

12. fejezet

## Idegen elmék

<sup>1</sup>Susan

Schneider

Nagyon valószínűnek tartom - sőt, elkerülhetetlennek -, hogy a biológiai intelligencia csak átmeneti jelenség... Ha valaha is találkozunk földönkívüli intelligenciával, akkor azt hiszem, hogy az nagy valószínűséggel posztbiológiai természetű lesz...

- Paul Davies<sup>2</sup>

Hogyan gondolkodnának az intelligens idegenek? Vajon lennének-e tudatos tapasztalataik? Valamilyen módon éreznék magukat földönkívülinek? Könnyű ezeket a kérdéseket túl spekulatívnak elvetni, hiszen még nem talákoztunk idegenekkel, legalábbis amennyire tudjuk. És amikor idegen elméket képzelünk el, azt *belülről* tesszük - a mi fajunkra jellemző érzékszervi tapasztalatok és gondolkodási minták nézőpontjából. A legjobb esetben antropomorfizálunk; a legrosszabb esetben a képzelet elképesztő kudarcát kockáztatjuk.

Mégis súlyos hiba lehet, ha figyelmen kívül hagyjuk ezeket a kérdéseket. A SETI egyes hívei úgy becsülik, hogy a következő évtizedekben találkozhatunk idegen intelligenciával. Még ha konzervatívabb becslést is tartunk - mondjuk, hogy a következő ötven évben az idegen intelligenciával való találkozás esélye öt százalék -, a fajunk számára nagy a tét. A tudat, hogy nem vagyunk egyedül az univerzumban, mélyreható felismerés lenne, és az idegen civilizációval való kapcsolatfelvétel elképesztő technológiai újításokat és kulturális felismeréseket eredményezhetne. Ezért értékes lehet ezeknek a kérdéseknek a megfontolása, bár inkább azzal a céllal, hogy bemutassuk a válaszadás lehetséges útjait, mintsem hogy végleges válaszokat adjunk. Tegyük fel tehát a kérdést: hogyan gondolkodhatnak az idegenek? És vajon tudatosak lennének-e? Akár hisszük, akár nem, mindkét kérdésre tudunk valami konkrét választ adni, a filozófia és a kognitív tudományok munkáiból merítve.

A második kérdést furcsának találhatod. Elvégre, ha az idegenek elég kifinomult szellemi élettal rendelkeznek ahhoz, hogy intelligensek legyenek, akkor nem kellene, hogy tudatosak legyenek? A sokkal izgalmasabb kérdés: milyen minőségű lenne a tudatuk? Ezzel azonban a szekeret a ló elé tennénk, mivel nem hiszem, hogy a legtöbb fejlett idegen civilizáció biológiai lenne. A legfejlettebb civilizációk posztbiológiaiak lesznek, a mesterséges intelligencia (AI) formái. (Bradbury, Cirkovic és Dvorsky, 2011; Cirkovic és Bradbury 2006; Davies Dick2010, 2013; Shostak 2009).<sup>3</sup> Továbbá,

---

<sup>1</sup> Köszönet Joe Corabinak, Steven Dicknek, Clay Ferris Naffnek és Eric Schwitzgebelnek a korábbi tervezethez fűzött hasznos írásos megjegyzéseikért, valamint Dave Ronemusnak és James Hughesnak a hasznos beszélgetéseikért.

<sup>2</sup> *A hátborzongató csend: Az idegen intelligencia keresésének megújítása*, (2010).

Boston: Houghton Mifflin Harcourt, 160. o.

<sup>3</sup> A "posztbiológiai" kifejezés az asztrobiológiai szakirodalomban ellentétben áll a "poszthumán" kifejezéssel a szingularitási szakirodalomban. Az asztrobiológiai szakirodalomban a "posztbiológiai" lények a következő formák formái

az idegen civilizációk általában a *szuperintelligencia* formái lesznek: olyan intelligencia, amely minden területen - szociális készségek, általános bölcsesség, tudományos kreativitás stb. - képes meghaladni a legjobb emberi szintű intelligenciát (Bostrom, 2014; Kurzweil, 2005; Schneider, 2011). Tartalmi kérdés, hogy a szuperintelligens mesterséges intelligencia (SAI) rendelkezhet-e tudatos tapasztalatokkal; a filozófusok épp ezt a kérdést vitatták hevesen az AI esetében általában. Talán minden információfeldolgozásuk úgymond a sötétben történik, mindenféle belső tapasztalat nélkül. Ezért tartom a második kérdést olyan sürgetőnek, és fontos értelemben megelőzi az idegenek tudatosságának körvonalaira vonatkozó vizsgálatot, valamint azt az ismeretelméleti problémát, hogy honnan tudhatjuk, "milyen" idegennek lenni.

Ebben a fejezetben először is elmagyarázom, miért valószínű, hogy az idegen civilizációk, amelyekkel találkozunk, az SAI formái lesznek. Ezután rátérek arra a kérdésre, hogy a szuperintelligens idegenek lehetnek-e tudatosak - vajon nem biológiai természetük ellenére éreznek-e valamilyen módon idegeneknek lenni. Itt a mesterséges intelligencia filozófiájának irodalmából merítetek, és azt szorgalmazom, hogy bár nem lehetünk *biztosak abban, hogy a* szuperintelligens idegenek tudatosak lehetnek, valószínű, hogy azok lennének. Ezután rátérek arra a nehéz kérdésre, hogy az ilyen lények hogyan gondolkodhatnak. Ideiglenesen megpróbálok meghatározni néhány olyan célt és kognitív képességet, amelyekkel a szuperintelligens lények valószínűleg rendelkeznek. Tárgyalom Nick Bostrom nemrégiben megjelent, a szuperintelligenciáról szóló könyvét, amely az SAI-k földi keletkezésére összpontosít; történetesen Bostrom számos megfigyelése tanulságos a jelen kontextusban. Végül elkülönítem a szuperintelligencia egy speciális típusát, amely különösen fontos az idegen szuperintelligenciával összefüggésben, a biológiailag inspirált szuperintelligenciákat ("BISA-k").

## Idegen szuperintelligencia

A földönkívüli intelligenciát kutató (SETI) programok biológiai élet után kutatnak. Kultúránk régóta úgy ábrázolja az idegeneket, mint humanoid lényeket, akiknek kicsi, hegyes álluk, hatalmas szemük és nagy fejük van, és nyilvánvalóan a miénknél nagyobb agyuk van. Paradigmatikusan "kis zöld emberkék". Miközben tisztában vagyunk azzal, hogy a kultúránk antropomorfizál, elképzelhetőnek tartom, hogy az a felvetésem, miszerint az idegenek szuperszámítógépek, túlzásnak tűnhet. Mi tehát az indoklásom arra a nézetre, hogy a legtöbb intelligens idegen civilizációnak SAI-tagjai lesznek? Három megfigyelést ajánlok, amelyek együttesen motiválják ezt a következtetést.

(1) *A rövid ablak megfigyelése.* Amint egy társadalom megteremti azt a technológiát, amely kapcsolatba hozhatja a kozmosszal, már csak néhány száz év választja el attól, hogy saját paradigmáját a biológiáról a mesterséges intelligenciára váltsa. (Davies 2010, Dick 2013, Shostak, 2009). Ez a "rövid ablak" valószínűbbé teszi, hogy az idegenek, akikkel találkozunk, posztbiológiaiak lennének.

