

## Mesterséges gondolkodás és emberi értelem

Mesterséges intelligencia (MI) – már pusztán a kifejezés több értelmet rejthet magában: jelentheti az értelem utánzatának, valamiféle műintelligenciának az elkészítését, de mesterségesen létrehozott valódi – az emberéhez hasonló vagy vele azonos – intelligenciát is. A számítógépes megvalósítás tekintetében a gondolkodás gépi szimulációját vagy egyenesen a gép gondolkodását. Az ilyen irányú kutatások egyik kezdeményezője és főszereplője, Alan Turing nem tartja fontosnak ezt a különbséget, hanem éppen az utóbbi, „túlságosan semmitmondó” kérdés („Tudnak-e a gépek gondolkodni?”) helyett javasolja a róla elnevezett híres próbát. Jóslata szerint a (múlt) évszázad végére lehetnek az „imitációs játékok” olyan jól játszó számítógépek, hogy az átlagos kérdező elég nagy eséllyel nem tudja helyesen megállapítani, emberrel vagy géppel van-e dolga. Akkorra az emberek gondolkodása és a szavak mindennapi használata annyira megváltozik, hogy bátran beszélhetünk majd gépi gondolkodásról, míg a cikk írásának idején (1950) ez még „veszedelmes volna” (Turing 1965. 120, 134). A saját maga által a próbával szemben felhozott, de nyilvánvalóan a (sokak számára nagyon is baljóslatúan sokatmondó) kérdés által motivált ellenvetések közül talán a következő a két leggyakoribb: (1) a gondolkodás szavatolja az ember *felsőbbrendűségét* a többi teremtménnyel (állatokkal, gépekkel) szemben; (2) a gépnek nincs *ön-tudata*, maga nem érzi, hogy gondolkodik, nem lehet saját működésének tárgya. Ezekre próbál felelni a maga módján a jelen írás is.

Hasonló „filozófiai pragmatizmus” mellett foglal állást a terület egy másik úttörője, Herbert A. Simon, mondván, hogy nem győzték meg azok az érvek, amelyek az emberhez viszonyítva a számítógépek képességeinek ilyen-olyan korlátait igyekeznek bizonyítani, de hajlandó felfüggeszteni az ítéletét, amíg ez utóbbiakat teljesen meg nem ismerjük. Ez vonatkozik arra, hogy van-e a gépeknek tudata vagy érzelme. Ugyanakkor „technológiai dimenzió” szerint határozottan radikálisnak vallja magát: „a számítógépek el tudnak majd látni bármilyenű olyan feladatot, mint az ember. Véleményem szerint a számítógépek már most [1977] képesek olvasni, gondolkodni, tanulni és alkotni”. (Simon 1982. 17–18.) Abban a kis taxonómiában helyezi el így magát, amely az egymással

szemben álló, különféle tekintetben radikális és konzervatív („bal- és jobboldali”) nézeteket vázolja fel, mielőtt a számítógépek hatásának elemzéséhez kezd – ahhoz a „sötét foltéhoz”, amely „saját reményeink és aggodalmaink tükre” (uo. 13–15).

Ilyen aggodalmakat, illetve reményeket fogalmaz meg például Roger Penrose és Daniel C. Dennett, mégpedig nem pusztán a szimulációra, hanem kifejezetten arra a kérdésre vonatkoztatva, hogy vajon a számítógép működése eredményeképp ténylegesen *létrejön-e* a gondolat. Penrose fő törekvése cáfolni vagy legalább egyelőre megkérdőjelezni azt az állítást, hogy egy algoritmus végrehajtása előidézheti, kiválthatja a tudatosságot (Penrose 1993. 33, 435, 475–476). Érzés, öntudat, ítélőképesség, (matematikai) belátás, (művészi) ihlet tartozik ehhez a „szellemi minőséghez”, melynek meggyőződése szerint tartalmaznia kell valami nem-algoritmikus, és amelyhez a kvantumgravitáció majdani elmélete fog magyarázatul szolgálni (uo. 32–33, 438–441, 444–453). Ugyanezeket Dennett az evolúcióból származtatván, minthogy a természetes szelekciót algoritmikus folyamatként értelmezi, kisebb, „félíg megtervezett, önújratervező” szubrutinok eredményének tartja, elménket pedig nagyméretű komplex gépnek. Képességeinek „végső soron ’mechanikus’ magyarázattal kell rendelkezniük. A makróktól származunk és makrókból állunk” (Dennett 1998. 50–64, 222, 398–399, 473–476). Ennek megfelelően legközelebbi szövetségésének véli azt, ami ellen Penrose ágál: „a mesterséges intelligencia területe a darwini gondolat közvetlen leszármazottja” és „szilárdan Darwinnal tart” (uo. 28. 218. l.). Mindenesetre mindkét szerző számára a tudatosság magyarázatát nyújtja vagy nyújthatná az, ami *előidézi* azt: Penrose számára egy (még nem létező) fizikai elmélet, Dennett-nél az evolúció (szerinte sokszorosán igazolt) algoritmus.

A következőkben először speciális problémamegoldó – nevezetesen sakkozó – programok esetében járok utána annak a kérdésnek, hogy mi is valójában a számítógépben az, amit *gondolatnak* tekinthetünk. Olyan tárgyról van szó, amely a kezdetekkor a mesterséges intelligencia kutatására kiválóan alkalmas feladatnak mutatkozott (maga Turing és Simon is készített ilyen programot az 50-es években). Mára mint úgyszólván túlteljesített feladat látszólag elvesztette érdekességét e kutatás számára. Éppen ezért azonban biztonsággal állítom, hogy ennek a fő kérdéseket tekintve letisztult terepnek a vizsgálata a fentiekkel ellentétes következtetéseket implikál az ismertetett „aggodalmakkal és reményekkel” kapcsolatban egyaránt. Az így nyert tanulságokat azután megkísérlem tágabb értelemben is érvényesíteni, a számítógép elvi felépítésére és működésére. Ennek alapján végül a *mesterséges értelem* problémájának, sőt jelentésének olyan felfogása mellett érvelek, amely nem valamely tudomány eredményeire vagy akár alapeszméjére épít, hanem klasszikus filozófiai szerzők – Descartes, Locke, Kant – közvetlen gondolataira, melyek egyúttal megoldást kínálnak az itt körvonalazott „antinómiára” is.

## I. A MUSLICA LELKE

A sakkjáték több szempontból már abban az időben alkalmas kezdetnek és kísérleti terepnek látszott az általános célú számítógépek által elvégezhető feladatok kutatásához, amikor ezek még csak elméletben léteztek. A sakkprogramozás elismerten alapvető cikkében Claude Shannon a játék diszkrét struktúrája, jól meghatározott szabályai és célja mellett hivatkozik arra is, hogy a feladat nem éppen triviális, de nem is túlságosan nehéz ahhoz, hogy valamilyen többé-kevésbé kielégítő eredményt érjünk el, ami különféle elvek és ötletek kipróbálását és tesztelését (*trial and error*) követeli meg. Hozzáteszi továbbá, hogy mivel a sakkozásról általában feltételezzük, hogy gondolkodást igényel, „e probléma megoldása vagy arra fog kényszeríteni bennünket, hogy megengedjük a gépesített gondolkodás lehetőségét, vagy arra, hogy tovább korlátozzuk a ’gondolkodásról’ alkotott fogalmunkat” (Shannon 1950. 1–2). A mesterséges intelligencia kutatásának kezdeti szakaszában valóban úgy tartották, hogy abban szorosan összekapcsolódik a két cél: emberi gondolkodást igénylő feladatokat végrehajtani képes gépek létrehozása, illetve magának az emberi gondolkodásnak a megértése. Ehhez az elképzeléshez illeszkedve érdemelte ki a sakkprogramozás az „MI *Drosophilája*” címet a gyümöleslégy genetikában betöltött szerepének analógiájára. A későbbiek ismeretében azonban látszólag annak vált példájává, hogy mennyire elválhat egymástól az említett két cél és kutatási irány. Az általános nézet szerint ugyanis a sakkprogramok játékereje túlnyomórészt a technikai fejlődés révén és nagyon kevésbé a gépek sakktudása miatt szárnyalta túl fél évszázad alatt a legkiválóbb nagymestereket.

Az alábbiakban először ezt a bevett történetet vázolom föl nagyon „durván”. Azután pedig nem kevésbé határozottan igyekszem majd mellett érvelni, hogy a nézet jórészt hamis, és rámutatni arra, milyen döntő szerepe van az emberi (sakk)tudásnak a gépek rendkívüli teljesítményében, továbbá milyen következtetést érdemes ebből levonnunk.

### 1. A *brute story*

A sakk véges játék abban az értelemben, hogy a 64 mezőn a legfeljebb 32 báb nyilván véges sok variációban helyezkedhet el. Ebből adódik egy fontos szabály segítségével, hogy a játék úgy is véges, ha a kezdőállásból (alapállásból) létrejöheto játszmaváltozatokat tekintjük. A versenyszabályzat ugyanis kimondja, hogy ha egy játszmában háromszor ugyanaz a hadállás áll elő, beleértve azt is, hogy ugyanaz a fél van lépésen, akkor a lépésre következő döntetlent igényelhet („háromszori tükörkép” – ténylegesen nem ritka a gyakorlatban). Az ilyen eseteket döntetlen végződésnek tekintve tehát elkerüljük a végtelen ciklus lehetőségét.

A második szempontú végesség már elvi lehetőséget ad a tökéletes játszma vezetésre a *minimax algoritmus* segítségével. Az összes lehetséges játszma lefolyását egy az alapállásból mint gyökérből kiinduló és minden állásban a lehetséges lépések szerint elágazó fa ágai szemléltetik. A játszmák alakulásának lehetséges száma a megtett lépésekkel exponenciálisan nő ugyan (az alapállásban 20 félét léphetünk, egy tipikus középjáték-állásban 30-nál többet, a végjátékban kevesebbet), de a fenti megfontolásból tudhatjuk, hogy minden elágazás-sorozatnak van végpontja. Mégpedig háromféle végpont lehet, melyekhez egy-egy számot rendelünk (*kiértékelő függvény* a hozzárendelés neve): a matt kaphat valamilyen pozitív vagy negatív számot világos, illetve sötét győzelme esetén, a döntetlen végződés (patt vagy „tükörkép”) nullát. „Tökéletes” játékon azt értjük, hogy minden állásban a lehető legjobb lépést választjuk, vagyis a fa bármelyik csomópontjában az elágazások közül azt, amelyik a lépésre következő fél számára a három lehetséges szám közül a legkedvezőbb, azaz győzelmet vagy legalább döntetlent biztosít. A legjobb választás világos lépése esetén nyilván az egy lépéssel elérhető állások értékeinek maximuma, sötétnél a minimuma (innen az algoritmus elnevezése). Minthogy véges fánkon elméletileg adottak a végpontok, a „levelek” értékei, ezektől visszafelé haladva a teljes fa kiértékelhető egészen a gyökérig, így adván meg a sakkjáték végső megoldását.

