

TANULMÁNYOK

HERNÁD ISTVÁN

Universite d'Aix Marseille II., Franciaország

NEOKONSTRUKTIVIZMUS: A KOGNITÍV TUDOMÁNYOK EGYSÉGESÍTŐ KORLÁTJA¹

BEVEZETÉS

Mi a kognitív tudomány? A „kogníció” pusztán a gondolkodás szinonimája. Akkor a kognitív tudomány a gondolkodás tudománya lenne? De hát akkor meg mi is a gondolkodás? S mi utal arra, hogy egy külön tudományos diszciplínát hasít ki magának? Talán a kognitív tudomány nem is külön tudományos diszciplína, hanem olyan kutatási terület, amellyel minden olyan szakma foglalkozik, amely érdeklődik a megismerés iránt. Ha egy ilyen interdiszciplináris modellt alkalmazunk arra, hogy mi is a kognitív tudomány, akkor néhány nem túl sikeres előfutárt öröklünk, olyanokat, mint a kibernetika és az általános rendszerelmélet. E dolgozatban az elméletalkotás egy módszertani korlátját — neokonstruktivizmus — fogom javasolni, mint a kognitív tudomány egy másik egységesítési lehetőségét. E korlát megfogalmazását előkészítendő először leírom, mi is a konstruktivizmus.

KONSTRUKTIVIZMUS A MATEMATIKÁBAN

A matematikában van egy mozgalom, melyet hol finitizmusnak, hol intuicionizmusnak, hol konstruktivizmusnak neveznek. A konstruktivistákat nem elégíti ki az ellentmondás segítségével történő igazolás. Számukra

1 Fordította Pléh Csaba. Harnad, S. Neoconstructivism: a unifying theme for the cognitive sciences című dolgozatának jelentősen átdolgozott változata, az eredeti lelőhelye: SIMON, T. and SCHOLES, R. (Eds.), *Language, mind and brain*, Hillsdale: Erlbaum, 1982, 1—11. pp.

például nem elég egy matematikai objektum létének igazolásához az, ha kimutatjuk, hogy ha nem létezne, ez ellentmondáshoz vezetne (ez az úgynevezett apagogikus bizonyítás). Ehelyett azt kérik, hogy mutassanak rá a kérdéses tárgyra (osztenzív bizonyítás), vagy adják meg konstruálásának explicit szabályait. A konstruktivistáknak hasonló sztenderdjeik vannak arra, hogy mi is egy függvény vagy egy eljárás; konstruktív eljárás az, mely határozott kimenetelt eredményez véges számú explicit lépés után (HEYTING, 1971).

Egy olyan függvény gyakran idézett példája, amely nem konstruktíven meghatározott (lásd FRAENKEL és mts., 1973, 266. o.), s amely ezáltal a konstruktivista számára nem elfogadható, az a függvény, melynek értéke 1, ha π decimális kiterjesztése során megjelenik mondjuk az 123456789... számsor, értéke pedig 0, ha ez a számsor sosem jelenik meg. Ezzel a definícióval az a gond, hogy pillanatnyilag nem tudjuk, hogy ez az adott sor vagy bármely más önkényes számsor ténylegesen megjelenik-e a π tízes kibontása során, s így nem tudjuk, hogy valaha is le áll-e ennek a keresése. Az olyan keresésen alapuló kimenetel, mely lehet, hogy sosem áll le, meghatározatlan, s ezért a kérdéses függvény nem konstruktív.

A Turing gépek és rekurzív függvények elméletén keresztül bensőséges kapcsolat van a konstruálhatóság és a komputálhatóság között (lásd DAVIS, 1985; KLEENE, 1969). A komputálhatóságról később még szólni fogok, pillanatnyilag azonban a matematikai konstruktivizmusnak az analógiában később szükségessé váló egyetlen tulajdonsága az, hogy megválaszolható egy olyan explicit kérdés, hogy létezik-e valami, vagy hogy jól definiált-e egy eljárás. Nem lehet átvinni a következetesség vagy ellentmondásmentesség korlátját a matematikából, illetve ennek sajátos szerepét. Amikor ugyanis egy konstruktivista a nem konstruált objektumokat tekintve agnosztikusnak vallja magát, mivel ezek egyedüli létbizonyosága az ellentmondásmentesség és a kizárt harmadik törvénye lenne, azért azt elismeri, hogy az apagogikus bizonyításmentet kétségkívül nem önkényes, s a nem konstruktivista matematikusok formális tevékenysége egyáltalán nem üres.

