

---

*Akadémiai székfoglaló*

---

## DARWIN, SKINNER, TURING ÉS AZ ELME\*

HERNÁD ISTVÁN

University of Southampton  
E-mail: harnad@ecs.soton.ac.uk

*A székfoglaló előadás a darwini eszmerendszer provokatív jellegét elemzi a pszichológiában s tágabban a társadalomtudományban. Az ellenállás Darwinmal szemben jórészt vallási volt. A dolgozat megvizsgálja, hogyan próbált a darwinista gondolkodás, valamint az algoritmikus magyarázatok Skinner és Turing munkáiban egyre világosabb magyarázatot adni a viselkedésre. A dolgozatban elemzett Darwin/Skinner/Turing-típusú magyarázat azonban mindig vak, viselkedési magyarázat. Megmagyarázza annak a mögöttes oksági mechanizmusát, hogy mit teszünk. Nem magyarázza meg azonban érzéseinket. Ebben az értelemben a vallásos gondolkodók felvetette kérdéskör még mindig velünk van.*

Kulcsszavak: *Darwinizmus; operáns tanulás; algoritmus; érzés; vallás*

Darwin abban különbözik Newtontól és Einsteintől, hogy elméletének megértéséhez nincs szükség különösebben elmélyült vagy kifinomult gondolkodásra, s talán az elmélet megalkotása maga sem igényelt ilyet. A newtoni mechanikával és az Einstein-féle relativitáselmélettel ellentétben bármely iskolásnak könnyen elmagyarázható, hogy az élőlények darwini értelemben vett túlélő és szaporodó gépek. Felépítésük véletlenek sorozatának és e véletlenek következményeinek összegződéseként alakult ki. Az élőlények testének szerkezetét meghatározó genetikai állományban véletlenszerű változások következnek be, s ha e változások hozzásegítik hordozóikat ahhoz, hogy sikeresebbek legyenek a túlélésben és a szaporodásban, mint őseik vagy ellenfeleik, akkor a változások szükségképpen továbbadódnak, s így elterjedtebbek lesznek a következő generációkban. Ami sikeresebben él túl és szaporodik, az sikeresebben túlél és szaporodik. Ez a tautológikus erő alakított ki bennünket.

\* Székfoglaló előadás a Magyar Tudományos Akadémián, 2002. május 27-én. Fordította: Gervain Judit

A darwini elképzelés tehát nem nagyfokú intellektuális komplexitása miatt fontos, nem is széleskörű érvényessége miatt. Hiszen míg Newton és Einstein elmélete eddigi tudásunk szerint az egész univerzumra vonatkozik, addig Darwiné mindössze kicsiny bolygónk egy aprócska szeletére. Mivel azonban a „fontosság” emberi mérce, az ember pedig része e kicsiny bolygó azon aprócska szeletének, amelyre Darwin elképzelése érvényes, így az – számunkra legalábbis – mindenképpen van olyan fontos, mint Einsteiné vagy Newtoné. Ezért lehetséges az, hogy bár a darwini modell nagy vonalaiban már a XIX. században elkészült, részleteit még ma, a XXI. században is folyamatosan dolgozzák ki.

A Darwin elméletét övező ellenérzés nem intellektuális természetű. Nem az az oka, hogy Darwint nehéz lenne megérteni. Sokkal inkább az, hogy vannak, akik azt szeretnék, hogy ne legyen igaz. Az ellenérzéseket nagyrészt a vallásosság szülte, hiszen a vallásoknak megvolt a maguk magyarázata az emberiség eredetére. De még ennél is lényegesebb ok volt, hogy az elmét – vagy még inkább a lelket – olyan nem anyagi létezőnek hitték, amely kívül esik a túlélő és szaporodó gépeket formáló, teljességgel materialista darwini erők uralmán.

A vallás és a darwinizmus közötti ellentétet nem boncolgatjuk most tovább. Mindössze röviden megjegyezzük, hogy a vallásnak tulajdonképpen nem volt sok félnivalója, hiszen a darwinizmus végül is nem volt képes teljesebb vagy helytállóbb magyarázatot adni a lélekről. Mint ahogyan látni fogjuk, a darwinizmus éppoly kevésbé bizonyult alkalmasnak a mentális állapotok létezésének, céljának vagy eredetének magyarázatára, mint bármely más fizikai elmélet, például Newtoné vagy Einsteiné.