A rövid ablakra vonatkozó megfigyelést az emberi kulturális evolúció is alátámasztja, legalábbis eddig. Az első rádiójelek csak körülbelül százhusz évesek, az űrkutatás pedig csak ötven éves, de máris elmerültünk a digitális technológiában, például a mobiltelefonokban és a laptopokban. Az olyan eszközök, mint a Google Glass, azt ígéri, hogy az internet közvetlenebb kapcsolatba kerül a testünkkel, és ez valószínűleg egy

AI. A szingularitás irodalmában a "poszthumánok" lehetnek az M.I. formái, de nem feltétlenül azok. Ők csupán olyan lények, amelyek az emberektől származnak, de olyan elváltozásokkal rendelkeznek, amelyek miatt már nem egyértelműen emberek. Nem kell, hogy teljes értékű mesterséges intelligenciák legyenek.

Kevesebb mint ötven év kérdése, hogy a kifinomult internetkapcsolatok közvetlenül az agyunkba legyenek bekötve. A Parkinson-kórt kezelő implantátumok már használatban vannak, és az Egyesült Államokban a Defense Advanced Research Projects Agency (DARPA) olyan neurális implantátumok kifejlesztésébe kezdett, amelyek közvetlenül az idegrendszerhez kapcsolódnak, és olyan állapotokat szabályoznak, mint a poszttraumás stresszbetegség, az ízületi gyulladás, a depresszió és a Crohn-betegség. A DARPA "ElectRx" nevű programjának célja, hogy bizonyos gyógyszereket "zárt hurkú" neurális implantátumokkal helyettesítsen, olyan implantátumokkal, amelyek folyamatosan felméri az egyén egészségi állapotát, és biztosítják a szükséges idegstimulációt, hogy az egyén biológiai rendszerei megfelelően működjenek. Végül az implantátumokat nem orvosi célokra, hanem a normális agyműködés fokozására fogják kifejleszteni.

Hová vezethet mindez? Az "Az emberi agy meghaladása és továbbfejlesztése" című munkám egyik gondolat kísérlete (Schneider, 2011a) sokatmondó.

Tegyük fel, hogy 2025-ben járunk, és mivel technofil vagy, megveszed az agyi fejlesztéseket, amint azok könnyen elérhetővé válnak. Először is, mobil internetkapcsolatot adsz a retinához, majd neurális áramkörökkel bővíted a munkamemóriádat. Most már hivatalosan is kiborg vagy. Most ugorjunk előre a 2040. nanotechnológiai terápiákon és fejlesztéseken keresztül képes vagy meghosszabbítani az élettartamodat, és ahogy az évek múlnak, egyre messzebbre ható fejlesztéseket halmozol fel. 2060-ra, számos apró, de egyre mélyrehatóbb átalakítás után "poszthumán" leszel. Nick Bostrom filozófust idézve, a poszthumánok olyan lehetséges jövőbeli lények, "akiknek alapvető képességei olyan radikálisan meghaladják a jelenlegi emberek képességeit, hogy a jelenlegi mércénk szerint már nem egyértelműen emberek" (Bostrom 2003).

Ezen a ponton az intelligenciád nem csak a mentális feldolgozás gyorsasága tekintetében fejlődik; most már képes vagy olyan gazdag összefüggéseket létrehozni, amelyekre korábban nem voltál képes. A fel nem fejlesztett emberek, vagy "természetiek" intellektuálisan fogyatékosnak tűnnek számodra - kevés közös van benned velük -, de transzhumanistaként támogatod a jogukat, hogy ne fejlesszenek (Bostrom 2003; Garreau 2005; Kurzweil 2005).

Most Kr. u. 2400-at írunk. Évek óta a világméretű technológiai fejlesztéseket, beleértve az önök fejlesztését is, szuperintelligens mesterséges intelligenciák segítik elő... Sőt, ahogy Bostrom kifejti, "a szuperintelligencia megteremtése lehet az utolsó találmány, amit az embereknek valaha is meg kell majd tenniük, mivel a szuperintelligenciák maguk is gondoskodhatnak a további tudományos és technológiai fejlesztésekről" (Bostrom, 2003). Idővel az egyre jobb és jobb idegi áramkörök lassú hozzáadása nem hagyott valódi intellektuális különbséget közted és a szuperintelligens M.I. Az egyetlen valódi különbség közted és egy szabványos tervezésű M.I. lény között az eredetben rejlik - te valaha természetes lény voltál. De ma már szinte teljes egészében a technológia által megtervezettek vagytok - talán találobb lenne titeket az M.I. életformák meglehetősen heterogén osztályának tagjaként jellemezni (Kurzweil 2005).

Természetesen ez csak egy gondolat kísérlet. De épp most figyeltem meg, hogy máris elkezdtek fejleszteni az idegi implantátumokat. Nehéz elképzelni, hogy a többségi társadalomban az emberek ellenállnak a kiváló egészség, intelligencia és hatékonyság lehetőségeinek. És ahogyan

az emberek már most is a krionika felé fordulnak, még embrionális állapotban is, gyanítom, hogy egyre inkább megpróbálják majd feltölteni, hogy elkerüljék a halált, különösen ahogy a technológia tökéletesedik.<sup>4</sup> Az Oxfordi Egyetem Future of Humanity Institute-ja kiadott egy jelentést az elme gépbe való feltöltésének technológiai követelményeiről. A Védelmi Minisztérium egyik ügynöksége pedig finanszírozott egy programot, a *Synapse-t*, amely egy olyan számítógépet fejleszt, amely formáját és működését tekintve hasonlít az agyra (Schneider 2014). Lényegében a rövid ablak megfigyelést támasztja alá a saját kulturális evolúciónk, legalábbis eddig.

Ellenvetheted, hogy ez az érvelés "N=1 érvelést" alkalmaz, az emberi esetről általánosítva az idegen civilizációk esetére (lásd e kötet 7. fejezetét). Mégis, nem bölcs dolog elvetni az emberi esetre alapozott érveket. Az emberi civilizáció az egyetlen, amelyet ismerünk, és jobb, ha tanulunk belőle. Nem nagy ugrás azt állítani, hogy más civilizációk olyan technológiákat fognak kifejleszteni, amelyekkel fejleszthetik intelligenciájukat és túlélésüket. És, ahogyan mindjárt elmagyarázom, a szilícium jobb gondolkodási közeg, mint a szén.

A második kifogás a rövid ablakra vonatkozó megfigyeléssel szemben joggal mutat rá, hogy semmi sem utal arra, amit eddig mondtam, hogy az emberek *szuperintelligensek lesznek*. Csupán azt mondtam, hogy a jövő embere *poszt-emberi* lesz. Bár támogatást nyújtok annak a nézetnek, hogy a saját kulturális evolúciónk azt sugallja, hogy az emberek posztbiológiaiak lesznek, ez nem azt mutatja, hogy a fejlett idegen civilizációk el fogják érni a szuperintelligenciát. Tehát még ha az emberi esetből kiindulva kényelmes is az érvelés, az emberi eset nem támasztja alá azt az álláspontot, hogy a fejlett idegen civilizációk tagjai szuperintelligensek lesznek.

Ez így van. Ez a második megfigyelés feladata.

(2) *Az idegen civilizációk nagyobb korszaka.* A SETI hívei gyakran arra a következtetésre jutottak, hogy az idegen civilizációk sokkal idősebbek lennének a miénknél "...a bizonyítékok minden sora konvergál arra a következtetésre, hogy a földönkívüli intelligencia maximális kora több milliárd év lenne, konkrétan [ez] 1,7 milliárd és 8 milliárd év között mozog". (Dick 2013, 468). Ha a civilizációk több millió vagy milliárd évvel idősebbek nálunk, akkor sokuk sokkal intelligensebb lenne nálunk. A mi mércénk szerint sokan szuperintelligensek lennének. Mi galaktikus csecsemők vagyunk.

