ugyanolyan marad.

Az ember evolúciója esetében felmerül az a megoldás, hogy míg a főemlősöknél nem beszélhetünk csoport-szelekcióról, addig ez a szelekciós szint az emberi csoport esetében már jelen van (vagy volt). Természetesen, ha biológiailag korrekt elméletet akarunk alkotni, akkor a csoport megjelenését is folytonos, graduális folyamatként kell felfognunk, nem egyik pillanatról a másikra pattant ki az erős csoportkohézió, hanem folyamatosan egyre erősebb csoportkötelékek alakultak ki. A főemlősök esetében is igen nagy fokát találjuk a csoportkohézióknak, de jóval kisebbet, mint az embernél.

További problémát jelent, hogy a mai társadalomban nem igazán találunk példát – az említett szekták kivételével – olyan csoportokra, amelyekben valóban olyan erős lenne a belső kohézió, és oly nagy a többi csoporttól való elhatárolódás, az izoláció, hogy Sober és Wilson kritériumai szerint valóban csoport-szelekcióról lehessen beszélni. Ezt az ellenvetést védi ki Csányi Vilmos elmélete, amely azt hangsúlyozza, hogy a csoport-szelekciós modell a ma létező társadalmakra nem érvényes, a mai társadalom éppen az egymással semmilyen kapcsolatot nem tartó, izolált csoporttársadalmak szétesése révén jött létre.<sup>267</sup> Az izolált csoporttársadalmakban az egyén bár-mikor feláldozta magát a csoport túléléseért. Az egyed-csoport viszony itt még valóban olyan volt, mint a sejt és a szervezet közötti kapcsolat. A csoportok azonban fokozatosan átfedésbe kerültek, megszűnt az izoláció. Míg azonban a csoportok mára már megszűntek, megmaradt az ember igen erős biológiai diszpozíciója arra, hogy csoportot alkosson. Ezért lehet az emberből igen könnyen a csoporthoz való nagyfokú lojalitást, illetve más csoportok elleni gyűlöletet kiváltani.

Ha igaz, hogy az emberi csoport esetében létezik vagy legalább létezett csoport-szelekció, akkor más megvilágításba kerül a kultúra és a kulturális evolúció problémája is. Ezen a ponton lehet végre definiálni a kultúra fogalmát (vagy legalábbis egy minimáldefiníciót adni): kultúrának nevezzük az egyed feletti szerveződési szint, a csoport replikálódó struktúráját. Ha ezt elfogadjuk, akkor könnyebben kezelhető az a kérdés, hogy létezik-e csimpánzkultúra, hiszen ekkor ez nem egy kétesélyes eldöntendő probléma, hanem fokozat kérdése: a csoport-szelekció a csimpánznál még csak csökevényes formában van jelen (ha egyáltalán csoport-szelekcióról lehet beszélni esetükben). A csoport bizonyos strukturális sajátosságai replikálódnak, és megmaradnak annak ellenére, hogy a csoportot alkotó egyedek cserélődnek, meghalnak, és helyükre újak születnek. Ez a változatlan, replikálódó struktúra azonban a csimpánzoknál még igencsak egyszerű. A gyógynövények keresése vagy a pálmadió lapos kövének (üllőjének) alátámasztása például ilyen replikálódó elemeknek tekinthetők, de nincs meg az a bonyolult, generációról generációra átadódó rendszer, amely az emberi kultúrát alkotja.

A csoportközpontú modell teljesíti a kultúra genetikai meghatározottságával szemben a fejezet elején támasztott követelményeket is. Egyfelől nem teljesen független a kultúra a génállománytól, hiszen az új evolúciós egység elemei, az emberek genetikailag meghatározott lények, nem bújhatnak ki a bőrükből. Másfelől azonban a kultúrát nem teljesen határozzák meg a gének, hiszen egy új szerveződési szint, a kultúra szintje jött létre, amely ugyanúgy saját evolúciós szabályoknak engedelmessékedik, mint az egyének közötti szelekció szintje.

Láttuk, hogy a mémelmélet esetében a reprezentációk elméről elmére adódása, tehát a mémek terjedése drasztikusan más elmemoddellen – jelesül az implementációfüggetlenség modelljén – alapult, mint az egyes elemek felépítése. Meg kell tehát vizsgálni, hogy a csoportközpontú modellben nem szembesülünk-e ugyanilyen problémákkal. Kérdés tehát, hogy a csoport megjelenése az elmeműködés milyen változásával járt együtt.

Sober és Wilson szerint a csoport-szelekció egyik feltétele, hogy a csoportot alkotó egyedek eléggé hasonlítanak egymáshoz. Ez az embercsoport esetére vonatkoztatva elsősorban nem biológiai vagy alkati hasonlóságot jelent, hanem sokkal inkább azt, hogy hasonló dolgokat gondoljanak. A csoport nagyfokú kohéziójának kialakításához tehát arra volt szükség, hogy a csoport tagjai hasonló mentális reprezentációkkal rendelkezzenek a világról. Ehhez viszont ki kellett küszöbölni az egyes emberek mentális reprezentációinak különbségeit, egységesíteni kellett a reprezentációkat. A személyes tapasztalatok különbségeitől kellett megfosztani a reprezentációkat, a reprezentációt felépítő apró eltéréseket kellett elsimítani, egyszóval implementációfüggetlenné kellett tenni őket.

Furcsának tűnhet, hogy miután az egész könyv során az implementációfüggetlenség ellen érveltem, végül mégiscsak előkerül ez a fogalom. Nincs azonban szó ellentmondásról, hiszen az emberi elme alapvetően nem implementációfüggetlen szerkezetét kellett az erősebb csoportkohézió miatt implementációfüggetlenné tenni, ez

<sup>265</sup> Lásd különösen Wilson 1989.

<sup>266</sup> A biológia filozófiájának egyik kulcsproblémája ez. Mind filozófiai, mind biológiai szempontból a legszínvonalasabb áttekintését Robert N. Brandon adja, ő vezeti be az elhomályosítás fogalmát is a szelekció szintjeinek vizsgálatába (Brandon 1996, különösen p. 132f.).

<sup>267</sup> Csányi 1994, Csányi 1999.

a változás azonban nem jelentette azt, hogy valóban megszűnt minden kapcsolat a megvalósító közeggel és az alacsonyabb szintű reprezentációkkal, egyszerűen csak egy implementációfüggetlen, alapvetően nyelvi jellegű háló került a továbbra is implementációfüggő reprezentációk tetejére.

Olyan ez, mint amikor a bukszust kúpalakúra nyírják, két bukszus belsejében az ágak másként helyezkednek el, de a külső alak mégis ugyanolyan.<sup>268</sup> Reprezentációink is hasonlóak, de megvalósító közegük igencsak különböző, és ez magára a reprezentációra is kihat. A kúp alak ugyanúgy mesterséges keretekbe kényszerítése a növénynek, mint ahogy az implementációfüggetlen máz is mesterséges keretekbe kényszerítése a reprezentációnak. Ez az implementációfüggetlenedés azonban soha nem teljes, emiatt nem jó a funkcionalizmus és a mémelmélet modellje. A megvalósító közegtől és az egyszerűbb reprezentációktól független reprezentáció csak külső máz, amely alatt ott a reprezentáció gazdag belső struktúrája. Mindvégig megmarad azonban a reprezentációk implementációfüggő jellege, csak néha a másik ember jobb megértése és a csoporton belüli kommunikáció kedvéért ezt nem vesszük figyelembe. Voltaképpen ál-implementációfüggetlenségről kellene beszélni.

A funkcionalizmus és a mémelmélet tehát nem tévedett, amikor az implementációfüggetlenséget az elme fontos tulajdonságának tekintette. Az implementációfüggetlenség azonban egyrészt kizárólag az emberi elme sajátja, másrészt még az emberi elme számára is csak hasznos segédkonstrukció implementációfüggetlenségnek tekinteni a reprezentációkat, mert ekkor gördülékenyebb a kommunikáció a többi csoporttaggal. Az elmén belüli evolúció és az implementációfüggetlenség tézise tehát, amelyről korábban azt állítottam, hogy egymást kizáró elméletek, itt összekapcsolódik: az elmén belüli evolúció végpontja az implementációfüggetlenség, amely azonban csak az emberi kommunikációhoz szükséges segédeszköz.

Ha viszont ez igaz, akkor a mémelméletet a csoportközpontú modell speciális eseteként lehet felfogni. Egyrészt az implementációfüggetlen mémek átadásának Dawkins által adott modellje jól jellemzi az emberi csoport kialakulásának végállapotát, ahol igen nagy csoportkohézió mellett implementációfüggetlen reprezentációk – mémek – adása és kapása folyik. Ez a modell – és ezt fontos hangsúlyozni – csak a végállapot, az idevezető utat, a csoport kialakulását Dawkins nem vizsgálja. Az implementációtól való függetlenedés folyamatát nem elemzi, csak az implementációfüggetlenség végállapotát. Másrészt Dawkins csak az érem egyik oldalát látja. Csak az elme reprezentációival, illetve azok terjedésével foglalkozik, azt a struktúrát, amelybe mindez illeszkedik, és mindezt lehetővé teszi, tehát a csoportot nem veszi figyelembe.