De mielőtt még rátérnénk arra, amit a darwinizmus nem képes megmagyarázni, tekintsük át, mi az, amit meg tud. Vitathatatlan, hogy a darwinizmus a biológiai struktúrák létezésének, céljának és eredetének legjobb és legteljesebb magyarázata. A testméret és -alkat, a szárnyak, az uszonyok és a lábak egytől egyig úgy érthetőek meg a leginkább, ha feltételezzük, hogy sikeresen járultak hozzá bizonyos élőlények túléléséhez és szaporodásához egy adott környezetben, az ott fennálló feltételek mellett.

A biológiai struktúrák mellett a darwinizmus a biológiai funkciókat is érthetővé teszi, hiszen ez utóbbi elválaszthatatlan az előzőtől. A szárnyak, uszonyok és lábak azért segítik a túlélést és a szaporodást, mert használni lehet őket – repülésre, úszásra és futásra. E viselkedésszerű funkciók egy része veleszületett, előhuzalozottan megtalálható az agyban, mint például a reflexek. De általában nem kifizetődő az evolúció számára, sőt egyáltalán nem is lehetséges, hogy előre kitaláljon minden mozdulatot, amelyre egy élőlénynek élete során esetleg szüksége lehet. Következésképpen a testfelépítéssel ellentétben a viselkedésszerű működések kialakításában az élőlény környezetének is gyakran szerepe van a tanulás folyamata révén. (Valójában az egyedfejlődés során több ponton, „epigenetikus” folyamatok formájában megéri a testfelépítés kialakítását is a környezetre hárítani.)

„Baldwini” evolúciónak nevezzük azt a folyamatot, amelynek során a viselkedésszerű képességek valós időben, valódi élettapasztalatok alapján sajátítódnak el, nem pedig a genomban előre kódolt utasítások szerint alakulnak ki, mint az anatómiai jellemzők. Az élőlény pusztán egyes dolgok korai és gyors elsajátításának

képességét és hajlamát örökli. A kiskacsának nem veleszületett képessége, hogy az anyját felismerje és kövesse: mindössze arra van hajlama, hogy az első mozgó alakot kövesse, amelyet meglát, és megtanulja azt az anyjával és saját fajtájával azonosítani. Így nemcsak kiskorában követi azt, de felnőttként is hasonlóakkal próbál párizni.

A baldwini evolúció, amely bizonyos dolgok célzott elsajátításának hajlamát alakítja ki, nem más, mint egyfajta „előkészített” tanulás, amely általában meghatározott kontextusokra és tárgyakra összpontosít, és a fejlődés során egy úgynevezett kritikus periódusban a legfogékonyabb. A nyelv elsajátításának emberi képessége gyakran emlegetett példa az előkészített tanulásra. Elképzelhető, hogy egy faj nem is képes másmilyen tanulásra, mint kisebb vagy nagyobb mértékben előkészítettre. A „tabula rasa”, melyre bármi felírható, talán csak a filozófusok fantáziájában létezik. Mindazonáltal úgy tűnik, hogy az ember általános, nem specifikus tanulásra is képes, s más fajok is rendelkeznek hasonló készséggel. Így tehát Darwin mellett Skinner is belép a képbe.

Skinner a nevéhez fűződő operáns tanulást gyakran hasonlította az evolúcióhoz. Akárcsak a darwini evolúció esetén, ahol is adaptív „következményeik”, azaz túlélési és szaporodási sikerességük alapján („természetes”) szelekció válogat a véletlenszerű genetikai változatok között, a skinneri tanulás a véletlenszerű viselkedésváltozatok között szelektál, azok kedvező, azaz „megerősítést” kiváltó „következményei” szerint. Az élőlények agyát a darwini evolúció úgy alakítja ki, hogy ami adaptív az élőlény számára, az jutalomnak *éreződjék*.