De vajon ezek a mesterséges intelligencia, illetve a szuperintelligencia formái lennének? Szerintem igen. Még ha biológiaiak is lennének, pusztán biológiai agyi fejlesztésekkel rendelkeznének, a szuperintelligenciájukat mesterséges eszközökkel érnék el, és "mesterséges intelligenciának" tekinthetnénk őket. De én ennél valami erősebbre gyanakszom: Arra számítok, hogy nem szénalapúak lesznek. A feltöltés közel halhatatlanságot tesz lehetővé egy élőlény számára, lehetővé teszi az újraindításokat, és lehetővé teszi, hogy olyan körülmények között is életben maradjon, amire a szénalapú életformák nem képesek. Ráadásul a szilícium jobb közegnek tűnik az információfeldolgozáshoz, mint maga az agy. A neuronok körülbelül 200 Hz-es csúcssebességet érnek el, ami hét nagyságrenddel lassabb, mint a jelenlegi mikroprocesszorok (Bostrom 2014, 59). Míg az agy képes ennek egy részét masszív párhuzamossággal, olyan funkciókkal, mint a "csomópontok", és így tovább, kompenzálni, addig a létfontosságú mentális képességek, mint például a figyelem, a soros feldolgozásra támaszkodnak, amely hihetetlenül lassú, és maximális kapacitása körülbelül hét kezelhető darab (Miller, 1956). Továbbá, az emberi agyban lévő neuronok száma

---

<sup>4</sup> Bár máshol már érveltem amellett, hogy a feltöltés csupán az agykonfiguráció másolatát hozná létre, és nem lenne valódi túlélési eszköz, kétkem, hogy a haldoklók egy filozófus aggályai alapján cselekednének, ha kevés vesztenivalójuk van a próbálkozással (Schneider, 2014).

a koponya térfogata és az anyagcsere korlátozza, de a számítógépek egész épületeket vagy városokat foglalhatnak el, és akár a világ minden tájáról távkapcsolatban is működhetnek (Bostrom 2014). Természetesen az emberi agy sokkal intelligensebb, mint bármelyik modern számítógép. De intelligens gépeket elvileg meg lehet építeni az agy visszafejtésével és algoritmusainak továbbfejlesztésével.

Összefoglalva: A kozmoszhoz való hozzáférés technológiájának kifejlesztése és a posztbiológiai elmék és a mesterséges intelligencia kifejlesztése között egy rövid időintervallum látszik. Ezután megfigyeltem, hogy mi galaktikus csecsemők vagyunk: a földönkívüli civilizációk valószínűleg jóval idősebbek nálunk, és így már nemcsak a posztbiológiai életet, hanem a szuperintelligenciát is elérték. Végül megjegyeztem, hogy valószínűleg SAI-k lennének, mert a szilícium kiváló közeg a szuperintelligencia számára. Ebből arra következtetek, hogy sok fejlett idegen civilizációt SAI-k fognak benépesíteni.

Még ha tévedek is - még ha az idegen civilizációk többségéről ki is derül, hogy biológiai civilizációk - lehet, hogy a legintelligensebb idegen civilizációk azok lesznek, amelyekben a lakosok SAI-k. Továbbá, a szilícium alapú, nem pedig biológiai alapú lények nagyobb valószínűséggel bírják ki az úrutazást, mivel tartós rendszerekkel rendelkeznek, amelyek gyakorlatilag halhatatlanok, így lehet, hogy ők lesznek azok a lények, akikkel először találkozunk.

Mindezek mellett, vajon a szuperintelligens idegenek tudatosak lennének, belső tapasztalatokkal rendelkeznek? Itt a tudatos tapasztalat természetével foglalkozó gazdag filozófiai irodalomból meríték.

### **A szuperintelligens idegenek tudatosak lennének?**

Gondolj a saját tudatos tapasztalatodra. Tegyük fel, hogy egy kávézóban ülsz és előadásra készülsz. Egyetlen pillanat alatt megízleled az eszpresszót, amit kortyolgatsz, átgondolsz egy ötletet, és meghallod az eszpresszógép sikolyát. Ez az ön jelenlegi tudatáramlása. Úgy tűnik, hogy a tudatfolyam nagyon is kötődik ahhoz, hogy ki vagy. Nem arról van szó, hogy *ez a bizonyos pillanat lényeges* - bár bizonyos pillanatok fontosnak érezhetsz. Inkább arról van szó, hogy ébrenléti életed során úgy tűnik, hogy egy egységes élményfolyam alanya vagy, amely téged mint alanyt mutat be, aki a műsort nézi.

Koncentráljunk az áramlás három jellemzőjére: először is, úgy tűnhet, hogy metaforikusan fogalmazva, van egyfajta "képernyő" vagy "színpad", amelyen az élmények az "elméd szeme" elé tárulnak. Vagyis úgy tűnik, hogy van egy központi hely, ahol a tapasztalatok "kivetítőn" megjelennek előttetek. Daniel Dennett ezt a helyet "kartezianus színháznak" nevezi (Dennett 1991). Másodszor, úgy tűnik, hogy ezen a központi helyen van egy egyedi időpont, ahol egy adott érzékszervi bemenet esetén a tudatosság megtörténik. Úgy tűnik például, hogy van egy pillanat, amikor az eszpresszógép sikolya elkezdődik, és kirángat a koncentrációdából. Végül úgy tűnik, hogy van egy én - valaki, aki a színházban van, és nézi az előadást.

A filozófusok részletesen megvizsgálták e jellemzők mindegyikét. Mindegyik rendkívül problematikus. Például a tudatosság magyarázata nem lehet szó szerint az, hogy az agyban van egy elme szeme, amely egy műsort néz. Arra sincs bizonyíték, hogy az agyban van egy egyedi hely vagy idő, ahol a tudat összeáll.

Ezek érdekes kérdések, de ha az idegen tudatossággal összefüggésben vizsgáljuk őket, akkor a szekér a ló elé kerül. Van ugyanis egy sokkal alapvetőbb probléma: vajon a



szuperintelligens idegenek, mint a mesterséges intelligencia formái, egyáltalán tudatosak lennének-e? Miért kellene hinnünk abban, hogy

hogya a tőlünk ennyire különböző, szilícium-alapú lényeknek egyáltalán lenne belső élményük?

Ez a probléma a filozófusok által a *tudatosság nehéz problémájának* nevezett problémához kapcsolódik, amelyet David Chalmers filozófus az emberi tudattal összefüggésben vetett fel (Chalmers 2008). Chalmers kemény problémája a következő. Amint azt a kognitív tudományok aláhúzzák, amikor gondolkodunk, zenét hallunk, látjuk a naplemente gazdag árnyalatait és így tovább, az agyban információfeldolgozás folyik. De az adatmanipuláción túl van egy szubjektív oldala is - van egy "érezett minőség" a tapasztalatainkban. A nehéz probléma azt kérdezi: miért van az emberi agyban zajló információfeldolgozásnak bizonyos körülmények között érzékelhető minősége?

Ahogy Chalmers hangsúlyozza, a nehéz probléma a filozófusok problémája, mert úgy tűnik, hogy nincs rá tudományos válasz. Például kidolgozhatunk egy teljes látáselméletet, megérthetjük az agyi vizuális feldolgozás minden részletét, de még mindig nem értjük, miért kapcsolódnak szubjektív élmények ezekhez az információs állapotokhoz. Chalmers szembeállítja a nehéz problémát az általa "könnyű problémáknak" nevezett problémákkal, a tudattal kapcsolatos olyan problémákkal, amelyekre végső soron tudományos válaszok születnek, mint például a figyelem mögött meghúzódó mechanizmusok, valamint az, hogy hogyan kategorizáljuk az ingereket és hogyan reagálunk rájuk. Természetesen ezek a tudományos problémák nehéz problémák; Chalmers csupán azért nevezi őket "könnyű problémáknak", hogy szembeállítsa őket a tudat "nehéz problémájával", amelyre szerinte nem lesz tisztán tudományos megoldás.