Csak hogy a fa hatalmas, és ez a megoldás pusztán matematikai egzisztenciát jelent. Mégis a leírt algoritmus adja a gyakorlatban is tűrhetően sakkozó számítógép programjának alap gondolatát. Számítsunk ki a vizsgálandó pozícióból annyi elágazást, a teljes fának akkora részét, amelyet a számítógép sebessége megadott idő alatt lehetővé tesz (a sakkállások és a lépések gépi reprezentációja önmagában nem probléma, habár a módja nagyon is fontos lehet a hatékonyság vagy a különféle célokra alkalmazhatóság szempontjából). A részfa végpontjai természetesen általában nem lesznek valamilyik fél javára eldőlt vagy döntetlen állások, ezért Shannon közelítő értékeket ad, megtartva azt, hogy a pozitív számok világos, a negatívak sötét előnyét mutatják, de finomabb felbontásban jelenítve meg a felek esélyeit. Ezen a ponton be kell lépnie a sakkról alkotott emberi ismereteknek, mégpedig számokra fordítva. A számba veendő tényezők közül az első a táblán található „anyagi” helyzet, vagyis a különféle figurák fajtája és száma. Itt Shannon átveszi a sakkozó által alkalmazott hozzávetőleges súlyozást az egyes bábok értékét illetően, és a leggyengébbnek, a gyalognak megszabott egység azóta az állásértékelésnek mintegy mértékegységévé vált. Ehhez viszonyulnak ugyanis nála az állás további, „pozíciós” tényezőinek adott kisebb (tört) súlyok. Illusztráció gyanánt felvázol egy ilyen függvényt, amely figyelembe veszi a figurák mozgékonyosságát, a gyalogszerkezetet és a királyállás biztonságát is, de a cikk függelékében jőpár további szempontot is felsorol és felhívja a figyelmet a tapasztalatok és kísérletek tág terére e tényezők és megfelelő súlyozásuk megállapításához. Az erre az emberi tudást hasznosító értékelőfüggvényre épülő minimax algoritmust Shannon „A típusú stratégiának” nevezi (Shannon 1950. 5–6, 10–11, 17).

Ez a megjelölés láthatóan az első megközelítés kezdetlegességének is szól, hiszen – mint írja – a program így még túl kevés változatot vizsgálna az exponenciális növekedéshez mérve, és túlságosan sok fölösleges, mert (emberi) szemmel láthatóan rossz lépésre vesztegetné az idejét. Eleve a stratégia javításaként adja elő a *B* típust, amely a sakkozók egy másik képességét is mintául veszi. Mind a mesterek gyakorlata, mind az erre vonatkozó pszichológiai vizsgálatok azt mutatják ugyanis, hogy egy sakkállásban csupán néhány igazán szóba jöhető lépés van, csak ezeket veszik számba a sakkozók, így juthatván a számolásban viszonylag nagyobb mélységig – s természetesen minél erősebb játékosok, annál jobb minőségben végzik el ezt a szelekciót. Az ilyen plauzibilis lépések megtalálása és a hozzá vezető módok elengedhetetlenek látszottak a jó sakkprogramhoz, egyúttal pedig a MI eredeti intencióinak megfelelően újabb ponton kapcsolta össze a gondolkodó gép alkotásának célját az emberi gondolkodással. A sakkprogramozás kezdeti szakaszát a szelektív, tudásalapú keresés *ad hoc* módjai jellemezték – meglehetősen szerény eredménnyel.

A 70-es évek közepén azonban beköszöntött a „technológia korszaka”, amikortól a programozók elkezdték alkalmazni a már korábban felfedezett *alfa-béta* algoritmust (Simon–Schaeffer 1992. 9–10). Ezzel az eljárással (melyről a következőkben bővebben is szó lesz) a gép sokkal gyorsabban is megtalálhatja ugyanazt az optimumot, így lehetővé válik, hogy *akár*<sup>1</sup> kétszer olyan mélységig jusson a fa kutatásában, mint a minimax algoritmussal, miközben mégiscsak minden lépést figyelembe vesz („teljes szélességben”) az egyes elágazásoknál. A sakkban egyetlen hiba vagy kihagyott lehetőség döntően befolyásolhatja a végeredményt. Minthogy pedig gyakoriak az első lépésben anyagi veszteséggel („áldozattal”) járó, de később (matt, anyagi vagy pozícióelőny révén) megtérülő kombinációk, erősen korlátozza a program játékát, ha ezeket eleve kiszelektálja. Most, a teljes szélességben kereső „brute force” algoritmusokkal (melyet általában Shannon *A* típusú stratégiájával kapcsolnak össze) ez a veszély elhárul. Paradoxul hangzik, de az előbbi megfontolás alapján érthető, hogy az új, a számítógép „nyers erejét” jobban kihasználó programok nemcsak játékerőben izmosodtak, de gyakran találtak briliáns megoldásokat a „megértés”, sőt a „fantázia” képzetét keltve az értő szemlélőben (Frey–Atkin 1978. 186–187; Berliner 1982. 583–584).

A „ne az embert próbáljuk utánozni, hanem hagyjuk a gépet azt csinálni, amit a legjobban tud” jelszavát követve a számítási sebesség növelése utáni hajsza indult a szuperszámítógépek között, melyekhez immár speciális, sakkcélú hardverek, majd multiprocesszoros rendszerek is csatlakoztak. Az önmaga ellen, de a két „fél” számára különböző időkorláttal (kvázi két különböző sebességgel) játszó programok eredményességének tesztelésével mérhetővé vált, hogy például

<sup>1</sup> Ennek a szócskának az itteni, reklámszövegekre hajazó jelentésére és jelentőségére később fog fény derülni az algoritmus továbbfejlesztési kapcsán.

kétszeres sebességnövekedés a játékerő mekkora gyarapodásával jár a sakkozásban használt jól bevált (mondhatni a tudományos igényt is kielégítő) Élő-pontokban kifejezve. Jóllehet ez a pusztán sebességre támaszkodó növekmény csökkent, mikor a gépek sakkmeisteri szinten kezdtek játszani a 80-as években, mégis egy kimondottan erre a szemléletre épülő projekt érte el először a nagymesteri fokozatot az évtized végére. Majd ennek továbbfejlesztéseként győzte le egy hat játszmás meccsen (2 győzelem, 1 vereség, 3 döntetlen) 1997-ben Garri Kaszparov világbajnokot a Deep Blue nevű gép – vagy inkább a 30 processzor által vezérelt 480 sakk-chip, avagy keresőgép (Campbell–Hoane–Hsu 2002. 4).

Sokat mondó, ahogy a masina egyik tervezője és programozója, Hsu a mérkőzésre készülés időszakában úgy nyilatkozik, hogy csaknem „tudásmentes” gépet kellene készíteni, mert például a sakkozók tanácsai, minden szakértelmük mellett, saját előítéleteiket viszik bele az értékelőfüggvénybe; ennél a fontosabb jellemzők helyesebb súlyait kapjuk a sakkadatbázisok statisztikai kutatásával (Hsu 1991). Erősen alátámasztani láthatjuk tehát az afféle vélekedéseket, melyek szerint a Deep Blue „kevesebbet tud magáról a sakkról, mint egy kezdő emberi játékos néhány órai tanulás után”, hiszen a sakkállások alig pár tucat egyszerű, jól ismert jellemzőjét vizsgálja a kiértékelő függvényében, amely ezért joggal tekinthető akár pusztán egyetlen kognitív sémának, szemben Kaszparov sok tízezernyiével. A sakkozó gépek építésének félévészázados folyamata pedig így volna összegezzhető: „'nagymennyiségű állás' konkrét mennyisége éppen mostanra érte el azt a szintet, amikor a sok tízezernyi kognitív séma fölé tud kerelkedni” (Mérő 2001. 252–254).

S ez még mindig hízogó volna az emberi intellektusra nézve annyiban, hogy fenntartja az emberi gondolkodás sajátos minőségi különbségének elképzelését. Mert például Dennett azt sugalmazza, hogy a sakkprogramok Arthur Samuel 50-es évektől fejlesztett öntanuló dámajáték-programjának nyomdokain haladnak, amely önmaga ellen játszva új és új „mutációkat” próbált ki és szelektálta ki a gyengébb változatokat (Dennett 1998. 465, 468, 225). Ilyenként, azaz lényegében csak a szabályos sakk játszására tervezettként használja ezeket párhuzamként Penrose matematikai belátásokra hivatkozó érvei ellen: az életbenmaradásra biológiailag tervezett algoritmusok produktumainak egyfajta „póttethetése” az, hogy matematikai igazságokat ismernek fel, és semmi meglepő nincs abban, ha e túlélőgépek számára az elméjüket létrehozó algoritmus megismerhetetlen (uo. 471–476).

Egyes filozófusoknak és tudósoknak ez az ember- és gépképe keretelméletként szolgálhat még a sakkoktatás didaktikai kérdéseire is. Nemrég az angol sakkszövetség „év könyve” díját nyerte el egy holland sakkedző, Willy Hendriks nemzetközi mester beszédes című műve: *Move first, think later*. A gondolatait uraló tudatos szubjektum „klasszikus” képét már azzal kikezdi a szerző, hogy az *agy* tevékenységéről beszél, mintha a sakkozó gondolkodásához mindenekelőtt a neurológia nyújtana hozzáférést. Ha a „mennyiség” – az ember számára kö-

vethetetlenül nagy számok – fejezik ki az elme lényegét, akkor itt a statisztika a kompetens, illetve a természetes szelekcióval magyarázható annak a tudásnak a kiépülése, amely megmagyarázza, hogy a lépések öntudatlanul „buggyannak ki” a memóriából. Ha egyszer nem tudunk másképp számot adni a döntéseinkről (amint azt a könyv több száz kitérő példáján igyekeznek illusztrálni), akkor ne csapjuk be magunkat a sakkal kapcsolatban felállított általános szabályokkal, formulákkal, maximákkal, melyeknek ő nemcsak a játék, de még a tanulás és felkészülés során sem juttatna szerepet. Inkább egyszerűen bízzunk a bennünk lakozó sakkozó „modulban” és „etessük a bestiát” konkrét állásokkal, egyedi szituációkkal, jó lépésekkel. Az utólagos (esetleg pszichológiai) belemagyarázás, az általánosságok és nyelvi megfogalmazásaik nem többek, mint „a letűnt pre-computer kor metafizikai relikviái”. Valóban, a számítógép nemcsak mint az igazságot osztó „órákulum” bukkan fel néhol, hanem Hendriks kifejezetten épít rá nézeteinek alátámasztásaként. Ő is azt az általános képet vázolja fel, miszerint a gép – nagyon kevés és durva tudással – a változatok milliárdjainak segítségével hozza meg a jó választásait, s ez teszi az agy működésével analóggá. Hangsúlyozza viszont azt a különbséget, hogy a gép esetén a nagy számok ellenére elvben végig lehetne követni bármelyik (vagy az összes) lépésváltozatot, míg az emberi (értsd: agyi) mennyiséget így nem tudjuk felderíteni. Ebből aztán egészen elragadó következtetést von le: az emberi sakkozó előtt, ha egy állásra néz, a lépések szinte „automatikusan” merülnek fel, míg a számítógép valóban gondolkodik – jóllehet ez a legutóbbi, tiszteletreméltó kifejezés nem olyan, amihez szerzőnknek gusztusa volna ragaszkodni, inkább átpasszolja azt az MI kutatóinak (Hendriks 2012. 244–245, 146, 97, 168–173).

