

## BEJELENTKEZÉS: Mesterséges intelligencia

### PEIRCE ÉS AZ AI\* FILOZÓFIÁJA

James H. Fetzer

**Összefoglaló:** A gondolkodás, a mentalitás és az intelligencia természetével kapcsolatos történelmi álláspontok filozófiai értékelése, amely Descartes, Turing, Newell és Simon nézeteivel kezdődik, de Haugeland, Fodor, Searle és más jelentős tudósok munkásságát is felöleli. Az alapkérdések a szintaxis, a szemantika és a pragmatika közötti különbségtételre vonatkoznak, ahol a fizikai számítógépeket a jeleket manipuláló vagy szintaxis-feldolgozó mechanizmusoknak tűnik a legmegfelelőbbnek. Alternatív beszámolók születtek arról, hogy mi kell ahhoz, hogy gondolkodó dolog legyen, többek között, hogy Turing-gépek, szimbólumrendszerek, szemantikai motorok és szemiotikai rendszerek, amelyek a Charles S. Peirce-i értelemben vett jelek használatára képesek. A reprezentációk természetével és a mentális algoritmusok létezésével kapcsolatos elméletek azt sugallják, hogy az elmék szemiotikai rendszerekként való elmélete előnyben részesítendő az alternatíváival szemben, ahol a digitális számítógépek még elme nélkül is "intelligens gépeknek" minősülhetnek.

**KULCSSZÓSZÓK:** gondolkodó dolgok, szimbólumrendszerek, szemantikus motorok, szemiotikai rendszerek, jelek és elmék, tudat és megismerés.

#### Történelmi háttér

A számítástechnikai gépek megjelenése előtt a mentalitás és a gondolkodás természetével kapcsolatos elméletek felállítása elsősorban a filozófusok hatáskörébe tartozott, akik közül történelmileg talán a legnagyobb hatást René Descartes (1596-1650) gyakorolta, akit általában "a modern filozófia atyjának" neveznek. Descartes egy *ontikus* (vagy ontológiai) tézist állított fel arról, hogy az elmék a világ jellemzőiként milyen dolgok, és egy *episztemikus* (vagy ismeretelméleti) tézist arról, hogy az ilyen jellegű dolgokat hogyan lehet megismerni. Descartes szerint, aki a dualizmus egy olyan formáját képviselte, amelyben az elme és a test egymást kölcsönösen kizáró kategóriák, az "elmék" olyan dolgok, amelyek képesek gondolkodni, ahol az elmékhez való hozzáférés az "introspekciónak" néven ismert képesség segítségével biztosítható, amely egyfajta befelé irányuló észlelés a személy saját mentális állapotairól.

Descartes megközelítése még a 20. században is óriási hatást gyakorolt, amikor a digitális számítógépek fejlődése kezdte megragadni azok fantáziáját, akik a gondolkodó dolgok természetének tudományosabb és kevésbé szubjektív felfogását

keresték. A legfontosabb újításokat Alan Turing (1912-54) brit matematikus vezette be, akit (posztumusz) a "számítógépek atyja" címre jogosítottak fel. tudomány", sőt a mesterséges intelligencia (vagy "AI") is nehezen tagadhatónak tűnik. Turing

legfontosabb kutatásai a matematikán belüli bizonyítás korlátaira vonatkoztak, ahol azt javasolta, hogy a kiszámítható (olyan matematikai problémák, amelyek megoldása logikai szabályok véges alkalmazásai alapján elérhető) határai megegyeznek azokkal, amelyek egy bizonyos fajta problémamegoldó gépezet segítségével megoldhatók.

Az ilyen jellegű dolgok, amelyeket *Turing-gépeknek* nevezünk, egy tetszőlegesen hosszú, szegmentált szalagból és egy olyan eszközből állnak, amely négy műveletre képes a szalagon, nevezetesen: jelet tenni, jelet eltávolítani, a szalagot egy szegmenssel előre és egy szegmenssel hátrafelé mozgatni. (A szalag állapotát a műveletsorozat alkalmazása előtt "bemenetnek", a szalag állapotát az alkalmazás után "kimenetnek", az utasítássorozatot pedig "programnak" nevezhetjük). E gépek szemszögéből nézve nyilvánvalóvá vált, hogy vannak olyan matematikai problémák, amelyekre nem léteznek véges vagy kiszámítható megoldások. Hasonló eredményekre, amelyek a hatékony eljárásokat a kiszámítható problémákhoz kapcsolják, egyidejűleg jutott az amerikai logikus, Alonzo Church.

## **A Turing-teszt**

Church munkája tisztán matematikai feltevéseken alapult, míg Turing munkája egy nagyon speciális típusú gépre hivatkozott, amely absztrakt modellt nyújtott a "(digitális) számítógépeket" megfelelően meghatározó eljárások fizikai megtestesítésére, és megalapozta a számítástechnika elméletét. Turing azzal érvelt, hogy az ilyen eljárások korlátokat szabnak az emberi gondolkodásnak, és ezáltal a program és az elme fogalmát egy olyan gép formájában kombinálta, amely elvileg sokféle fizikai megvalósításra lehet képes. Munkája így vezette be az *elme számítástechnikai felfogását*, amely megfordítja a gépeket elméletlennek tekintő karteziánus felfogást azáltal, hogy magát az elmét is különleges gépfajtává változtatja,

ahol a számíthatóság határai határozzák meg a gondolkodás határait.

Turing állítása, miszerint ő a mesterséges intelligencia atyja, a *Turing-tesztnek* nevezett eljárás bevezetésén alapul, amelyben egy vagy több egyfajta dolgot állítanak szembe egy vagy több másfajta dologgal. Egy olyan partijátékot adaptálva, amelyben egy férfi és egy nő versenyezhet abban, hogy a férfi képes-e megtéveszteni egy versenyzőt, hogy az összetéveszse őt a nővel (olyan kontextusban, amely nem árulja el a játékot), azt javasolta, hogy egy embert állítsanak szembe egy élettelen géppel (amely megfelelő programmal és kommunikációs móddal van felszerelve). Ha tehát egy beszélgetőpartner nem tudott különbséget tenni közöttük a feltett kérdésekre adott válaszok alapján, akkor ezeket a rendszereket (az általa vélt) *intelligencia tekintetében* egyenlőnek (vagy ekvipotensnek) kell tekinteni (Turing 1950).

Ez három különböző szempontból is figyelemre méltó előrelépést jelentett a karteziánus elképzelésekhez képest. Először is, a gondolkodó dolog homályos fogalmán javított a Turing-gép pontos fogalmának bevezetésével, mint olyan eszköz, amely egy program irányítása alatt képes a jelek manipulálására. Másodsor, az elme/test probléma megoldását sugallta, amely szerint a hardver olyan a szoftverhez, mint a test az elméhez, és amely kevésbé metaforikus és tudományosabb volt, mint az elmével rendelkező testek fogalma. Harmadszor, a gondolkodó dolgok létezésére való következtetést alátámasztó empirikus bizonyítékokért az introspekció helyett egy viselkedési kritériumra hivatkozott, ami az elme tanulmányozását sokkal kevésbé szubjektívnek tüntette fel.

## **Fizikai gépek**

Így Descartes elképzelése az emberi elméről mint gondolkodó dolgokról attól függ, hogy valóban vannak-e gondolataik, ami nem biztos, hogy így van, ha nem tudatosak (mondjuk, alszanak, elkábítva vannak, vagy más módon képtelenek a gondolkodásra), mivel gondolkodó dolgokként való létezésük ekkor nem lenne önvizsgálat tárgya, és ez

alátámasztja a (h1) hipotézist:

(h1) Az (öntudatos) emberi elmék gondolkodó dolgok (Descartes);

Analóg módon, Turing felfogása, hogy ezek a gépek gondolkodó dolgok, attól függ, hogy a

a védjegyek manipulálására való képesség gyakorlása elegendő feltétel a birtokláshoz intelligencia, amely összehasonlítható az emberekével, ami a (h2) hipotézisre utal: (h2) A

jeleket manipuláló Turing-gépek intelligenciával rendelkeznek (Turing);

ahol az *intelligencia* és a *mentalitás* azonosítása alátámasztja azt a következtetést, hogy a megfelelően programozott és megfelelően működő Turing-gépek ember alkotta gondolkodó dolgoknak, vagy John McCarthy kifejezésével élve "mesterséges intelligenciának" minősülhetnek.

A Turing-gépek azonban, mint idealizált eszközök, amelyek olyan tulajdonságokkal rendelkeznek, amelyekkel a fizikai rendszerek nem rendelkezhetnek, beleértve a tetszőleges hosszúságú, szegmentált szalagokat (vagy "memóriákat") és a tökéletes teljesítményt, absztrakt entitások. Mivel nem léteznek térben/időben, képtelenek bármilyen kauzális hatást gyakorolni a térben/időben lévő dolgokra, még akkor is, ha a definíció szerint pontosan a szándéknak megfelelően teljesítenek (Fetzer 1988). A különbségtétel analóg a számok és a számjegyek közötti különbségtétellel, ahol a számok olyan absztrakt entitások, amelyek nem léteznek térben/időben, míg a számjegyek, amelyek őket jelölik, a következők: "A számok nem léteznek térben/időben.

fizikai dolgok, amelyek valóban léteznek a térben/időben. A római számoknak, arab számoknak és hasonlóknak meghatározott helyük van meghatározott időben, meghatározott alakjuk és méretük van, bejönnek és eltűnnek, és egyik sem igaz a számokra mint időtlen és változatlan absztrakt entitásokra.

Ezek a "gépek" mindazonáltal legalább részben megvalósíthatók fizikai dolgokként, különböző módon, különböző anyagok felhasználásával, például 0-ak és 1-ek digitális sorozataival, "be" vagy "ki" kapcsolókkal, vagy nagyobb és kisebb feszültséggel.

Egyesek vákuumcsövekből, mások tranzistorokból, megint mások szilíciumchipekből készülhetnek. Ezek aztán fizikai dolgok példányai lesznek, a magukfajta dolgok véges tulajdonságaival. Egyikük sem működik pontosan úgy, ahogyan az a szándékukban áll, pusztán definíció kérdése: mindegyikükben megvan a hibás működés és a változó teljesítmény lehetősége, mint a repülőgépekben, gépkocsikban, televíziókészülékekben és más fizikai eszközökben. Emlékezetük



meghatározott fizikai tulajdonságok határozzák meg, mint például a regiszterek mérete; és bár több memória hozzáadásával növelhetők, egyik sem végtelen.

## Symbol Systems

Míg Isten (egyres elképzelései) egy időtlen és változatlan gondolkodó dolog példajaként szerepelhetnek, az ilyen entitások létezése túlmutat az empirikus és tudományos vizsgálatok hatókörén. A számítástechnikán belül a Turing-féle megközelítés legszélesebb körben elfogadott és nagy hatású adaptációja az Alan Newell és Herbert Simon által kidolgozott *fizikai szimbólumrendszer-felfogás* volt, ahol a szimbólumrendszerek olyan fizikai gépek - esetleg emberek -, amelyek fizikai szimbólumstruktúrákat dolgoznak fel az időben (Newell és Simon 1976). Ezek a digitális gépek speciális fajtái, amelyek soros feldolgozó (vagy von Neumann) gépeknek minősülnek. Így Turing koncepcióját egy fizikai gép hipotézis segítségével valósítják meg (h3),

(h3) A szimbólumokat manipuláló fizikai számítógépek intelligensek (Newell és Simon);

ahol, mint Turing esetében, az "intelligens dolog" kifejezés ugyanazt jelenti, mint a "gondolkodó dolog".