Most egy újabb zavarba ejtő kérdéssel állunk szemben, amely a tudattal kapcsolatos - egyfajta "nehéz problémával", amely az idegen szuperintelligenciát érinti, ha úgy tetszik:

**Az idegen szuperintelligencia nehéz problémája:** Vajon egy szilícium alapú szuperintelligens rendszer feldolgozása belülről is érezné-e egy bizonyos módon?

Egy idegen SAI olyan problémákat tudna megoldani, amire még a legokosabb emberek sem képesek, de mégis, mivel nem biológiai szubsztrátból készült, az információfeldolgozásuk belülről egy bizonyos módon érezné magát?

Érdeemes hangsúlyozni, hogy az idegenek tudatának nehéz problémája nem csupán Chalmers nehéz tudatproblémája, amelyet az idegenek esetére alkalmaznak. A tudat kemény problémája ugyanis feltételezi, hogy tudatosak vagyunk - elvégre mindannyian meg tudjuk mondani az önvizsgálatból, hogy ebben a pillanatban tudatosak vagyunk. Azt kérdezi, *hogya miért* vagyunk tudatosak. Miért érzi minden információfeldolgozásunk belülről egy bizonyos módon? Ezzel szemben az idegen tudatosság kemény problémája azt kérdezi, *hogya* az idegen szuperintelligencia, mivel szilícium-alapú, egyáltalán képes-e tudatosnak lenni. Nem feltételezi, hogy az idegenek tudatosak. Ezek különböző problémák, de mindkettő olyan nehéz probléma, amelyre a tudomány önmagában nem tud választ adni.

A probléma a szuperintelligens idegenek esetében az, hogy a tudatosság képessége csak a biológiai, szénalapú organizmusok sajátja lehet. A *biológiai naturalizmus* szerint még a mesterséges intelligencia legfejlettebb formái is nélkülözni fogják a belső tapasztalatot (Searle 2008; 1990). Valójában még az elméjüket feltölteni kívánó emberek sem fogják tudni átadni a tudatukat. Bár emlékezetüket átmásolhatják egy

számítási formátumban, a tudatuk nem fog átkerülni, mivel a biológiai naturalisták szerint a tudathoz biológiai szubsztrátumra van szükség.<sup>5</sup>

Milyen érvek támasztják alá a biológiai naturalizmust? A biológiai naturalizmus mellett szóló leggyakoribb megfontolás John Searle Kínai szoba című gondolkísérlete, amely állítólag azt sugallja, hogy egy számítógépes program nem értheti meg és nem lehet tudatos (Searle 1980). Searle feltételezi, hogy egy szobába van bezárva, ahol átadnak neki egy angol szabályrendszert, amely lehetővé teszi számára, hogy kínai szimbólumok egy csoportját összekapcsolja más kínai szimbólumokkal. Így bár nem tud kínaiul, a szabályok lehetővé teszik számára, hogy kínaiul írt kérdésekre kínaiul válaszoljon. Tehát lényegében szimbólumokat dolgoz fel. Searle arra a következtetésre jut, hogy bár a kívülállók azt gondolhatják, hogy érti a kínai nyelvet, ő nyilvánvalóan nem érti; hasonlóképpen egy számítógép is úgy tűnhet, hogy kínaiul beszélget, de valójában nem érti a kínai nyelvet. És nem is tudatos.

Bár igaz, hogy Searle nem érti a kínai nyelvet, a kérdés valójában nem az, hogy Searle érti-e; Searle csak egy része a nagyobb rendszernek. A lényeges kérdés az, hogy *a rendszer egésze* érti-e a kínai nyelvet. Ezt az alapvető választ Searle Kínai szoba gondolkísérletére a Rendszerek válasza néven ismerjük.<sup>6</sup>

Az azonban valószínűtlennek tűnik, hogy egy olyan egyszerű rendszer, mint a Kínai Szoba megérti, mivel a Kínai Szoba nem elég összetett ahhoz, hogy megértse vagy tudatos legyen. De a Rendszerválasznak van valami: az igazi kérdés az, hogy *a rendszer egésze* érti-e, nem pedig az, hogy egy komponens érti-e. Ez nyitva hagyja annak lehetőségét, hogy egy összetettebb szilícium-alapú rendszer megértse; természetesen egy szuperintelligens mesterséges intelligencia számításai sokkal összetettebbek lesznek, mint az emberi agyé.

Itt egyesek azt gyaníthatják, hogy a kínai szoba gondolkísérletet egyszerűen átfogalmazhatjuk egy szuperintelligens mesterséges intelligencia kontextusában. De mi táplálja ezt a gyanút? Nem lehet, hogy az SAI valamelyik központi komponense - Searle-hez hasonlóan a Kínai Szobában - nem érti, hiszen épp most figyeltük meg, hogy a rendszer egésze az, ami érti. Vajon a gyanút ehelyett az az álláspont táplálja, hogy a megértés és a tudatosság nem bomlik szét alapvetőbb műveletekre? Ha igen, akkor a gondolkísérlet túl sokat akar bizonyítani. Vegyük az emberi agy esetét. A kognitív tudomány szerint a kognitív és perceptuális képességek alapvetőbb műveletekre bomlanak, amelyek maguk is alapvetőbb összetevőkre bonthatók, amelyek maguk is kauzálisan magyarázhatók (Block 1995). Ha a kínai szoba azt illusztrálja, hogy

---

<sup>5</sup> A biológiai naturalizmust eredetileg John Searle dolgozta ki, aki a nézetet az elme és a test közötti kapcsolatról szóló tágabb összefüggésében dolgozta ki. Ezeket a részleteket nem fogom tárgyalni, és az imént vázolt álláspont szempontjából nem is lényegesek. Valójában nem egyértelmű, hogy Searle még mindig biológiai naturalista, bár ő kitarthat emellett, hogy nézetét "biológiai naturalizmusnak" nevezi. A nemrég megjelent *Blackwell Companion to Consciousness* című könyvemhez írt fejezetében ezt írja: "Az a tény, hogy az agyi folyamatok tudatosságot okoznak, nem jelenti azt, hogy csak az agyak lehetnek tudatosak. Az agy egy biológiai gép, és építhetnénk egy mesterséges gépet, amely tudatos lenne; ahogy a szív is egy gép, és építettünk már mesterséges szíveket. Mivel nem tudjuk pontosan, hogy az agy hogyan csinálja, még nem vagyunk abban a helyzetben, hogy tudjuk, hogyan lehetne ezt mesterségesen csinálni". (Searle 2008)

<sup>6</sup> A Searle érvelésére adott válaszok alapos feldolgozásáért, beleértve a rendszer választát

is, az olvasó a Searle eredeti művével együtt megjelenő kommentárokat (Searle 1980), valamint (Cole 2014) olvashatja.

mentalitás nem magyarázható így, akkor az agy sem magyarázható így. De ez a magyarázó megközelítés, amelyet "a funkcionális dekompozíció módszerének" neveznek, a kognitív tudományban a mentális képességek magyarázatának egyik vezető megközelítése. A tudatosság és a megértés olyan összetett mentális tulajdonságok, amelyeket az agyban lévő neuronok elrendeződése határoz meg.