AZ ALULDETERMINÁLTSÁG ÉS A TUDOMÁNY EGYSÉGESÍTÉSE

A tudománynak is következetesnek kell lennie, nemcsak formális értelemben, hanem abban az értelemben is, hogy nem mond ellent az adatoknak. Sajnos azonban még mindig túl nagy azoknak a dolgoknak az univerzuma, melyek önmagukban következetesek s összhangban vannak a rendelkezésre álló adatokkal. Nem olyan ez, mint a matematikai igazságok világa. A

híres eldönthetlenségi eredmények esetleges kivételével (DAVIS, 1965) a tudomány igazságai „aluldetermináltak”. Az empirikus adatok mindig számos elmélettel összeilleszthetők, úgyhogy további korlátokat, például a gazdaságosságot vagy az általánosságot kell bevezetni, hogy leszűkítsük a lehetőségeket és segítsük a rivális elméletek közti választást. A gazdaságosság (*parsimonia*) azt az elméletet részesíti előnyben, mely minden egyéb azonosan tartása mellett a legkisebb számú paramétert feltételezi. Az általánosság azt az elméletet preferálja, mely minden egyéb azonosan tartása mellett a legnagyobb számú és legsokfélébb adatról ad számot. A gazdaságosságot és az adathalmaz nagyságát elvileg kvantifikálni lehet. A modern komplexitáselmélet (CHAITIN, 1975; RABIN, 1977) információs és kompjutációs elveket adott a gazdaságosság kvantifikálására. Nem ilyen könnyű azonban eldönteni, hogy mik egy elmélet adatterének *határai* — vagyis, hogy mennyire sokrétű kell legyen az elmélet. A tudományban van egy határozott egységesítő vonal, mely egyetlen egységes elmélet segítségével a lehető legnagyobb számú adatot szeretné megmagyarázni, még akkor is, ha szakmákon és diszciplináris határokon vág keresztül. Nehéz tehát egy elméleti program definiálása során előre megadni a határokat. Világos azonban, hogy valamilyen egységesítő témára szükség van, ha másért nem, hát azért, hogy az elméletek egyetértsenek abban, hogy ugyanazzal a problémával foglalkoznak.

A logikai pozitivizmust már nem tartják a tudomány járható filozófiai vagy módszertani alapjának. Ennek fő oka az, hogy lehetetlen teljesen elválasztani egymástól az elméletet és az adatokat. Azt, hogy mit tartunk adatnak s hogyan definiáljuk őket, már nagymértékben megfertőzte elméleti előfeltevés-rendszerünk. Nem világos, hogy milyen mértékig befolyásolja ez a probléma hétköznapi tudományos gyakorlatunkat. A probléma számítógépes analógja az, hogy nem lehet teljesen elválasztani egymástól a számítógépes modellt annak specifikus számítógépes implementációjától, megvalósításától. Mindkét esetben (elmélet/adatok, modell/implementáció) a gyakorlati, pragmatikus következmények hiánya nagyon csökkenti a probléma súlyát. A gyakorlati jelentéktelenség felhangja mellett is igaz az, hogy az adatok elméletterhelt volta egy újabb érv lehet amellett, hogy nem lehet abszolút határokat húzni a diszciplinák között abból a szempontból, hogy milyen adattípusokat tartunk egy diszciplína natív adatainak. Ezért aztán egy tudományos vállalkozás egységesítésének egyetlen megmaradó jelöltje a belső egységesítés lehetősége. Az elmélet szempontjából vett belső egységesítés, szemben az adatokkal.

Viszonylag nem sokkal ezelőttig a pszichológia területe és a biológia is határozottan ambivalens, sőt ellenséges volt az elméletekkel szemben. Úgy gondolták, hogy a „kemény” adatok magukért beszélnek, s minden egyéb csak spekuláció (lásd például SKINNER, 1984a; HARNAD, 1984; SKINNER, 1984b). A behaviorizmus ellenérzése az elmélettel szemben két alapon nyugodott. Az egyik szerintem érvényes, a másik nem. Az évek során ezek szétbogozhatatlanul összekeveredtek, de mégis szét kell választanunk őket. Az érvényes alap az volt, hogy minden animizmust el kell vetnünk, mint tudományosan inadekvátat. Ebbe beletartozott minden olyan adatgyűjtés és elmélet-ellenőrzés elvetése, mely szubjektív introspekciót használt, s minden törekvés arra, hogy a magyarázatot fizikai helyett mentális okokra vezessék vissza. A behaviorizmus helyesen vette észre, hogy a viselkedés, s talán az idegi aktivitás is, a pszichológia egyetlen objektív adatterülete — és hogy a szubjektív élménytartalmak sem adatokat, sem elméletet nem adnak.