## *2. Az értelmes történet*

Meddig jutott a sakkozó gépek játékeréje az ominózus párosmérkőzés óta? Vezető nagymesterek (időnként igen súlyos) vereségei mellett a kimondott világelcsők, Kaszparov és Kramnyik a 2000-es évek elején háromszor is döntetlent értek el, de az utóbbi 4-2 arányú veresége (2006) után már alig találunk komoly példát hasonló játszmákra. A kimondani is szörnyű előnyadós partikat (ahol a gép a kezdőállásban gyalog, tiszt vagy néhány lépés előnyt ad az embernek) gyorsan elfelejtve megemlítem azt a nagymesterversenyt 2009-ből, amelyen egy mobiltelefon produkált világbajnoki szintű eredményt. A sakkozók játékeréjét az egymás elleni versenyeredmények alapján folyamatosan nyilvántartó listán a világ legjobbjai 2800 Élő-pont fölé jutnak. A gépek szintén egymás elleni verseny- és teszteredményein alapuló listákon a legelső manapság 3100–3200 pont körül járnak, mely pontozást természetesen eredetileg az előbbihez igazítottak, amikor még gyakoriak voltak az ember-gép párharok. Nincs is különösebb okunk kételkedni az erőviszonyok realitásában amiatt, hogy újabb játszmák ezt nem

hitelesítik; az Élő-féle pontszámítással 300 pont fölötti különbség már amúgy is nehezen mérhető közvetlenül, lévén az erősebbik fél elvárása vérszesen közelít a 100% felé.

Mit is várhatnánk mást, ha a számítógépek gyorsasága az elmúlt másfél évtizedben talán ezerszeresére is nőhetett? Inkább az a kérdés, hogy vajon az előbbi megalázó bekezdést miért nem az előző „brute” szakaszhoz csatoltam. Nos az igazság az, hogy a felsorolt eredményeket (egy kivétellel) a kereskedelemben kapható személyi számítógépeken érték el. Intézmények (egyetemek, kutatóintézetek, cégek) a Deep Blue után már nem építenek hasonló szupergépet erre a célra – érthetően: a feladat elvégeztetett. A soha nem látott és nem is álmódott sakkteljesítmények az asztalainkon zajlanak. Csakhogy éppen ez az, ami fölveti az igazi kérdést. A sakkprogramok sebességét ugyanis a másodpercenként megvizsgált állások számával mérik. Márpedig ebben a Deep Blue-nál klasszissal jobb mai programok több százszor *lassabbak*, mint a monstrum, a vele legalábbis pariban levő zsebgép meg éppenséggel tízezerszer. Itt az ideje úgy mondani el újra a történetet, hogy végre értelmet is vigyünk bele. Látni fogjuk, hogy a sakkprogramozás alapvető módszerei vagy kifejezetten támaszkodnak az *emberi* sakkozás elveire, vagy legalábbis azokhoz hasonló gondolatokat rejtenek.<sup>2</sup>

A fordulópontot, ahonnan a sakkgépek készítése állítólagosan a „brute force” keresés felé vette az irányt, az *alfa-béta algoritmus* felfedezése jelentette. Az eljárás eszméje az, hogy ha egy M lépést meggyőzően „megdönt” egy M' válasz lépés, vagyis M' vizsgálata és kiértékelése azt mutatja, hogy M rosszabb, mint valamelyik korábban már megvizsgált és kiértékelte alternatívája, akkor M-et biztosan nem érdemes választani és fölösleges az ellenfél további lehetséges válaszait (M' alternatíváit) kutatni, ezeket az ágakat nyugodtan „levághatjuk” a fáról. Ezt az informális megfogalmazást nem igyekszem most formálissal kiegészíteni, mert nagyon is igazat adok John McCarthy-nak, az algoritmus első javasolójának és névadójának: „Az alfa-béta vágás jellemzi az emberi játékot, de ezt a korai sakkprogramozók nem vették észre [...]. Mi, emberek nemigen vagyunk jók az általunk használt heurisztikák azonosításában” (McCarthy 2006). Szigorúan szólva nem heurisztikus eljárásról van szó, mert bizonyíthatóan ugyanazt az eredményt adja, mint a minimax algoritmus. Mégis a gyakorlati szempont, a kötött idő melletti minél jobb játék a programozó gondolkodását a tisztán matematikai gondolkodástól a döntési helyzetben lévő emberéhez közelíti.

Még „életszerűbb” az algoritmusnak az a továbbfejlesztése, amikor nem az egyes szinteken már kiszámított értékek adják azokat a határokat, melyek alá egyik illetve másik fél nem lesz hajlandó menni („alfa-béta ablak”), hanem

<sup>2</sup> Az alább szóba kerülő eljárások megértésében és értelmezésében (akárcsak az eredmények fenti összefoglalásában) nagyban támaszkodtam a <<http://chessprogramming.wikispaces.com>> weboldalra is, Frey–Atkin 1978, Simon–Schaeffer 1992, Marsland–Björnsson 1997, Hsu 1999 és Schröder 2004 mellett.



egy előre elvárt intervallumon belül keresi a kölcsönös optimumot (*aspiration window*). A sakkozónak, mikor előre kalkulál, az állás megítélése alapján van egy bizonyos követelése, mondjuk úgy gondolja, nem lesz hajlandó olyan változatba „belemenni”, amely gyalog- vagy hasonló súlyú pozícióvesztéssel jár; és ugyanezt várja az ellenféltől is, avagy – a saját szemszögéből fogalmazva – boldogan választaná azt az elágazást, ami számára ilyen nyereséggel jár, és ezért nem számolja azt tovább. Előfordul persze az álláshoz tartozó ablak túl szűkre szabása, és kiderülhet, hogy a legjobb játék esetén az értékelés átlépi valamelyik határt (jobban vagy rosszabbul állunk, mint gondoltuk). Ilyenkor meg kell ismételnünk az elemzést tágabb, vagy az eredeti eljárás szerinti teljesen nyitott ablakkal. Ez a módszer ezzel az idővesztéssel együtt is hatékony a számítógépeknél (figyelembe véve az ablak méretével való játék, finomhangolás lehetőségét is) és alighanem elengedhetetlen „vágás” az embernél.

A vágásra annál jobb esély van egy adott szinten, minél korábban vizsgáljuk a jó lépéseket (optimális esetben a legjobbat legelőször), így a hátralévők majd gyorsabban kihullanak. Az algoritmus még akkor is jelentős megtakarítást biztosít a minimaxhoz képest, ha teljesen véletlenszerűen rendezzük sorba az állás szabályos lépéseit, de nyilvánvalóan sokkal hatékonyabb a lehetőleg legjobb sorrend esetén. Ehhez viszont már valóban heurisztikus módszerek kellenek, amelyben természetesen a sakkról való tudás is fontos szerepet kap. Egészen természetes például a vizsgálandó lépések listáján előre helyezni egy gyalogátváltozást vezérré vagy egy ütést, kivált ha egy kisebb értékű figura üthet egy magasabbat. Idesorolhatjuk a sakkadást is, ami ha nem is feltétlenül jó lépés, de biztosan „olesó”, amennyiben csak igen korlátozott számú szabályos választ enged meg, így kevésbé növeli a fa méretét. De „csendes lépés” (nem ütés vagy sakkadás) is lehet olyan gyakran előforduló cáfolat egy adott szinten, amely miatt érdemes tárolni és – ha legális – hamar kipróbálni (M társait gyorsan elintézheti mondjuk ugyanaz az M’ „fenyegetés”).

A sorrend helyes megállapításának jelentőségét mutatja az *iteratív mélyítés* időgazdálkodás szempontjából biztonságos, bár elsőre nagyon költségesnek látszó eljárásának használata. Ha ugyanis a program az állás legsekélyebb, 1 féllépésnyi (hagyományosan a fehér-fekete lépéspárt is nevezik ’lépésnek’ a sakkban) vizsgálatától kezdve fokozatosan, mindig egy féllépéssel mélyebben végzi el a teljes alfa-béta keresést, akkor az engedélyezett idő lejárta esetén is viszonylag jó lépést tesz. Ugyanakkor, ha a mindenkor megelőző szinten kapott értékelés szerint *sorbarendezi* az elemzendő lépéseket (összehangolva a fentiekhez hasonló heurisztikákkal), az jócskán megtérül a további vágások számában. Egy bizonyos mélységű elemzés után a mindkét fél részéről legerősebbnek ítélt lépések sorozatára, a „főváltozatra” egy az algoritmust tovább gyorsító módszer épül (*Principal Variation Search*), amely megintcsak nem idegen az emberi sakkozástól. Nagyon is jellemző, hogy a sakkozó akár a számolás végeztével a lépés előtt, akár a leginkább várt lépés után először a főváltozat lépésein fut végig ellenőrzésképpen: van-e esetleg

valamelyik fél számára jobb folytatás. A programban ilyenkor a főváltozat (adott szinten) egyetlen lépésén kívül a többi minimális ablakkal, vagyis a feltételezés szerint minél rövidebben vizsgálják. Ez a szinte agresszív, az egyéb lépések gyengébb voltát gyorsan igazolni szándékozó módszer különösen támaszkodik a lehetséges folytatások jól megállapított sorrendjére.