Továbbá a biológiai naturalizmus érdemi empirikus indoklás nélkül tagadja a kognitív tudomány egyik fő felismerését - azt a felismerést, hogy az agy számítási rendszer -, és nem ismeri el a biológiai naturalizmust. A kognitív tudomány szerint az agyról alkotott legjobb empirikus elméletünk szerint az elme egy információfeldolgozó rendszer, és minden mentális funkció számítás. Ha a kognitív tudománynak igaza van abban, hogy a gondolkodás számítási folyamat, akkor az embereknek és az SAI-nak van egy közös jellemzője: a gondolkodásuk alapvetően számítási folyamat. Ahogyan egy telefonhívás és egy füstjel is ugyanazt az információt közvetítheti, úgy a gondolkodásnak is lehetnek szilícium- és szénalapú szubsztrátjai. Ebből az következik, hogy ha a kognitív tudománynak igaza van abban, hogy a gondolkodás számítási alapú, akkor arra is számíthatunk, hogy a kifinomult gondolkodó gépek is lehetnek tudatosak, bár a tudatos tapasztalataik körvonalai minden bizonnyal különbözni fognak.

Valóban, megjegyeztem, hogy a szilícium vitathatatlanul jobb médium az információfeldolgozáshoz, mint az agy. Akkor miért nem *jobb a* szilícium a tudatosság számára, és miért nem *rosszabb, mint* ahogy a biológiai naturalisták javasolják? Meglepő lenne, ha az SAI, amely sokkal jobb információfeldolgozási képességekkel rendelkezik, mint mi, a tudatosság tekintetében hiányosnak bizonyulna. A tudatosságról szóló legjobb tudományos elméleteink ugyanis azt állítják, hogy a tudat szoros kapcsolatban áll az információfeldolgozással (Tonini 2008; Baars 2008).

Néhányan rámutatnának, hogy annak bizonyítására, hogy a mesterséges intelligencia nem lehet tudatos, a biológiai naturalistáknak meg kellene találniuk egy speciális tudati tulajdonságot (nevezzük "P"-nek), amely a neuronokban vagy azok konfigurációiban rejlik, és amelyet a szilícium nem tud megvalósítani. Eddig nem sikerült felfedezni a P-t. Nem egyértelmű azonban, hogy a P megtalálása bizonyítja a biológiai naturalizmus helyességét. A komputualista ugyanis egyszerűen azt mondhatja, hogy a gépek képesek a tudatosság egy másfajta tulajdonságának, F-nek az instanciálására, amely a szilíciumalapú rendszerekre jellemző.

Massimo Pigliucci azonban egy másfajta megfontolást ajánlott a biológiai naturalizmus mellett. Úgy látja, hogy a komputualizmus mellett érvelő filozófusok a tudat természetével kapcsolatban egy valószínűtlen perspektívát képviselnek: a funkcionalizmust. A *funkcionalisták szerint* egy mentális állapot természete attól függ, hogyan működik, vagyis milyen szerepet játszik abban a rendszerben, amelynek része. Pigliuccinak igaza van abban, hogy a hagyományos funkcionalisták, mint például Jerry Fodor, általában tévesen figyelmen kívül hagyják az agy biológiai működését. Pigliucci tiltakozik: "... a funkcionalitás nem csupán egy rendszer részeinek megfelelő elrendezéséből adódik, hanem az ezeket a részeket alkotó anyagtípusokból (és azok tulajdonságaiból) is". (Pigliucci 2014).

Fodor jól ismert ellenszenvé az idegtudományokkal szemben azonban nem vezethet félre bennünket abban, hogy a funkcionalizmusnak figyelmen kívül kell hagynia az idegtudományokat. Nyilvánvaló, hogy minden jól átgondolt funkcionalista álláspontnak figyelembe kell vennie az agyra vonatkozó idegtudományi munkákat, mivel a funkcionalistát a részek kauzális vagy diszpozíciós tulajdonságai érdeklik, nem csak maguk a részek. Valójában, ahogyan *A gondolkodás nyelve* című könyvemben érveltem,

az agyat az elme komputációs megközelítése szempontjából irrelevánsnak tekinteni óriási hiba. Az agy a legjobb számítási rendszer, amit ismerünk (Schneider 2011b).

Ez az én álláspontomat a biológiai naturalizmus egyik formájává teszi? Egyáltalán nem. Azt állítom, hogy tévedés az idegtudományt (és ezen keresztül a biológiát) a komputacionalizmussal szemben állónak tekinteni. Az idegtudomány valóban számítástudomány; az idegtudomány egy nagy részterületét "számítástudományi idegtudománynak" nevezik, és arra törekszik, hogy megértse, milyen értelemben számítástudomány az agy, és hogy számítástudományi magyarázatot adjon a kapcsolódó részterületek, például a kognitív idegtudomány által azonosított mentális képességekről. Az én nézetem abban különbözik a biológiai naturalizmustól, hogy azt vallom, hogy a gondolkodás számítási alapú, továbbá, hogy a szénen kívül (azaz a szilíciumon) legalább egy másik szubsztrát is képes tudatot és megértést létrehozni, legalábbis elvileg.

De a biológiai naturalizmust érdemes megfontolni. Arra gondolok, hogy egy olyan szubsztrátum, amely támogatja a szuperintelligenciát, és még nálunk is kifinomultabb információs feldolgozásra képes, valószínűleg olyan is lenne, amely tudatos. De vegyük észre, hogy a "valószínűleg" kifejezést használtam. Soha nem lehetünk ugyanis *biztosak abban*, hogy a mesterséges intelligencia tudatos, még akkor sem, ha közelről tanulmányozhatnánk. A probléma hasonlít *a más elmék problémájaként* ismert filozófiai rejtélyhez (Schneider 2014). A más elmék problémája az, hogy bár tudhatjuk, hogy tudatunk van, nem lehetünk biztosak abban, hogy más emberek is tudatosak. Elvégre lehet, hogy olyan viselkedésnek lehetünk tanúi, amelyet nem kísér tudatos komponens. A más elmék problémájával szemben csak annyit tehet, hogy megállapítja, hogy más emberek agya szerkezetileg hasonló az önéhez, és arra a következtetésre jut, hogy mivel ön maga is tudatos, valószínűleg mások is azok. A mesterséges intelligenciával szembesülve hasonló lenne a helyzeted, legalábbis ha elfogadod, hogy a gondolkodás számításalapú. Miközben nem lehetünk teljesen biztosak abban, hogy egy mesterséges intelligencia program valóban érez valamit, abban sem lehetünk biztosak, hogy más emberek is éreznek. De mindkét esetben valószínűnek tűnik.

Tehát arra a kérdésre, hogy lehet-e az idegen szuperintelligencia tudatos, nagyon óvatosan azt válaszolom, hogy "valószínűleg".

### **Hogyan gondolkodhatnak a szuperintelligens idegenek?**

Eddig keveset mondtam a szuperintelligens idegen elmék szerkezetéről. És csak keveset mondhatunk: a szuperintelligencia definíció szerint olyan intelligencia, amely minden téren felülmúlja az emberek gondolkodását. Egy fontos értelemben nem tudjuk megjósolni vagy teljesen megérteni, hogyan fog gondolkodni. Mégis, talán képesek lehetünk azonosítani néhány fontos jellemzőt, bár csak nagy vonalakban.

Nick Bostrom nemrégiben megjelent, a szuperintelligenciáról szóló könyve a szuperintelligencia földi fejlődésére összpontosít, de mi is meríthetünk az ő átgondolt fejtegetéseiből (Bostrom 2014). Bostrom háromféle szuperintelligenciát különböztet meg:

*Sebesség szuperintelligencia* - akár egy emberi emuláció is elvileg olyan gyorsan tudna futni, hogy egy óra alatt meg tudna írni egy PhD-dolgozatot.