Egyúttal ez a találkozási pontja is a darwini pszichológia két „mechanizmusának”. Az egyiket, a „távoli” vagy „eredeti” mechanizmust azok a körülmények hozták létre, amelyek az élőlény őseit túlélni és szaporodni segítették annak eredeti környezetében. A másik, úgynevezett „közeli” mechanizmus ezzel szemben azt befolyásolja, hogyan viselkedik élete során mai környezetében az élőlény. A távoli evolúció alakítja tehát ki mindazt, ami miatt ha kiskacsa vagyok, hajlamos leszek az általam először megpillantott mozgó tárgyat követni, ha pedig embergyermek, akkor a hallott szavakat utánozni, a cukrot finomnak, a kinint viszont rosszízűnek érezni. Ami miatt viszont kiskacsaként az anyámat követem, emberként az anyanyelvemet beszélem, megtanulok megdolgozni az ételért és szabályoknak engedelmeskedni a pénzbüntetés vagy a börtön elkerülése végett, az a közeli tanulás és tapasztalat. A távoli mechanizmusok a genotípus evolúciós idejében működnek, a közeli mechanizmusok ezzel szemben a fenotípus életidejében.

Ezzel azonban két probléma is van. Egyrészt az, hogy míg a darwini evolúció (Mendel, Watson és Crick segítségével) többé-kevésbé definiálni tudta távoli mechanizmusait, addig a skinneri tanulás közeli mechanizmusa nem ismert. De ez talán nem is lenne olyan nagy gond, hiszen számtalan feltételezett jelölt van, a neurális hálóktól kezdve a statisztikai és gépi tanuláson át egészen a távoli mechanizmusokból átvett genetikai algoritmusokig.

De még mindig ott van a második probléma, miszerint a darwini evolúciónak előbb (kedvező adaptív következményeik miatt) bizonyos dolgokat a jutalom *érzéssel* kell felruházni, hogy azután azok a skinneri tanulás folyamatában közeli megerősítőkké válhassanak. Mindazonáltal a darwini evolúció szigorú értelemben

véve csak azt tudja megmagyarázni, miért alakítunk ki hajlamot a cukorevésre – leegyszerűsítve azért, hogy vércukorszintünket fenntartsuk –, azt viszont már nem, miért is *érezné* ezt jutalomnak (ebben az esetben jóízűnek) az élőlény. Valójában Darwin semmit sem tud mondani arról, miért érződik egyáltalán bármi is valamilyennek.

Ezt másként is megfogalmazhatjuk. Ahhoz, hogy a skinneri megerősítés működni tudjon, egyáltalán nincs szükség érző élőlényekre. A jutalom egyszerűen az, ami arra készíti az élőlényt, hogy még egyszer megtegye a jutalmat kiváltó cselekvést. Lehet olyan gépet tervezni, amely skinneri módon tanul. Ez az a pont, ahol Turing társául szegődik Darwinnak és Skinnernek az elme megértésében.

Turing nem biológiai, hanem mérnöki szempontból közelített az elméhez. Hogyan lehet intelligens gépeket készíteni – olyanokat, amelyek gondolkodnak? Ez természetesen felveti azt a kérdést, mi az intelligencia és a gondolkodás. Válaszként Turing egy elég természetesen adódó tesztet javasolt, amely „Turing-teszt” néven vált ismertté. A próba két feltételt foglal magába. Nevezük „jelöltnek” azt a szóban forgó rendszert, melyről el akarjuk dönteni, hogy intelligens-e. Első feltételként a jelöltnek meg kell tudnia mindazt *tenni*, amit az intelligens, gondolkodó rendszerek tudnak. Ha nem, megbukott. Ha ezt sikerül teljesítenie, akkor lép életbe a második feltétel, nevezetesen az, hogy a jelölt mindent az emberekkel azonos módon, tőlük olyannyira megkülönböztethetetlenül tegyen, hogy ne bírjuk észrevenni a különbséget.