Ezen a ponton érdemes visszatérni a csoportközpontú megközelítés korábban már említett problémájára, nevezetesen, hogy ma már egyáltalán nem létezik az a nagy csoportkohézió, amely a csoport szelekcióhoz, illetve az attól elválaszthatatlan mentális változásokhoz szükséges. Sober és Wilson szerint a csoport szelekció két feltétele az elég nagyfokú hasonlóság a csoport egyedei között és az elég nagyfokú különbség a csoportok között, más szóval a szinkronizáció és az izoláció. E két kritérium egyike sem jellemző a mai társadalomra, de az elmúlt néhány száz év európai társadalmaira sem. Ha tehát védhetővé akarjuk tenni a csoportközpontú megközelítést, akkor Csányit követve azt kell állítanunk, hogy a csoport szelekció az emberré válás egy bizonyos, jó hosszú szakaszára, az izolált csoporttársadalmak érájára volt jellemző, a csoport szelekció ma már nem, vagy csak csökevényes formában létezik.<sup>269</sup>

Régen tehát egymással semmilyen kommunikációt nem folytató, zárt csoportok léteztek, ahol az egyén valóban heteronóm helyzetben volt a csoporthoz képest: nem léteztek saját céljai, érdeke a csoport érdekéből származott. Nem a saját túlélése volt az igazán fontos számára, hanem az, hogy a csoport megmaradjon. Épp ezért bármikor feláldozta magát a csoport fennmaradásáért. Nem kellett döntéseket hoznia, hiszen a csoportnorma egyértelműen meghatározta cselekedeteit. Egy csavar volt a gépezetben.

Az ilyen csoport szerkezetre ma már alig találunk példát. A még létező törzsi társadalmak ugyan hasonlóan épülnek fel, mint ahogy a szekták is, de az izoláció fenntartása, a külső hatásoktól való függetlenség, amely a régi csoporttársadalmakban adott volt, itt a legkomolyabb kiküszöbölendő tényező. Egy középkori falu vagy egy ókori kisállam jobban megközelíti ezt a csoportmodellt. A struktúra replikálódott, a csoporttagok cserélődtek, anélkül hogy bármi is változott volna. Az izoláció azonban már itt sem volt teljes, az igazi csoporttársadalmak valószínűleg jóval a történelmi idők előtt léteztek.

E csoportok azonban azóta átfedésbe kerültek, megszűnt az izoláció, és a mai társadalom embere nem kap készen valamilyen csoportnormát. Nem egyértelműen következik számára a csoport érdekéből, hogy mit kell tennie, megjelent a választás, a döntés kényszere; az embernek sokszor két egymással átfedésbe kerülő csoport és csoportnorma között kell választania. E változást megörökítő lenyomatoknak lehet tekinteni a nagy görög tragédiákat, amelyekben legtöbbször éppen az okozza a konfliktust, hogy az egyetlen cselekvési lehetőség és norma helyett, amely a csoporttársadalmat jellemezte, már több választható csoportnorma áll. Antigone választás elé kerül: kövesse a Kreon által diktált társadalmi rendet, avagy temesse el testvérét, Polüneikészt. Ilyen konfliktus a csoporttársadalmak esetében elképzelhetetlen lett volna: Antigone döntését egyértelműen meghatározta volna a csoport, amelynek része.

<sup>268</sup> Quine bukszus-példáját megint nem eredeti értelmében használom. Cf. Quine 1960. p. 8.

<sup>269</sup> Ez az elmélet persze éppen ezért empirikusan nem tesztelhető, inkább hipotézisnek tekinthető, amely mellett azonban több erős érv is szól, mint látni fogjuk.

A csoporttársadalomban tehát az egyén kompakt gondolati rendszereket kapott készen, ma viszont már nem áll rendelkezésre ilyen egyértelműen meghatározott, készen kapott csoporteszme. A dawkinsi elmélet ebből a szempontból is a csoportközpontú modell speciális esetének tekinthető. Mint láttuk, a mémelmélet szerint a mémek egyenként, külön-külön rohamozzák meg az elmét. Ezzel szemben a csoportközpontú modellben, amelyben minden csoporttagnak ugyanazt a kultúrát, ugyanazt a csoportstruktúrát kellett elsajátítania, a reprezentációk komplex rendszerét kellett egyik generációról a másikra, egyik emberről a másikra átadni. Ha azonban igaz, hogy ez a modell csak az izolált csoporttársadalmakra jellemző, amelyek már jó ideje szétestek, akkor a mémelmélet annyiban jó közelítést ad a mai helyzetről, hogy ma már nem kapjuk készen a reprezentációk komplex rendszerét a kultúrától. Ma, sok kultúra metszéspontjában, különböző kultúrákból származó törmelékek, rendszerbe nem illeszkedő forgácsok rohamozzák meg elménket, és nekünk e törmelékekből kell – amennyire tudunk – olyan rendszert formálnunk, amely legalább távolról emlékeztet azoknak a rendszereknek a kompakt-ságára, amelyeket valaha készen kaptunk a csoporttársadalomtól.

Az evolúciós pszichológia azt állította, hogy a mai emberi elme nem a mai környezethez, hanem egy többszázéves környezethez adaptálódott. Fontos azonban hangsúlyozni azt az evolúciós pszichológiai gondolatot is, hogy ez a környezet nem puszta a szavanna fáit jelenti, hanem legalább ilyen súllyal a társadalmi környezetet is.<sup>270</sup> A csoporttársadalmak tehát szétestek, de az ember elméje, amely a csoporttársadalmak követelményeihez adaptálódott, ma is hasonló az akkorihoz. Az embernek tehát biológiailag, evolúciósan rögzített preferenciája, hogy csoportkötélékben éljen, csak hogy ezt a preferenciát a mai csoportok nem elégítik ki, hiszen ma már nem igazán léteznek rögzített, izolált csoportok. Ezért lehet az emberekből könnyen összetartó csoportot formálni, mint ezt szociálpszichológiai kísérletek egész sora mutatja. Az ember már nem csavar a gépezetben, mivel nincsenek nagy gépezetek, de nagyon szeretne az lenni, és ennek érdekében konstruál is magának olyan gépeket, amelyek csavarja lehet. A szekták nagy népszerűsége éppen ezzel magyarázható, a szektatagnak nem kell döntéseket hoznia, minden cselekedetét egyértelműen meghatározzák a szekta normái, érdekei és szabályai.<sup>271</sup> Az ember megint csavar lehet egy gépezetben.

A nacionalizmus különböző formái is éppen azon alapulnak, hogy egy csoportnormából, a nemzeti szellemből következik a hazafi minden cselekedete. Ahogy Rousseau írta a lengyel alkotmány tervezetében: „Mindenekelőtt lengyelek legyetek, s csak azután emberek.”<sup>272</sup> Szintén szembeötlő, hogy milyen nagy erkölcsi nimbusza van annak, ha az egyén feláldozza magát a hazáért, amikor tehát a csoport, a haza fontosabb, mint az egyén túlélése. Nem meglepő, hogy a szektákhoz hasonlóan itt is óriási jelentősége van az izolációnak, a világ többi részétől való elhatárolódásnak; erre – közelmúlt délszláv konfliktusáig bezárólag – sajnos, sok-sok történelmi bizonyítékot ismerhetünk.

Elrugaszkodottabb példa, de valószínűleg ugyanezen mozgatórugók működésének következménye a nagy romantikus szerelem kétszemélyes csoportja, ahol szintén egy nagyobb egység – a kétszemélyes csoport – részévé válik az ember, és nem kell önálló döntéseket hoznia, hiszen a nagyobb egység egyértelműen meghatározza, mit kell tennie, és mit kell gondolnia.<sup>273</sup> „Minden, amit együtt élünk át, egymásba fonódik, [...] az én kezem az övé is, az ő keze az enyém is, az én nevem az övé is, az ő neve az enyém is. [...] Már nem két ember élt ott, csak egy ember, öntudatlan, tökéletes jóérzésben.” – írja Goethe a *Vonzások és választásokban*.<sup>274</sup>

A szinkronizáció, az egyéni különbségek kiküszöbölése is nagymértékben jellemzi a kétszemélyes csoportot, mint ahogy az izoláció is. Világosan megfigyelhető, hogy szinte minden olyan művészeti alkotásban, amelynek tárgya a nagy, világrengető, romantikus szerelem, jelen van az izoláció, a világtól és a többi embertől való elkülönülés is. Egy kézenfekvő példa Jean-Luc Godard *Bolond Pierrot* című filmje, ahol Ferdinand és Marianne a természetben, minden embertől távol próbál élni – persze a valódi izoláció elérhetetlensége miatt ez az elvonulás végül nem túl sikeres.