A "szimbólumrendszerek" szavakkal kapcsolatban van egy kétértelműség a szimbólumokat feldolgozó rendszerek és az általuk feldolgozott szimbólumrendszerek tekintetében, ahol Newell és Simon nagyobb figyelmet fordított a gépek által feldolgozott szimbólumrendszerekre, mint a szimbólumokat feldolgozó rendszerekre. De kétségtelen, hogy természetesnek vették, hogy az ezeket a szimbólumokat feldolgozó rendszerek fizikaiak. Ezért ettől a ponttól kezdve fontos megkülönböztetni a "Turing-gépeket" mint absztrakt entitásokat és a "digitális számítógépeket" mint az ilyen gépek fizikai megvalósításait, ahol a digitális számítógépek, de nem a Turing-gépek, véges memóriával és a hibás működés lehetőségével rendelkeznek. Newell és Simon a

számítógépekre mint fizikai gépekre összpontosított, ahol a számítógépek által manipulációnak alávetett "jelek" státuszát igyekeztek tisztázni.

Ezeket az általuk "szimbólumoknak" nevezett fizikai minták halmazaként értelmezték, amelyek más, általuk "kifejezéseknek" (vagy "szimbólumstruktúráknak") nevezett minták összetevőiben fordulhatnak elő. Az alfanumerikus (ábécés és numerikus) karakterek (például ASCII vagy EBCDIC) halmazaihoz képest a kifejezések a szimbólumok sorozatai, amelyeket karakterek sorozataként értelmeztek. "Szimbólumrendszereik" mint fizikai gépek, amelyek szimbólumokat manipulálnak, így a (h4) hipotézis által megfogalmazott értelemezéshez szükségesnek és elégségesnek minősülnek:

(h4) (Az a) szimbólumrendszer szükséges és elégséges az intelligenciához (Newell és Simon);

ami még a Turing-gépek mint absztrakciók és a szimbólumrendszerek mint fizikai dolgok közötti különbségtől eltekintve is sokkal erősebb állításnak bizonyul, mint (h2) vagy akár (h3). Ezek a hipotézisek nem jelentik azt, hogy minden gondolkodó dolognak digitális számítógépnek vagy Turing-gépnek kell lennie. A (h2) és a (h3) feltételezés egyaránt összhangban van olyan gondolkodó dolgok létezésével, amelyek nem digitális számítógépek vagy Turing-gépek. De a (h4) nem teszi lehetővé olyan gondolkodó dolgok létezését, amelyek nem digitális gépek.

## **A kínai szoba**

A hipotézisek (h1)-től (h2)-ig (h3)-ig és talán (h4)-ig tartó fejlődése úgy tűnik, hogy jelentős előrelépést jelent Descartes elképzeléséhez képest, különösen a Turing-teszttel kombinálva, mivel ezek nemcsak az elme természetét tisztázzák, és megvilágítják az elme és a test kapcsolatát, hanem még azt is megmagyarázzák, hogyan lehet tudni más elmék létezéséről, ami az ontikus és episztemikus tézisek erőteljes kombinációja, amely úgy tűnik, hogy támogatja a mesterséges intelligencia kilátásait. Amint a számítástechnikai gépeket az emberi lényekéhez hasonló teljesítőképességgel tervezik,

helyénvaló lenne ezeknek az élettelen entitásoknak a gondolkodó dolgok mentális tulajdonságait tulajdonítani. Legalábbis úgy tűnt, amikor John Searle filozófus kritikát fogalmazott meg a mesterséges intelligencia kilátásaival kapcsolatban.

amely "a kínai szoba" néven vált ismertté, és mindezt kétségbe vonta (Searle 1980).

Searle egy gondolkísérletet javasolt két személy részvételével, nevezzük őket "C"-nek és "D"-nek,

az egyik (C) folyékonyan beszél kínaiul, a másik (D) nem. Tegyük fel, hogy C-t bezárják egy zárt helyiségbe, ahová papírlapokon jelek sorozatát küldik, amire C úgy válaszolhat, hogy más papírlapokon más jelek sorozatát küldi. Ha a küldött jelek a kérdések kínaiul voltak, és az elküldött jelek kínaiul voltak a válaszok, akkor minden bizonnyal úgy tünne, mintha a szoba lakója tudna kínaiul, ahogyan - hipotézis szerint - tud is. De tegyük fel, hogy ehelyett D-t bezárják ugyanabba a szobába egy olyan táblával, amely lehetővé teszi számára, hogy a beküldött jelek sorozataira válaszul elküldhető jelek sorozatait keresse meg. Ha nagyon ügyes lenne ebben a tevékenységben, akkor a teljesítménye megegyezhetne a kínaiul tudó C teljesítményével, még akkor is, ha D a hipotézis szerint nem tud kínaiul.

Searle érvelése megsemmisítő ellenpélda volt a Turing-tesztre, amely természetesnek veszi, hogy a teljesítmény hasonlósága az intelligencia hasonlóságára utal. A kínai szoba forgatókönyvében ugyanazok a "bemenetek" ugyanazokat a "kimeneteket" eredményezik, de az ezeket előállító folyamatok vagy eljárások nem azonosak. Ez azt sugallja, hogy különbséget kell tenni a "szimulációk" között, amikor a rendszerek *szimulálják egymást*, amikor ugyanazokból a bemenetekből ugyanazokat a kimeneteket adják, és a "replikációk" között, amikor a rendszerek *replikálják egymást*, amikor ugyanazokból a bemenetekből ugyanazokat a kimeneteket adják ugyanazokkal a folyamatokkal vagy eljárásokkal. Ezen a nyelven Searle megmutatja, hogy még ha a Turing-teszt elégséges is a bemeneti/kimeneti viselkedés (szimulációk) összehasonlításához, nem elegendő a az ezeket a kimeneteket eredményező folyamatok vagy eljárások összehasonlítására (replikációk).

## **Gyenge AI**

Searle kritikájának ereje akkor válik nyilvánvalóvá, ha megkérdezzük, hogy melyik forgatókönyv, a C vagy a D, hasonlít jobban egy számítógép teljesítményére, amely egy programot hajt végre, amely megvalósulhatna

mint egy automatizált keresőtáblázat: a számítógép a jelek sorozatai formájában beérkező adatokat a programja alapján más jelek sorozatai formájában történő kimenetekké dolgozza fel. Ezért helyénvalónak tűnik az összehasonlítást kiterjeszteni egy harmadik foratókönyvre is, nevezzük azt "E"-nek, ahol egy megfelelően programozott számítógép ugyanazokat a bemeneteket veszi, és ugyanazokat a kimeneteket adja ki. Mert ahogyan D teljesítménye szimulálhatja C teljesítményét, még ha D nem is tud kínaiul, úgy E teljesítménye is szimulálhatja D teljesítményét, még ha E nem is rendelkezik mentalitással. A szimuláció pusztán összefüggései tehát túl gyengének tűnnek ahhoz, hogy megállapítsuk, hogy a rendszerek intelligenciájukhoz képest egyenlőek.

Searle különbséget tett az általa "erős mesterséges intelligencia" és a "gyenge mesterséges intelligencia" között is, ahol a *gyenge mesterséges intelligencia* azt állítja, hogy a számítógépek hasznos eszközök az elme tanulmányozásában, különösen hasznos modellek (vagy szimulációk) létrehozásában, az *erős mesterséges intelligencia* viszont azt állítja, hogy amikor programokat hajtanak végre, a számítógépek megfelelően elmének (vagy replikációnak) minősülnek. A gyenge mesterséges intelligencia tehát egy episztemikus álláspontot képvisel a számítógép-alapú modellek vagy szimulációk értékével kapcsolatban, míg az erős mesterséges intelligencia egy ontikus álláspontot képvisel az olyan dolgokkal kapcsolatban, amelyek ténylegesen az elme példányai. Feltehetően az erős mesterséges intelligencia feltételezi a gyenge mesterséges intelligenciát, mivel az elmék tényleges példányai megfelelő alanyok lennének az elme tanulmányozására. A gyenge mesterséges intelligencia ellen természetesen gyakorlatilag senki sem tiltakozik, míg az erős mesterséges intelligencia számos okból ellentmondásos marad.

Ez nem jelenti azt, hogy hiányoznak a szenvedélyes szószólók. A mesterséges intelligencia egyik legérdekesebb bevezetőjét Eugene Charniak és Drew McDermott

közösen írták (Charniak és McDermott 1985). A szerzők már az első fejezetben úgy határozzák meg a "mesterséges intelligenciát", mint a mentális képességek tanulmányozását számítási modellek segítségével. Ennek az álláspontnak a tarthatósága kétségtelenül azon a hallgatólagos előfeltevése múlik, hogy a mentális képességek számítási folyamatok alapján működnek, amit valóban explicitté tesznek azzal, hogy hasonlóképpen azt a posztulációt állítják, hogy amit az agyak csinálnak, "valamilyen szinten egyfajta számításként gondolható el" (Charniak és McDermott 1985,

p. 6). A "gyenge" és az "erős" mesterséges intelligencia közötti döntő különbségtétel azonban a következőktől függ



hogyan az agyak valóban számítógépnek minősülnek-e, nem pedig az, hogy azt gondolják-e róluk, hogy azok.

## **Erős AI**

Tovább mennek, amikor azt állítják, hogy "a mesterséges intelligencia kutatásának végső célja az, hogy egy ember vagy, még szerényebben, egy állat". Általános felfogásuk szerint a mesterséges dolgok megalkotásának meg kell ragadnia biológiai társaik kulcsfontosságú tulajdonságait, legalábbis a bemenet, a feldolgozás és a kimenet fajtái tekintetében. Így az általuk figyelembe vett "bemenetek" közé tartozik a látás (látvány) és a beszéd (hangok), amelyeket a tanulás, a következtetés, a magyarázat és a tervezés belső moduljai dolgoznak fel, amelyek keresési és rendezési mechanizmusokat foglalnak magukban. Ezek a beszéddel és a motoros képességekkel kombinálva beszéd (hangok) és viselkedés (mozgások) formájában megjelenő "kimeneteket" eredményeznek, amelyeket néha "robotikának" neveznek. A döntő kérdés tehát az, hogy ezek a "robotok" az emberhez hasonlóan viselkednek-e (agyatlan) szimulációként, vagy ehelyett (tudatos) másolatokat testesítenek meg.

Figyelmük arra összpontosul, hogy mi történik a "fekete dobozban" az inger és a válasz között, ahol az elmével rendelkezők *belső reprezentációkra* támaszkodnak és azokat használják.

mint olyan rendszerek állapotai, amelyek a világ különböző aspektusait írják le vagy más módon reprezentálják. Valójában a szempontok némelyike lehet magának a rendszernek a belseje, és így saját belső állapotait, mint saját aspektusainak belső reprezentációit képviseli. De bár az öntudatot és az öntudatosságot gyakran az intelligencia vagy a mentalitás fontos fajtáinak tekintik, úgy tűnik, hogy ezek nem nélkülözhetetlenek ahhoz, hogy általában véve intelligenciával vagy mentalitással rendelkezünk, szemben a meghatározott típusú intelligenciával vagy mentalitással.