A második feltételt „Turing-megkülönböztethetlenségnek” nevezzük, és tulajdonképpen viselkedésbeli megkülönböztethetlenséget jelent. Turing megfigyelése abban állt, hogy bár előre nem tudjuk megmondani, mi is az intelligencia, a gondolkodás vagy az elme, társas „tudatolvasó” képességünk miatt nagyon ügyesen következtetünk mások mentális állapotaira – természetesen a viselkedésük alapján. Nincs itt szó semmiféle telepátiairól. Könnyűszerrel meg tudjuk ítélni, hogy egy kő, folyó, hegy, autó vagy magnó *nem* rendelkezik elmével, vagyis velünk ellentétben nem gondolkodik és nem viselkedik intelligensen.

Vegyük észre, hogy az előbb megnevezett jelöltek között, amelyekről elég biztosan meg tudjuk állapítani, hogy nincs elméjük, nem szerepel egyetlen élőlény sem. Bár az egyedüli jelöltek, amelyekről elmeolvasó képességünk teljes bizonyossággal meg tudja állapítani, hogy rendelkeznek elmével, azok épp saját fajtársaink, legközelebbi rokonaink viselkedése majdnem ugyanilyen meggyőző, de még emlős, sőt gerinces társainkkal kapcsolatban is elég biztosak vagyunk. Komoly kétségeink az alsóbbrendű gerinctelenek, a mikroorganizmusok és a növények esetén támadnak.

Turingot azonban nem az élőlények intelligenciája, hanem a mesterséges intelligencia érdekelte. Honnan tudhatjuk egy ember alkotta jelöltről, hogy az gondolkodik? Ha Turing filozófus lett volna, nyilván azt állította volna, hogy ezt csak akkor lehetne biztosan tudni, ha mi magunk *lennénk* az adott jelölt. És mivel az egyetlen jelölt, aki az ember lehet, az saját maga, így csak magunkról tudjuk biztosan, hogy van elménk – ahogyan erre Descartes felhívja a figyelmünket. De – s ezt szintén Descartes jegyzi meg – teljes bizonyosságra csak a matematikában tehetünk szert. Minden más téren kevesebb is be kell érnünk. Az intelligencia tekintetében

mi legyen ez a kritérium? Turing javaslata szerint ugyanaz, mint amit egymással szemben is használunk.

A többi embernél viselkedésükből és cselekedeteikből következtetünk arra, hogy van elméjük. Ezenkívül befolyásol még a megjelenésük is, de tudjuk, hogy ez valójában nem lényeges. Az eredeti Turing-teszt éppen ezért a külső megjelenésből eredő hatások kizárása érdekében csak a nyelven alapult. Semmi kétség, a nyelv része az emberi viselkedésnek, még hozzá igen általános és nagy hatású része, még ha nem is az egyetlen, s talán még csak nem is a legalapvetőbb eleme. Mindenesetre Turingnak valószínűleg igaza van abban, hogy ha olyan jelöltet tudnánk létrehozni, amely verbális képességeit tekintve megkülönböztethetetlen lenne tőlünk – mondjuk egy életen át tudnánk vele e-mailezni, anélkül, hogy valaha is felmerülne bennünk a gyanú: az illető nem – magunkhoz hasonló – emberi levelezőtárs –, nem lenne több okunk kétségbe vonni, hogy a jelöltünk elmével rendelkezik, mint egy hús-vér levelezőtárs esetében.

De az is igaz, hogy a nyelvi képességek nem a semmiből bukkannak elő, hanem más fajoknál is meglévő, nem verbális képességekre, például a szenzomotoros rendszerre és a tanulási képességekre építenek. Ezért, bár Turing eredeti tesztje pusztán nyelvi természetű volt, azóta világossá vált, hogy egy valóság-hű Turing-tesztnak robotikusnak kell lennie.