Sem csoportkötélékben, sem anélkül nem érezzük tehát jól magunkat: csoportkötélék nélkül azért nem, mert evolúciósan úgy lettünk kialakítva, hogy csoporttagok legyünk, a csoportok viszont, amelyekben ma tagok lehetünk, az izoláció hiánya miatt nem tűnnek annyira magától értetődőnek, mint ez a stabil csoportkötélékhez szükséges lenne. A csoporttársadalmak embere nem kérdőjelezte meg a csoport szokásait és normáit, egyszerűen azért, mert nem ismert mást. Ma viszont, amikor sokféle normát és szokást ismerünk, nehéz totális lojalitással viseltetni bármilyen csoport iránt. Tehát sem csoportban, sem anélkül nem jó: hasonlít ez Kant társiatlan társiasság fogalmára: az ember társadalmi lény, mondja Kant, de ugyanakkor nem érzi magát teljesen otthon a társadalomban.<sup>275</sup>

A csoportok szétesése tette lehetővé azt a képességet is, amit ma racionális gondolkodásnak nevezünk.<sup>276</sup> Az izolált csoportban csak egyetlen, feltétlen lojalitást igénylő eszmerendszer létezett, és ezt – mivel ez volt az egyetlen eszmerendszer, amelyet a csoporttag ismert – nem lehetett megkérdőjelezni, kritikálni. Ahogy azonban

<sup>270</sup> Cf. Barkow–Cosmides–Tooby 1992.

<sup>271</sup> Cf. Hostetler 1963.

<sup>272</sup> Jean-Jacques Rousseau lengyelalkotmány-tervezetét cf. Rousseau 89, róla cf. Lukowski 1994.

<sup>273</sup> Cf. Wallace–Hartley 1988. Durkheim-értelmezését e tárgyban.

<sup>274</sup> Vas István fordítása.

<sup>275</sup> Kant 1980, p. 62.

<sup>276</sup> Ez persze csak az egyik lehetséges magyarázat az emberi racionalitás kialakulására.

a csoportok és a csoportnormák átfedésbe kerültek, megjelent a szkepszis, az egyértelmű igazságok megkérdőjelezése, amely lehetővé tette a racionális gondolkodást.

Ezen a ponton nem lehet elkerülni, hogy valamilyen értékijelentést tegyünk arról, hogy ezt a változást, a csoporttársadalmak szétesését miként is kell értékelnünk.<sup>277</sup> Itt, a könyv végén én egy viszonylag optimista értelmezést javaslok. A csoporttársadalmak embere mai szemmel igen furcsa életet élt. Igaz, hogy nem kellett a döntéshozás néha valóban nehéz feladatát magára vállalnia, de cserében nem volt több, mint akarat nélküli tárgy, a csoport nagy gépezetének egyik alkatrésze. Mechanikusan hajtotta végre egy nagyobb egység, a csoport akaratát. A kultúra tehát hiába tudott függetlenedni – legalábbis részben – a génektől, a genetikai meghatározottság helyére a csoport általi meghatározottság lépett. Nem a génjeim határozták meg, hogy mit teszek, hanem a csoportnorma. Cseberből vederbe.

Ezzel szemben ma, a csoportok szétesése után meglehetősen egyedi helyzet állt elő. Ma már sem a gének nem határozzák meg teljes mértékben, mit tegyünk, sem a csoport érdeke. Először fordul tehát elő, hogy nem állunk semmilyen kontroll – sem a gének, sem a csoport kontrollja – alatt, függetlenül, szabadon hozhatjuk meg döntéseinket. A csoport által biztosított elvesztett bizonyosságért pedig talán kárpótol ez az evolúciós újdonság, a szabadság érzése.

<sup>277</sup> Csányi például egyértelműen negatívumként, a „biológiai optimumtól” való eltávolodásként értékeli ezt a folyamatot. Cf. Csányi 1999, különösen az utolsó fejezet.

## IX. Bibliográfia

- Adams, Paul R. (1998): Hebb and Darwin. *Journal of Theoretical Biology* 195 pp. 419–438.
- Armstrong, D. (1968): *A Materialist Theory of Mind*. London: Routledge.
- Baars, B. J. (1988): *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- Baldwin, J. M. (1896): A New Factor in Evolution. *American Naturalist* 30 pp. 441–451. pp. 536–553.
- Barkow, J. H.–Cosmides, L. M.–Tooby, J. (1992, eds.): *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.
- Beckermann, A.–Floh, H.–Kim, Jaegwon (1992, eds.): *Emergence or Reduction?* Berlin: De Gruyter.
- Bereczkei Tamás (1998): *A belénk íródott múlt*. Budapest–Pécs: Dialóg Campus Kiadó. Berlin, Isaiah (1996): *Az emberiség göcsörtös fája*. Budapest: Európa Könyvkiadó.
- Bickerton, Derek (1981): *Roots of Language*. Ann Arbor: Maroma.
- Bickerton, Derek (1990): *Species and Language*. Chicago: Chicago University Press.
- Bickerton, Derek (1992): A reprezentációs rendszerek kezdetei. *Café Babel* nos. 3–4. pp. 83–95.
- Bickle, John (1998): *Psychoneural Reduction: the New Wave*. Cambridge, MA.: The MIT Press. Block, Ned (1978): Troubles with Functionalism. *Minnesota Studies in the Philosophy of Science* 9 pp. 261–325.
- Block, Ned (1983): Mental Pictures in Cognitive Science. *Philosophical Review* 93 pp. 499–542.
- Block, Ned (1986): Advertisement for a Semantics for Psychology. In: French, P.–Uehling, T. Jr.–Wettstein, H. (eds.): *Studies in the Philosophy of Mind*. Minneapolis: University of Minnesota Press.
- Block, Ned (1995): On a Confusion about a Function of Consciousness. *Behavioral and Brain Sciences* 18 pp. 227–287.
- Block, Ned–Fodor, Jerry A. (1972): What Psychological States are Not? *Philosophical Review* 81 pp. 159–181.
- Block, Ned–Flanagan, Owen–Giizeldere, Gitven (1997, eds.): *The Nature of Consciousness*. Cambridge, MA.: The MIT Press.
- Boesch, C. (1993): Toward a New Image of Culture in Wild Chimpanzees? *Behavioral and Brain Sciences* 16 pp. 514–573.
- Boesch, C.–Boesch, H. (1990): Tool Use and Tool Making in Wild Chimpanzees. *Folia Primatologica* 54 pp. 86–99.
- Boyd, Richard (1980): Materialism Without Reductionism: What Physicalism Does Not Entail. In: Block, Ned (ed.): *Readings in the Philosophy of Psychology. Volume I*. Cambridge: Cambridge University Press.
- Boyd, R.–Richerson, P. J. (1985): *Culture and the Ecolutionary Process*. Chicago: Chicago University Press.
- Brandon, Robert N. (1996): The Levels of Selection. Hierarchy of Interactors. In: Brandon, Robert N.: *Concepts and Methods in Evolutionary Biology*. Cambridge: Cambridge University Press, pp. 124–143.
- Brandon, Robert N.–Antonovics, Janis (1996): The Coevolution of Organism and Environment. In: Brandon, Robert N.: *Concepts and Methods in Evolutionary Biology*. Cambridge: Cambridge University Press, pp. 161–178.
- Brentano, Franz (1983): A pszichikai és fizikai jelenségek közötti különbségről. In: Pléh Csaba (ed.): *Pszichológiatörténeti szöveggyűjtemény I*. Budapest: Tankönyvkiadó, pp. 157–163.
- Bunge, Mario (1980): *The Mind–Body Problem*. Oxford: Pergamon.
- Burge, Tyler (1979): Individualism and the Mental. *Midwest Studies in Philosophy* 6 pp. 73–121.
- Burge, Tyler (1986): Individualism and Psychology. *Philosophical Review* 95 pp. 3–80.
- Buss, D. M. (1994): *The Evolution of Desire: Strategies of Human Mating*. New York: Basic Books.
- Buss, D. M. (1995): Evolutionary Psychology: A New Paradigm for Psychological Science. *Psychological Inquiry* 6 pp. 1–30.
- Bühler, Karl (1922): *Die geistige Entwicklung des Kindes*. Jena: Fischer.
- Byrne, Richard–Withen, Andrew (1988, eds.): *Machiavellian Intelligence*. Oxford: Clarendon Press.
- Campbell, Donald (1974): Evolutionary Epistemology. In: Schilpp, Paul A. (ed.): *The Philosophy of Karl Popper*. LaSalle: Open Court, pp. 413–63.
- Carruthers, Peter (1996): *Language, Thought, and Consciousness*. Cambridge: Cambridge University Press.
- Carruthers, Peter–Smith, Peter K. (1996, eds.): *Theories of Theories of Mind*. Cambridge: Cambridge University Press.
- Cavalier–Smith, T. (1987): The Origin of Eukariotic and Archaeac–terial Cells. *Annals of the New York Academy of Sciences* 503 pp. 17–54.
- Chalmers, David (1996): *The Conscious Mind*. Oxford: Oxford University Press.
- Changeux, Jean–Pierre (1985): *Neuronal Man: The Biology of Mind*. New York: Pantheon.
- Cheeny, Dorothy L.–Seyfarth, Robert, M. (1990): *How Monkeys See the World*. Chicago: University of Chicago Press.
- Cheney, Dorothy L.–Seyfarth, Robert M.–Silk, J. B. (1995) The Response of Female Baboons (*Papio cynocephalus ursinus*) to Anomalous Social Interactions: Evidence for Causal Reasoning? *Journal of Comparative Psychology* 109 pp. 134–141.
- Chisholm, Roderick M. (1955): Sentences about Believing. *Proceedings of the Aristotelian Society* 56 pp. 125–148.
- Chisholm, Roderick M. (1967): Intentionality. In: Edwards, P. (ed.): *The Encyclopedia of Philosophy* 3–4. New York: Mac–Millan.
- Chomsky, Noam (1966): *Cartesian Linguistics*. New York: Harcourt.
- Chomsky, Noam (1985): *Generatív grammatika*. Budapest: Európa.
- Chomsky, Noam (1995): *Mondattani szerkezetek – Nyelv és elme*. Budapest: Osiris.
- Churchland, Paul (1984): *Matter and Consciousness*. Cambridge, MA.: The MIT Press.
- Churchland, Paul (1986): *Neurophilosophy*. Cambridge, MA.: The MIT Press.
- Clark, A. (1991): In Defense of Explicit Rules. In: Ramsey, W.–Stich, S. P.–Rumelhart, D. E. (eds.): *Philosophy and Connectionst Theory*. Hillsdale, NJ: Erlbaum, pp. 115–128.
- Clark, A. (1994): *Associative Engines*. Cambridge, MA.: MIT Press