Ha visszatekintünk Shannon két stratégiájára, azt kell látnunk, hogy a *B* típus megvalósításához pontosan egy adott állásban lehetséges lépések értékének *sorrendjét* tartja meghatározandónak, például egyre csökkenő számot adva a kényszerítő lépéseknek (sakkadás, ütés, figuratámadás), fejlődő lépéseknek, védőlépéseknek és végül a többinek. E rend szerint kell kutatnunk a további változatokat, mégpedig – tekintettel a szelekció kockázatára – első lépésben *minden* lépést vizsgálva, majd a számolásban mélyebb szintekre jutva egyre inkább a lista elejére szorítkozni (Shannon 1950. 13–14). Tehát, jöllehet a *B* típust valóban a szelekció különbözteti meg az *A*-tól, de ez a vizsgálandó lépések sorrendjének megválasztását is jelenti és kiszelektálást csak a gyökértől távolabbi szinteken.

Ez így van a mai programokban is: a biztonságos teljes szélességű (minden lépést vizsgáló), akkoriban persze elképzelhetetlen mélységű alfa-béta elemzést még különböző – valóban szelektív – *kiterjesztések* követnek, amelyek igen eltérő (és legtöbbször sakkbeli ismereteket felhasználó) heurisztikákon alapulnak. A keresési mélység horizontjának kitolása bizonyos pontokon egyenesen kényszer. Mi sem mutatja jobban, hogy Shannon az *A* típust csak egy kezdeti megközelítésnek tekintette, mint hogy azt azon a „rendkívül eredménytelen” módon működként írja le, ahol mindig egy előre rögzített mélységig hatol az elemzés. Ez ugyanis nem veszi figyelembe, hogy a sakkban értelmetlen olyan állás finom értékelése (a kiértékelő függvény alkalmazása), amely aktuálisan sok „taktikai” lehetőséget tartalmaz (például az említett kényszerítő lépéseket); ilyenkor feltétlenül tovább kell számolni egy stabilabb, „nyugodt” állásig (Shannon 1950. 12–13). *Quiescence search*nek nevezik a programokban ezt az elengedhetetlen funkciót. Nincs tehát olyan sakkprogram, amely ne Shannon tanulmányának („emberi”) *B* stratégiája nyomdokain járna, miközben az *A* típusra valóban tökéletesen illik a *brute force* kifejezés (habár ő maga nem így használja a cikkben).

Ugyanakkor nem is szoltunk még arról a témáról, amit elismerten a sakk-tudáshoz kapcsolnak, nevezetesen az *állásértékelésről*. Holott természetesen az alfa-béta algoritmus – minden javításával – végső soron erre épül: a „mélyről feljövő” értékek határozzák meg a kalkuláció egyes döntéseit és végül legjobbnak ítélt lépéssorozatot. A sakkállás egyes jellemzőit számokra fordító és összegző kiértékelő függvény elsöre nagyon eltérni látszik az emberi intuíciótól, megérzéstől, „sakkérzéktől”. Ráadásul Shannon az egyes strukturális elemek jelentőségét kimondottan didaktikus *maximák és elvek* segítségével javasolja megbecsülni. Valóban nagyon merevnek hatnak például a sakkprogramok különböző figurák számára készített táblái, amelyek a báb elhelyezkedése (a centrumtól vagy az ellenséges táborból) való távolság függvényében pluszpontokat ad, eset-

leg levon büntetésül. Nemkülönben az olyan általa is felsorolt speciálisabb szabályok számszerűsített változatai, mint hogy a bástyák jól állnak a nyílt vonalakon, a visszamaradt, izolált vagy dupla gyalogok gyengék, a védtelen király gyengéséget jelent (a végjátékig). Hiszen ha van egyöntetűen elfogadott „maxima” az általános elveket illetően, akkor az így hangzik: „bármely sakkelvnek ellene vethetők egyes ellenpéldák”.

Csak hogy ezt ismét Shannontól idéztem! A sakkozók efféle elvei számtalan játszma alapján alkotott *empirikus általánosítások*, és természetesen csakis statisztikai érvényességgel bírnak (Shannon 150. 5). Ilyenként alkalmasak arra, hogy különböző „súlyok”, fokozatok adják vissza őket; ahogy általában az érzetek, úgy az itt felmerülő „jóság”, „gyengeség”. „erősség” is mérés tárgya lehet. Az állítólagos kivételeket pedig az összegzés kezeli: dupla gyalogom gyengésége nem szűnik meg attól, hogy a másik oldalon a királyomat támadják, hanem lényegtelenné válik, az utóbbi tényező egyszerűen felülírja – úgy, ahogy annak egy jó értékelőfüggvényben történnie kell.

A „nagyon gyors, de nagyon buta számítógép” legendája rögtön átíródik, ha szemügyre vesszük a Deep Blue értékelőfüggvényét. Az 1996-os 4–2-re elvesztett meccs visszavágójára készülve teljesen újratervezték a kiértékelő függvényt, amely immár 8000 féle tulajdonságot vagy „mintát” regisztrált, értékelt és használt az összegzéshez, különféle sakkelméleti fogalmak szerint csoportosítva és kezelhetővé téve „a bőség zavarát”. Az eddig említetteken túl olyan speciálisabb fogalmakról van szó, mint kötés, szabadgyalog, blokádnál, egy figura elfogása vagy korlátozása, előőrs, király- vagy vezérszárnyai gyalogtöbblet, bástya a 7. soron, futópár, ellenkező színű futók, fejlődési előny stb. (Campbell–Hoane–Hsu 2002. 3–8, 18–22, 26–28; Hsu 1999. 71, 79, 81). Egészen különleges bepillantást enged ebbe a munkába (és eredményébe) az az interjú, amelyben Hsu, aki korábban a „majdnem tudásmentes” gép építését propagálta, nemcsak általánosságban vallja meg, hogy Kaszparov ellen készülve fel kellett ismerniük, a sakk tudás bővítése elengedhetetlen, hanem a legmegdöbbentőbb példákon demonstrálja ezt. A mérkőzés két (a gép szempontjából) győztes játszmájának kulcsmozzanatáról meséli el, hogy milyen döntő szerepet játszott bennük a csapatot segítő Joel Benjamin nagymester tanácsai nyomán korrigált értékelőfüggvény (Hsu 2005. 3–5, 9).<sup>3</sup>

<sup>3</sup> Az egyikben arról van szó, hogy ha a térelőnyben levő fél kinyithat egy vonalat, akkor érdemes ezzel várni és „fenyegetni” vele akár a bástyák megduplázásával a vonalon, amire az ellenfélnek nincs elég helye. Valóban így játszott a gép a második játszmában. A hatodik játszma kritikus állásával, ahol a gép – Kaszparov várakozásával ellentétben – a 7. lépésben beáldozta a huszárját egy gyalogért és a király elleni támadásért, a megnyitáselméleti felkészülés során (két hónappal a meccs előtt) volt alkalmuk találkozni és a királybiztonság paramétereit tesztelni. Elemeztek ugyanis egy akkoriban játszott ember-gép partit, amelyben az „anyagias” masina nem élt *pontosan ugyanazzal* a lehetőséggel.

Ma pedig már a sakkozók és sakkszakírók munkájának olyan gépi támogatása is felmerül, amely nem pusztán lépéseket és számszerű értékeléseket nyújt, hanem – szakértői rendszer gyanánt – az adott állások és döntések mögött rejlő *pozíciós* tényezőkről is kommentárt képes szolgáltatni. Az alapgondolat az, hogy a program a legjobb lépésváltozatok (vagy épp a hibák) esetén azt keresi, van-e az értékelőfüggvénynek olyan összetevője, amely az illető lépéssorozat kiértékelésében a megnövekedett (vagy lecsökkent) pontszámért különösen felelősé tehető (Guid 2010. 69–83). A helyzet látszólag paradox, de mindenképpen nagyon tanulságos. A sakkjátzmák és megnyitások leírt elemzésében már évtizedekkel ezelőtt megjelent az a szigorú, tudományosnak és talán „gépiesnek” mutakozó mód, hogy csak a fontos lépésváltozatokat és eltéréseket adják meg egészen minimális szöveggel vagy csak bizonyos bevett szimbólumokkal jelezve, ki áll jobban. Ez az, ami tökéletesen megfelel a Willy Hendriks által is képviselt ideológiának. Most azonban azt kell látnunk, hogy a fő szövetségesének vélt számítógép egyenesen hátbatámadni készül: azzal fenyeget, hogy a számokat a sakkállást jellemző fogalmakra és elvekre lefordítva, a „pre-computer era” tanítóinak módjára fogja oktatni a sakkozót. Ami viszont a számunkra pillanatnyilag legfontosabb: mindez éppen azért volna lehetséges, mert az értékelőfüggvény a jellegzetes esetekben *vissza* képes adni a programozó és szakértő társai által benne elrejtett, előbb emberi, majd gépi nyelven *megfogalmazott* sakk-gondolatokat.

### 3. Következtetések

A fenti rövid összefoglalás lehetőséget teremt arra, hogy feloldjunk néhány olyan dichotómiát, amelyet a sakknak az MI kutatásban betöltött szerepéért aggódók hajlamosak feltételezni. Egyike ezeknek a 'keresés *versus* tudás' témaköre, ahol a két fogalom az emberi sakktudást, illetve a brute force számítást mint két szélsőséget vagy két tengelyt jelölné, melyeket szerintük egymáshoz közelíteni kellene (Donskoy–Schaeffer 1990; Simon–Schaeffer 1992; Winkler–Fürnkranz 1998). Ha a programok sakktudását kizárólag a kiértékelő függvényre értenénk, akkor is láthatjuk, hogy még a brute force szimbólumának tekintett óriásgép is messzemenően e kíváncsi felé mutat. De még fontosabbnak tartottam rámutatni arra, hogy magának a keresésnek a legalapvetőbb algoritmusai és az azt hatékonytá tevő heurisztikák mennyire megfelelnek alapeszméiket tekintve a sakkozók gondolkodásának – ha nem is lényeges itt, hogy mennyire gyökereznek kifejezetten abban. Bátran mondhatjuk, hogy a sakkprogramokban *a keresést a sakktudás vezérli*.

Egy másik állítólagos ellentét egyfelől a mérnöki-programozói tevékenység, a versenysikerek hajszolása és az üzleti érdekek, illetve másfelől a (kognitív) tudomány között merülne fel. Ez a képzet egyeseket arra az ítéletre vezet, hogy

„pusztán” sikeres sakkgépek és programok készítése nem lehet az MI eredeti célja, mert nem vonja szükségképpen magával, hogy jobban megértjük a gondolkodó (sakkozó) emberi elmét (Donskoy–Schaeffer 1990. 1–7; Winkler–Fürnkranz 1998. 10; McCarthy 1997). A sakkozás szempontjából mindenesetre kiemelkedő jelentősége van annak, hogy a gépek minél erősebben játszanak. Így támogatják leginkább a sakk tudományának művelését, segítenek felderíteni, mi rejlik egy-egy állásban.<sup>4</sup> Ehhez pedig éppen a mérnöki és programozói munka vezetett el.