Tesztjével Turing megadta a közeli mechanizmusok magyarázatának módszerét. Közel mechanizmusaink azok, amelyek segítségével át tudunk menni a Turing-teszten. Következésképpen a mesterségesintelligencia-kutatás – különösen pedig a robotika – nem más, mint a közeli mechanizmusok vizsgálata. Talán nem is olyan különös, hogy a biológia találkozik a mérnöki tervezéssel. A „vak órásmester” metafora, amelyet Richard Dawkins használt az evolúció távoli mechanizmusaira, szintén ebbe az irányba mutat. Túlélő, szaporodó gépek vagyunk. Génjeink a túlélésben és a szaporodásban való sikerességük alapján változnak meg és választódnak ki – automatikusan és vakon. Nem valamilyen elképzelt céllal alkottak meg bennünket. A túlélést és a szaporodást *mi magunk csináljuk*. A vak órásmestert tehát ugyanazok a szempontok vezérlik, mint a Turing-tesztet, a viselkedési képességek.

Mit kell tudnia egy jelöltnek ahhoz, hogy átmenjen a Turing-teszten? Túl általános lenne azt feltenni, hogy „mindent, amit mi tudunk”. Mert mit is tudunk mi? Első ránézésre úgy tűnik, hogy „eligazodunk” a környezetünkben. Valójában azonban – jóllehet az önálló helyváltoztatás a vártnál nagyobb fejtörést okozott a robotikának – igazi kihívást a környezetünkben lévő tárgyakon végzett *cselekedeteink* jelentenek. Mint ahogyan azt sem elegendő egyszerűen kijelenteni, hogy túlélünk és szaporodunk, hiszen ha ez pusztán a táplálkozásra és a párzásra korlátozódna, akkor is szembe találnánk magunkat a kategorizáció problémájával: *miféle* dolgokat együnk vagy ne együnk meg? (A párzás hasonló problémájától most eltekintek.)

Ahogyan William James költőien megfogalmazta, az újszülött tapasztalati világa „vibráló, zúgó zűrzavar” (ATKINSON et al., 1994). Tegyük most egy pillanatra félre, milyen érzés e tapasztalati inputot érzékelni, s vonatkoztassuk Turing kérdését arra a mechanizmusra, amely az észleleti bemenetet ehető és nem ehető dolgokra válogatja szét.

Mint már korábban megjegyeztük, Darwintól kapunk némi előzetes segítséget: kategóriáink egy részét „készen kapjuk” az evolúciótól. A béka légydetektora már születéskor be van állítva azokra az inputokra, amelyeket érdemes hosszú nyelvnek kinyújtásával elkapnia. Természetesen az efféle detektorok távoli eredete magyarázatot igényel, s ez feltehetően igencsak hasonló lenne ahhoz, amelyet a sokkal nagyobb számú előkészítetlen, tapasztalati úton szerzett kategória létezésére adunk.

Legalapvetőbb Turing-képességünk tehát talán éppen az, hogy kategóriákat tudunk elsajátítani, azaz képesek vagyunk próba-szerencse alapon a helyes vagy helytelen kategorizáció következményei által adott skinneri visszacsatolás eredményeként megtanulni, mely dolgok fontosak a túlélés szempontjából, és hogy azokat hogyan használjuk fel. Ehhez pedig az kell, hogy feloldjuk valahogyan a jamesi zűrzavart, s rájövünk, mi melyik kategóriába tartozik. Általánosan véve meg kell tanulnunk az input bizonyos jellemzőit észrevenni, másokat figyelmen kívül hagyni. Ez néha könnyű, például ha csak kétfajta dolgot, mondjuk zebrákat és zsiráfokat kell elkülöníteni. Máskor azonban nehéz, például amikor rákos sejtekről vagy a kiscsibék neméről van szó.

Születtek kezdeményezések a tapasztalati kategóriális tanulás modellezésére, de nincs köztük olyan, amely megütné az ember, vagyis a Turing-teszt szintjét. Egyelőre még pusztán játékmódell. Vajon olyan modellt kellene építenünk, amely az összes emberi kategóriát próba-szerencse alapon tanulja meg? Talán nem, hiszen bizonyítékaink vannak arra, hogy mi sem csinálunk mindent így. Most nem a Darwintól „ingyen” kapott, kész kategóriákra célzok. Létezik egy harmadik mód is, amely nemcsak más fajoktól választ el bennünket, de a Turing-teszt eredeti, kizárólag verbális formájára is visszautal. Ha a skinneri próba-szerencse alapú tanulást az azt irányító megerősítéssel együtt a kategóriák robotikus elsajátításának nevezhetjük, akkor e másik módszert hívhatjuk nyelvinek. A kategóriák saját tapasztalat alapján való elsajátítása helyett – amely nehéz, időigényes, így költséges és gyakran veszélyes is – úgy is meg lehet tanulni, hogy mi micsoda, ha valaki, aki már ismeri a kategóriát, szóban jellemzi azt.