A mentalitásnak vagy intelligenciának különböző fajtái lehetnek - például matematikai, verbális és művészi -, de feltehetően ezek bizonyos alapvető vagy közös tulajdonságokkal rendelkeznek.

Úgy tűnik, aligha lehet kétséges, hogy ha a mesterséges gépeket mentális képességeik tekintetében az emberhez hasonlónak akarjuk minősíteni, akkor az alábbiakra van szükségük.

azonos vagy hasonló képességekkel rendelkeznek a belső reprezentációk használatára és manipulálására, legalábbis bizonyos meghatározott - feltehetően alfanumerikus - feladatok tekintetében. Ugyanazokat vagy hasonló külső inputokat (vagy "ingereket") kell fogadniuk, ugyanazokkal vagy hasonló "mentális" mechanizmusokkal kell feldolgozniuk, és ugyanazokat vagy hasonló "mentális" mechanizmusokat kell létrehozniuk.

külső kimenetek (vagy "válaszok"). Míg Charniak és McDermott arra törekszik, hogy mesterséges állatot építsen, a mesterséges intelligencia közösség egésze kétségtelenül beérné egy mesterséges gondolkodó dolog megalkotásával, feltételezve, hogy lehetséges az egyiket a másik nélkül létrehozni.

## **Népi pszichológia**

Feltételezhető, hogy az összehasonlítás tárgyát képező különböző rendszerek azonos vagy hasonló, ok-okozati szempontból releváns háttérfeltételek között működnek. Senki sem feltételezné, hogy egy elromlott alaplapú számítógépnek ugyanazokat a kimeneteket kell produkálnia ugyanazokból a bemenetekből, mint egy hasonló, hardverhiba nélküli számítógépnek, még akkor sem, ha ugyanazokkal a programokkal vannak feltöltve. Hasonlóképpen, senki sem feltételezné, hogy például egy törött karú embernek ugyanazt a viselkedést kellene tanúsítania, mint egy törött karral rendelkező embernek.

ugyanarra az ingerre (mondjuk egy labda, amely egyenesen felé jön, miközben a lelátón ül egy meccsen), mint egy másik személy, akinek nincs törött karja. Ez azonban nem jelenti azt, hogy nem hasonló reprezentációk segítségével dolgoznak fel hasonló ingereket.

Az emberi lények bonyolult mechanizmusok, függetlenül attól, hogy a mesterséges intelligencia számára fontos értelemben "gépnek" minősülnek-e vagy sem. Úgy tűnik,

hogyan az emberi viselkedést befolyásoló, ok-okozati szempontból releváns tényezők teljes skálája magában foglalja az indítékokat, a meggyőződéseket, az etikát, a képességeket, a képességeket és a lehetőségeket (Fetzer 1996). Az azonos vagy hasonló motívumokkal és meggyőzésekkel rendelkező, de erkölcsileg különböző személyek például várhatóan eltérő viselkedést tanúsítanak olyan körülmények között, ahol az etika különbséget tesz, még akkor is, ha hasonló képességekkel rendelkeznek, és nem akadályozzák őket e képességek gyakorlásában. Mint mindannyian tudjuk, az emberek végtelen órákat töltenek azzal, hogy arra törekedjenek, hogy

magyarázzák és megjósolják mások és önmaguk viselkedését az ilyen jellegű, okozati szempontból releváns tényezők keretrendszerének alkalmazásával, amelyet "népi pszichológiaként" ismerünk.

Kétségtelen, hogy ha a tudományos elméletek megfelelőségének feltételei - mint például a nyelvezet világossága és pontossága, a magyarázatok és előrejelzések alkalmazási köre, az empirikus alátámasztás mértéke, valamint az eredmények elérésének gazdaságossága, egyszerűsége vagy eleganciája - szempontjából vizsgáljuk, a népi pszichológia nagyfokú empirikus támogatottságot élvez, mivel képes az esetek széles körét az elvei körébe vonni. E látszólagos siker egy része azonban annak a nyelvezetnek a kissé homályos és pontatlan jellegéből adódhat, amelyen alapul, ahol úgy tűnik, hogy lehetőség van a felülvizsgálatra és a finomításra, hogy növeljék vagy korlátozzák alkalmazási körét. Egyes hallgatók mégis a teljes eltörlése mellett érvelnek.

### **Eliminatív materializmus**

Paul Churchland például azt állítja, hogy a népi pszichológia nemcsak hiányos, hanem pontatlan is, mivel belső állapotaink és mentális tevékenységeink "téves ábrázolása". Odáig megy, hogy azt javasolja, hogy az idegtudományok fejlődésének nem egyszerűen a népi pszichológia finomításához, hanem teljes felszámolásához kellene vezetnie (Churchland 1984, 43. o.). A Churchland által felkarolt modell tehát a "flogisztion" eltávolításának mintáját követi a kémia nyelvéből és a "boszorkányok" eltávolításának mintáját a pszichológia nyelvéből. Azt állítja tehát, hogy többek között a *motívumok* és a *hiedelmek* kategóriái is arra rendeltetnek, hogy a hasonló sorsra jutnak az idegtudományok fejlődésével. Churchland elismeri, hogy nem tudja garantálni, hogy ez bekövetkezik, ahol a tudomány története ebben az esetben inkább csak tükrözi a következő tényeket

a néplélektani elvek némi kiigazítása vagy egyes fogalmainak elhagyása.

A mélyebb probléma, amellyel az eliminatív materializmus szembesül, azonban ugyanaz a probléma, amellyel a redukcionizmus klasszikus formái szembesülnek, nevezetesen, hogy hozzáférés nélkül

egyrészt az agyi állapotokat az elmeállapotokkal, másrészt az elmeállapotokat a viselkedési hatásokkal összekapcsoló információkhoz, lehetetlen lenne az agyi állapotokból a viselkedési hatásokra való előrejelző következtetéseket levezetni. Ha ezek a viselkedési hatások ráadásul bizonyos körülmények között a viselkedésre irányuló diszpozíciók megnyilvánulásai, akkor valószínűtlennek tűnik, hogy egy "érett" idegtudomány el tudná érni a céljait, ha nem lenne képes az agyi állapotokat a diszpozíciók révén a viselkedési hatásokkal kapcsolatba hozni, mert akkor nem lenne alapja annak, hogy az elmeállapotokat az agyi állapotokhoz, az agyi állapotokat pedig az emberi viselkedéshez kapcsoljuk.

A féltékenység (az ellenségesség, az őszintétlenség és így tovább) mint a viselkedésünket befolyásoló oksági tényezők esetében a néplélektani sémában, ha fel akarjuk fedezni azokat az agyi állapotokat, amelyek ezen elmeállapotok mint a féltékenység (az ellenségesség és így tovább) viselkedésére irányuló diszpozíciók mögött állnak, meghatározott feltételek mellett, amelyek magukban foglalják egyéb belső állapotainkat, akkor az emberi viselkedés szigorú tudományát úgy lehetne kifejleszteni, hogy keresünk és felfedezünk néhány mögöttes agyi állapotot, ahol a viselkedésre irányuló diszpozíciók megfelelően (feltehetően törvényszerűen) kapcsolódnak ezen agyi állapotokhoz. Néha az agyi állapotoknak olyan hatásai lehetnek az emberi viselkedésre, amelyeket nem az elmeállapotok közvetítenek, mint például a agykárosodás vagy szellemi fogyatékoság. A neurológiailag normális alanyok esetében az elmeállapotok képesek kapcsolatot teremteni az agyi állapotok és azok viselkedésre gyakorolt hatása között.

### **Feldolgozási szintaxis**

A számítási koncepció által nyújtott erőforrások kiaknázására törekvő filozófusok uralkodó megközelítése azonban inkább annak finomítása irányába mutatott, hogy mi

kell ahhoz, hogy elmével rendelkezünk, mintsem hogy az elme, a test és a viselkedés közötti kapcsolatról beszéljünk. Bár elismerik, hogy ezek az összefüggések elengedhetetlenek bármely beszámoló megfelelőségéhez, elsősorban arra a kilátásra összpontosítottak, hogy a nyelv és a mentalitás megfelelően jellemezhető lenne a Turing-gépek által megkövetelt általános jellegű, tisztán formális megkülönböztetések alapján - a fizikai formák, méretek és a testek relatív helyzete.



az általuk manipulált jelek - amikor azokat alfanumerikus karakterekként értelmezik, amelyek szavakat, mondatokat és más mondatkombinációkat alkotnak, mint egy nyelv elemeit.

Jerry Fodor például megfigyelte, hogy a nyelv és a mentalitás számításalapú felfogása azt a tézist vonja maga után, hogy "... a mentális folyamatok csak azon mentális reprezentációk formális (nem szemantikai) tulajdonságaihoz férnek hozzá, amelyek felett definiálják őket" (Fodor 1980, 307. o.). Kifejti a gondolatok formája (szintaxis) és tartalma (szemantika) közötti kapcsolatot, azt állítva, hogy (a) a gondolatok csak akkor különböznek tartalmilag, ha azonosíthatók különböző reprezentációkkal, de anélkül, hogy magyarázatot adna arra, hogy (b) hogyan lehetséges, hogy bármely konkrét gondolat azonosítható bármely konkrét reprezentációval, mely problémára másutt "a gondolat nyelve" néven kínál megoldást. De minden olyan beszámoló, amely azt állítja, hogy ugyanaz a szintaktika mindig ugyanazzal a szemantikával rendelkezik, vagy hogy ugyanaz a szemantika mindig ugyanazzal a szintaktikával rendelkezik, egyrészt a többértelműség, másrészt a szinonimia problémáiba ütközik.

Mindazonáltal a számítógépes koncepciók legerősebb változatai hajlamosak arra, hogy elkerüljék a szemantikával való törődést, és ehelyett a szintaxis központi szerepére összpontosítanak. Stephen Stich úgy vezette be *az elme szintaktikai elméletét (STM)*, mint amely agnosztikus álláspontot képvisel a tartalommal kapcsolatban, és nem ragaszkodik sem ahhoz, hogy a szintaktikai állapot típusoknak (mint a szintaktika megismételhető mintáinak) ne legyen tartalma, sem ahhoz, hogy a szintaktikai állapotjelzőknek (a szintaktikai állapot típusok konkrét példányainak) ne legyen tartalma: "Egyszerűen hallgat az egész kérdésről. . . (T)az STM tulajdonképpen azt állítja, hogy a pszichológiai elméleteknek nincs szükségük arra, hogy tartalmat vagy

más

szemantikai tulajdonságok" (Stich 1983, 186. o.). Az STM ezáltal a (h7) hipotézis mellett

kötelezi el magát: (h5) A szintaxist feldolgozó fizikai számítógépek rendelkeznek

elmével (STM);

ami kezdetben sokkal erősebbnek tűnhet, mint a (h3). De Newell és Simon

"szimbólum" fogalmát formálisan definiálják, és "szimbólumrendszereik" is

számítógépek.

Mindkét megközelítés azzal a veszéllyel jár, hogy a "gondolkodó dolgokat" az agyatlan gépekkel azonosítják.