Másrészt a programok alapvető működési elveinek a vizsgálata azt mutatja, hogy helyesen tekintve a sakkprogram készítése nagyon is hozzájárulhat a sakkozó gondolkodásának pontosabb megértéséhez. Nem azzal, amit produkál, hanem azzal, ahogy a program készítője vagy annak segítője (esetleg az azon meditáló) *reflektál* az algoritmusok értelmére, jelen esetben azok sakkelméleti vagy gyakorlati jelentésére. Az introspekciót megbízhatatlannak tartják a tudomány számára, de az általa sugallt módszerek kezdetektől meghatározó szerepet kapnak a sakkprogramozásban, vagy ahol ez nem felderíthető, ott is gyakran visszaigazolja az alapgondolatot a sakkozó önreflexió. A sakkozó gondolkodásának utólagos úgymond „önbecsapásai” a programok eljárásainak előzetes megfogalmazásait alkotják, mely *általánosságokat* (*pace* Hendriks) a programozó egy bizonyos *nyelvre* (programozási nyelvre) fordítva közöl a legeminensebb tanítvánnyal, a géppel. Azt állítom tehát, hogy a sakkprogramok készítői nagyon is sokat hasznosítanak a „metafizikai relikviákból”, akár a maguk sakktanulmányaiból, akár a sakkozóktól szóban kapott tudás révén, és a saját („technikainak” mondott) felfedezéseiket is többnyire meg lehet fogalmazni és érthetővé tenni a sakk nyelvén.<sup>5</sup>

<sup>4</sup> Egy megjegyzést megérdemel „maga a dolog”: a példaként használt kisdud játék (jó esetben egy „muslica”) kimerít(het)etlensége. A múlt századnak a húszas és a hetvenes éveiben is volt egy-egy hulláma a „remihalál” miatti sopánkodásnak, vagyis annak a nézetnek, hogy a sakkelmélet kimerülöben van és a legmagasabb szinten a játszmák egyre nagyobb hányada döntetlenül végződik (*remis*), a másodhúzóknak, azaz sötétnek pedig különösen nehéz nyernie. (A világbajnoki döntők kiváltképp igazolni látszanak az elméletet) Nos, a minden valaha élt embernél sokkal erősebb sakkprogramok egymás közötti – megalkuvást nem ismerő, ellenben kiválóan mérhető – mérkőzései nagyjából a következő statisztikát mutatják (a szokásos időellenőrzés mellett, ami lépésként átlagosan három perc): a döntetlen játszmák aránya 40–45%, és kb. 10%-kal többször nyer világos, mint sötét.

<sup>5</sup> Attól tartok, a programozók általában nem tulajdonítanak ennek akkora jelentőséget, a sakkozók pedig a ténnyel sincsenek tisztában. Pedig igenis tanulhatnának nemcsak jó lépéseket a géptől, hanem a programok alapgondolataiból is. Ha mást nem, mint azt, hogy ezeket tőlük vették és fogalmazták át „protokollá” és számokká a programozók, az sem kevés, sőt talán a legtöbb! Mert azt erősítheti a sakkozóban, hogy ne tágítson az általános fogalmaktól és elvektől: ezek vezették az „emberfeletti” teljesítményekhez az eszeveszett (pontosabban önmagában esztelen) sebességet is. Itt szeretnék köszönetet mondani egyrészt Sziva Erika nemzetközi nagymesternőnek, többszörös olimpikonnak az inspirációért (Hendriks-könyv), másrészt Horváth Gyulának, a Pandix nevű 2013-as világbajnoki harmadik, az egyforma hardveren futó szoftverek között második (icga.uvt.nl) sakkprogram alkotójának a konzultációért.

A sakkozásra szűkítve a problémát, a shannoni dilemma – mely szerint vagy „megengedjük a gépesített gondolkodás lehetőségét, vagy [...] tovább korlátozzuk a 'gondolkodásról' alkotott fogalmunkat” (Shannon 1950. 1–2) – eldőlni látszik az első alternatíva javára. Nem kell azonban azt hinnünk, hogy a nevezett „tiszteletreméltó” fogalom ezzel a kiterjesztésével devalválódna, vagy minket magunkat valamiféle veszteség érne. *Saját gondolatainkat* látjuk gépesítve, azaz megláthatjuk, mire képesek *elkeink* vasfegyelemmel, valamint hatalmas számítási kapacitással és pontossággal megerősítve. Ezen a vonalon próbálunk még egyet lépni.

## II. A GÉP ELMÉJE ÉS AZ EMBER ÉRTELME

A gondolkodás, de még inkább az intelligencia fogalmával valamilyen igen összetett és „magasrendű” tevékenységet kapcsolunk össze, ezért is megrázó, ha azt kell látnunk, hogy egészen mechanikus, „gépies” működések is produkálhatnak hasonlót. Az előzőekben egyik e fogalmak alá sorolt szellemi tevékenységünkről igyekeztem megmutatni, hogy gépiesített változatában az alapgondolásokat tekintve mennyire irányadó a sakkozó gondolkodása. Bizonyára lehetséges a mesterséges intelligencia más alkalmazásai és egyéb gépi problémamegoldás esetén is hasonló, az adott terület emberi tudására reflektáló gondolatmenetet követni és annak szerepét kimutatni. Innen továbblépni olyan keményebb diók felé illene, mint például a nyelvhasználat, alakfelismerés vagy a hétköznapi józan ész, melyek – a sakkal ellentétben – a való világba ágyazott problematikák. Én azonban könnyű célpontot választva ellenkezőleg, visszafelé, a „leggyengébb MI” irányába hátrálva magára a számítógép működési elveire vonatkoztatva szeretnék a fentihez hasonló kísérletet tenni. Filozófiailag ugyanakkor ez a lépés a radikálisabb, ami abban is megnyilvánul, hogy a gondolkodás és az értelem nem összetettként vagy magasrendűként, hanem egyszerűként és alapvetőként jelenik majd meg.

A mechanikus gondolkodás állítólagos paradoxonát jelenti, hogy a szimbólumok vagy gondolatok manipulálásából álló okoskodás csak akkor járhat fizikai hatással, ha ehhez feltételezünk egy a két elvet eleve összekapcsoló belső manipulátort, egy misztikus „*homunculust*” – ami viszont a végtelen regresszió rémével fenyeget (Haugeland 1989. 40–41, 117; Horst 2009). Mármost nagyon is lehetséges, hogy a (gépi vagy emberi) gondolkodás *magyarázatának* igénye efféle paradoxonnal terhes, és valóságos misztérium, hogy vajon mi is tudná kielégíteni. Egészen más viszont a helyzet a gondolkodás *létrehozásával*. Érdeemes mindenekelőtt szem előtt tartanunk Turing következő szavait:

A digitális számológépek alap gondolatát úgy lehet megértetni, ha elmondjuk, hogy ezek a gépek rendeltetészerűen minden olyan műveletet el tudnak végezni, amit emberi számoló el tud végezni. [...] Az élő számolók a valóságban emlékeznek arra, hogy mit kell tenniük. Ha valaki a géppel utánoztatni akarja azt, hogy az emberi számoló hogyan viselkedik valamely komplex művelet során, meg kell kérdezni, hogyan végezte el a dolgát, és azután a választ le kell fordítani egy utasításgyűjtemény formájában. Az utasításgyűjtemények konstruálását általában „*programozásnak*” nevezik (Turing 1965. 125, 127–128).

Turing a legtermészetesebb módon hivatkozik mintaként az emberi számolóra mind az egyes műveletek elvégzését, mind azok rendjét illetően, mégpedig mint olyanra, aki ezekről számot is tud adni. Egy „kis számolót” kell tehát létrehozni belső manipulátor gyanánt, amely „emlékezetében” őrzi és folyamatosan figyelemmel kíséri a feladatát, valamint végre is hajtja vagy hajtja valamilyen részegységével. Ha a szóhasználat esetleges is, mégis sokatmondó e vezérlőegységgel azonosított gép „belső állapotairól” beszélni (uo. 133). Nyilván az adattároló képviseli a külső környezetet, amely persze szintén a géphez tartozó memória.

E mesterséges dualitás nyomán nem veti fel senki a test-elme metafizikai problémáját, eleve adottnak vesszük az adatok forgalmát és kezelését. Még inkább elvenni látszik az elme életét, hogy a Neumann-féle felépítésben a programokat is adatként tárolják: adat és a manipulálására vonatkozó utasítás egynemű a hordozója tekintetében. Így azután – John Haugeland szavaival – a „magabiztos és mindent leigázó huszadik századi materializmus” nem szorul rá a jelentésekre kifejezetten *figyelő akarat* képességére vagy a „fennkölt” *transzcendentális énre* – az idealizmus (vagy a dualizmus) legfeljebb egy kitérő az MI származástörténetében (Haugeland 1989. 44, 41).

Pedig e szubjektumközpontú, vagy inkább a szubjektumból kiinduló filozófiáknak nem kell inkább szembesülniük valamiféle „ontológiai szakadék” problémával, mint a számítógépnek. Az észleléstől a tapasztalaton át az erkölcsi és esztétikai megítélésig minden ítélet hordozói, „anyaga” egyöntetűen az „ideák”, vagyis a képzetek. „[A]z ember uralma saját értelmének e kicsiny világában” nem több, de nem is kevesebb, mint az adott anyag manipulálása, beleértve a „reflexió ideáit” is (Locke 2003. 124). Másfelől a számítógép (tárolót, bemenő adatokat is tartalmazó) „kicsiny világán” belül konstruált manipulátor, a működő program rugalmasságát is éppen az adja, hogy műveleti utasításai „eszmei dolgokként” (számokként: a gép „képzeteiként”)<sup>6</sup> a memóriában vannak tárol-

<sup>6</sup> Pontosabban szólva a számítógép nem számokat kezel, hanem olyan elektromágneses alapú jeleket, amelyek végső soron számokon kívül bármi mást is jelölhetnek, minél fogva általános célú szimbólum-manipuláló rendszernek tekinthető (Simon 1980. 6264; Simon 1982. 6566).

va, így a gép a saját működését szabályozó utasításokat is módosíthatja (Neumann 2003. 279–280).

Továbbra is a gépeknél maradvá, de az embert nem a saját szeme elől tévesztve arra szeretnék rámutatni, hogy a Turing- vagy a Neumann-féle gép veleje, azaz pontosan egy ilyen belső manipulátor, egészen jól megfelel a megidézett, de gondosan elkerült filozófiák intencióinak és kifejtett alapelveinek.