Könnyű megfeledkezni arról, mekkora evolúciós előnnyel jár e képesség. A nem emberi fajok sok erőfeszítéssel tanulják meg a kategóriákat, például úgy, hogy fajtársaik viselkedését utánozzák, de arra nem képesek, hogy kifejezetten tanítsák egymást, ahogyan a mi fajunk teszi. Számítógépes szimulációk tanúsága szerint azok a fajok, amelyek képesek kategóriákat verbálisan elsajátítani, néhány generáció alatt túlélési és szaporodási sikerességükben jóval meghaladják a csak robotikusan tanuló fajokat. A játékmódell „nyelve”, ha ezt annak lehet nevezni, nagyon gyorsan fejlődik, és rövid idő alatt a kategóriaelsajátítás uralkodó eszközévé válik.

De nem az egyetlen eszközévé. Nemcsak azért, mert az új kategóriákat definiáló vagy leíró régebbi kategóriáknak legalább egy részét mindenképpen a régi, robotikus tanulás segítségével kellett korábban létrehozni, hanem azért is, mert a nyelvi kategóriákat valószínűleg minden szinten fel kell időnként „frissíteni” közvetlen szenzomotoros tapasztalatokkal. Elméletben lehetséges, hogy a kategórianevek kisszámú, közvetlenül tapasztalati meghatározott készletéből kiindulva a többit pusztán nyelvi meghatározások és magyarázatok alapján is le lehetne vezet-

ni. De még a logikát és a matematikát sem kizárólag szimbolikus úton tanuljuk meg. Mindig szükségünk van konkrét példákön szerzett közvetlen tapasztalatra is.

Hajlamosak vagyunk rögtön elfogadni ezt, saját élményeink alapján magától értetődőnek tartván, hogy az elvont nyelvi leírás, még ha minden szó jelentését pontosan ismerjük is, önmagában nem elég. Hiszen tudásunkat a közvetlen tapasztalat lényegileg gazdagítja. Egy kép látványa mindig többet ér ezer vagy tízezer szónyi leírásnál.

Ez így igaz, de ne feledjük, hogy most kategóriák, azaz dolgok *fajtaínak* elsajátításáról van szó, nem pedig egyedi tárgyakról – a maguk végtelen sok részletével. A fajták megtanulásához elengedhetetlen az egyedi példák irreleváns részleteinek figyelmen kívül hagyása.

Erre persze mondhatnák, hogy még a fajták felismeréséhez, megértéséhez és emlékezetben tartásához is fontos, hogy ne csak jellemzőik elvont leírását és azonosításuk szabályait halljuk, hiszen a kategorizáció éppúgy szenzomotoros képesség, mint a teniszezés, így egyiket sem lehet pusztán nyelvi instrukciók alapján végezni.

Természetesen ez is igaz, de a háttérben olyan feltételezés húzódik meg, amely nagy meggyőző ereje ellenére akár helytelen is lehet. Ez a feltételezés pedig az, hogy nem elegendő elvont leírást hallani valamiről ahhoz, hogy tudjuk, hogyan is néz ki a dolog valójában, ténylegesen látnunk kell. Hiszen végül is az a durva kép, amelyet a verbális leírás fest, szintén a valós észleleti tapasztalaton alapszik.

Ezzel pedig visszajutottunk ahhoz a problémához, amely miatt a vallásos hívők elleneztek és még ma is ellenzik Darwint. A darwini, skinneri, turingi magyarázat ugyanis buta, viselkedéses magyarázat. A *cselekedeteink* mögött meghúzódó oksági mechanizmusokat adja meg. Azt nem, mit *érezünk*. A skinneri jutalomnak nem kell jó *érzést* (egyáltalán: semmilyen sem kell) kiváltania ahhoz, hogy a viselkedés megerősítésének mechanisztikus funkcióját ellássa.