## **Szemantikus motorok**

A manipulációs szabályokkal rendelkező jelrendszerek példái az úgynevezett *formális rendszereknek*, amelyek tanulmányozása a tiszta matematika területére tartozik. Ha ezeket a formális rendszereket értelmezni lehet, különösen a fizikai világ tulajdonságai és tárgyai tekintetében, akkor tanulmányozásuk az alkalmazott matematika területére esik. Az informatikán belül vita folyt arról, hogy a tudományágnak a tiszta matematika vagy az alkalmazott matematika mintájára kell-e modelleznie magát (Colburn et al., 1993). De bármi legyen is a vitában részt vevő felek érdeme, aligha lehet kétséges, hogy a puszta jelmanipuláció, még a szintaxis feldolgozásának álcája sem elegendő a gondolkodó dolgokhoz. A gondolatoknak ugyanúgy van tartalma, mint formája, ahol nem túlzás azt állítani, hogy a gondolkodás tekintetében a tartalom dominál a forma felett.

Az STM, amely a szintaxis-feldolgozást elegendőnek teszi a mentalitás birtoklásához, így túl erősnek tűnik, de egy gyengébb változat is igaz lehet. A szintaxis feldolgozásának képessége ehelyett szükséges lehet a mentalitáshoz, ahogyan azt a (h3) hipotézis is sugallja, ha Newell és Simon "szimbólumait" helyesen manipulálható jelekként értelmezzük. Így a (h5) egy hihetőbb változatának a (h6) helyett a (h6) feltételezést kellene fenntartania:

(h6) A (tudatos) elmék fizikai számítógépek, amelyek szintaxist dolgoznak fel;

ahol a szintaxis olyan jelekből és a manipulációjukra vonatkozó szabályokból áll, amelyek megfelelnek az őket értelmessé tevő korlátozásoknak. Mivel azonban a jelek bármely véges sorozatának végtelen sok lehetséges értelmezése létezik, egy bizonyos értelmezés (vagy értelmezések osztálya) "a szándékolt értelmezés" néven specifikációt

igényel. A jelek csak bizonyos értelmezésekhez képest minősülhetnek szintaxisnak, amelyekhez képest a jelek értelmet nyernek.

Ebből a szempontból egy (megfelelően működő) számítástechnikai gépet a következőképpen lehet minősíteni

*automatikus formális rendszerként működik*, amikor egy programot hajt végre, de csak akkor válik értelmessé, amikor a szintaxisa kielégíti a szándékolt értelmezés korlátait.

Valóban, egy olyan automatikus formális rendszert, ahol "a szemantika követi a szintaxist", Daniel Dennett "szemantikai motornak" nevezett. Ez alátámasztja azt az állítást, amelyet egyesek a kognitív tudomány alapgondolatának neveztek - hogy *az intelligens lények szemantikai motorok*, vagyis olyan automatikus formális rendszerek, amelyek alapján következetesen értelmet alkotnak (Haugeland 1981, p. 31). A (h6) tehát minősítést igényel, hogy az értelmezés szerepét a (h7)-ként való

értelmezésbe foglalja: (h7) A szemantikai motorok szükségesek és elégségesek az intelligenciához;

ahol, mint Newell és Simon esetében, az "intelligens dolgok" egyben "gondolkodó dolgok" és "(tudatos) elmék" is, amelyeket fizikai számítógépekként értelmeznek, amelyek egy értelmezés alapján szintaxist dolgoznak fel. A probléma a szintaxis és a szemantika megfelelő "párosítása".

## **Kognitív tudomány**

John Haugeland hangsúlyozta, hogy a (tudatos) elmék szemantikus motorokként való felfogása többé-kevésbé egyenrangúvá teszi a kognitív pszichológiát és a mesterséges intelligenciát, ahol az emberek és az intelligens gépek egyszerűen ugyanazon mögöttes jelenségek különböző megnyilvánulásainak bizonyulnak. Valóban, jegyzi meg, beláthatjuk, hogy miért

ebből a szempontból a mesterséges intelligencia a pszichológia különösen tiszta és absztrakt formájának tekinthető. Ugyanazokat az alapvető struktúrákat vizsgáljuk,

de a mesterséges intelligenciában minden releváns paraméter közvetlen kísérleti kontroll alatt áll (a programozásban), anélkül, hogy bármilyen kusza fiziológia vagy etika útban lenne (Haugeland 1981, 31. o.).

Ez kiegészíti a nyelv és az elme komputációs felfogását, amely szerint a gondolkodás érvelés, az érvelés számolás, a számolás számítás, és a számítás számítás.

a kiszámíthatóság határai határozzák meg a gondolkodás határait, ahogyan Turing elképzelte. Haugeland szerint legalább két stratégia létezik a kognitív tudomány támadására,

így megértett, ahol (az általa "üres héj" stratégiának nevezett) stratégia azt állítja, hogy bármennyire is sikeresen teljesít egy szemantikus motor, a megértés még mindig hiányzik belőle; és (az általa "szegény helyettesítő" stratégiának nevezett) stratégia azt állítja, hogy a puszta szemantikus motorok soha nem fognak olyan jól teljesíteni, mint az emberi lények. A második tehát azt állítja, hogy egyetlen szemantikus motor sem lesz képes átmenni a Turing-teszten, míg az első azt állítja, hogy még ha át is menne, akkor sem minősülne megfelelően gondolkodó dolognak. Tekintettel a szimuláció és a replikáció közötti különbségtételre, a második azt állítja, hogy egyetlen szemantikus motor sem fog valaha is felérni az emberi lényekkel a szimulációs képesség tekintetében, az első pedig azt állítja, hogy a szemantikus motorok soha nem érik el a replikációval való egyenértékűséget.

Nem meglepő, ha a digitális gépekről kiderül, hogy szemantikus motorok, amelyek szemantikája követi a szintaxisukat, mivel a programozók pontosan ennek biztosítására tervezik a programokat. Így helyesebb lenne azt sugallni, hogy legalábbis ebben az esetben a szintaxis követi a szemantikát, mivel az általuk használt jelek és a rajtuk végrehajtott műveletek pontosan ezt a célt szolgálják. Természetesen nem minden programozó egyformán ügyes abban, hogy ilyen eredményeket érjen el, amint azt a programok fordítási és végrehajtási folyamata - beleértve a futtatást és a hibakeresést is - megmutatja. Az emberi lények esetében azonban a helyzet bonyolultabb, mivel a "megismerés tudományán" belül nem lenne helyénvaló Istenre mint intelligens tervezőre hivatkozni, ami azt sugallja, hogy az ember evolúció vagy tanulás révén vált szemantikai motorokká.

## **A gondolkodás nyelve**

Jerry Fodor (1975) egy olyan érvelést terjesztett elő, amely feltételezi a veleszületett nyelv létezését, amely fajspecifikus, és amellyel minden neurológiailag normális emberi lény rendelkezik. Ő ezt *mentális nyelvnek* (vagy "a gondolkodás nyelvének") nevezi. Azt állítja, hogy az egyetlen módja annak, hogy



egy nyelv megtanulása az adott nyelven előforduló mondatok igazságfeltételeinek megismerése: ". . . az L (egy nyelv) megtanulása azt jelenti, hogy megtanuljuk, hogy 'Px' akkor és csak akkor igaz, ha x G minden helyettesítési esetre. De vegyük észre, hogy ennek megtanulása csak egy olyan organizmus számára jelentheti a P megtanulását (megtanulni, hogy mit jelent a P), amely már megértette a G-t" (Fodor 1975, 80. o.). Tekintettel arra a kellemetlen választásra, hogy az alacsonyabb szintű nyelvek jelentésének meghatározására szolgáló, egymás után egyre gazdagabb és gazdagabb metanyelvek végtelen hierarchiája és egy nem tanult alapnyelv között kell választani, Fodor egy veleszületett és veleszületett gondolkodási nyelv létezése mellett dönt.

A tanult nyelv és a gondolkodás nyelvének összekapcsolása szemantikai motorokká teszi az embert, amit a (h8) hipotézis a következőképpen adhat vissza:

(h8) Az ember szemantikus motor, amelynek gondolati nyelve van (Fodor).

Fodor azonban hibát követ el érvelésében, amikor figyelmen kívül hagyja annak lehetőségét, hogy a nyelvtanulás által feltételezett előzetes megértés *nem nyelvi jellegű lehet*. A gyerekek megtanulnak mellbimbót szopni, labdázni és zsírkrétával rajzolni. jóval azelőtt, hogy tudnák, hogy amit csinálnak, az "mellbimbókkal", "golyókkal" vagy "zsírkrétákkal" kapcsolatos. Az ilyen jellegű dolgokkal való interakció során cselekvési és gondolkodási szokásokat sajátítanak el az ilyen jellegű dolgok tényleges és lehetséges viselkedésével kapcsolatban. A különböző típusú dolgokra vonatkozó cselekvési és gondolkodási szokások *fogalmak*. Amint ezt a nem nyelvi megértést elsajátították, az ezek leírására szolgáló nyelvi diszpozíciók elsajátítása viszonylag problémamentesnek tűnik (Fetzer 1990).

Fodor koncepciójának egyik figyelemre méltó vonása, hogy a gondolkodás

veleszületett és veleszületett nyelve olyan szemantikai gazdagsággal rendelkeznek, hogy ennek az alapnyelvnek kellően teljesnek kell lennie ahhoz, hogy a történelmi fejlődés bármely szakaszában (múlt, jelen és jövő) bármely természetes nyelv (francia, német, szuahéli és hasonló) közötti összefüggéseket fenntartsa. Ez azt jelenti, hogy a mentális nyelvnek nemcsak a mindennapi nyelvhasználat alapját kell biztosítania.

szavakat, mint például az angolban a "mellbimbó", a "labda" és a "zsírkréta", hanem a fejlettebb fogalmakra, mint például a "sugárhajtás", a "gyermekbénulás elleni vakcina" és a "színes televízió", mivel máskülönben a gondolkodás nyelve nem tudná betölteni a neki szánt szerepet. E felfogás kevésbé hihető következményei közül kiderül, hogy mivel minden embernek ugyanaz a veleszületett nyelve van, amelynek minden egyes példányában teljesnek kell lennie, a különböző nyelvek közötti sikertelen fordítások és a nyelv időbeli evolúciója elvileg egyaránt lehetetlen, ami nehezen védhető álláspont.

### **Formális rendszerek**

Fodor megközelítése Noam Chomsky munkásságának kiterjesztését jelenti, aki régóta képviseli a veleszületett és fajspecifikus szintaxis koncepcióját, amelyhez Fodor szemantikát csatolt. Chomsky munkásságának nagy része a *kompetencia* és a *teljesítmény* közötti különbségtételre épült, ahol a különböző nyelvhasználók egyébként azonos nyelvtani viselkedése közötti különbségeket körülményes különbségekkel kell magyarázni, például a fiziológiai állapotokban vagy a pszichológiai kontextusban. Elvileg minden nyelvhasználó rendelkezik egy olyan *(korlátlan) számítási kompetenciával*, amely véges alapból végtelen sok mondat konstruálható olyan rekurzív eljárások alkalmazásával, mint amelyeket Church és Turing tanulmányozott a hatékony eljárásokról szóló klasszikus munkájában.