Ezt az érvényes elméletellenes pozíciót a *mentalizmus* elvetéseként kell értékelnünk. Mentalizmuson a viselkedés magyarázatának azt az elméletét kell érteni, mely szerint a viselkedésről az ad számot, hogy mi megy végbe az elmében. A behaviorizmus azonban túláltalánosította a mentalizmus elvetését abból a szempontból, hogy kezelhetetlennek tartotta azt, hogy mi is megy végbe a fejünkben. És elfogadhatatlannak deklarálta mindazokat az elméleti következtetéseket, melyek megfigyelhetetlen eseményekre, folyamatokra utalnak. Az internalizmus minden formájának ez a kiiktatása az elmélet elvetését eredményezte, s egy súlyosan inadekvát tudományfilozófián alapult, mely egyebek mellett a fizikát is csak megfigyelhető dolgokra korlátozná, nem tenné lehetővé, hogy túllépjen a kinematikán. E felfogásban a fizikában se lehetnének elektronok, kvarkok, mezők vagy hasonlók. Még azoknak a funkcionális elveknek a kialakulását sem tette volna lehetővé, melyekre ahhoz lett volna szükség, hogy a pszichológiát azzá a viselkedéstudománnyá tegyék, amit éppen a behaviorizmus képzelt el.

Sok szempontból tehát a kognitív tudomány pusztán egy végre elmélettel is felruházott behaviorizmus. Sajnos azonban az internalizmus két formáját (mentális és kraniális) még ma is összemoszák. Együtt azzal az új engedéllyel, hogy feltételezhetünk a fejünkben zajló folyamatokat, egy új mentalizmus is megjelenik, mely túlinterpretálja ezeket a kikövetkeztetett, a fejünkön belüli folyamatokat, s odáig megy el, hogy úgy értelmezi

ezeket, mint amelyek arról szólnak, hogy mi megy végbe az elmében (HARNAD, 1993b).

A pszichológiának egy olyan egységesítő korlátra van szüksége, amely 1) az elméleti lehetőségek világát kezelhető méretűre csökkenti le, és 2) ad némi utalást arra, hogy milyen tudomány is a sajátosan pszichológiai tudomány, és legfontosabbként 3) objektíven meghatározza, mi számít elméletnek, és mi nem e területen.

A NEOKONSTRUKTIVIZMUS A KOGNITÍV TUDOMÁNYBAN

A terminológiai változások néha megtisztítják a levegőt a szubsztanciális változások számára. A „pszichológia” teljesen jó név volt egy tudomány számára, de túl sok ember számára túl sokfélét kezdett jelenteni. A laikusoknak valami olyasmi volt ez, amit „használni” lehet, s a tied lehet, hogy „jó” vagy „rossz”. A klinikusoknak főleg a pszichopatológiával kapcsolódott. A filozófusok számára az volt, ami akkor marad, ha a filozófiából kivonjuk az ontológiát, az ismeretelméletet, az etikát és a logikát. Sok nyelvész számára ez biztosította a lényegében veleszületett nyelvi képesség végső behangolását (lásd HARNAD, 1976). Alsóévesek nemzedékei számára pedig „a viselkedés tudománya” volt, labirintusban szaladgáló patkányok kapcsolódtak hozzá, s a gondolkodás és az érzelem tabu szó volt.