- Clark, A. (1996): *A megismerés építőkövei*. Budapest: Osiris Könyvkiadó. Eredetileg: *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. Cambridge, MA.: The MIT Press, 1989.
- Code, Alan (1991): Aristotle, Searle and the Mind–Body Problem. In: LePore, Ernest–Van Gulick, Robert (eds.): *John Searle and His Critics*. Oxford: Blackwell. pp. 105–113.
- Code, Alan–Moravcsik, Julius (1992): Explaining Various Forms of Living. In: Nussbaum, M.–Rorty, A. (eds.): *Essays on Aristotle's 'De Anima'*. Oxford: Oxford University Press.
- Cosmides, Leda M.–Tooby, John (1997): *Evolutionary Psychology: A Primer*, <http://www.psych.ucsb.edu/research/cep/primer.html>
- Crick, Francis (1989): Neural Edelmanism. *Trends in Neuroscience* 12 no. 7. pp. 240–248.
- Csányi Vilmos (1979): *Az evolúció általános elmélete*. Budapest: Akadémiai.
- Csányi Vilmos (1988): *Evolúciós rendszerek*. Budapest: Gondolat.
- Csányi Vilmos (1994): *Viselkedés, gondolkodás, társadalom: etológiai megközelítés*. Budapest: Akadémiai.
- Csányi Vilmos (1999): *Humánétológia*. Budapest: Vince Kiadó.
- Cummins, Robert (1975): Functional Analysis. *Journal of Philosophy* 72 pp. 741–765.
- Cummins, Robert (1983): *The Nature of Psychological Explanation*. Cambridge, MA.: The MIT Press.
- Dahlbom, Bo (1993, ed.): *Dennett and his Critics*. Oxford: Blackwell.
- Darwin, Charles (1859): *The Origin of Species*. London: Dent (magyarul: *A fajok eredete*. Budapest: Akadémiai, 1955)
- Davidson, Donald (1980): Mental Events. In: Davidson, Donald: *Essays on Actions and Events*. Oxford: Oxford University Press. pp. 207–225.
- Davidson, Donald (1984): Thought and Talk. In: Davidson, Donald: *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press, pp. 155–170.
- Davidson, Donald (1987): Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association* 60 pp. 441–458.
- Davidson, Donald (1991): Epistemology Externalized. *Dialectica* 45 pp. 191–202.
- Davidson, Donald (1993): Thinking Causes. In: Heil, John–Mele, Alfred (eds.): *Mental Causation*. Oxford: Clarendon, pp. 3–18.
- Davies, Martin–Humphreys, G. (1993, eds.): *Consciousness*. Oxford: Blackwell.
- Davies, Martin–Stone, Tony (1995, eds.): *Folk Psychology and the Theory of Mind Debate. Core Readings*. Oxford: Blackwell.
- Dawkins, Richard (1986): *Az önző gén*. Budapest, Gondolat.
- Dawkins, Richard (1989): *A hódító gén*. Budapest: Gondolat.
- Dawkins, R.–Krebs, J. R. (1978) Animal Signals: Information or Manipulation? In: *Behavioural Ecology: an Evolutionary Approach*. Oxford: Blackwell. pp. 282–309.
- Deacon, Terrence (1997): *The Symbolic Species*. New York: W. W. Norton.
- Dennett, Daniel C. (1983): Intentional Systems in Cognitive Ethology: The “Panglossian Paradigm” Defended. *Behavioral and Brain Sciences* 6 pp. 343–390. (Magyarul: In: Dennett 1998b. pp. 97–145.)
- Dennett, Daniel C. (1987): Beyond Belief. In: Dennett, Daniel C: *The Intentional Stance*. Cambridge, MA.: The MIT Press.
- Dennett, Daniel C. (1991): *Consciousness Explained*. Boston: Little Brown.
- Dennett, Daniel C. (1993): Labeling and Learning. Commentary on Clark and Karmiloff–Smith. *Mind and Language* 8 pp. 540–548.
- Dennett, Daniel C. (1996a): Az igazhívők: az intencionális stratégia és sikerének forrásai. In: Pléh Csaba (ed.): *Kognitív tudomány*. Budapest: Osiris–Láthatatlan Kollégium, pp. 152–168.
- Dennett, Daniel C. (1996b): Szövegek, emberek és más készítmények értelmezése. *Holmi* 8 no. 2. pp. 251–265.
- Dennett, Daniel C. (1996c): *Micsoda elmék*. Budapest: Kulturtrade.
- Dennett, Daniel C. (1998a): *Darwin veszélyes ideája*. Budapest: Typotex. (Eredetileg: Dennett, Daniel C: *Darwin's Dangerous Idea*. New York: Touchstone, 1995.)
- Dennett, Daniel C. (1998b): *Az intencionalitás filozófiája*. Budapest: Osiris–Gond.
- Descartes, René (1994a): *Elmélkedések az első filozófiáról*. Budapest: Atlantisz.
- Descartes, René (1994b): *A lélek szenvedélyei*. Budapest: Ictus.
- Descartes, René (1996): *A filozófia alapelvei*. Budapest: Osiris.
- Donald, Merlin (1991): *Origins of the Modern Mind*. Cambridge, MA.: Harvard University Press.
- Dretske, Fred (1981): *Knowledge and the Flow of Information*. Cambridge, MA.: The MIT Press.
- Dretske, Fred (1986): Misrepresentation. In: Bogdán, Radu (ed.): *Belief*. Oxford: Oxford University Press.
- Dretske, Fred (1993): Conscious Experience. *Mind* 102 no. 406. pp. 263–283.
- Dretske, Fred (1995): *Naturalizing the Mind*. Cambridge, MA.–London: The MIT Press / Bradford Books.
- Dreyfus, H. L. (1972): *What Computers Can't Do*. New York: Harper and Row.
- Dreyfus, H. L. (1993): *What Computers Still Can't Do*. Cambridge, MA.: The MIT Press.
- Dreyfus, H. L.–Dreyfus, S. E. (1986): *Mind over Machine*. New York: Harper and Row.
- Dreyfus, H. L.–Dreyfus, S. E. (1988): Making the Mind vs. Modelling the Brain. *Daedalus* 117 no. 1. pp. 15–43.
- Dunbar, Robin (1996): *Grooming, Gossip and the Language*. London: Faber and Faber. Edelman, G. M. (1987): *Neural Darwinism: The Theory of Neuronal Group Selection*. New York: Basic Books.
- Edelman, G. (1990): *The Remembered Present: A Biological Theory of Consciousness*. New York: Basic Books.
- Edelman, G.–Weinshall, D. (1989): *A Self-organizing Multiple-view Representation of 3D Objects*. Cambridge, MA.: MIT AI Laboratory and Center for Biological Information Processing.
- Eibl-Eibesfeldt, I. (1989): *Human Ethology*. New York: Aldine de Gruyter.