Fodor és Zenon Plyshyn (1988) például a nyelvhasználók által előállított mondatokra vonatkozóan olyan feltételeket fogadnak el, amelyek szerint a szintaktikai egészek szemantikai tartalma a *jelentés kompozicionalitásának elveként* a szintaktikai részek szemantikai tartalmának függvénye, és a molekuláris reprezentációk a *rekurzív generálhatóság elveként* más molekuláris vagy atomi reprezentációk függvényei. Ezek a feltételek nyilvánvaló következményei a strukturálisan atomos és strukturálisan molekuláris reprezentációk közötti különbségtételnek, mint a formális modellre épülő

gondolkodási nyelv előfeltételei.

rendszerek, mint például a mondattani kalkulus. A formális rendszerek elvei - akár automatizáltak, akár nem - azonban átvihetők, vagy nem, absztrakt kontextusból fizikai kontextusba, nem utolsósorban azért, mert a fizikai rendszerek, beleértve a digitális gépeket is, korlátozott kapacitásúak.

A végtelen szalaggal rendelkező és tévedhetetlen teljesítményű Turing-gépek egyértelműen absztrakt idealizációk a véges memóriával rendelkező, hibásan működő digitális gépekhez képest. A személyek és a számítógépek fizikai tulajdonságai határozottan különböznek az automatizált formális rendszerekétől, ami az absztrakt idealizáció egy másik esete. Ehhez képest a digitális gépek és az emberek látszólag nem rendelkeznek többel, mint *(korlátozott) komputációs kompetenciával* (Fetzer 1992). Így a formális rendszerek olyan tulajdonságai, mint például a Kurt Godel által megállapított magasabb rendű logikában a befejezetlenségi bizonyítás, amelyekről feltételezhető, hogy korlátokat szabnak a mentális folyamatoknak, és amelyek felkeltették a tudósok, köztük J. R. Lucas (1961) és Douglas Hofstadter (1979) érdeklődését, úgy tűnik, csekély jelentőséggel bírnak a megismerés természetének megértése szempontjából. A formális rendszerek hasznosak az érvelés modellezésében, de az érvelés a gondolkodás speciális esete. És ha meg akarjuk érteni a gondolkodás természetét, akkor inkább a gondolkodó dolgokat kell tanulmányoznunk, mint a formális rendszerek tulajdonságait. A gondolkodó dolgok és a formális rendszerek nem ugyanazok.

## **Konnekcionizmus**

A különböző cselekvési szokások és elméleti szokások elsajátítása saját élettörténetünk hatására tehát hajlamos meghatározni azt a képességünket, hogy a különböző tapasztalatokat megfelelő fogalmakkal foglaljuk össze. Minden neurológiailag normális

emberi lénynek ugyanaz a veleszületett lehetősége a nyelvtanulásra és más mentális képességek elsajátítására, de ténylegesen csak meghatározott környezeti feltételek mellett sajátítja el azokat. Fodor befogadja a nyelvtanulást, de kizárja a részleges nyelveket és a nyelvi evolúciót. Ahhoz, hogy egy metanyelv rendelkezzen a szükséges erőforrásokkal ahhoz, hogy minden alacsonyabb szintű nyelvet magába foglaljon,

önkényesen kiválasztott, akkor rendelkeznie kell az erőforrásokkal ahhoz, hogy mindet magába foglalja. És ugyanígy egy veleszületett nyelv csak akkor képes tetszőlegesen kiválasztott alternatív nyelvek tanulására, ha rendelkezik az összes nyelv megtanulásához szükséges (veleszületett és veleszületett) erőforrásokkal.

A *konnekcionizmus* néven ismert megközelítés az elmét az agy realisabb modelljei segítségével próbálja megérteni, amelyek a tanulást és az evolúciót is képesek figyelembe venni (Rumelhart et al. 1986). Az emberi agyban lévő neuronok száma a jelek szerint a

$12$   
10(vagyegy trillió) nagyságrendű, ahol a kapcsolatok száma, hogy ezek a neuronok  
 $15$   
a többi neuronnal létrehozható körülbelül 1000. Ez azt jelenti, hogy az agynak 10  
(vagy egy kvadrillió) lehetséges állapota van, amelyeket az aktivációs minták különböző elrendezései határoznak meg. E minták közül néhánynak az aktiválódása feltételezhetően mások aktiválódását is előidézi, mint olyan veleszületett tulajdonságok, amelyek nélkül normális agy nem létezhet, míg mások szerzett tulajdonságokként jöhetnek létre, amelyek nélkül normális agyak nem létezhetnek. A nyelvtanulás képessége (az emberileg tanulható nyelvek körén belül), például veleszületett, míg a francia, német vagy szuahéli nyelvtudás szerzett.

Egyes agyi állapotok más belső állapotokkal kölcsönhatásba lépnek, hogy átmenetet teremtsenek az egymást követő agyi állapotokba, míg mások a motoros viselkedések meghatározott sorozatait aktiválják, amelyek közé tartozhat a beszéd is. Az ezeket a hatásokat befolyásoló tényezők közé tartozik minden olyan tulajdonság, amely különbséget tesz ezen átmenetekben, és amely a neurális kontextust alkotja. Az aktiváció ezen mintázatai iránti érdeklődésünk a beszédhez és más viselkedésekhez való ok-okozati hozzájárulásukból ered, amint azt fentebb említettük. A motívumok és a meggyőződések a viselkedést és a beszédet befolyásoló energetizáló és irányító

tényezők változói, míg az etika egy speciális alcsoporthoz tartozik, amely a megfelelő feltételek hiányában hajlamos gátolni bizonyos viselkedés- és beszéd típusokat. Bizonyos képességeknek sajátos moduláris neurális helyeik lehetnek.

## **Mentális hajlamok**



Roger Penrose felvetette, hogy a gondolkodás kvantumjelenség lehet, és ezáltal *nem algoritmikusnak* minősülhet (Penrose 1989, 437-439. o.). Ennek a kilátásnak az a jelentősége, hogy az algoritmusokat általában olyan függvényekként értelmezik, amelyek egy bizonyos tartományon belüli egyes értékeket egy bizonyos tartományon belüli egyes értékekre képeznek le. Ha a mentális folyamatok algoritmikusak (függvények), akkor *determinisztikusnak* kell lenniük, abban az értelemben, hogy ugyanaz a mentális állapotú ok (teljesen meghatározott) mindig ugyanazt a mentális állapotú hatást vagy viselkedési választ eredményezi. Mivel a kvantumjelenségek nem determinisztikusak, ha a mentális jelenségek kvantumfolyamatok, akkor nem függvények - még csak nem is részleges függvények, amelyek esetében, ha egy tartományon belüli egyes értékek véletlenül éppen hogy meg kell adni, a megfelelő tartományban léteznek egyes értékek, de előfordulhat, hogy a vonatkozó változók tartományának és tartományának egyes értékei nincsenek megadva.

Azokat a rendszereket, amelyeknél minden olyan tulajdonság jelenléte vagy hiánya, amely a kimenetel szempontjából különbséget tesz, teljesen meghatározott, "zártnak", míg azokat, amelyeknél néhány, a kimenetel szempontjából különbséget jelentő tulajdonság jelenléte vagy hiánya nem meghatározott, "nyitottnak" nevezzük. A determinisztikus és (ebben az esetben) a valószínűségi oksági folyamatok közötti különbségtétel az, hogy zárt rendszerek esetében a *determinisztikus* oksági folyamatok esetében ugyanaz az ok (vagy a feltételek teljes halmaza) változatlanul (vagy univerzális erősséggel  $u$ ) ugyanazt a hatást idézi elő, míg a *valószínűségi* oksági folyamatok esetében folyamatok esetében ugyanaz az ok változóan ( $p$  valószínűségi erősséggel) egy vagy egy másik hatás a lehetséges kimenetek ugyanazon rögzített osztályán

belül. Egy polónium

atom bomlásának valószínűsége például 3,05 perc alatt  $1/2$ .

218

Annak megállapítása, hogy egy rendszer, például egy

polóniumato

m,zárt rendszer-e

vagy

semnehéz ismeretelméleti problémákat vet fel, amelyek az emberi lények esetében

még súlyosabbak, pontosan azért, mert azok sokkal összetettebb oksági rendszerek.

Ráadásul a valószínűségi rendszereket meg kell különböztetni az (úgynevezett)

*kaotikus rendszerektől*, amelyek determinisztikus rendszerek, "akut érzékenységgel

a

kezdeti feltételek", ahol e feltételek legkisebb változása is előidézhethet korábban nem várt hatásokat. Egy űrszondát vezérlő több százezer sornyi kódban például egyetlen rossz karakter, egyetlen rosszul elhelyezett vessző megjelenése okozta, hogy a Mariner 1, az Egyesült Államok első bolygóközi űrszondája letért a pályájáról, majd meg kellett semmisíteni.

## **A keretprobléma**

Úgy tűnik, hogy legalább három olyan kontextus van, amelyben a valószínűségi ok-okozati összefüggés fontos lehet az emberi viselkedés szempontjából, nevezetesen: az érzékszervi adatok neurális aktivációs mintázatokká történő feldolgozása; az egyik aktivációs mintázat és a másik közötti átmenet; valamint a hangok és más mozgások viselkedési válaszként történő előállítása. Mindhárom típusú folyamatot valószínűségi vagy kaotikus determinisztikus folyamatok irányíthatják, és ezért nehezebb megmagyarázni vagy megjósolni, még akkor is, ha a vizsgált rendszer fajtája történetesen ismert. Ezek az aggályok a mentális modellek vagy a világról alkotott reprezentációk tanulmányozásával összefüggésben is felmerülnek, pontosabban azzal kapcsolatban, hogy mi van a *a keretprobléma* néven ismert, amelyet Charniak és McDermott a következőképpen ír le: A keretproblémának nevezzük azt az igényt, hogy explicit módon következtetni kell arra, hogy egy állapot nem változik az idő múlásával. Ez azért probléma, mert szinte minden állapot nem változik meg egy esemény során, és a gyakorlati rendszerekben óriási számban lesznek ilyenek, ami nem célszerű kifejezetten foglalkozni vele. Ez a nagy halmaz egy "keretet" alkot, amelyen belül kis számú változás történik, innen a kifejezés. (Charniak és McDermott 1985, 418. o.)

Míg a keretprobléma számos különböző jellemzésre alkalmasnak bizonyult - ezek sokasága megtalálható például Ford és Hayes (1991) munkájában -, a probléma egyik fontos aspektusa az, hogy a tudásbázis milyen mértékben teszi lehetővé az előrejelzést.

és a rendszerek magyarázata, amikor ezek a rendszerek nem ismertek, hogy nyitva vagy zárva vannak.

Ebből a szempontból a keretprobléma még a klasszikus *indukciós problémának* is tűnik, amellyel akkor találkozunk, amikor a múltra vonatkozó információk alapján próbáljuk megjósolni a jövőt, és amelyet David Hume (1711-76), egy jelentős befolyással bíró skót filozófus azonosított. Hume megfigyelte, hogy nincsenek deduktív garanciák arra, hogy a jövő hasonlítani fog a múltra, mivel logikailag lehetséges marad, hogy bármennyire is egységesen társultak az egyik fajta események egy másik fajta eseményekhez, ezután sem biztos, hogy így lesz. Ha azonban a természeti törvények az időben megmaradnak, akkor zárt rendszerek esetében - változatlanul vagy valószínűségi alapon - pontosan megjósolhatónak kell lennie, hogy ezek a rendszerek hogyan fognak viselkedni  $t^*$  -  $t$  időintervallumokban, amennyiben az ilyen típusú rendszerek teljes kezdeti feltételei és törvényei ismertek.