### 1. A tökéletes kartezianus

Descartes az analitikus geometria megalkotásával vagy az állati (speciálisan az emberi) test gépként kezelésével bizonyára helyet kap ezen a családfán, de leggyakrabban a gépi értelem lehetetlenségéről tett kijelentését idézik ez ügyben, amellyel mintha mindjárt el is metszené a maga ágát. Nem valamilyen különleges tevékenység vagy szerv az, ami szerinte az embert kiemeli az állatok és a gépek közül, hiszen sok mindenben felülmúlhatják azok őt, hanem egy „egyetemes eszköz, amely minden esetben feltalálja magát”, nevezetesen az ész. Másik érve is az emberek közötti egyáltalán nem rendkívüli, hanem egészen általános, de csak rájuk jellemző képességre támaszkodik: az értelmes nyelvhasználatéra (Descartes 1992. 64–66). Ezzel valójában Descartes csak az MI „erős” változatának kényes pontjára tapint rá, és csupán „erkölcsileg” (tehát nem logikailag vagy metafizikailag) mondja lehetetlennek az egyetemes eszközt különösekk aggregátumaként létrehozni.

Am sokkal tanulságosabb azt a fát vizsgálni, amely a tudomány épületét szimbolizálja, mégpedig a gyökerénél. Itt, a metafizikában (első filozófiában) a matematikán pallérozott gondolkodás a bizonyosság igényét a *cogito*-érv által kielégítő igazságkritériumban találja meg azt a szilárd alapot, amelyre majd – a matematikai tudomány „igazi hasznaként” is – mechanikai mesterségeknél „valami magasabb” is építhető (uo. 20). Igen, az igaz és a hamis megkülönböztetése az életvezetéshez ad biztos fogódzót, „hogy az értelem az élet minden helyzetében meg tudja mutatni az akaratnak, mit kell választania” – s ezt egy gépi értelemtől alig várhatjuk (Descartes 1980. 98; Descartes 1992. 23, 65). De az a *képesség*, amelynek helyes használata vezethet ehhez, és amely az ember formáját és természetét alkotja, önmagában „egész mindenkiben”, s nem tökéletesíthető. Akár a józan értelem vagy ész ez a képesség, akár (szűkebben) az akarat, melynek folytán lehet az embert „Isten képeinek nevezni, és Istenhez hasonlónak tartani”, ez a fokozhatatlanság, állítás és tagadás dichotómiája az, ami Descartes emberfogalmában egybecseng a megvalósult számítógépek alapfogolataival (Descartes 1992. 15; Descartes 1994. 71–72).

Amennyiben a descartes-i elmefogalmat tartjuk szem előtt, úgy az ember a számítógépet a maga képére és hasonlatosságára teremtette. A gondolkodást Descartes két általános módusra vezeti vissza, az észlelésre és az akarat tevé-



kenységére, a választásra. A gépnél ezeknek első megközelítésben valamilyen adat beolvasása, illetve a következő feladat végrehajtása (esetleg adat kiírása a tárolóba) felelne meg. Ugyanakkor a beolvasott adat a gép aktuális belső állapotával (mondhatnánk: a külső és belső észlelés) együtt tökéletesen meghatározza a következő lépést. Ez azonban a descartes-i szabadságfogalomnak nem feltétlenül mond ellent, csak annyit jelent, hogy a gép – hardver vagy szoftver hibáktól eltekintve – sohasem „semleges”, hanem az ember gondoskodik arról, hogy folytonosan az „evidens belátás” állapotában legyen (vö. Descartes 1994. 73).

Természetesen nem arról van szó, hogy az ember Isten, aki elmét teremtett. Inkább arról, hogy úgy tekinthetjük: saját elméje ideális használata, „a legnagyobb és legkiválóbb emberi tökéletesség” számára alkotott egy annak megfelelő testet. Olyat, amelybe nemcsak a világos és elkülönített belátások követésének „hajlandósága” van eleve beépítve, hanem a nem-tévedés „habitus” is (uo. 77): programhiba esetén a gép leáll, „tartózkodik az ítéletalkotástól” és akarata nem terjed túl az értelmén. Így nincs szükség arra, amire Descartes kénytelen olyan nagy hangsúlyt fektetni az ember testéből fakadó ellenállás és az érzékeiből származó előítéletek miatt, tudniillik a szabadság helyes használatának begyakorlására, megfelelő szokások kialakítására.

„A gépek tökéletes karteziánusok. Csakis azokkal a diszkrét és meghatározott információdarabkákkel tudnak bánni, melyeket Descartes ’világos és elkülönített ideáknak’ nevez” – mondja Hubert Dreyfus. Ezzel együtt szintén idézi néhány oldallal később a mindennapi élet „információfeldolgozásának” gépi szimulálását megkérdőjelező descartes-i passzust. Szerinte ez jól mutatja annak a „tisztetreméltó, de némiképp idejétmúlt” elmeképnek a kudarcát, vagy legalábbis korlátait, amely éppen a digitális számítógépek megjelenésével kapott új erőre. Eszerint „*a gondolkodásnak felbonthatónak kell lennie egyszerű, meghatározott műveletekre*” (Dreyfus 1965. 66, 69, 48–49).

Roger Penrose is Descartes nevével rémisztgeti az MI kutatóit az algoritmusokból eredeztetett értelem elképzelése miatt, ami szerinte együtt jár a „dualizmussal” (az értelem–test probléma MI változatával), amelybe – szándékukkal ellentétben – azzal esnek, hogy az algoritmusoknak fizikai megtestesüléstől független létezését tulajdonítanak. Akár pusztán az emberi értelemben léteznek ezek a fogalmak, akár valamilyen fizikai jelben, amely azonban csak értelmezés révén válik algoritmussá, mindenképpen előzetesen létező értelemre van szükség, s így „a kör bezárulna” (Penrose 1993. 37, 457).

A számítógép elemi lépéseire gondolva azonban nyugodtan élhetünk az utóbbi feltevések bármelyikével a körbenforgás veszélye nélkül, mert ezek a legaltalanosabb, „univerzális” értelemfogalomra vonatkoznak, nem arra a speciálisra, amely „egy elegendően összetett algoritmus megtestesülésén keresztül találja meg létezését” (Penrose 1993. 457). Lehet, hogy ezt célozza az MI-kutatás, most viszont csak a „gyűjtőszerkezettel” foglalkozunk – kétségkívül az egyszerű alkatrészekre bontva és későbbre halasztva a lőtávolság korlátainak kérdését.

A matematikailag rendben lévő, azaz „(le)futó” program persze a mi szempontunkból lehet igencsak rossz, ha nem azt csinálja, amit, vagy nem úgy, ahogy szeretnénk, esetleg éppen szerencsétlen módon nem áll le. Kérdés ezért, *hogyan jutunk* az algoritmusokban is megtestesülő *clara et distincta* belátásokhoz. Már maga a híres *módszer* fogalma arra érdemesíti Descartes-ot, hogy ő legyen „az MI nagyatyja” és Hobbes inkább a számítógépé (vö. Haugeland 1989. 23); mert más az a kérdés, illetve válasz, hogy mi a gondolkodás? – számolás (Hobbes), és más az, hogy mit kell vele tennünk? – a „rend” szerint vezetni (Descartes 1996. 15, 28–29, 31–32, 46).

Akár ősatyát is mondhatunk, amennyiben Descartes nem pusztán az emberi értelmet utánzó vagy akár azt egy másfajta anyagba másoló, hanem magát az értelmet létrehozó „mesterség” úttörője, a *reflexióé*. Schmal Dániel nemrég megjelent könyve hívja fel a figyelmünket arra a konstrukciós tevékenységre, amellyel Descartes a kísérleti körülmények megtervezéséhez hasonlóan teremti meg a *res cogitans*szal kapcsolatos reflexiók és az elme elkülönített fogalma számára a filozófiai vizsgálódás terepét. Ugyanis „a *cogito* kezdőpontja nem adódik természetes módon az elmélkedő számára”, gondolkodásunk „az elmélkedések artificiális közegében mutatja meg saját természetét, a *cogito* tehát ebben az értelemben mesterségesen ’előállított’ kezdet” (Schmal 2012. 12–13, 276–279). „A mesterséges értelem” tehát nemcsak imitált vagy akár esetlegesen létrehozott értelmet jelent, hanem a descartes-i elme képességeként az értelem szükséges tulajdonságát fejezi ki: *az értelem – mesterséges*. A reflexió mesterséges eljárása ugyanakkor semmiféle tudományos magyarázatot nem nyújt az értelemről.

## 2. Az eredendő szintézis

Hasonló választ kaphatunk a másik – terminusa révén – megszólított vagy legalább megidézett filozófustól is. Kant az általában vett értelem (avagy a gondolkodás) képessége analízisének vezérfonalán jut el az értelmi kategóriák felfedéséhez, Descartes módszerét és Locke kutatásait követve, amelyek során az elme saját műveleteire reflektálva szerzi ezek ideáit (Kant 2004a. 116, 118, 127; Locke 2003. 108). Még általánosabban pedig valóban „a szubjektum öntevékenységének aktusa” az értelem „legfelső pontja” (Kant 2004a. 144–146). Ám Kant kifejezetten és gyakran figyelmeztet, hogy ezt *logikai funkcióként* kell érteni, és a „transzcendentális én” nem valami „magasabbra” utal, hanem a minden tapasztalatnak is alapul szolgáló öntudatra. Az elme mélyén cselekvő „Én” (a tudomány számára talán nevetséges) *fogalma* alkotja alapját törvénynek és társadalomnak, erkölcsnek és felelősségnek, de még a problémamegoldás értelmének is – nemcsak a transzcendentálfilozófia, hanem a mesterséges intelligencia egyik atyja, nevezetesen Marvin Minsky szerint is (Minsky 1985).

Mármost úgy látszik, ilyen „eredendő szintetikus egységre” a számítógép architektúrájának is szüksége van. Minden egyes művelet végrehajtásával „együtt járhatna” a ’dolgozom’ vagy a ’vagyok’ „képzete” (vö. Kant 2004a. 145), sőt, ha úgy tetszik, együtt is jár az órajel formájában. Arra nincs szükség, hogy a gép kinyilvánítsa ezt a készenlétet, ahogy a ’gondolkodomnak’ sem kell ténylegesen kísérenie minden képzetemet. De a gép konstruktőre és programozója csak ezáltal tudja uralni teremtményének tevékenységét. Azzal, hogy a gondolkodás egyszerű és meghatározott műveletekké analizálásának elvét Dreyfus „aszszociacionista feltevésnek” nevezi és Descartes-ét Lucretius és Hume nevével együtt ahhoz kapcsolja (Dreyfus 1965. 48), már eleve hallgatólagosan, de téveesen dönt az itt szóban forgó *szintézis* módjáról is. A Turing vagy a Neumann-féle gép elemi műveleteit nem a véletlen vagy valami „gyengéd késztetés” (Hume 1976. 34) terelgeti, hanem azok egyenesen az „a priori szintézis” gondolatának megtestesülései. A tudatosság állítólagos misztériuma valóban nem „jön létre” a gépben, hanem éppenséggel ez az, ami működteti azt logikailag<sup>7</sup> – mint *funkció*, mint a sokaság egysége a legelemibb szinten.