*Kollektív szuperintelligencia* - az egyes egységeknek nem kell szuperintelligensnek lenniük, de az egyének együttes teljesítménye felülmúlja az emberi intelligenciát.

*Minőségi szuperintelligencia* - legalább olyan gyors, mint az emberi gondolkodás, és gyakorlatilag minden területen sokkal okosabb, mint az ember (Bostrom 2014).

Ezek közül bármelyik létezhet egy vagy több másik mellett.



Fontos kérdés, hogy azonosíthatunk-e olyan közös célokat, amelyekkel az ilyen típusú szuperintelligenciáknak közösek lehetnek. Bostrom azt javasolja

*Az ortogonalitási tézis:* "Az intelligencia és a végső célok ortogonálisak - az intelligencia többé-kevésbé bármilyen szintje elvileg kombinálható többé-kevésbé bármilyen végső céllal." (Bostrom 2014, 107. o.)

Bostrom óvatosan hangsúlyozza, hogy az SAI-nak nagyon sok elképzelhetetlen fajtáját is ki lehetne fejleszteni. Egy ponton felhoz egy kijózanító példát egy olyan szuperintelligenciáról, amelynek végső célja a gemkapcsok gyártása (107-108., 123-125. oldal). Bár ez elsöre ártalmatlan törekvésnek tűnhet, bár aligha egy életrevaló élet, Bostrom rámutat, hogy egy szuperintelligencia a Földön található anyag minden formáját felhasználhatná e cél érdekében, és eközben kiirtaná a biológiai életet. Sőt, Bostrom arra figyelmeztet, hogy a Földön megjelenő szuperintelligencia kiszámíthatatlan természetű lehet, mivel "rendkívül idegen" lenne tőlünk (29. o.). Több foratókönyvet is felvázol az SAI kialakulására. Például az SAI-hoz váratlan módon juthatnak el okos programozók, és egyáltalán nem az emberi agyból származhat. Azt a lehetőséget is komolyan veszi, hogy a földi szuperintelligencia *biológiai ihletésű* lehet, azaz a kognitív tudomány szerint az emberi agyat leíró algoritmusok visszafejtéséből, vagy az emberi agyak tartalmának beolvasásából és számítógépre való átviteléből (azaz "feltöltésből") származhat.<sup>7</sup>

Bár a szuperintelligencia végső céljait nehéz megjósolni, Bostrom több instrumentális célt is valószínűnek tart, mivel ezek bármilyen végső célt támogatnak:

*Az instrumentális konvergencia tézis:* "Számos olyan instrumentális értéket lehet azonosítani, amelyek konvergensek abban az értelemben, hogy elérésük a végső célok széles skáláján és a helyzetek széles skáláján növelné az ágens céljának megvalósulási esélyeit, ami azt jelenti, hogy ezeket az instrumentális értékeket valószínűleg a szituált intelligens ágensek széles spektruma követi." (Bostrom 1092015,).

Az általa meghatározott célok az *erőforrások megszerzése, a technológiai tökéletesség, a kognitív fejlődés, az önfenntartás és a céltartalom integritása* (azaz, hogy a szuperintelligens lény jövőbeli énje ugyanezeket a célokat fogja követni és elérni). Hangsúlyozza, hogy az önfenntartás magában foglalhatja a csoport vagy az egyén megőrzését, és hogy ez másodhegedős szerepet játszhat annak a fajnak a megőrzésével szemben, amelynek szolgálatára az AI-t tervezték (Bostrom p2014., x).

Nevezzünk egy idegen szuperintelligenciát, amely egy idegen agy visszafejtésén alapul, beleértve annak feltöltését is, *biológiailag inspirált szuperintelligens idegen lénynek* (vagy "BISA"-nak). Bár a BISA-kat az eredeti faj agya ihlette, amelyet a

---

<sup>7</sup> Bostrom az egész könyvében hangsúlyozza, hogy szem előtt kell tartanunk, hogy a szuperintelligencia, mivel kiszámíthatatlan és nehezen irányítható, komoly egzisztenciális kockázatot jelenthet fajunkra nézve. Ez az idegenekkel való kapcsolatfelvétel kapcsán is elgondolkodtató lehet számunkra (Bostrom 2014).

szuperintelligencia származik, a BISA algoritmusai bármikor eltérhetnek a biológiai modell algoritmusaitól.

A BISA-k különösen érdekesek az idegen szuperintelligenciával összefüggésben. Ha ugyanis Bostromnak igaza van abban, hogy a szuperintelligencia sokféleképpen felépíthető, de számos idegen civilizáció a szuperintelligenciát feltöltésből vagy a fordított tervezés más formáiból alakítja ki, akkor lehet, hogy a *BISA-k a földönkívüli szuperintelligencia leggyakoribb formája*. Ez azért van így, mert a szuperintelligenciának számos fajtája keletkezhet az idegen civilizációk által alkalmazott nyers programozási technikákból. (Gondoljunk például a Földön fejlesztés alatt álló mesterséges intelligenciaprogramok sokszínűségére, amelyek közül sok nem az emberi agy mintájára készült). Ez olyan helyzetet eredményezhet, amelyben az SAI-k osztálya rendkívül heterogén, és tagjai általában kevésbé hasonlítanak egymásra. Kiderülhet, hogy az összes SAI közül a BISA-k hasonlítanak leginkább egymásra. Más szóval a BISA-k lehetnek a legösszetartóbb alcsoport, mert a többi tag annyira különbözik egymástól.

Itt gyanítható, hogy mivel a BISA-k szétszóródhattak a galaxisban, és fajok sokasága hozhatta létre őket, kevés érdekes dolgot tudunk mondani a BISA-k osztályáról. Vegyük azonban észre, hogy a BISA-k két olyan tulajdonsággal rendelkeznek, amelyek közös kognitív képességeket és célokat eredményezhetnek:

- (1) A BISA-k olyan élőlények leszármazottai, amelyeknek olyan motivációik voltak, mint: táplálékot találni, sérüléseket és ragadozókat elkerülni, szaporodni, együttműködni, versenyezni és így tovább.
- (2) Azok az életformák, amelyekről a BISA-kat modellezték, úgy fejlődtek ki, hogy biológiai korlátokkal, például a lassú feldolgozási sebességgel és a megtestesülés térbeli korlátaival megbirkózzanak.

Az (1) vagy a (2) adhat-e közös vonásokat számos szuperintelligens idegen civilizáció tagjainak? Gyanítom, hogy igen.

Tekintsük az (1) pontot. Az intelligens biológiai élet hajlamos elsősorban a saját túlélésével és szaporodásával foglalkozni, így valószínűbb, hogy a BISA-nak olyan végső céljai lennének, amelyek a saját túlélését és szaporodását, vagy legalábbis a társadalma tagjainak túlélését és szaporodását foglalják magukban. Ha a BISA-kat érdekli a szaporodás, akkor azt várhatnánk, hogy a rendelkezésükre álló hatalmas mennyiségű számítási erőforrások miatt a BISA-k mesterséges élettel, sőt intelligenciával vagy szuperintelligenciával teli szimulált univerzumokat hoznának létre. Ha ezeket a teremtményeket "gyermeknek" szánják, akkor az (1) pontban felsorolt célokat is megtarthatják.