## **Elmék és agyak**

Mivel a konnekciónizmus az idegcsomópontok aktiválódási mintázataira, nem pedig az egyes csomópontokra, mint az agy reprezentációként működő és a viselkedést befolyásoló jellemzőire apellál, úgy tűnik, hogy számos fontos szempontból javít a számításalapú elképzeléseken, beleértve az ismerős minták érzékelési kiegészítését a hiányzó részek kitöltésével, az új minták felismerését még a korábban ismeretlen esetek vonatkozásában is, a "kegyes leépülés" néven ismert jelenséget és a mentalitás kapcsolódó megnyilvánulásait (Rumelhart et al. 1986, 18-31. o.). A legfontosabb különbségek közé tartozik, hogy a konnekcionista "agyak" képesek az úgynevezett *párhuzamos feldolgozásra*, ami azt jelenti, hogy a (szekvenciális) Turing-gépekkel ellentétben képesek egyszerre több adatfolyam (egyidejű) feldolgozására.

Ez a különbség természetesen kiterjed a fizikai számítógépekre is, amelyek

elrendezhetők az adatok egyidejű feldolgozására, de mindegyikük maga is szekvenciális processzor marad. A párhuzamos feldolgozás előnyei jelentősek, különösen az alábbiak szempontjából

evolúció, ahol például a ragadozók szagának és hangjának érzékelése a ragadozók látványa előtt adaptív előnyökkel járna. Ezen túlmenően a tanulás általában úgy értelmezhető, mint a csomópontok meghatározott mintáira vonatkozó aktivációs küszöbök növelésének vagy csökkentésének folyamata, ahol a klasszikus és az operáns kondicionálás olyan folyamatokként fogadható el, amelyek az aktivációs minták között asszociációt hoznak létre, és hasonló ingerfeltételek mellett valószínűbbé (vagy kevésbé valószínűvé) teszik azok előfordulását, ahol egyes minták aktiválása általában beszédet és más viselkedést vált ki.

Azok, akik továbbra is meg akarják védeni a számítógépes felfogást, azt állíthatják, hogy még ha belső reprezentációik el is vannak osztva, az ember szemantikus motor (h9):

(h9) Az emberek szemantikus motorok, elosztott reprezentációkkal;

de az indoklás egyre kevésbé hihető, és a mechanizmus - például egyre több "független, de összehangolt" soros processzor - egyre inkább ad hoc jellegűnek tűnik. Az eliminatív materializmussal kapcsolatban felmerült okok miatt azonban bármilyen sikeres is legyen a konnekcionizmus mint az agy elmélete, nem tud magyarázatot adni a testek és az elmék közötti kapcsolatra az elme védhető felfogása nélkül, amelynek meg kellene magyaráznia, hogy a szimbólumrendszerek és a szemantikus motorok miért nem gondolkodó dolgok.

## **A teljes Turing-teszt**

A viselkedési bizonyítékok és a mentálisztikus hipotézisek összekapcsolásának nehézségei olyan problémákat vetnek fel, amelyek azt a lehetőséget vetik fel, hogy a Turing-teszt egyszerűen nem elég messzire megy. Stevan Harnad például egy

cikksorozatban vizsgálta annak jellegét és következményeit, amelyben hangsúlyozza annak fontosságát, hogy az egyébként tisztán szintaktikai karakterláncokat szemantikai tartalommal töltsük meg, ha értelmes szimbólumoknak akarjuk őket tekinteni, és nem pusztán értelmetlen jeleknek, ahogyan azt egy megfelelő elmélet megköveteli. A szükségszerűség, hogy egy



a szimbólumok jelentéssel való ellátására alkalmas mechanizmust, amelyet "szimbólum-alapozási problémának" nevez (Harnad 1990). Harnad ezért a Turing-teszt (TT) új és továbbfejlesztett változatát kínálja a teljes Turing-teszt (TTT) formájában, amely a nem verbális és a verbális viselkedést is magában foglalja, ahol a szimbólumokat az általuk megjelenített viselkedés "megalapozza".

Harnad megközelítésében a karteziánus dualizmus és a Turing-féle behaviorizmus keveredik. Mivel egy robot tetszőlegesen hasonló verbális és nem verbális viselkedést mutathat, mint egy emberi lény - egy valódi "gondolkodó lény" - , és mégsem rendelkezik mentalitással, a nyilvános viselkedésünk és a privát elménk között tartósan áthidalhatatlan szakadék tátong:

Ahogy a Searle [Chinese Room] érvével szembeni immunitás sem garantálhatja a mentalitást, úgy a földhözragadság sem teheti ezt meg. Csak azzal az ellenvetéssel szemben nyújt védettséget, hogy a szimbólum és a szimbólum tárgya közötti kapcsolat csak a [külső] megfigyelő elméjében létezik. Egy TTT-vel megkülönböztethetetlen rendszernek még mindig nem lehet elméje; még mindig lehet, hogy nincs benne értelem. Sajnos ez egy ontikus állapot a számunkra episztemikusan örökre elérhetetlen dolgok: Nem lehetünk bölcsebbek. (Harnad 1993, 30. o.)

Harnad tehát nemcsak a Turing-tesztet utasítja el, mint a mentalitás elégséges feltételét

(h10): (h10) A Turing-tesztet (TT) kiálló rendszerek mentalitással (Turing) rendelkeznek;

de elutasítja saját teljes Turing-tesztjét is, mint a mentalitás elégséges feltételét

(h11): (h11) A teljes Turing-tesztet (TTT) teljesítő rendszerek mentalitással

rendelkeznek;

ami miatt arra a következtetésre jut, hogy a más elmék problémája - hogy rajtunk kívül másnak is van-e elméje - soha nem oldható meg. Örökké tudatlanok maradunk.

Harnad azonban talán alábecsüli a rendelkezésünkre álló episztemikus erőforrásokat. A *legjobb magyarázatra való következtetésként* ismert érvelési séma olyan módszertant kínál, amely túlléphet az ő elképzelésein. Ez a megközelítés magában foglalja az egyik tag kiválasztását

a rendelkezésre álló alternatívák halmazát, mint azt a hipotézist, amely a legjobb magyarázatot adja a rendelkezésre álló bizonyítékokra. Azok a hipotézisek, amelyek a rendelkezésre álló bizonyítékok nagyobb részét magyarázzák, előnyben részesítendők azokkal szemben, amelyek kevesebbet magyaráznak. Azokat, amelyek nem egyeztethetők össze a bizonyítékokkal, hamisnak kell elvetni. Azokat, amelyek előnyben részesíthetők, ha elegendő bizonyíték áll rendelkezésre, szintén elfogadhatók. Az elfogadható hipotézisek lehetnek hamisak, ami az ilyen jellegű következtetést hibássá teszi, de a rendelkezésre álló alternatívák közül a legracionálisabbak maradnak. Ha a mentalitás létezésének hipotéziséről kiderül, hogy jobb magyarázatot ad a rendelkezésre álló bizonyítékokra, mint az alternatívái, akkor még mindig indokolt lehet az elfogadása.

### **Szemiotikus rendszerek**

Az elme *szemiotikai* (vagy jelhasználó) rendszerként való felfogása alternatívát kínál a számítási elméletekhez képest, amely úgy tűnik, hogy jól illeszkedik az agy konneccionista modelljéhez. Nem számítási keretet biztosít az elme természetének, az elme és a test kapcsolatának, valamint más elmék létezésének vizsgálatához. E megközelítés szerint *az elmék* olyan dolgok, amelyek esetében valami valamilyen szempontból valami mást helyettesíthet (Fetzer 1990; 1996; 2001). Maga a szemiotikai viszony, amelyet az amerikai filozófus, Charles S. Peirce dolgozott ki, triadikus (vagy háromhelyű), amennyiben a jelek és használói közötti *oksági* viszonyt foglalja magában, egy (döntő fontosságú) *megalapozási* viszony a jelek és az általuk képviselt dolgok között, valamint egy *értelmező* viszony a jelek, az általuk képviselt dolgok és a jelek használói között.

A szemiotika elméletének három ága van: a *szintaxis* a jelek közötti kapcsolatok tanulmányozása, valamint annak vizsgálata, hogy hogyan kombinálhatók új jelek létrehozására, a *szemantika* a jelek közötti kapcsolatok és a jelek által jelzett dolgok tanulmányozása, a *pragmatika* pedig a jelek, a jelek által jelzett dolgok és a jelhasználók közötti kapcsolatok tanulmányozása. Az elmék különböző fajtái ezután a következők alapján osztályozhatók

a jelek fajtái, amelyeket képesek használni, mint például az *ikonok*, amelyek hasonlítanak arra, amit képviselnek (hasonlóak az alakjuk, méretük és hasonlóak); az *indexek*, amelyek okai vagy hatásai annak, amit képviselnek (hamu, tűz és füst), és a *szimbólumok*, amelyek pusztán szokásosan kapcsolódnak ahhoz, amit képviselnek (szavak, mondatok és dolgok), mint a mentalitás ikonikus, indexikus és szimbolikus fajtái.

A jelentések azon lehetséges és tényleges viselkedés összességével azonosíthatók, amelyet a jelhasználó a jel jelenlétében a *kontextus* függvényében mutathat, ami a jelhasználók által a jelekkel való találkozás során hozott motívumok, meggyőződések, etikai elvek, képességek és adottságok kombinációja. Az idegi aktiváció mintái pedig belső jelekként működhetnek, ahol (minden és csakis) a gondolkodó dolgok szemiotikai rendszerek, (h12):

(h12) A gondolkodó dolgok, beleértve az embereket is, szemiotikai rendszerek.

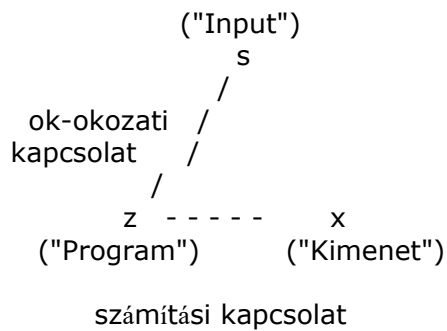
Ez a megközelítés megmagyarázza, hogy mit jelent a jelek egy osztályához viszonyítva tudatosnak lenni, ahol egy rendszer akkor *tudatos az* adott típusú jelek tekintetében, ha képes az adott típusú jelek felhasználására, és nem gátolja e képesség gyakorlásában. És ez alátámasztja *a megismerés* olyan hatásként való felfogása, amelyet a jelek és a jelhasználók közötti kölcsönhatás hoz létre (esetleg valószínűségi alapon), ha megfelelő kauzális közelségben vannak.