- Elman, J. L.–Bates, E. A.–Johnson, M. H.–Karmiloff–Smith, Anette–Parisi, D.–Plunkett, K. (1996): *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA.: The MIT Press.
- Fahlman, S. E.–Lebiere, C. (1990): The Cascade–Correlation Learning Architecture. In: Touretzky, D. S. (ed.): *Advances in Neural Information Processing Systems. Vol. 2*. Los Angeles: Morgan Kaufman, pp. 38–51.
- Feldman, J. A.–Ballard, D. (1982) Connectionists Models and their Properties. *Cognitive Science* 6 pp. 205–254.
- Flanagan, Owen (1992): *Consciousness Reconsidered*. Cambridge, MA.: The MIT Press.
- Fodor, Jerry, A. (1968): The Appeal to Tacit Knowledge in Psychological Explanation. *Journal of Philosophy* 65 pp. 627–640.
- Fodor, Jerry, A. (1974): Special Sciences: the Disunity of Science as a Working Hypothesis. *Synthese* 28 pp. 97–115.
- Fodor, Jerry, A. (1975): *The Language of Thought*. Cambridge, MA.: Harvard University Press.
- Fodor, Jerry, A. (1980): Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology. *Behavioral and Brain Sciences* 3 pp. 63–110.
- Fodor, Jerry, A. (1983): *The Modularity of Mind*. Cambridge, MA.: The MIT Press.
- Fodor, Jerry, A. (1987): *Psychosemantics*. Cambridge, MA.: The MIT Press.
- Fodor, Jerry, A. (1990): *A Theory of Content and Other Essays*. Cambridge, MA.: MIT Press.
- Fodor, Jerry (1994): *The Elm and the Expert. Mentalese and its Semantics*. Cambridge, MA.: Bradford Book/MIT Press.
- Fodor, Jerry, A. (1996a): Összefoglalás az elme modularitásához. In: Pléh Csaba (ed.): *Kognitív tudomány*. Budapest: Osiris–Láthatatlan Kollégium, pp. 197–206.
- Fodor, Jerry, A. (1996b): Fodor kalauza a mentális reprezentációhoz: az intelligens nagynéni segédlete. In: Pléh Csaba (ed.): *Kognitív tudomány*. Budapest: Osiris–Láthatatlan Kollégium, pp. 64–86.
- Fodor, Jerry, A. (1998): The Trouble with Psychological Darwinism. *London Review of Books* 20 no. 2.
- Fodor, Jerry A.–Pylyshyn, Zenon (1981): How Direct is Visual Perception? Some Reflections on Gibson's „Ecological Approach”. *Cognition* 9 pp. 139–196.
- Fodor, Jerry A.–Phylyshyn, Zenon (1988): Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition* 28 pp. 3–71.
- Forrai Gábor (1998): Hitek, vágyak és szemantikai tartalmuk. In: Pléh Csaba (ed.): *Megismeréstudomány és mesterséges intelligencia*. Budapest: Akadémiai, pp. 95–112.
- Frege, Gottlob (1953): *The Foundations of Arithmetics*. Oxford: Oxford University Press.
- Frege, Gottlob (1980): Jelentés és jelölet. In: Frege, Gottlob: *Logika, szemantika, matematika*. Budapest: Gondolat.
- Gardner, H. (1985): *The Mind's New Science*. New York: Basic Books.
- Gergely, György–Csibra, György (1998): A mentális viselkedésmagyarázatok teleológiai gyökere: Egy fejlődéslelektani hipotézis. In: Pléh Csaba (szerk.): *Megismeréstudomány és mesterséges intelligencia*. Budapest: Akadémiai Kiadó.
- Gergely, G.–Nádasdy, Z.–Csibra, G.–Bíró, S. (1995): Taking the Intentional Stance at 12 Month of Age. *Cognition* 56 pp. 243–259.
- Gibson, James J. (1966): *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.
- Gibson, James J. (1979): *An Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Godfrey–Smith, Peter (1994): A Modern History Theory of Functions. *Nous* 28 pp. 344–362.
- Gómez, Juan–Carlos (1996): Non–human Primate Theories of (Non–human Primate) Minds: Some Issues Concerning the Origins of Mind–reading. In: Carruthers, Peter–Smith, Peter K. (eds.): *Theories of Theories of Mind*. Cambridge: Cambridge University Press, pp. 330–343.
- Gould, Stephen Jay (1977): *Ontogeny and Philogeny*. Cambridge, MA.: Belknap.
- Gould, Stephen Jay (1990): *A panda hüvelykujja*. Budapest: Európa. Gould, Stephen Jay (1996): *Full House*. New York: Crown.
- Gould, Stephen Jay (1997): Evolution: The Pleasures of Pluralism. *New York Review of Books* June 26.
- Gould, Stephen Jay–Lewontin, R. (1979): The Sprandels of San Marco and the Panglossian Paradigm. *Proceedings of the Royal Society B205* pp. 581–598.
- Gould, Stephen Jay–Vrba, Elisabeth S. (1982): Exaptation—a Missing Term in the Science of Form. *Paleobiology* 8 pp. 4–15.
- Gould, Stephen Jay–Eldredge, N. (1993): Punctuated Equilibrium Comes of Age. *Nature* 366 pp. 223–227.
- Gregory, Richard L. (1974): *Az értelmes szem*. Budapest: Gondolat.
- Griffiths, P. E. (1993): Functional Analysis and Proper Functions. *British Journal for the Philosophy of Science* 44 pp. 409–422.
- Hare, R. M. (1952): *The Language of Morals*. Oxford: Clarendon Press.
- Harman, G. (1973): *Thought*. Princeton: Princeton University Press.
- Harman, G. (1982): Conceptual Role Semantics. *Notre Dame Journal of Formal Logic* 23 pp. 242–256.
- Hamad, Stevan [Hernád István] (1994): A szimbólumlehorgonyzás problémája. *Magyar Pszichológiai Szemle* 48–49 pp. 365–383.
- Haugeland, John (1996): A kognitívizmus jellege és kézenfekvő volta. In: Pléh Csaba (ed.): *Kognitív tudomány*. Budapest: Osiris–Láthatatlan Kollégium, pp. 35–63.
- Havelock, Eric A. (1963): *Preface to Plato*. Cambridge, MA.: Harvard University Press.
- Heil, John–Mele, Alfred (1993, eds.): *Mental Causation*. Oxford: Clarendon.
- Hellman, Geoffrey (1985): Determination of Logical Truth. *Journal of Philosophy* 82 pp. 607–616.
- Hendler, J. A. (1989a): Hybrid System. *Connection Science* 1 no. 2. Hendler, J. A. (1989b): Marker–passing over Microfeatures: Towards a Hybrid Symbolic/Connectionist Model. *Cognitive Science* 13 pp. 79–106.
- Heyes, C. M. (1994): Reflection on Self–recognition in Primates. *Behavioral and Brain Sciences* 16 pp. 524–525.
- Hinton, G. E. (1984): *Distributed Representations*. Pittsburgh: Carnegie–Mellon University Press.