Az tehát, hogy egy kép többet ér ezer szónál, egyáltalán nem azon múlik, hogy a képet *látjuk*. Hiszen a gépnek ezt a bemenetet is ugyanúgy le kell fordítania és fel kell dolgoznia. Már a mai egyszerű, elme nélküli, vak számítógépekkel is demonstrálni tudjuk, mennyiben több egy kép szavak ezreinél. Ugyanez igaz a nyelvi leírásokra, amelyek a közvetlen érzéki bemenetnél kevésbé informatívak. Mindez ugyanúgy igaz vak, elme nélküli zombikra, mint ránk, látó, tudatos emberekre. A köztük lévő különbséget sem Darwin vak órásmestere, sem a Turing-teszt nem képes kimutatni.

Így – bár nem hiszem, hogy okunk lenne vallásos örvendezésre vagy empirikus feltételezések megtételére – nyugodtan elmondható, hogy a darwini–turingi világ teljességgel összeegyeztethető a halhatatlan lélek létezésével.

## IRODALOM

- ATKINSON et al. (1994) *Pszichológia*. Ford. Csibra Gergely. Osiris Kiadó, Budapest
- CANGELOSI, A., HARNAD, S. (2001) The Adaptive Advantage of Symbolic Theft Over Sensorimotor Toil: Grounding Language in Perceptual Categories. *Evolution of Communication* (Special Issue on Grounding)  
<http://cogprints.soton.ac.uk/documents/disk0/00/00/20/36/index.html>
- DAWKINS, R. (1986) *Az önző gén*. Gondolat Kiadó, Budapest
- DAWKINS, R. (1994) *A vak órásmester*. Akadémiai Kiadó, Budapest
- HARNAD, S., ed. (1987) *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press, New York  
<http://cogprints.soton.ac.uk/documents/disk0/00/00/15/71/index.html>
- HERNÁD I. (1996) A szimbólumlehorgonyzás problémája. *Magyar Pszichológiai Szemle*, 32–33, 365–383.  
<http://cogprints.soton.ac.uk/documents/disk0/00/00/06/15/index.html>
- HARNAD, S. (2000) Minds, Machines, and Turing: The Indistinguishability of Indistinguishables. *Journal of Logic, Language, and Information*, 9 (4), 425–445. (special issue on „Alan Turing and Artificial Intelligence”)  
<http://cogprints.soton.ac.uk/documents/disk0/00/00/16/16/index.html>
- HARNAD, S. (2002) Turing Indistinguishability and the Blind Watchmaker. In Fetzer, J. (ed.) *Evolving Consciousness*. 3–18. John Benjamins, Amsterdam  
<http://cogprints.soton.ac.uk/documents/disk0/00/00/16/15/index.html>
- HARNAD, S., STEKLIS, H. D., LANCASTER, J. B., eds (1976) Origins and Evolution of Language and Speech *Annals of the New York Academy of Sciences*. 280.  
<http://cogprints.soton.ac.uk/documents/disk0/00/00/08/66/index.html>
- SKINNER, B. F. (1984) Selection by consequences. *Behavioral and Brain Sciences*, 4, 477–510.
- TURING, A. M. (1950) Computing Machinery and Intelligence. *Mind*, 49, 433–460.  
<http://cogprints.ecs.soton.ac.uk/archive/00000499/>

## DARWIN, SKINNER, TURING AND THE MIND

HARNAD, STEVEN

*This inaugural address at the Hungarian Academy of Sciences analyses the provocative nature of Darwinian thought in psychology and in social science at large. The resistance to Darwin was mainly religious. The paper discusses how Darwinian thought and algorithmic explanation by Skinner and Turing tried to give clearer and clearer explanations for behavior. However, the Darwin/Skinner/Turing explanation analyzed here is all a blind, behavioral explanation. It explains the underlying causal mechanisms of what we can do. But it does not and cannot explain what we feel. In this sense, the challenge emphasized by religious thinkers is still with us.*

Key words: *Darwinism; operant learning; algorithms; feelings; religion*