### **Kritikus különbségek**

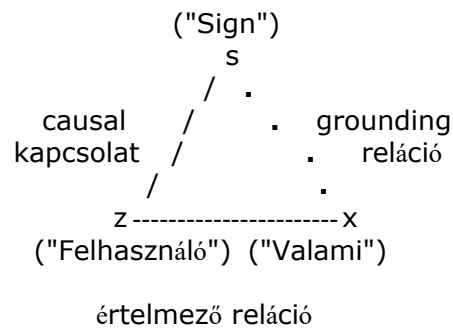
A szemiotikai rendszerek és a számítógépes beszámolók közötti legfontosabb különbségek közül az egyik legfontosabb ezen a ponton válik nyilvánvalóvá, mivel a

mentalitás szemantikai dimenzióját az ilyen típusú rendszerek definíciója magában foglalja. Figyeljük meg például a szimbólumrendszerek és a szemiotikai rendszerek közötti különbséget az 1. és 2. ábrán, ahol

a szemiotikai rendszerek olyan megalapozó kapcsolatot tükröznek, amely a szimbólumrendszerekből hiányzik, a következők szerint:



1. ábra. Szimbólumrendszerek.



2. ábra. Szemiotikus rendszerek.

Ez a különbség még akkor is fennáll, ha ezek a rendszerek azonos eljárásokkal dolgozzák fel a védjegyeket. Egy adóbevallást feldolgozó számítógép ugyanazokat a kimeneteket adhatja ugyanazokból a bemenetekből, de ezek a rendszer számára nem jelentenek semmit, mint például a jövedelem, a levonások vagy a fizetendő adók. Különbséget kell tenni a rendszer számára értelmes jelek és a rendszer felhasználói számára értelmes jelek között. Működhetnek jelekként a felhasználóik számára, és nem működhetnek jelekként e rendszerek számára.

A "szimbólumokat" a szemiotikai rendszerek ezen értelmében ezért világosan meg kell különböztetni a szimbólumrendszerek értelmében vett "szimbólumoktól", amelyek jelentés nélküli jelek lehetnek, nehogy összetéve a Newell és Simon értelmében vett szimbólumrendszereket a (szimbólumokat használó) szemiotikai rendszerekkel, ahogyan azt John McCarthy tette (McCarthy 1996, 12. fejezet). Ez a számítógépes rendszerek és a gondolkodó dolgok közötti (mondhatni) *statikus különbséget* tükrözi. Egy másik

az, hogy a digitális gépek algoritmusok kauzális megvalósításaként programok irányítása alatt állnak, ahol az "algoritmusok" viszont hatékony döntési eljárások. A hatékony döntési eljárások teljesen megbízhatóan, az esetek megfelelő osztályain belül olyan problémamegoldásokat hoznak létre, amelyek változatlanul helyesek, és ezt véges számú lépésben teszik. Ha ezek a gépek algoritmusok irányítása alatt állnak,

de az elme nem, akkor

van *egy dinamikus különbség*, amely lehet, hogy finomabb, de nem kevésbé fontos.

Valóban, sokféle gondolkodás létezik - az álmoktól és álmodozásoktól kezdve az emlékezeten és az észlelésen át a hétköznapi gondolkodásig -, amelyek nem felelnek meg az általunk felállított korlátoknak.



hatékony döntéshozatali eljárásokkal. Ezek nem megbízható problémamegoldó eljárások, és nem kell, hogy véges számú lépésben végleges megoldást adjanak a problémákra. Úgy tűnik, hogy a gondolatok közötti átmeneteket befolyásoló oksági kapcsolatok jobban függenek élettörténetünktől és a kapcsolódó érzelmektől (a pragmatikus kontextusainktól), mint önmagukban a szintaxistól és a szemantikától. Még ugyanaz a tábla, például egy piros lámpa egy kereszteződésben, a táblát használó mindenkori kontextusának függvényében értelmezhető ikonként (mert hasonlít más piros lámpákra), indexként (mint meghibásodott forgalomirányító eszköz) vagy szimbólumként (ahol a járművezetőknek be kell húzniuk a féket és teljesen meg kell állniuk). Bárki más ugyanabban a kontextusban, feltehetően ugyanúgy értelmezte volna ezt a táblát.

## **A hermeneutikai kritika**

Akár a szemiotikai felfogás érvényesül végül, akár nem, a jelenlegi kutatások egyre nyilvánvalóbbá teszik, hogy a mentalitás megfelelő bemutatásának számos olyan aggályt kell kielégítenie, amelyeket a Hubert Dreyfus (1979) által megfogalmazott hermeneutikai kritika vetett fel. Dreyfus nemcsak a reprezentáció atomista felfogása ellen emelt kifogást, amely a jelentés kompozicionalitásának és a rekurzív generálhatósági téziseknek az alapjává vált, amelyeket Fodor és Pylyshyn magáévá tett, hanem a testek mint jelentéshordozók szerepének fontosságát is hangsúlyozta, különösen a világgal való interakciók révén, nagyon is Harnad, sőt Peirce szellemében, akivel sok közös vonása van. Maga a gondolat, hogy olyan mesterséges gondolkodó dolgokat hozzunk létre, amelyek nem fonódnak össze elválaszthatatlanul az ő testek és a világgal való interakcióra alkalmasak, így egyre valószínűtlenebbé válik.

Valójában ma már egyértelműnek tűnik, hogy a Turing-gépek, a digitális kom- és a

digitális számítógépek közötti különbségek...

és az emberi lények még a fent említetteken is túlmutatnak, ahol a tudat és a megismerés szemiotikai felfogása például a hibázás képességét kínálja fel a mentalitás általános kritériumaként, ahol a hibázás magában foglalja a hiba vállalását, és ahol a hiba elkövetése

valamit, hogy valami mást jelképezzen, de rosszul teszi, ami a helyes eredmény.

Ebből a szempontból három legfontosabb különbség látszik, nevezetesen:

	(Kivonat) Turing-gépek	(Fizikai) Digitális számítógépek	(Tényleges) Emberi lények
Végtelen kapacitások:		Nem	Nem
Meghibásodásra hajlamos:	Nincs	Igen	Igen
Képes hibázni:	Nem	Nem	Igen

3. ábra. A dolgok három, egymástól jelentősen különböző fajtája

Még ha eltekintünk is a reprezentáció egy sajátos elméletétől, amely a gépek által manipulálható jelek jelentését hivatott megmagyarázni, az 1. ábrából nyilvánvalónak tűnik, hogy ez három, egymástól határozottan különböző dolog, ahol a gondolkodó dolgok nem hasonlítanak a gépekre.

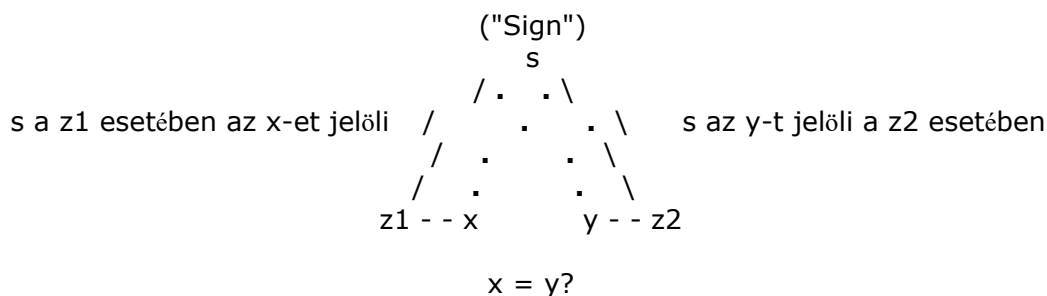
Végső soron természetesen az elmeelmélet megfelelősége az általa biztosított, az agyat, az elmét és a viselkedést összekapcsoló jelentéselmélet megfelelőségétől függ. A döntő szempontnak az tűnik, hogy akár a testek és az elmék determinisztikus, kaotikus vagy valószínűségi rendszerek, az elméletnek teljesen kauzális magyarázatot kell adnia arról, hogy az elmék által használt jelek hogyan befolyásolják e rendszerek viselkedését, ami elegendő ahhoz, hogy az adatok legjobb magyarázataként alátámassza a mentalitás létezésére való következtetést. Az egyik mód, ahogyan ez bekövetkezik, azokból a különböző módokból derül ki, ahogyan az érzetek befolyásolják a viselkedést, ahol a kutya megugatta a bokrot, amikor fúj a szél, *mert azt* tévesen idegennek vélte; ahol Mary a kopogtatás hangjára az ajtóhoz sietett, *mert azt* hitte, hogy a barátja jött; vagy ahol Bob lelassított, amikor a lámpa pirosra váltott, *mert* tudta, hogy be kell húznia a féket, és meg kell állítani az autót.

## **Egyezmények és kommunikáció**

Mivel a különböző felhasználási módok különböző jeleket használhatnak azonos jelentéssel és azonos

különböző jelentésű jeleket, sőt az is lehetséges, hogy egy jelhasználó olyan módon használja a jeleket, amelyek összességükben nem azonosak bármely más használóéval. Ez azt jelenti, hogy a nyelvről alkotott társadalmi elképzelések, amelyek szerint a magánnyelv lehetetlen, a szemiotikai rendszerek szempontjából nem megalapozottak. Egy elhagyatottan egy lakatlan szigeten magára hagyott személy például azzal múlathatja az időt, hogy kidolgozott osztályozási rendszert alkot a sziget növény- és állatvilágára. Még ha ez a jelrendszer ezért egyedi jelentőséggel is bírna az adott felhasználó számára, ez a jelrendszer feltehetően mégis megtanulható lenne abban az értelemben, hogy nincs ok arra, hogy miért ne lehetne megtanítani másoknak. Egyszerűen csak az lenne a helyzet, hogy soha nem volt.

A kommunikációs helyzetekben - legyen szó szóbeli, írásbeli vagy egyéb kommunikációról - a különböző jelhasználók általában akkor járnak sikerrel, ha a jeleket azonos módon használják, vagy olyan mértékben, amilyen mértékben ugyanazokat a dolgokat értik alattuk. A felmerülő kérdés az, hogy ugyanaz az s jel ugyanazt az x dolgot jelenti-e a különböző z1 és z2 jelhasználók számára bizonyos körülmények között:



4. ábra. *Kommunikációs helyzetek.*

Ha z1 és z2 különböző nyelveket beszél, például angolul és németül, a fordítás sikerét nehéz megállapítani. De akkor is nehéz lehet, ha nagyon hasonló hangokhoz olyan jelentések társulnak, amelyek nem biztos, hogy minden felhasználó számára ugyanazt

jelentik.

Vannak olyan körülmények, amelyek között előnyben részesíthetjük a jelzéseink bizalmas kezelését. Maga Turing például a II. világháború idején sikeresen feltörte az Enigma rejtjelezőt, így az angolok megérthették a németek kódolt üzeneteit. Egyéb

körülmények azonban ösztönzik az azonos jelek azonos módon történő használatát, mint például egy olyan közösség esetében, amelynek tagjai közös célokkal és célokkal rendelkeznek. Az állami iskolák rendszereit például általában azzal a céllal finanszírozzák, többek között, hogy a közösség tagjait az általuk használt nyelv közös megértésére neveljék, ami elősegíti a köztük lévő kommunikációt és együttműködést. Az olyan nagy nemzetek, mint az Egyesült Államok, mérhetetlenül sokat profitáltak abból, hogy "olvasztótégelyként" működnek, ahol sok országból származó emberek találkoznak, és az angol nyelvre való támaszkodás egyesíti őket, amelynek hiányában ez az ország, akárcsak mások, a balkanizálódás felé hajlana.