- Hinton, G. E.–Dagan, P.–Frey, B.–Neal, R. M. (1995): The Wake-Sleep Algorithm for Self-organizing Neural Networks. *Science* 268 pp. 1158–1160.
- Hobbes, Thomas (1970): *Leviatán*. Budapest: Helikon.
- Hostetler, John A. (1963): *Amish Society*. Baltimore: John Hopkins University Press.
- Huffman, M. A.–Wrangham, R. W. (1994): Discovery of Medicinal Plant Use by Chimpanzees in the Wild. In: Wrangham, R. W.–McGrew, W. C.–de Waal, F. B.–Heltne, P. G. (1994): *Chimpanzee Cultures*. Cambridge, MA.: Harvard University Press.
- Hull, David L. (1989): *The Metaphysics of Evolution*. Albany: SUNY Press.
- Hume, David (1976): *Értekezés az emberi természetről*. Budapest: Gondolat.
- Hume, David (1995): *Tanulmány az emberi értelemről*. Budapest: Nippon.
- Hurford, J. R.–Studdert–Kennedy, M.–Knight, C. (1998): *Approaches to the Evolution of Language*. Cambridge: Cambridge University Press.
- Jackson, F. (1986): What Mary Didn't Know. *Journal of Philosophy* 83 no. 5. pp. 291–295.
- Jacob, Francois (1986): *A lehetséges és a tényleges valóság*. Budapest: Európa Könyvkiadó.
- James, William (1880): Great Men, Great Thoughts, and the Environment. *The Atlantic Monthly* 46 no. 276. pp. 441–59.
- James, William (1890): *Principles of Psychology*. New York: Henry Holt.
- Kampis, George (1991): *Self-modifying Systems in Biology and Cognitive Science*. Oxford: Pergamon.
- Kampis György (1999): Dennett és az evolúció. *Replika* no. 26. pp. 25–44.
- Kant, Immanuel (1980): Az emberiség egyetemes történetének eszméje világpolgári szemszögből. In: Kant, Immanuel: *A valóság a pusztán ész határain belül és más írások*. (Budapest:) Gondolat.
- Kant, Immanuel (1997): *A tiszta ész kritikája*. Szeged: Ictus.
- Karmiloff–Smith, Anette (1992): *Beyond Modularity*. Cambridge, MA.: The MIT Press.
- Karmiloff–Smith, Anette (1996): Túl a modularitáson: a kognitív tudomány fejlődésméleti megközelítése. In: Pléh Csaba (ed.): *Kognitív tudomány*. Budapest: Osiris–Láthatatlan Kollégium, pp. 254–282.
- Katz, Jerrold J. (1990): *The Methaphysics of Meaning*. Cambridge, MA.: The MIT Press.
- Kim, Jaegwon (1984a): Concepts of Supervenience. *Philosophy and Phenomenological Research* 45 pp. 153–176.
- Kim, Jaegwon (1984b): Epiphenomenal and Supervenient Causation. *Midwest Studies in Philosophy* 9 pp. 257–270.
- Kim, Jaegwon (1987): 'Strong' and 'Global' Supervenience Revisited. *Philosophy and Phenomenological Research* 48 pp. 315–326.
- Kim, Jaegwon (1989): The Myth of Nonreductive Materialism. *Proceedings and Addresses of the American Philosophical Association* 63 no. 3. pp. 31–47. Kim, Jaegwon (1990): Explanatory Exclusion and the Problem of Mental Causation. In: Villanueva E. (ed.): *Information, Semantics and Epistemology*. Oxford: Blackwell. pp. 36–56.
- Kim, Jaegwon (1993): Can Supervenience and 'Non-Strict Laws' Save Anomalous Monism? In: Heil, John–Mele, Alfred (eds.): *Mental Causation*. Oxford: Clarendon, pp. 19–26.
- Kim, Jaegwon (1994): Supervenience. In: Guttenplan, Samuel (ed.): *A Companion to the Philosophy of Mind*. Oxford: Blackwell. pp. 575–588.
- Kim, Jaegwon (1998): *Mind in a Physical World*. Cambridge, MA.: The MIT Press.
- Kitchner, Philip (1982): *Abusing Science*. Cambridge, MA.: The MIT Press.
- Kovács Ilona (1991): Egy tudományos vízió. *Pszichológia* 11 pp. 75–112.
- Kripke, Saul (1982): *Wittgenstein on Rules and Private Language*. Cambridge, MA.: Harvard University Press.
- Lamarck, Jean–Baptiste (1809): *Zoologie philosophique*. Paris.
- Leibniz, Gottfried Wilhelm (1986): *Válogatott filozófiai írások*. Budapest: Európa.
- LePore, Ernest–Loewer, Barry (1987): Mind Matters. *Journal of Philosophy* 84 pp. 630–642.
- LePore, Ernest–Van Gulick, Robert (1991, eds.): *John Searle and His Critics*. Oxford: Blackwell. Leslie, Alan M. (1987): Pretense and Representation in Infancy: The Origins of the "Theory of Mind". *Psychological Review* 94 no. 4. pp. 412–426.
- Lewis, David (1972): Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy* 50 no. 3. pp. 249–258.
- Lewis, David (1980): Mad Pain and Martial Pain. In: Block, Ned (ed.): *Readings in the Philosophy of Psychology*. Cambridge, MA.: Harvard University Press.
- Lewontin, Richard–Rose, Steven–Kamin, Leon (1984): *Not in our Genes: Biology, Ideology and Human Nature*. New York: Pantheon.
- Lieberman, Philip (1984): *The Biology and Evolution of Language*. Cambridge, MA.: Harvard University Press.
- Lieberman, Philip (1998): *Eve Spoke*. New York: W. W. Norton. Lima, S. D.–Corrigan, R. L.–Iverson, G. K. (1994, eds.): *The Reality of Linguistic Rules*. Amsterdam: Benjamins.
- Llull, Ramon et al. (1993): *Doctor Illuminatus: A Ramon Hull Reader*. Princeton: Princeton University Press.
- Locke, John (1979): *Értekezés az emberi értelemről*. Budapest: Akadémiai.
- Locke, John (1986): *Értekezés a polgári kormányzatról*. Budapest: Gondolat.
- Lukowski, Jerry (1994): Recasting Utopia: Montesquieu, Rousseau and the Polish Constitution of 1791. *Historical Journal* 37 pp. 65–88.
- Lumsden, Charles J.–Wilson, Edward O. (1981): *Genes, Mind and Culture: The Coevolutionary Process*. Cambridge, MA.: Harvard University Press.
- Lycan, W. G. (1987): *Consciousness*. Cambridge, MA.: The MIT Press.
- Lycan, W. G. (1990, ed.): *Mind and Cognition*. Cambridge, MA.: Blackwell.
- Macdonald, C.–Macdonald, G. (1995, eds.): *Connectionism: Debates on Psychological Explanation*. Oxford: Blackwell.
- Mach, E. (1927): *Az érzetek elemzése*. Budapest: Franklin.
- Mackie, J. L. (1985): Mind, Brain and Causation. In: Mackie, J. L.: *Selected Papers*. Oxford: Clarendon, pp. 131–144.