A kognitív tudomány talán elég új ahhoz, hogy saját identitást alakítson ki. Annak az identitásnak a jelöltje, amire a neokonstruktivizmus kritériumot javasolom. A pszichológiában már van valami „konstruktivizmusnak” nevezett dolog. A konstruktivisták azok a perceptuális elméletalkotók (ROCK, 1980; ULLMAN, 1980), akik szembeállítják magukat az azzal a felfogással, melyet GIBSON (1979) képvisel, aki szerint a percepció „direkten” jelenik meg. A gibsoniánusok szerint, mivel az ingerlésben benne van minden olyan információ, mely az adaptív perceptuomotoros tevékenység irányításához *szükséges*, ennek az információnak *elégésnek* is kell lennie a teljesítményhez, s az idegrendszer pusztán passzívan „felveszi” azt. A konstruktivisták tagadják ezt, szerintük az idegrendszernek valamilyen aktív feldolgozást kell végeznie („konstrukció”), hogy adaptív teljesítményt nyújtson. Cikkemnek nem témája, hogy e sajátos vitában hol is van az igazság. Arra mutatnék csak rá, hogy sem a gibsoniánusok, sem a konstruktivisták nem túl „konstruktívak” itt abban az értelemben, ahogyan azt a matematikusok használják. A matematikában a konstrukció eldöntő tevékenység. Nem vitatkozhatunk arról, hogy a pi decimális kiterjesztése során szerepel-e egy meghatározott számsor. Ténylegesen meg

kell mutatnunk ezt, vagy véges módszert kell adnunk arra, hogy hogyan lehet megtalálni.

Mi lehetne ennek a döntő próbának a megfelelője a kognitív tudományban? Van egy viszonylag új konstruktív játék, mely ezt végzi: a digitális számítógép. De miért kell komolyan vennünk egy illékony, holnapra már jelentéktelenné váló technológiát? Hisz nem olyan rég az embereket még órákhoz vagy gőzgépekhez hasonlították? A válasz az, hogy a számítógérek tényleg lehet, hogy jöttek, aztán mennek, a számítási elmélet (KLEENE, 1969) és az információelmélet (SHANNON és WEAVER, 1949) azonban itt maradnak, olyan örökök, mint a többi platonai örök igazság. S a számítógép, mint egy elvont számítási szerkezet (Turing gép) implementációja lesz az az összerakós játék, amit én javasolni szeretnék. Nem az a leegyszerűsítő, s nyilván rossz gondolat ez, mely szerint az emberek IBM-számítógépek, hanem az, hogy ha konstruktív gondolataink vannak arról, hogy milyenek is az emberek, akkor ennek ellenőrzésére a legjobb mód az, ha formalizáljuk őket, s megnézzük, hogy működnek-e egy számítógéppel. Nemcsak a pszichológiai elmélet, hanem minden tudományos elmélet jellemzője az, hogy leírható úgy is, mint input/output viszonyok modellje. A bemenetet, az inputot a peremfeltételek és a kísérleti előkészületek adják, s az elméletnek be kell jósolnia a kimenetet, a kísérleti megfigyeléseket. Elvileg minden fizikai elmélet számítógépesen ellenőrizhető. Amikor kézzel oldunk meg egy differenciál egyenletrendszert s tényleges értékeket helyettesítünk be, az elméletíró egy számítógép szerepét játssza. Az asztrofizika és a statisztikus mechanika bizonyos komplex problémáinál azonban már számítógépet használnak annak meghatározására, hogy milyen jóslások vannak az elméletből valós bemeneti adatok esetén.

Ha a számítógépesen tesztelhető elméletekre történő korlátozódást „neokonstruktivizmusnak” nevezzük (hogy elkülönítsük a matematikai és perceptuális konstruktivizmusoktól), akkor minden fizikai tudomány máris nyilvánvalóan neokonstruktív (többnyire triviálisan, abban az értelemben, hogy a fizika neokonstruktív jellegét bizonyítandó nincs szükségünk valódi számítógépekre, hanem, akár csak a matematikában, papír és ceruza is elég lesz). A neokonstruktivizmus tehát nem képes egyedi identitást adni a kognitív tudománynak, csak arra a kérdésre ad választ, hogy mi számít kognitív elméletnek. Elméleti kritériummal felvértezve azonban, a kognitív tudományokat elég jól jellemzi adattömegük, mely nagyjából úgy írható le, mint minden emberi és állati, valamint ember- és állatszerű teljesítmény és képesség. A kognitív tudományok azonban láthatóan igencsak

használnak az asztrofizikára és a statisztikus fizikára, jobban, mint a fizika más területeire, mivel a szervezetek teljesítménye, eltérően a billiárdgolyók viselkedésétől, a számítógép segítségével nélkül nem magyarázható. A kognitív tudomány legrealistább modellje talán az elméleti mérnöktudomány: annak tudománya, hogy hogyan lehet elérni, hogy szervezetek azt csinálják, amire mi képesek vagyunk, s ezáltal egy funkcionális magyarázat-javaslatot készítünk arra, hogy *hogyan* csináljuk mi.