Ettől az elemi egységek nem feltétlenül „buták”. Egy hosszú és mély számo-lás eredményét összegző „zseniális” vezéráldozat meglépésére vonatkozó végső utasítás egyszerű részei összesűrítve tartalmazzák a keresés által létrehozott, az adott állásra vonatkozó, valamint már a keresésben is felhasznált általános tudást. Megoldítva: ez utóbbi „kontextus” teszi a végső műveleteket „nagygyá”. Ebben az értelemben a számítógép igenis „tűri” a *kétértelműséget* (vö. Dreyfus 1965. 30. skk); már pusztán a változók használatával, de a tárolt programnak is alapja, hogy ugyanaz, ami utasítás tárgyaként adat, az maga is utasítás lehet. Mindez arra is rávilágít, hogy ha az itt mozgásba hozott adatokat feleltetjük meg a képzelet és az emlékezet tartalmainak, akkor jó (mozgó)képet kapunk a teljesen transzparens „kartezianus” elméről, ahol a pillanatnyilag tudattalan tartalmak biztosan elérhetők, sőt valamikor ténylegesen tudatosak voltak (Lásd Locke 2003. 39, 158).

Mégis nehéz elfogadni, hogy a számítógép gondolkodik, különösen a tudatossághoz kapcsolódó *érzés* miatt. Ezt a gépről nem feltételezzük, itt a legélesebben elválni látszik a szimuláció a valóságtól. Ráadásul ha a gépnek értelmet, öntudatot, érzést tulajdonítunk, akkor úgy vélhetjük, használatukkal erkölcsi felelősséget veszünk magunkra és olyan problémákkal szembesülünk, „mint amilyenek más emberekkel vagy állatokkal való kapcsolatainkban fellépnek” (Penrose 1993. 24).

Turing szerint viszont mi emberek is csak „a gyakorlatban udvariasan megállapodunk” abban, hogy mindenki gondolkodik, illetve érez (Turing 1950. 139). A kartezianus „tökéletesség” ennek a megállapodásnak a motivációira is fényt vet. Azok a képzetek, amelyeknél a tiszta értelmi ismeretek korlátaikkal talál-

<sup>7</sup> Ettől teljesen független az „erősítés” fizikai szempontja: az inger vagy az ingerek kombinálásának logikai tevékenysége „csak irányítja, vezérli” a forrás energiáját, adja azt elektromos áram vagy a neuron anyagcseréje (Neumann 2003. 174, 290–291, 310).

koznak, *gyakorlati szempontból* ítéltetnek meg mint hasznosak, boldogságunkat szolgálók vagy erényesek. Ilyenek először is az érzetek és az érzések, melyek végső soron a tárgyak hasznosságáról, illetve gondolataink örömteli voltáról tudósítanak (Descartes 1996. 72; Locke 2003. 134–135, 608–609). Ilyen Kantnál a szép feletti tetszés is, melyhez intellektuális és morális érdek fűződik, kiváltképpen pedig az erkölcsi törvény tudatából fakadó tisztelet érzése, amely a morális életvitel és nevelés legfőbb mozgatórugója (Kant 1997. 221–227; Kant 2004b. 89. skk). Ezeknek az érzéseknek és tudatállapotoknak tehát ismét *funkciójuk* van, amely semmiképpen sem abban áll, hogy sirassuk őket a számítógép térhódítása miatt, vagy pedig mégoly meggyőző gondolatkísérletekbe, netán még nem létező, de bizonyára szilárd (mert fizikai) elméletekbe menekítsük őket. John McCarthy joggal kérdőjelezi meg Penrose idézett könyvének *Epilógusát*, mely szerint a szuperszámítógép nem értené a „mit érez?” kérdését. Egy jövőbeli számítógép bizonyára fog tudni válaszolni azáltal, hogy képes lesz megfigyelni saját működési folyamatát. Pillanatnyi állapotáról és készülségéről kellene számot adnia, amely csak akkor hasonlítana a mi érzéseink leírásához, ha ennek programozását valamiért előnyösnek tartanák az alkotói (McCarthy 1990).

Ugyanezen az alapon Marvin Minsky elképzelhetőnek tartja, hogy gépek tudatosabbak legyenek nálunk és „gazdagabb mentális életük” legyen, különösen ha éppen ez teszi még okosabbá őket, hiszen nincs akadály, hogy erre „huzalozzuk” őket. Ezen a módon az MI hozzájárul ahhoz is, hogy többet tudjunk meg saját elménkről, a gondolkodásról, az érzésről (Minsky 1985). Herbert Simon egyenesen odáig megy, hogy az új találmány, a számítógép segítségével megtanulunk engedelmeskedni az „Ismerd meg önmagad!” régi parancsának. (Simon 1980. 6268). Az MI kutatói a józan ész okoskodásaira is gondolnak ezzel, egyes problémamegoldó „szakértők” alapelveire pedig bátran hivatkozhatnak is – ezt láttuk a sakknál. Most pedig azzal döbönt meg bennünket ez a fajta gép, hogy egyenesen szókratészi vagy sztoikus *erényeket* mutat. „Érzései” teljességgel reflektáltak és uraltak, hiszen eleve gondolatok – mi gondoltuk beléjük azokat.<sup>8</sup>

### III. KÖVETKEZMÉNYEK

A gondolkodó géppel kapcsolatos félelmeknek és bizakodásoknak egyaránt alapja az az elképzelés, hogy ha a gondolatot algoritmus hozza létre, akkor az mintegy felbomlik egyszerű, primitív összetevőkre és elveszíti eredeti méltóságát, avagy eredeti durvaságát – kinek mit. A felidézett klasszikus szerzők szerint

<sup>8</sup> Mind a sakkprogramok, mind a számítógép esetén a *létrehozásukban* szerepet játszó gondolatokról beszéltek, nem arról a jelenségről, amikor produktumaikba gondolunk, látunk, érzünk bele valamit.

viszont *az egyszerű gondolat az eredeti*, és ebben az alapgondolatban a legközvetlenebb szövetségesre lelhetnénk a számítógépben. Ahogy a sakkprogramban saját gondolatainkat, a vezérlőegység működésére reflektálva öntudatunk „eredendő szintézisét” ismerhetjük fel, ismét csak szinte tökéletes „kivitelben”. A simoni „technikai” értelemben nyugodtan mondhatjuk, hogy a gépnek öntudata van, érez, gondolkodik és lehet saját működésének tárgya. Ha ezek az „elmeállapotok” valami *funkciót* látnak el bennük, akkor miért tagadnánk meg tőlük ezeket a szavakat, amikor saját magunkra vonatkozóan is használjuk őket *ugyanebben* az értelemben is?

A kényes kérdéssel akkor találjuk magunkat szembe, ha egyes tudatállapotoknak nem funkciót, hanem magában való és csak az emberre jellemző értéket tulajdonítunk. Az említett rendszeralkotó filozófusok még kényesebb válasza szerint viszont nem az elme valami sajátos, magának való része adhat ilyen értéket, hanem az, ha a platóni lélekállam mintájára minden képesség *teszi a dolgát*, kimondottan az emberre jellemző pedig csak az lehet közülük, amelyik *irányítani* képes a többit. A gépek általában is teljesítik az első feltételt, a számítógép pedig a másikat is egészen jól – a részletek bonyolultsága tekintetében a tervezője számára is követhetetlenül. Így „még furcsább” az a kérdés, melyet Minsky az önmagukat fejlesztő gépek lehetősége kapcsán vet fel: „építsünk-e olyan gépeket, amelyek valamiképp 'jobbak' lehetnek nálunk?” (Minsky 1985). Mert nem is csak a specializált szakértő jelentésében beszélhetünk „jobbról”, hanem *morális* értelemben, amennyiben a gép mentes azoktól a testi gyengeségektől, amelyek miatt a hagyományos erkölcsi felfogás szerint gyakran hiába „látjuk a jót”, miközben a 'jó' *eszméje* a racionalista etikák szerint megfogalmazható, és megvalósítása kifejezetten épít a rendre, módszerre, szabályokra.

Itt azonban semmiképp sem hallgathatunk a legátfogóbb és legsúlyosabbnak ható ellenvetésről. A gépről mindvégig lényegében mint az emberi cselekvés eszközéről beszéltem, márpedig a technikának ez az „instrumentális és antropológiai meghatározása” Heidegger mellett (vagy az ő nyomán) sokan mások szerint is elégtelen. Sokkal inkább szokás a korábban megfogalmazott gondolat fordítottjáról beszélni: a gépek formálnak minket a saját képükre és hasonlóságukra. A modern technika „által a valóságos többé vagy kevésbé felfogható módon mindenütt állománnyá válik”, és végül már az embert „magát is állományszámba lehet csak venni. Másfelől éppen az ily módon fenyegetett ember terpeszkedik el a földön annak uraként. Ezáltal úrhodik el annak a látszata, hogy minden, ami föllelhető, annyiban áll csak fenn, amennyiben emberi csinálmány” (Heidegger 2004. 112, 122, 124, 126). „Minden funkcionál. Éppen az az ijesztő, hogy funkcionál, és a funkcionálás egyre további funkcionálásra törekszik, és hogy a technika egyre jobban elszakítja az embert a földtől és gyökértelessé teszi” – mondja Heidegger a vele készült nevezetes interjúban. Arra a kérdésre pedig, hogy mi tölti be most a „halálán lévő” filozófia helyét, amely-

től mindig segítséget várhattunk a világ folyamatainak megértéséhez és döntéseinkhez, kurtán így felel: „A kibernetika” (Heidegger 1995. 12–13).