- Marcus, Gary F. (1998): Can Connectionism Save Constructivism? *Cognition* 66 pp. 153–182.
- Marr, David (1982): *Vision*. San Francisco: W. H. Freeman.
- Marras, Ausonio (1993): Psychophysical Supervenience and Nonreductive Materialism. *Synthese* 95 pp. 275–304.
- Maturana, Humberto, R.–Varela, Francisco, J. (1980): *Autopoiesis and Cognition*. Boston: D. Reidel.
- Maturana, Humberto, R.–Varela, Francisco, J. (1987): *The Tree of Knowledge*. London: New Science Library.
- Maynard Smith, John (1958): *The Theory of Evolution*. Cambridge: Cambridge University Press.
- Maynard Smith, John–Szathmáry Eörs (1997): *Az evolúció nagy lépései*. Budapest: Scientia.
- McClelland, J. L.–Rumelhart, D.E. (1986a): *Parallel Distributed Processing. Vol. 2*. Cambridge, MA.: MIT Press.
- McClelland, J. L.–Rumelhart, D.E. (1986b): Amnesia and Distributed Memory. In: McClelland, J. L.–Rumelhart, D.E.: *Parallel Distributed Processing. Vol. 2*. Cambridge, MA.: MIT Press.
- McGinn, Colin (1991): *The Problem of Consciousness*. Oxford: Blackwell.
- McGrew, W. C. (1994): *Chimpanzee Material Culture*. Cambridge: Cambridge University Press.
- McGrew, W. C. (1998): Non–Human Culture and the Animal Mind. *Evolving the Human Mind Conference*, 1998. June 24–27., Sheffield.
- McLaughlin, Brian P. (1993): On Davidson’s Response to the Charge of Epiphenomenalism. In: Heil, John–Mele, Alfred (eds.): *Mental Causation*. Oxford: Clarendon, pp. 27–40.
- Merleau–Ponty, Maurice (1945): *La phénoménologie de la perception*. Paris: Gallimard.
- Millikan, Ruth G. (1984): *Language, Thought and Other Biological Categories*. Cambridge, MA.: MIT Press.
- Millikan, Ruth G. (1989): Biosemantics. *Journal of Philosophy* no. 86. pp. 281–297.
- Millikan, Ruth G. (1993): *White Queen Psychology and other Tales for Alice*. Cambridge, MA.: MIT Press.
- Mills, Susan K.–Beatty, John H. (1979): The Propensity Interpretation of Fitness. *Philosophy of Science* 46 pp. 263–286.
- Minsky, M.–Papert, S. (1969): *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA.: The MIT Press.
- Mithen, Steven (1996): *The Prehistory of the Modern Mind*. London: Phoenix.
- Montesquieu (1981): *Perzsa levelek*. Budapest: Európa Könyvkiadó.
- Nagel, Ernest (1961): *The Structure of Science*. New York: Harcourt.
- Nagel, Ernest (1974): What is it Like to Be a Bat? *Philosophical Review* 83 pp. 435–450.
- Nánay Bence (1996): Új divat a tudatfilozófiában: a konnekciónizmus. *BUKSZ* 8, no. 3. pp. 262–269.
- Nánay Bence (1997a): Szavak és reprezentációk. *Magyar Filozófiai Szemle* nos. 5–6. pp. 805–826.
- Nánay Bence (1997b): A filozófia és a természettudományok találkozása: a kognitív tudomány. *BUKSZ* 9 no. 2. pp. 148–157.
- Nánay Bence (1997c): James J. Gibson affordanciaelmélete és a kortárs percepciókutatás. *Pszichológia* no. 1. pp. 53–76.
- Nánay Bence (1998a): A szupervencia evolúciója. In: Pléh Csaba (ed.): *Megismeréstudomány és mesterséges intelligencia*. Budapest: Akadémiai, pp. 83–93.
- Nanay, Bence (1998b): La natura della riduzione. (trans. Sandra Cavallo) In: *MONTAG IV*. Roma: Edizione Fahrenheit 451. pp. 139–146.
- Nánay Bence (1999): Evolúció és elme. *Magyar Pszichológiai Szemle LIV* no. 2. pp. 215–228.
- Neander, K. (1991): Functions as Selected Effects. *Philosophy of Science* 58 pp. 168–184.
- Nelson, R. J. (1976): Mechanism, Functionalism, and the Identity Theory. *Journal of Philosophy* 73 pp. 364–386.
- Neuman János (1978): *A számológép és az agy*. Budapest: Magvető.
- Neurath, Ottó (1972): A fizikaiizmus szociológiája. In: Altrichter Ferenc (ed.): *A Bécsi Kör filozófiája*. Budapest: Gondolat, pp. 551–574.
- Newell, A. (1980): Physical Symbol Systems. *Cognitive Science* no. 4. pp. 251–283.
- Newell, A. (1990): *Unified Theories of Cognition*. Cambridge, MA.: Harvard University Press.
- Newell, A.–Simon, Herbert (1972): *Human Problem Solving*. Cambridge, MA.: Harvard University Press.
- Nussbaum, Martha–Rorty, Amélie Oksenberg (1992, eds.): *Essays on Aristotle’s ‘De Anima’*. Oxford: Clarendon Press.
- Nussbaum, Martha–Putnam, Hilary (1992): Changing Aristotle’s Mind. In: Nussbaum, Martha–Rorty, Amélie Oksenberg (eds.): *Essays on Aristotle’s ‘De Anima’*. Oxford: Clarendon Press, pp. 27–56.
- Nyíri Kristóf (1994): *A hagyomány filozófiája*. Budapest: T–Twins.
- Osherson, N. et al. (1990, eds.): *An Invitation to Cognitive Science*. Cambridge, MA.: The MIT Press.
- Papineau, David (1987): *Reality and Representation*. Oxford: Blackwell.
- Papineau, David (1993): *Philosophical Naturalism*. Oxford–Cambridge, MA.: Blackwell.
- Papineau, David (1996): *Precis of Philosophical Naturalism*. *Philosophy and Phenomenological Research* 56 no. 3. pp. 657–665.
- Paull, R. Cranston–Sider, Theodore R. (1992): In Defense of Global Supervenience, *Philosophy and Phenomenological Research* 52 pp. 833–854.
- Peirce, Charles Sanders (1975): A jelek felosztása. In: Horányi Özséb–Szépe György (eds.): *A jel tudománya*. Budapest: Gondolat.
- Pereboom, Derek–Kornblith, Hilary (1991): The Methaphysics of Irreducibility. *Philosophical Studies* 63 pp. 125–145.
- Perner, Josef (1991): *Understanding the Representational Mind*. Cambridge, MA./London: Bradford Books/The MIT Press.
- Petrie, Bradford (1987): Global Supervenience and Reduction. *Philosophy and Phenomenological Research* 48 pp. 119–130.
- Piaget, Jean (1962): *Play, Dreams and Immitation in Childhood*. London: Routledge and Kegan Paul.
- Piatelli–Palmarini, Massimo (1996): Evolúció, szelekció és megismerés. In: Pléh Csaba (ed.): *Kognitív tudomány*. Budapest: Osiris–Láthatatlan Kollégium.
- Pinker, Stephen (1997): *How the Mind Works*. New York: Norton.
- Pinker, Steven (1999): *A nyelvi ösztön*. Budapest: Typotex.
- Pinker, Steven–Bloom, Paul (1990): Natural Language and Natural Selection. *Behavioral and Brain Sciences* 13 pp. 707–784.
- Pinker, Steven–Mehler, J. (1988, eds.): *Connections and Symbols*. Cambridge, MA.: The MIT Press.

- Pinker, Steven–Prince, A. (1988): On Language and Connectionism: Analysis of Parallel Distributed Processing Model of Language Acquisition. *Cognition* 28 pp. 73–193.
- Pinker, Steven–Prince, A. (1994): Regular and Irregular Morphology and the Psychological Status of Rules of Grammar. In: Lima, S. D.–Corrigan, R. L.–Iverson, G. K. (eds.): *The Reality of Linguistic Rules*. Amsterdam: Benjamins, pp. 321–350.
- Pléh Csaba (1986): Az elme modularitása. *Magyar Filozófiai Szemle* 30 pp. 211–216.
- Pléh Csaba (1991): Szükséglet vagy divat? A megismeréstudományról. *BUKSZ* 3 pp. 227–235.
- Pléh Csaba (1992): Az asszociáció reneszánsza a kognitív pszichológiában. *Janus* 9 no. 2. pp. 12–22.
- Pléh Csaba (1995): Popper és a pszichológia. *Replika* 17–18, pp. 67–86.
- Pléh Csaba (1996, ed.): *Kognitív tudomány*. Budapest: Osiris–Látha–tatlan Kollégium.
- Pléh Csaba (1997, ed.): *A megismeréskutatás egy új útja: a párhuzamos feldolgozás*. Budapest: Typotex.
- Pléh Csaba (1998a): Az evolúciós pszichológia. *Magyar Tudomány* pp. 1054–1060.
- Pléh Csaba (1998b): *Bevezetés a megismeréstudományba*. Budapest: Typotex.
- Pléh Csaba (1998c): Ernst Mach és Daniel Dennett: A megismerés két evolúciós modellje. In: Pléh Csaba (ed.): *Megismeréstudomány és mesterséges intelligencia*. Budapest: Akadémiai, pp. 41–58.
- Pléh Csaba (1999): Az intencionalitás mint a kognitívizmus alapkategóriája. In: Pléh Csaba: *Hagyomány és újítás a pszichológiában*. Budapest: Osiris.
- Plotkin, Henry C. (1997): *Evolution in Mind*. New York: A Lane.
- Poggio, T.–Edelman, S. (1990): A Network that Learns to Recognise Three-dimensional Objects. *Nature* 343 pp. 263–266.
- Polanyi, Michael (1968): Life's Irreducible Structure. *Science* 160 pp. 1308–1312.
- Popper, Karl (1963): *Conjectures and Refutations*. London: Routledge.
- Popper, Karl (1972): *Objective Knowledge: An Evolutionary Approach*. Oxford: Clarendon.
- Popper, Karl (1996): *Megismerés, történelem, politika*. Budapest: Aduprint.
- Popper, Karl (1998): *Szüntelen keresés*. Budapest: Áron.
- Premack, D.–Woodruff, G. (1978): Do Chimpanzee Have a Theory of Mind? *Behavioral and Brain Sciences* 4 pp. 515–526.
- Putnam, Hilary (1967): Psychological Predicates. In: Capitan, W. H.–Merrill, D. D. (eds.): *Art, Mind and Religion*. Pittsburgh: University of Pittsburgh Press.
- Putnam, Hilary (1975a): The Meaning of 'Meaning'. In: Putnam, Hilary: *Mind, Language and Reality*. Cambridge: Cambridge University Press.
- Putnam, Hilary (1975b): *Mind, Language and Reality*. Cambridge: Cambridge University Press.
- Putnam, Hilary (1981): Brains in a Vat. In: Putnam, Hilary: *Reason, Truth and History*. Cambridge: Cambridge University Press.
- Pylyshyn, Zenon W. (1978): When is Attribution of Belief Justified? *Behavioral and Brain Sciences* 1 pp. 592–593.
- Pylyshyn, Zenon W. (1984): *Computation and Cognition*. Cambridge, MA.: MIT Press. Quine, W. V. O. (1960): *Word and Object*. Cambridge, MA.: The MIT Press.
- Quine, W. V. O. (1998): Naturalizált episztemológia. In: Forrai Gábor–Szegedi Péter (eds.): *Tudományfilozófia*. Budapest: Osiris.
- Ramsey, W.–Stich, S. P.–Rumelhart, D. E. (1991) (eds.): *Philosophy and Connectionist Theory*. Hillsdale, NJ: Erlbaum.
- Rock, Irwin (1983): *The Logic of Perception*. Cambridge, MA.: The MIT Press.
- Rosenberg, A. (1985): *The Structure of Biological Science*. Cambridge: Cambridge University Press. Rosenblatt, F. (1962): *Principles of Neurodynamics*. New York: Spartan.
- Rosenthal, D. (1986): Two Concepts of Consciousness. *Philosophical Studies* 49 pp. 329–359.
- Rousseau, Jean–Jacques (1963): *A társadalmi szerződés*. Budapest: Bibliotheka.
- Rousseau, Jean–Jacques (1989): *Oeuvres politiques*. Paris: Bordas.
- Rumelhart, D.–McClelland, J.L. (1986a, eds.): *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1*. Cambridge, MA.: MIT Press.
- Rumelhart, D.–McClelland, J.L. (1986b): On Learning the Past Tenses of English Verbs. In: McClelland, J. L.–Rumelhart, D.E. (eds.): *Parallel Distributed Processing. Vol. 2*. Cambridge, MA.: MIT Press.
- Rumelhart, D.–McClelland, J. L. (1987): Learning the Past Tenses of English Verbs: Implicit Rules or Parallel Distributed Processing? In: McWhinney, B. (ed.): *Mechanisms of Language Acquisition*. Hillsdale, NJ.: Erlbaum.
- Russell, Bertrand (1956): On Denoting. In: Marsh, R. C. (ed.) *Logic and Knowledge*. London: Allen and Unwin.
- Russell, Bertrand (1971): Descriptions. In: Rosenberg, J.–Travis, C. (eds.): *Readings in the Philosophy of Language*. New Jersey: Prentice–Hall.
- Ryle, Gilbert (1974): *A szellem fogalma*. Budapest: Gondolat.
- Ryle, Gilbert (1990): *Selected Papers*. Bristol: Thoemmes.
- Schauer, Frederick (1991): *Playing by the Rules*. Oxford: Clarendon.
- Searle, John R. (1983): *Intentionality*. Cambridge: Cambridge University Press.
- Searle, John R. (1992): *Rediscovery of the Mind*. Cambridge, MA.: The MIT Press.
- Searle, John R. (1993): The Failure of Computationalism. *THINK* 2 no. 1. Powers, D. M. W.–Flach, P. A. (eds.) *Special Issue on Connectionism versus Symbolism*.
- Searle, John R. (1995): *The Construction of Social Reality*. New York: Free Press.
- Searle, John R. (1996): Az elme, az agy és a programok világa. In: Pléh Csaba (ed.): *Kognitív tudomány*. Budapest: Osiris–Láthatatlan Kollégium, pp. 136–151.
- Sellars, Wilfred (1963): *Science, Perception and Reality*. New York: Humanities Press.
- Shastri, L.–Ajjanagadde, V. (1993): From Simple Associations to Systematic Reasoning: A Connectionist Representation of Rules, Variables and Dynamic Bindings Using Temporal Synchrony. *Behavioral and Brain Sciences* 16 pp. 417–494.