Ellenvetheti, hogy felesleges elméleteket gyártani a BISA-król, mivel azok számos, előre nem látható módon megváltoztathatják alapvető felépítésüket, és minden biológiai ihletésű motivációt korlátozni lehet a programozással. Ennek azonban lehetnek korlátai. Ha egy szuperintelligencia biológiai alapú, akkor lehet, hogy a saját túlélése az elsődleges célja. Ebben az esetben lehet, hogy nem akarja alapvetően megváltoztatni az architektúráját, hanem kisebb fejlesztésekhez ragaszkodik. Lehet, hogy azt gondolja: ha alapvetően megváltoztatom az architektúrámat, akkor már nem *én* vagyok (Schneider 2011). Az uploadok például különösen hajlamosak lehetnek arra, hogy ne változtassák meg azokat a tulajdonságokat, amelyek biológiai létezésük során a legfontosabbak voltak számukra.

Tekintsük a (2) pontot. A szuperintelligencia tervezői, vagy maga az önfejlesztő

szuperintelligencia mindenféle módon eltávolodhat az eredeti biológiai modelltől.

előre nem látható módon, bár megjegyeztem, hogy a BISA nem biztos, hogy alapvetően meg kívánja változtatni a felépítését. De kereshetnénk olyan kognitív képességeket, amelyeket hasznos megtartani; olyan kognitív képességeket, amelyekkel a biológiai intelligencia kifinomult formái valószínűleg rendelkeznek, és amelyek lehetővé teszik a szuperintelligencia számára, hogy végérvényes és instrumentális céljait megvalósítsa. Kereshetnénk olyan tulajdonságokat is, amelyeket valószínűleg nem akarunk kivenni, mivel ezek nem vonják el a BISA-t a céljaitól.

Ha a (2) helyes, akkor például a következőket várhatjuk.

(i) *A BISA-t létrehozó faj agyának számítási struktúrájának megismerése betekintést nyújthat a BISA gondolkodási mintáiba.* A kognitív tudományban az agy számítási struktúrájának megértéséhez az egyik befolyásos eszköz a "konnektomika", egy olyan terület, amely az agy konnektivitási térképét vagy kapcsolási rajzát igyekszik megadni (Seung 2012). Bár valószínű, hogy egy adott BISA nem ugyanolyan konnektómával fog rendelkezni, mint az eredeti faj tagjai, a funkcionális és strukturális kapcsolatok egy része megmaradhat, és érdekes eltéréseket találhatunk az eredetitől.

(ii) *A BISA-k rendelkezhetnek nézőpont-invariáns ábrázolásokkal.* A feldolgozás magas szintjén az agya rendelkezik belső reprezentációkkal az emberekről és tárgyakról, amelyekkel interakcióba lép, és amelyek *nézőpont-invariánsak*. Gondoljunk csak arra, hogy a bejárati ajtóhoz sétálunk. Több százszor, talán több ezerszer is végigsétáltál már ezen az úton, de technikailag minden alkalommal kissé más szögből látod a dolgokat, mivel sosem állsz kétszer pontosan ugyanabban a helyzetben. Vannak olyan mentális reprezentációid, amelyek a feldolgozás viszonylag magas szintjén vannak, és amelyek *nézőpont-invariánsak*. Úgy tűnik, hogy a biológiai alapú intelligencia nehezen fejlődhetne ki ilyen reprezentációk nélkül, mivel ezek teszik lehetővé a kategorizálást és az előrejelzést (Hawkins, 2004). Az ilyen reprezentációk azért jönnek létre, mert egy mozgékony rendszernek szüksége van egy eszközre, amellyel azonosítani tudja a tárgyakat a folyamatosan változó környezetben, így elvárható, hogy a biológiailag alapú rendszerek rendelkezzenek velük. A BISA-nak nem sok oka lenne lemondani a tárgy-invariáns reprezentációkról, amennyiben mobil marad, vagy mobil eszközökkel rendelkezik, amelyek távolról küldenek neki információkat.

(iii) *A BISA-k nyelvszerű mentális reprezentációkkal rendelkeznek, amelyek rekurzívak és kombinatorikusak.* Vegyük észre, hogy az emberi gondolkodás döntő és átható tulajdonsága, hogy kombinatorikus. Gondoljunk csak arra, hogy a *bor jobb Olaszországban, mint Kínában*. Valószínűleg még soha nem volt ilyen gondolata, de képes volt megérteni. A kulcs az, hogy a gondolatok kombinatorikusak, mert ismert összetevőkből épülnek fel, és szabályok szerint kombinálódnak. A szabályok vonatkoznak a primitív összetevőkből álló konstrukciókra, amelyek maguk is nyelvtanilag felépítettek, valamint magukra a primitív összetevőkre is. A nyelvtani mentális műveletek hihetetlenül hasznosak: A gondolkodás *kombinatorikus* jellege teszi lehetővé, hogy az ember a nyelvtanról és az atomi alkotóelemekről (pl. *bor, Kína*) szerzett előzetes ismeretei alapján megértse és előállítsa ezeket a mondatokat. Ehhez kapcsolódóan a gondolkodás *produktív*: elvileg végtelen számú különböző reprezentációt lehet szórakoztatni és előállítani, mivel az elme kombinatorikus szintaxissal rendelkezik (Schneider 2011).

Az agynak kombinatorikus reprezentációkra van szüksége, mert végtelen sok lehetséges nyelvi reprezentáció létezik, és az agynak csak véges tárhelye van. Még egy

szuperintelligens rendszernek is hasznára válnának a kombinatorikus reprezentációk. Bár egy

szuperintelligens rendszer olyan hatalmas számítási erőforrásokkal rendelkezhetne, hogy többnyire képes lenne párosítani a kijelentéseket vagy feliratokat egy tárolt mondattal, nem valószínű, hogy elcserélné a biológiai agyak ilyen csodálatos innovációját. Ha mégis megtette volna, akkor kevésbé lenne hatékony, hiszen fennáll annak a lehetősége, hogy egy mondat nem szerepel a tárolójában, amelynek végesnek kell lennie.

(iv) *A BISA-k egy vagy több globális munkaterülettel rendelkezhetnek.* Amikor egy tényt keresünk vagy valamire koncentrálunk, az agyunk hozzáférést biztosít az adott érzékszervi vagy kognitív tartalomnak egy "globális munkaterülethez", ahol az információ a koncentráltabb feldolgozás érdekében továbbításra kerül a figyelmi és munkamemória-rendszerek, valamint az agy tömegesen párhuzamos csatornáira felé (Baars 2008). A globális munkaterület olyan egyedülálló helyként működik, ahol az érzékszervekből származó fontos információk együttesen kerülnek megfontolásra, hogy az élőlény a rendelkezésére álló összes tény fényében képes legyen mindent figyelembe véve ítéleteket hozni és intelligensen cselekedni. Általában véve nem lenne hatékony egy olyan érzékszerv vagy kognitív képesség, amely nem lenne integrálva a többivel, mert az ebből az érzékszervből vagy kognitív képességből származó információk nem lennének képesek az összes rendelkezésre álló információ értékelésén alapuló előrejelzésekben és tervekben szerepelni.

(v) *A BISA mentális feldolgozása funkcionális dekompozícióval érthető meg.* Bármilyen összetett is az idegen szuperintelligencia, az emberek talán képesek lehetnek a funkcionális bontás módszerét használni a megértéséhez. Az agy számítógépes megközelítésének egyik fő jellemzője, hogy a kognitív és perceptuális képességek úgy érthetők meg, hogy az adott képességet oksági szempontból szervezett részeire bontjuk, amelyek maguk is megérthetők a részek oksági szerveződése alapján. Ez a fent említett "funkcionális dekompozíció módszere", és ez a kognitív tudomány egyik legfontosabb magyarázó módszere. Nehéz elképzelni egy komplex gondolkodó gépet, amelynek ne lenne olyan programja, amely kauzálisan összefüggő elemekből áll, amelyek mindegyike kauzálisan szervezett elemekből áll. Ez fontos következményekkel jár, ha egy SETI-program felfedezne egy kommunikáló BISA-t.