A KOGNITÍV TUDOMÁNY RIVÁLIS FELFOGÁSAI

A kognitív elmélet neokonstruktivista kritériumát néhány egyéb mai jelölttel is össze lehet vetni.¹ Először is a neokonstruktivizmus csak igen minimálisan *extenzionális* kritérium. Más szóval, nem törekszik arra, hogy listaszerűen felsorolja mindazon példákat, amely kognitívnek számít. Még az sem egészen újdonság, hogy az adatterületet mint organizztikus és organizmusszerű teljesítményt emeljük ki. Egy bizonyos fontos értelemben ugyanis a kognitivistákat továbbra is jellemezhetjük „behavioristaként”, abban az értelemben, hogy elismerik azt, hogy a tényleges adatpontok segítségével mindig csak a szervezet kimeneti és bemeneti jellemzőihez fog-nak eljutni. (Lehet, hogy ebben az értelemben minden tudós „behavioris-ta”.) A kognitivistákat jobban jellemezzük azonban, ha „rekonstruált” behavioristáknak nevezzük őket a következő értelemben.

a) A kognitivisták ambíciója a teljesítményadatokról számot adó neokonstruktív elmélet.

b) A kognitivisták felhasználhatják az introspekciós adatokat, az idegtudományból származó vagy bármilyen más forrásból származó adatokat, ha azok konstruktívan hozzájárulnak magyarázó elméletükhöz, s annak képességéhez a teljesítményadatokról való számot adáshoz. Ezen felül, mivel minden teljesítményről szeretnének számot adni, a kognitivistáknak nincs okuk arra, hogy ne kíséreljék meg modellálni a teljesítmény mögött álló szubjektív fenomenológiát, amit például szóbeli beszámolókból tudnak feltárni, különösen mivel a fenomenológia sikeres szimulációjának nagy valószínűséggel szerepe lesz a teljesítmény modellálásában. DENNETT (1993) ezt a hibrid megközelítést „heterofenomenológiának” nevezte.

1 A neokonstruktivizmussal kapcsolatos eredeti javaslatomtól (HARNAD, 1982a), mint a kognitív elmélet kritériumától teljesen függetlenül JOHNSON-LAIRD (1983) hasonló felvetést tett.

c) A kognitivisták nem korlátozódnak a nyers teljesítmény húzó-taszító dinamikájára: a teljesítmény és a teljesítményre való képesség (a kompetencia) magasabbszintű szabályszerűségei mind tiszteletreméltó adatok.

d) A kognitivisták felismerik, hogy milyen gyenge a behavioristák hagyományos fogalmi arzenálja, gondolva itt olyan fogalmakra, mint megerősítés és asszociáció (lásd HARNAD, 1984) a neokonstruktivista szempontból nézve. Ezek a fogalmak egyszerűen nem elégségesek azoknak a viselkedési képességeknek a generálására, melyeket magyarázni hivatottak. Valójában van is egy affinitás a behavioristák és a gibsoniánusok között abban az értelemben, hogy bár világos, hogy az adaptív teljesítményhez szükséges információt az ingerek, a jutalmak és a büntetések hordozzák, ez a tény önmagában (együtt közvetlen paramétereivel) nem lenne elégséges a teljesítmény „generálására” (ULLMAN 1980).

A kognitívizmus a teljesítmény önmagában megálló, generatív magyarázatához szükséges belső struktúrák és folyamatok feltárására törekszik. Talán erre a neokonstruktív behaviorizmusra gondolt WITTGENSTEIN (1967), amikor azt mondta, hogy a behavioristák a matematikai intuicionistákhoz hasonlítanak.

Összefoglalva tehát, az adatmező organizmikus teljesítményként való körülhatárolása lehet, hogy oly általános, hogy ha szükséges, ebbe beleérthetjük akár a billiárdgolyók interakcióit, vagy a tejszín feloldását egy csésze kávéban. Javaslatunk pusztán az, hogy a valós szervezetekhez hasonló teljesítményeket és teljesítménylehetőségeket kell kutatnunk és megkísérelnünk modellálni.