- Shephard, R. N.–Metzler, J. (1971): Mental Rotation of Three Dimensional Objects. *Science* 171 pp. 701–703.
- Shoemaker, S. (1975): Functionalism and Qualia. *Philosophical Studies* 79 pp. 271–315.
- Shweder, R. A. (1991): Cultural Psychology: What it is? In: Shweder, R. A.: *Thinking Through Cultures*. Cambridge, MA.: Harvard University Press.
- Skinner, B. F. (1953): *Science and Human Behavior*. New York: Macmillan.
- Smart, J. J. C. (1971): Reports of Immediate Experience. *Synthese* 22 pp. 346–359.
- Smolensky, Paul (1996): A konnekciónizmus helyes kezeléséről. In: Pléh Csaba (ed.): *Kognitív tudomány*. Budapest: Osiris, pp. 87–135.
- Smolensky, P.–Legendre, G.–Miyata, Y. (1992): *Principles for an Integrated Connectionist/Symbolic Theory of Higher Cognition*. Report 92–08. Institute of Computer Science, University of Colorado.
- Sober, Elliott–Wilson, David Sloan (1998): *Unto Others*. Cambridge, MA.: Harvard University Press.
- Sosa, Ernest (1984): Mind–Body Interaction and Supervenient Causation. *Midwest Studies in Philosophy* 9 pp. 271–281.
- Sosa, Ernest (1993): Davidson’s Thinking Causes. In: Heil, John–Mele, Alfred (eds.): *Mental Causation*. Oxford: Clarendon. pp. 41–50.
- Sperber, Dan (1985): Anthropology and Psychology: Towards an Epidemiology of Representations. *Man* 20 pp. 73–89.
- Sperber, Dan (1994): The Modularity of Thought and the Epidemiology of Representations. In: Hirschfield, Lawrence A.–Gelman, Susan A. (eds.): *Mapping the Mind: Domain Specificity in Cognition and Culture*. Cambridge: Cambridge University Press.
- Sperber, Dan–Origg, Gloria (1998): Issues in the Evolution of Human Language and Communication. *Evolving the Human Mind Conference*, 1998. June 24–27., Sheffield.
- Stalnaker, Robert (1984): *Inquiry*. Cambridge, MA.: The MIT Press. Stalnaker, Robert (1999): *Context and Content: Collected Papers on Intentionality in Speech and Thought*. Oxford: Oxford University Press.
- Stich, Steven (1983): *From Folk Psychology to Cognitive Science*. Cambridge, MA.: The MIT Press.
- Stich, Steven (1998): Darwin in the Madhouse. *Evolving the Human Mind Conference*, 1998. June 24–27., Sheffield.
- Stone, Tony–Davies, Martin (1995, eds.): *Mental Simulation: Evaluations and Applications*. Oxford: Blackwell.
- Thompson, John N. (1994): *The Coevolutionary Process*. Chicago: University of Chicago Press.
- Tomasello, Michael–Call, Josep (1997): *Primate Cognition*. Oxford: Oxford University Press.
- Turing, M. A. (1964): Számítógépek és gondolkodás. In: Szálai Sándor (ed.): *A kibernetika klasszikusai*. Budapest: Gondolat, pp. 120–160.
- Tye, M. (1995): *Ten Problems of Consciousness*. Cambridge, MA.: The MIT Press.
- Ullman, S. (1980): *Against Direct Perception*. New York: MacMillan.
- Varela, F.–Thompson, E.–Rosch, E. (1992) *The Embodied Mind*. Cambridge, MA.: Harvard University Press.
- Wallace, Ruth A.–Hartey, Sherley F. (1988): Religious Elements in Friendship: Durkheimian Theory in an Empirical Context. In: Alexander, Jeffrey C. (ed.): *Durkheimian Sociology: Cultural Studies*. Cambridge: Cambridge University Press.
- Walsh, D. M. (1996): Fitness and Function. *British Journal for the Philosophy of Science* 47 pp. 553–574.
- Weismann, August (1893): *The Germ Plasm: A Theory of Heredity*. London: Scott.
- Whyte, J. T. (1990): Success Semantics. *Analysis* 50 no. 3. pp. 148–157.
- Whyte, J. T. (1991): The Normal Rewards of Success. *Analysis* 51 no. 2. pp. 65–73.
- Williams, George C. (1966): *Adaptation and Natural Selection*. Princeton: Princeton University Press.
- Wilson, David Sloan (1989): Levels of Selection: An Alternative to Individualism in Biology and the Human Sciences. *Social Networks* 11 pp. 257–272.
- Wilson, David Sloan–Sober, Elliott (1994): Re–introducing Group Selection to the Human Behavior Sciences. *Behavioral and Brain Sciences* 77 pp. 585–608.
- Wilson, Edward O. (1975): *Sociobiology: A New Synthesis*. Cambridge, MA.: Harvard University Press.
- Wilson, Edward O. (1978): *On Human Nature*. Cambridge, MA.: Harvard University Press.
- Wilson, Robert A.–Keil, Frank (1999, eds.): *The MIT Encyclopedia of the Cognitive Science*. Cambridge, MA.: The MIT Press.
- Withen, Andrew (1996): When Does Smart Behaviour–reading Become Mind–reading? In: Carruthers, Peter–Smith, Peter K. (eds.): *Theories of Theories of Mind*. Cambridge: Cambridge University Press, pp. 277–292.
- Withen, Andrew–Byrne Richard W. (1997, eds.): *Machiavellian Intelligence II*. Cambridge: Cambridge University Press.
- Wittgenstein, Ludwig (1989): *Logikai–filozófiai értekezés*. Budapest: Akadémiai.
- Wittgenstein, Ludwig (1992): *Filozófiai vizsgálódások*. Budapest: Atlantisz.
- Wrangham, R. W.–McGrew, W. C.–de Waal, F. B.–Heltne, P. G. (1994): *Chimpanzee Cultures*. Cambridge, MA.: Harvard University Press.
- Wright, L. (1973): Functions. *Philosophical Review* 82 pp. 139–168.
- Wright, Robert (1994): *The Moral Animal: Why We Are the Way We Are: The New Science of Evolutionary Psychology*. New York: Random House.

Kávé Kiadó, 2000  
 Felelős vezető Nyíri András  
 Műszaki vezető Fraunhoffer Péter  
 Készült a Főnix nyomdában  
 Felelős vezető Soskóné Müller Judit  
 ISBN 963 9169 25 0 (kötött)  
 ISBN 963 9169 26 9 (füzött)  
 ISSN 1418-6624