Igen, pontosan ezzel áll szemben az az emberközpontú filozófia, amely szerint a technika „annak *művészetéhez* tartozik, hogy létrehozzunk valamit, amiről azt akarjuk, hogy legyen”, és amely nem riadhat vissza attól, hogy az embert állományszámba veszik, ha egyszer a „közönséges emberi értelem” maximájának tartja, hogy *mindenki más helyében* gondolkodjunk. Ugyanezen maximák közé tartozik, hogy *önállóan* gondolkodjunk, „az észnek sohasem szabad passzívnak lennie”, mert a passzivitásra való hajlam vezet az előítéletek legnagyobbikához: „ha valaki úgy jeleníti meg a természetet, mint amely nincs alávetve azoknak a szabályoknak, amelyeket az értelem a maga lényegi törvényével a természet alapjául tesz meg” (Kant 1997. 21, 219). Az értelem törvényeivel összhangban levő természethez illeszkedik az *azért* gondolkodó, mert ember *kigondolta* gép önálló gondolata is, minden funkciójával együtt. E funkciók egyike lehetne az a nem szándékolt, sőt váratlan, de legalábbis a számítógépek mindegyikénél közös funkció, hogy az embert segítse visszatalálni gyökereihez: *értelméhez* és a maga funkciójára vagy *posztjára és rendeltetésére* vonatkozó kérdéshez. Ugyanez áll magára a filozófiára is, hiszen pontosan „kibernetikának” kellene lennie, a *kormányzás mesterségének* – legalábbis annak az *Államnak* a törvényei szerint, amely „seholy a világon nincs” (Platón 1984. 648).

## IRODALOM

- Berliner, Hans 1981. An Examination of Brute Force Intelligence. *International Joint Conferences on Artificial Intelligence* 81. 581–587.
- Campbell, Murray – Joe Hoane – Feng-hsiung Hsu 2002. Deep Blue. *Artificial Intelligence* 134/1–2. 57–83. <<http://www.math-info.univ-paris5.fr/%7Eebouzy/Doc/PJR/DeepBlue.pdf>>, utolsó hozzáférés: 2013. augusztus 31.
- Dennett, Daniel C. 1998. *Darwin veszélyes ideája*. Ford. Kampis György – Kavetzky Péter. Budapest, Typotex.
- Descartes, René 1980. *Válogatott filozófiai művek*. Ford. Szemere Samu. Budapest, Akadémiai Kiadó.
- Descartes, René 1992. *Értekezés a módszerről*. Ford. Szemere Samu – Boros Gábor. Dabas, IKON Kiadó.
- Descartes, René 1994. *Elmélkedések az első filozófiáról*. Ford. Boros Gábor. Budapest, Atlantisz.
- Descartes, René 1996. *A filozófia alapelvei*. Ford. Dékány András. Budapest, Osiris.
- Donskoy, Mikhail – Jonathan Schaeffer 1990. Perspectives on Falling from Grace. In Tony Marsland – Jonathan Schaeffer (szerk.). *Computers, Chess, and Cognition*. New York, Springer-Verlag. <[http://webdocs.cs.ualberta.ca/~jonathan/publications/ai\\_publications/grace.pdf](http://webdocs.cs.ualberta.ca/~jonathan/publications/ai_publications/grace.pdf)>, utolsó hozzáférés: 2013. augusztus 31.
- Dreyfus, Hubert L. 1965. *Alchemy and Artificial Intelligence*. Santa Monica, California, The Rand Corporation.
- Frey, Peter W. – Larry Atkin 1978. Creating a Chess Player. An Essay on Human and Computer Chess Skill. *BYTE Magazine* 3/10. 182–191.

- Guid, Matej 2010. *Search and Knowledge for Human and Machine Problem Solving*. Ph.D. thesis, University of Ljubljana. <[http://eprints.fri.uni-lj.si/1113/1/Matej\\_Guid.disertacija.pdf](http://eprints.fri.uni-lj.si/1113/1/Matej_Guid.disertacija.pdf)>, utolsó hozzáférés: 2013. augusztus 31.
- Haugeland, John 1989. *Artificial Intelligence: The Very Idea*. Cambridge, MA – London, MIT Press.
- Heidegger, Martin 2004. Kérdés a technika nyomán. Ford. Geréby György. In Tillmann J. A. (szerk.) *A későújkor józansága II*. Budapest, Göncöl Kiadó. 111–133.
- Heidegger, Martin 1995. „Már csak egy Isten menthet meg bennünket...” (A *Der Spiegel* interjúja Heideggerrel 1966. szeptember 23-án). Ford. Krémer Sándor. In Vajda Mihály *Jegyzetek és szövegek Martin Heidegger Bevezetés a metafizikába* című művéhez. Dabas, IKON Kiadó. 7–15.
- Hendriks, Willy 2012. *Move First, Think Later. Sense and Nonsense in Improving Your Chess*. Alkmaar, New In Chess.
- Horst, Steven 2009. The Computational Theory of Mind <<http://plato.stanford.edu/entries/computational-mind/>>, utolsó hozzáférés: 2013. augusztus 31.
- Hsu, Feng-hsiung 1991. Designing an almost „knowledge-free” chess machine: „Expert inputs” are sometimes harmful. In *The Role of Chess in Artificial Intelligence Research. International Joint Conferences on Artificial Intelligence*. 1991. <<http://dli.iit.ac.in/ijcai/IJCAI-91-VOL1/PDF/084.pdf>>, utolsó hozzáférés: 2013. augusztus 31.
- Hsu, Feng-hsiung 1999. IBM’s Deep Blue Chess Grandmaster Chips. *IEEE Micro* 19/2. 70–81.
- Hsu, Feng-hsiung 2005. Oral History of Feng-Hsiung Hsu. Interviewed by: Dag Spicer. <[http://archive.computerhistory.org/projects/chess/related\\_materials/oral-history/hsu.oral\\_history.2005.102644995/hsu.oral\\_history\\_transcript.2005.102644995.pdf](http://archive.computerhistory.org/projects/chess/related_materials/oral-history/hsu.oral_history.2005.102644995/hsu.oral_history_transcript.2005.102644995.pdf)>, utolsó hozzáférés: 2013. augusztus 31.
- Hume, David 1976. *Értekezés az emberi természetről*. Ford. Bence György. Budapest, Gondolat Kiadó.
- Kant, Immanuel 2004a. *A tiszta ész kritikája*. Ford. Kis János. Budapest, Atlantisz.
- Kant, Immanuel 2004b. *A gyakorlati ész kritikája*. Ford. Papp Zoltán. Budapest, Osiris.
- Kant, Immanuel 1997. *Az ítélőerő kritikája*. Ford. Papp Zoltán. Szeged, Ictus.
- Locke, John 2003: *Értekezés az emberi értelemről*. Ford. Vassányi Miklós – Csordás Dávid, Budapest, Osiris.
- Marsland, Tony – Yngvi Björnsson 1997. From MiniMax to Manhattan. In *Deep Blue Versus Kasparov: The Significance for Artificial Intelligence*. AAAI Workshop. 31–36. <<http://www.ru.is/faculty/yngvi/pdf/MarslandB97.pdf>>, utolsó hozzáférés: 2013. augusztus 31.
- McCarthy, John 1997. AI as Sport. <<http://www-formal.stanford.edu/jmc/newborn/>>, utolsó hozzáférés: 2013. augusztus 31.
- McCarthy, John 2006. Human-level AI is harder than it seemed in 1955. <<http://www-formal.stanford.edu/jmc/slides/wrong/wrong-sli/wrong-sli.html>>, utolsó hozzáférés: 2013. augusztus 31.
- McCarthy, John 1990. Review of *The Emperor’s New Mind* by Roger Penrose. In *Bulletin of the American Mathematical Society* Volume 23, Number 2, October 1990. 606–616. <<http://www-formal.stanford.edu/jmc/reviews/penrose1/penrose1.html>>, utolsó hozzáférés: 2013. augusztus 31.
- Mérő László 2001. *Új észjárások*. Budapest, Tericum Kiadó.
- Minsky, Marvin 1985. Why people think computers can’t. In Denis P. Donnelly (szerk.) *The Computer Culture*. Cranbury NJ, Fairleigh Dickinson University Press. 1985. utolsó hozzáférés: 2013. augusztus 31.

- <<http://web.media.mit.edu/~minsky/papers/ComputersCantThink.txt>>, utolsó hozzáférés: 2013. augusztus 31.
- Neumann János 2003. *Válogatott írásai*. Vál. Ropolyi László. Ford. Tarján Rezsőné, Augusztinovics Mária, Szalai Sándor. Budapest, Typotex.
- Penrose, Roger 1993. *A császár új elméje. Számítógépek, gondolkodás és a fizika törvényei*. Ford. Gálfi László. Budapest, Akadémiai Kiadó.
- Platón 1984. *Állam*. Ford. Szabó Miklós. In *Összes művei II*. Budapest, Európa Könyvkiadó. 5–710.
- Schmal Dániel 2012. *A kezdet nélküli kezdet. Descartes és a kartezianizmus problémája*. Budapest, Gondolat Kiadó.
- Schröder, Ed 2004. How REBEL Plays Chess. <<http://members.home.nl/matador/Inside%20-Rebel.pdf>>, utolsó hozzáférés: 2013. augusztus 31.
- Shannon, Claude E. 1950. Programming a Computer for Playing Chess. *Philosophical Magazine* Ser. 7. 41/314. 256–275. <[http://archive.computerhistory.org/projects/chess/related\\_materials/text/2-0%20and%202-1.Programming\\_a\\_computer\\_for\\_playing\\_chess.shannon/2-0%20and%202-1.Programming\\_a\\_computer\\_for\\_playing\\_chess.shannon.062303002.pdf](http://archive.computerhistory.org/projects/chess/related_materials/text/2-0%20and%202-1.Programming_a_computer_for_playing_chess.shannon/2-0%20and%202-1.Programming_a_computer_for_playing_chess.shannon.062303002.pdf)>, utolsó hozzáférés: 2013. augusztus 31.
- Simon, Herbert A. 1980. Computers – Non-numerical computation. <[www.pnas.org/content/77/11/6264.full.pdf](http://www.pnas.org/content/77/11/6264.full.pdf)>, utolsó hozzáférés: 2013. augusztus 31.
- Simon, Herbert A. 1982. *A vezetői döntés új tudománya*. Ford. Csiky-Strauss Edit – Dr. Kovács Péter. Budapest, Statisztikai Kiadó Vállalat.
- Simon, Herbert – Jonathan Schaeffer 1992. The Game of Chess. In Robert J. Aumann – Sergiu Hart szerk. *Handbook of Game Theory with Economic Applications*. North Holland, Amsterdam, Elsevier Science Publisher. Vol I. 1–17.
- Turing, Alan M. 1965. Számológépek és gondolkozás. Ford. Tarján Rezsőné. In Szalai Sándor (szerk.) *A kibernetika klasszikusai*. Budapest, Gondolat Kiadó. 120–160.
- Winkler, Franz-Günter – Johannes Fürnkranz 1998. A Hypothesis on the Divergence of AI Research. *ICCA Journal*. 21/1. 3–13. <[http://members.aon.at/fgwinkler/PAPER/AI\\_chess.pdf](http://members.aon.at/fgwinkler/PAPER/AI_chess.pdf)>, utolsó hozzáférés: 2013. augusztus 31.