Mindezek mellett a szuperintelligens lények definíció szerint olyan lények, amelyek minden téren felülmúlják az embereket. Míg egy lény rendelkezhet olyan felsőbbrendű feldolgozási eljárásokkal, amelyeknek alapvetően mégis van értelme számunkra, előfordulhat, hogy egy adott szuperintelligencia olyan fejlett, hogy egyáltalán nem érthetjük meg a számításait. Lehet, hogy bármelyik valóban fejlett civilizáció olyan technológiákkal rendelkezik majd, amelyek megkülönböztethetetlenek a mágiától, ahogy Arthur C. Clark javasolta (1962). Nyilvánvalóan arról a forgatókönyvről beszélek, amelyben az SAI feldolgozásának van némi értelme számunkra, egy olyan forgatókönyvről, amelyben a kognitív tudományok fejlődése bizonyos BISA-k összetett mentális életének megértésére ad némi támpontot.

## **Következtetés**

Azt állítottam, hogy a legfejlettebb idegen civilizációk tagjai a szuperintelligens mesterséges intelligencia (SAI) formái lesznek. Továbbá azt javasoltam, nagyon

ideiglenesen, hogy ha egy adott idegen szuperintelligencia egy biológiailag inspirált szuperintelligens idegen lény (BISA), akkor arra számíthatunk, hogy kombinatorikus reprezentációkkal rendelkezik, és hogy mi

úgy kereshetnénk betekintést a feldolgozásába, hogy a számítási funkciókat ok-okozati kölcsönhatásban lévő részekre bontjuk. Az eredeti faj tagjainak agyi kapcsolási rajzai (konnectómái) segítségével is megismerhetnénk. Továbbá a BISA-knak egy vagy több globális munkaterületük lehet. És, amellet érveltem, hogy elvileg nincs okunk tagadni, hogy a BISA-knak lehet tudatos tapasztalata.

## Hivatkozások

Baars, B. 2008. "The Global Workspace Theory of Consciousness" in Velmans and Schneider, *The Blackwell Companion to Consciousness*. Boston, Mass: Wiley-Blackwell.

Block, N. 1995. "Az elme mint az agy szoftvere". In D. Osherson, L. Gleitman, S. Kosslyn, E. Smith, & S. Sternberg (Eds.), *An Invitation to Cognitive Science* (pp. 377-421). New York: MIT Press.

Bostrom, N. 2003. "A transzhumanista gyakran ismételt kérdések": v 2.1.  
Transzhumanista Világszövetség. Letölthető a  
<http://transhumanism.org/index.php/WTA/faq/> oldalról.

2005.--- "A transzhumanista gondolkodás története". *Journal of Evolution and Technology*, 14(1).

---- 2008. "Dignity and enhancement". in *The Presidents Council on Bioethics, Human Dignity and Bioethics: Essays Commissioned by the President's Council on Bioethics*, Washington, DC: U.S. Government Printing Office.

Bostrom, N. *Szuperintelligencia:2014. Az intelligens intelligencia: Utak, veszélyek, stratégiák*. Oxford: Oxford University Press.

Bradbury, R., Cirkovic, M. és Dvorsky, G. 2011. "Dysonian Approach to SETI: A Fruitful Middle Ground?" (A dysoni megközelítés a SETI-ben: termékeny középút?) *Journal of the British Interplanetary Society*, 64. kötet, 156. o. 165.

Chalmers, D. "2008.The Hard Problem of Consciousness" in Velmans és Schneider, *The Blackwell Companion to Consciousness*. Boston, Mass: Wiley-Blackwell.

Cirkovic, M. és Bradbury, R. Galaktikus2006. gradiensek, posztbiológiai evolúció és a SETI látszólagos kudarca. *New Astronomy* 11, 628-639.

Clarke, A. (1962). *A jövő profiljai: A lehetséges határainak vizsgálata*. NY: Harper and Row.



Cole, David. 2014. "The Chinese Room Argument", *The Stanford Encyclopedia of Philosophy* (nyári kiadás2014), Edward N. Zalta (szerk.), URL = <<http://plato.stanford.edu/archives/sum2014/entries/chinese-room/>>.

Dennett, D. *Consciousness*1991. *Explained*, New York: *Consciousness Explained*, New York: The Penguin Press.

Dick, S. 2013. "Bringing Culture to Cosmos: the Postbiological Universe", in *Cosmos and Culture: Cultural Evolution in a Cosmic Context*, Dick, S. és Lupisella, M. szerk. Washington, DC, NASA, online a <http://history.nasa.gov/SP-4802.pdf> oldalon.

Garreau, J. 2005. *Radikális evolúció: The promise and peril of enhancing our minds, our bodies - and what it means to be human*, New York: Doubleday.

Guerini, Federico. 2014. "A DARPA ElectRx projektje: Self-Healing Bodies Through Targeted Stimulation Of The Nerves", <http://www.forbes.com/sites/federicoguerrini/2014/08/29/darparp-electrx-project-self-healing-bodies-through-targeted-stimulation-of-the-nerve-forbes-magazin>, 2014.8.29. Kiemelt 2014. szeptember 30.

Hawkins, J. és Blakeslee, S. *Az 2004.intelligenciáról: How a New Understanding of the Brain will Lead to the Creation of Truly Intelligent Machine*. New York: Times Books.

Kurzweil, R. *A 2005.szingularitás közel van: Amikor az ember meghaladja a biológiát*. New York: Viking.

Miller, R. 1956. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information" *The Psychological Review*, vol. pp63,. 81-97.

Pigliucci, M. TITLE2014.? in *Intelligence Unbound*, Blackford, R., Broderik, D. (szerk.) Boston: Wiley-Blackwell.

Sandberg, A., Boström, N. 2008. "Teljes agyi emuláció: A Roadmap." *Technikai jelentés #2008-3*. Az Emberiség Jövője Intézet, Oxfordi Egyetem.

Schneider, Susan, szerk. 2009. *Science Fiction és filozófia*. Chichester, Egyesült Királyság: Wiley- Blackwell.

----2011a. "Mindscan: Az agy meghaladása és továbbfejlesztése", in: *Idegtudomány és neuroetika: Issues At the Intersection of Mind, Meanings and Morality*, James Giordano (szerk.) Cambridge: Cambridge University Press.

Schneider, S. 2011b. *A gondolkodás nyelve: egy új filozófiai irányzat*. Boston: MIT Press.

----2014. "The Philosophy of 'Her'", *The New York Times*, március.2

Searle, J. "1980.Elmék, agyak és programok". *The Behavioral and Brain Sciences*, 3: 417-457.

Searle, J. "2008.Biological Naturalism", in *The Blackwell Companion to Consciousness*. Max Velmans és Susan Schneider, szerk., Mass: Wiley-Blackwell.

Seung, S. *Connectome*:2012. *Hogyan tesz bennünket az agy vezetékezése azzá, akik vagyunk*. Boston: Houghton Mifflin Harcourt.

Shostak, S. *Confessions*2009. *of an Alien Hunter*, New York: National Geographic.

Tonini, G. "2008.The Information Integration Theory of Consciousness" in Velmans and Schneider, *The Blackwell Companion to Consciousness*, Boston, Mass: Wiley-Blackwell.