## **Más elmék**

Az elmék szemiotikai rendszerként való felfogása tisztázza és megvilágítja a különböző ok-okozati folyamatok sajátosan mentális aspektusait is. Amikor kauzális kapcsolatok lépnek fel (amikor az okok, mint például a bemenetek, hatásokat, mint például a kimenetek, idéznek elő), és ezek a bemenetek és kimenetek nem jelként szolgálnak egy rendszer számára, akkor *ingereknek* minősíthetők. Amikor a hatások az őket használó rendszerek számára a megalapozottságuk révén (mert az adott dolgokat az adott tekintetben képviselik) idéznek elő hatásokat, akkor ezeket helyesen *jelekként* lehet besorolni. Amikor pedig szemiotikai kapcsolatok jönnek létre (amikor az egyik felhasználó által használt jeleket egy másik felhasználó értelmezi) az őket használó rendszerek között, akkor tovább lehet őket *jelekként* osztályozni.

Néha az általunk küldött jelek szándékosak (sikeresek stb.), néha nem. Minden jelnek kell, hogy legyen inger, és minden jelnek kell, hogy legyen jel is, de fordítva nem.

Minden emberi lény, (más) állat és élettelen gép, amely képes jelek használatára, a szemiotikai felfogás szerint gondolkodó dolognak minősül. Ez a felismerés így megmagyarázza, hogy az álmok és az álmodozások, az emlékezet és az észlelés,

valamint a hétköznapi gondolkodás miatt mentális tevékenységek, míg a fogszuvasodás, a prosztatarák és a szabadesés ehhez képest miért nem azok. Akár a szemiotikai felfogás emelkedik ki a legadekvátabbként az alternatív felfogások közül, akár nem, nyilvánvalóvá vált, hogy egy adekvát beszámolóknak a következőkre kellene kiterjednie



legalábbis nagyon hasonlít rá, különösen a Turing-gépek, a digitális számítógépek és az emberek közötti alapvető különbségek figyelembevételével. Feltételezésem szerint az is nyilvánvalóvá vált, hogy az elmék nem gépek. Ha a gondolkodást mentális algoritmusok irányítanák, ahogyan az ilyen elméletek sugallják, akkor az elmék egyszerűen mechanikusan követnék az utasításokat, mint a robotok, és nem lenne szükségük éleslátásra, találgatásra vagy találgatásra. Talán azért tagadjuk, hogy nem vagyunk mások, mint robotok, mert mentális tevékenységeink sokkal többről szólnak. Valójában a gondolkodás néhány legjellegzetesebb aspektusa hajlamos elválasztani az elmét a gépektől.

A szimulációk egyértelműen túl gyengék, az *emulációk pedig*, amelyek ugyanazokból a kimenetekből ugyanazokkal a folyamatokkal ugyanazokat a bemeneteket állítják elő, és ugyanabból az anyagból készülnek, egyértelműen túl erősek. De a zátonyok alattomosak. David Chalmers például amellettt érvelt, hogy bizonyos rendszerek esetében a *szimulációk replikációk*, feltételezve, hogy ugyanazok a pszichofizikai törvényszerűségek működnek. Ha tehát egy  $t_1$  időpontban bekövetkező  $S_1$  kezdeti állapotból  $t_n$  időpontban egy  $S_n$  végállapotot eredményez, ahol a köztük lévő köztes lépések, mondjuk  $S_2$   $t_2$  időpontban és  $S_{n-1}$   $t_{n-1}$  időpontban, ugyanazok, akkor a velük törvényszerűen összefüggő tulajdonságoknak, mint például a tudatosság, velük együtt kell jönniük, még akkor is, ha más anyagból vannak (Chalmers 1996).

De ez csak akkor lesz igaz, ha az anyagbeli különbség nem befolyásolja maguknak a törvényeknek a működését. Azokban az esetekben, ahol igen, a *replikációkhoz emulációkra lehet szükség.*

## **Intelligens gépek**

Ha ezek a megfontolások megalapozottak, akkor az elmék ilyen irányú felfogása képes lenne megmagyarázni, hogy a szimbólumrendszerek és a szemantikus motorok miért nem gondolkodó dolgok. Vagy a gondolatok formájáról tudnak számot adni, de a tartalmukról nem, vagy magukról a gondolatok közötti átmenetekről nem tudnak számot adni. A Turing-gépek, amelyekkel kezdtük, még csak nem is fizikai dolgok, és nem is tudnak

fenntartani a véges elmék létezését, amelyek meghibásodhatnak és hibázhatnak.

Az agyoknak (nedves) neurális hálózatokként való konnekcionista felfogása döntő alapot nyújt az elme természetének újragondolásához, de kiegészítésre szorul az elme természetének egy nem komputációs jellegű bemutatásával. Az elme mint szemiotikai rendszer hipotézise az agy konnekcionista felfogásával ötvöződik, hogy

a tudatosság és a megismerés teljesen kauzális, nem komputeralapú elméletét támogatják.

Az ebből származó előnyök közül nem utolsósorban a mentalitásnak a biológiával és az evolúcióval összeegyeztethető bemutatása. A primitív organizmusoknak rendkívül elemi szemiotikai képességekkel kellett rendelkezniük, mint például a fényérzékenység a mozgást előidéző, zászlós sejtek segítségével. Ha a fény felé való mozgás elősegíti a túlélést és a szaporodást, akkor ez a viselkedés adaptív előnyökkel járna az ilyen egyszerű rendszerek számára. A genetikai mutáció, a természetes szelekció együttes hatására,

genetikai sodródás, szexuális szaporodás, szexuális szelekció, csoportos szelekció, mesterséges szelekció és géntechnológia, természetesen a biológiai evolúció, beleértve a saját fajunkat is, a mai napig folytatódik, és a szemiotikai rendszerek egyre összetettebb formáit hozza létre, olyan képességekkel, amelyek több hasonló jellegű és más, különböző jellegű jelek használatára képesek, az evolúció által lehetővé tett valószínűségi fejlődés következményeként.

Ahogy az ember által készített (száraz) neurális hálózatok konnekcionista rendszerei fejlődnek, nem lenne túl meglepő, ha elérnének egy olyan pontra, ahol megfelelően a *mesterséges gondolkodó dolgok* közé sorolhatók. Hogy ez a pont valaha is eljön-e, az a tudomány és a technológia fejlődésétől függ, amelyre a filozófusoknak nincs ráhatásuk. Bár a szimbólumrendszerek, sőt a szemantikus motorok koncepciója úgy tűnik, hogy nem képes megragadni a gondolkodó dolgok

jellegét, ez nem jelenti azt, hogy nem képesek megragadni az intelligens gépek jellegét. Amennyiben a gépek akkor minősülnek megfelelően "intelligensnek", ha képesek megbízhatóan feldolgozni összetett feladatokat, egyértelmű, hogy az *intelligens gépek* eljövetele már régen megérkezett.

A fogalmi zűrzavar egyszerűen abban állt, hogy az intelligens gépeket összekeverték a gondolkodó dolgokkal, amit a téma története mára lényegében tisztázott.

\* Ez a fejezet a "The Philosophy of AI and Its Critique" című fejezet kibővített és átdolgozott változata, Luciano Floridi, szerk., The Blackwell Guide to the Philosophy of Computing and Information (Malden, MA: Blackwell Publishers, megjelenés előtt). A szerző engedélyével.

## Hivatkozások

CHALMERS, D. (1996). A tudatos elme. New York, NY: Oxford University Press.  
CHARNIAK, E. és McDermott, D. (1985). Bevezetés a mesterséges intelligenciába.

Reading, MA: Addison-Wesley Publishing Company.

CHURCHLAND, P. (1984). Anyag és tudat. Cambridge, MA: The MIT COLBURN,

T., et al. (szerk.) (1993). Program Verification: Alapvető kérdések

az informatikában. Dordrecht, Hollandia: Kluwer Academic Publishers.

DRYFUS, H. (1979). Amire a számítógépek nem képesek: The Limits of Artificial

Intelligencia, átdolgozott kiadás. New York: Harper & Row.

FETZER, J. H. (1988). "Program Verification: The Very Idea". Communications

of the ACM 31, 1048-1063.

FETZER, J. H. (1990). Mesterséges intelligencia: Terjedelme és korlátai.

Dordrecht, Hollandia: Kluwer Academic Publishers.

FETZER, J. H. (1992). "A konnencionizmus és a megismerés: Why Fodor and

Pylyshyn are Wrong". In A. Clark and R. Lutz (eds.), Connectionism in Context

(pp. 37- 56). Berlin, Németország: Springer-Verlag.

FETZER, J. H. (1996). Filozófia és kognitív tudomány, 2. kiadás. St. Paul, MN:

Paragon House.

- FETZER, J. H. (2001). Számítógépek és megismerés: Why Minds are Not Machines.  
Dordrecht, Hollandia: Kluwer Academic Publishers.
- FODOR, J. (1975). A gondolkodás nyelve. Cambridge, MA: The MIT Press.
- FODOR, J. (1980). "A módszertani szolipszizmus mint kutatási stratégia a kognitív  
Pszichológia". In J. Haugeland (szerk.), Mind Design (pp. 307-338).  
Cambridge, MA: The MIT Press.
- FODOR, J. és Pylyshyn, Z. (1988). "A konnekciónizmus és a kognitív  
architektúra: A Critical Analysis". Cognition 28, 3-71.
- FORD, K. M. és Hayes, P. (szerk.) (1991). Reasoning Agents in a Dynamic World:
- HARNAD, S. (1990). "A szimbólumföldelés problémája". Physica D 42, 335-346.
- HARNAD, S. (1993). "Szimbólumok földelése az analóg világban neurális hálókkal:  
Egy hibrid modell". THINK 2, 12-20.
- HAUGELAND, J. (1981). "Szemantikus motorok: An Introduction to Mind Design". In  
J. Haugeland (szerk.), Mind Design (pp. 1-34). Cambridge, MA: The MIT Press.
- HOFSTADTER, D. (1979). Godel, Escher, Bach: Egy örök aranyfonat. New York:  
Basic Books.
- LUCAS, J. R. (1961). "Elmék, gépek és Godel". Philosophy 36, 112-127.
- MCCARTHY, J. (1996). A mesterséges intelligencia kutatásának védelme. Stanford,  
CA: CSLI Lecture Notes.
- NEWELL, A. és Simon, H. (1976). "A számítástechnika mint empirikus vizsgálat:  
Symbols and Search". In J. Haugeland (szerk.), Mind Design (pp. 35-66). Cambridge,  
MA:  
The MIT Press, 1981.
- PENROSE, R. (1989). A császár új elméje. New York: Oxford University Press.
- RUMELHART, D.,  
et al. (1986). Párhuzamos elosztott feldolgozás, Vols. 1. és 2. kötet. Cambridge,

MA: The MIT Press.

SEARLE, J. (1980), "Elmék, agyak és programok". Behavior and Brain Sciences

3, 417-457.

STICH, S. (1983). A népi pszichológiától a kognitív tudományig: Az ügy ellen

Hit. Cambridge, MA: The MIT Press.

TURING, A. M. (1950). "Számítógépek és intelligencia". Mind LIX, 433-460.