Felvethető az az ellenvetés, hogy a két lábon járás nem különösebben kognitív teljesítmény, bár nyilván organizmikus teljesítmény. Válaszunk erre az, hogy módszertani választás az, hogy elméletalkotásunkban alulról felfelé, vagy felülről lefelé haladjunk-e, s nyugodtan vehetjük biztos alapnak az alsó szinteket, ha ez semmi veszélyt nem jelent. De a gibsoniánusok számára például elég nehéz volt kihagyni egyenleteikből a helyváltoztatást még akkor is, amikor olyan komplex vizuális alakfelismerési teljesítményekről próbáltak számot adni, melyekről senki sem kérdőjelezné meg, hogy kognitívak-e (GYR, WILLEY és HENRY, 1979). Ezen felül, sokkal nehezebb lenne igazolni azt, hogy a kogníció fenomenális mezejét mondjuk a nyelvre, vagy a nyelvszerű képességekre korlátozzuk, mielőtt bármiféle (neokonstruktív) elképzelésünk lenne arról,

hogy mi is a nyelv és a nyelvszerű teljesítmény, s hol végződnek ezek, s hol kezdődik az alacsonyabbszintű teljesítmény.¹

Akárcsak azokban az egyházakban, ahol szétvált az alsó és a felső papság, kialakulhatnak „felső” és „alsó” kognitivisták iskolák, ahol a felső iskolák úgy tartják, hogy az alsó pusztán a teljesítmény „vegetatív” vonatkozásait modellálja, s egyáltalán nincs relevanciája a kognitív oldalra. Gyakorlati módszertani kérdés, hogy vajon az efféle szakadásokat bátorítanunk kell-e, s hogy vannak-e ezeknek a pusztá ideológiáival szembeállított logikai alapjaik is. Valójában még az organizmikus és az idegrendszeri „teljesítmény” közötti különbség sem biztos, hogy produktív a kognitív tudomány számára, ha ezt túlerőltetjük (HARNAD, 1991, 1994).

Vannak a kognitívnek egymással versengő intenzionális meghatározásai is. Vagyis olyan meghatározások, melyek elkülönítő jegyeken alapulnak. Ezek egy része olyan nem teljesen jól értett fogalmakon alapszik, mint a reprezentáció (PYLYSHYN, 1984) és az intenzionalitás (SEARLE, 1983). A „reprezentáció” valami olyasmi, melynek a jelentését a pszichológusoknak és a filozófusoknak még tisztázniuk kell. Tudjuk, hogy olyan fogalmakkal kapcsolódik, mint a referálás, a leírás és a jelentés, s valahogyan a szavak, képek és szimbólumok, valamint a nekik megfelelő tárgyak közötti viszonyokkal függ össze. Még az is lehet, hogy (akár negatívan is) valami köze van a reprezentáció fogalmához, ahogy azt a művészetben és az irodalomban használják. (Vagyis a hagyományos „tükrözés” fogalmához is!) Kétségkívül olyan fogalomról van szó azonban, melyet a filozófusoknak még nem sikerült mindenkit kielégítő módon körvonalazniuk. Az sem világos, hogy a kognitív tudománynak itt vajon várnia kellene-e a filozófusok megoldására, ahelyett, hogy azt nézné meg, milyen messze tud jutni olyan kevésbé elkötelező fogalmak segítségével, mint a kódolás, belső struktúra és folyamat fogalma. Ha ma a „kognitívat” azonosítanánk a „reprezentációssal”, akkor a vak vezetné a világtalant.

Az „intencionalitás” (vagy „valamire vonatkozás”) egy másik rivális jelölt. Különböző módokon értelmezik. Vagy azt értik rajta, ami közös az olyan aktusokban és az állapotokban, mint „hinni”, „gondolni”, „akarni”, „elvárni”, „tudni” és így tovább, szembeállítva olyan aktusokkal és állapo-

¹ HARNAD, 1989, 1990a, 1993 tárgyalja a modularitás kérdését, s azt, hogy hogyan kell a szimbolikus működéseket nem szimbolikus működésekben „lehorgonyozni”. A főemlősök nyelvi képességeivel kapcsolatos problémák egy része (HARNAD és mások, 1976) jól illusztrálja, hogy milyen problémák merülnek fel, ha külső határvonalakat szeretnénk húzni olyan teljesítmények között, melyeket neokonstruktív módon még nem értünk.

tokkal, mint „sétálni”, „ülni” és „enni”. Ez a közös vonás valahogyan szoros összefüggésben van azzal a ténnyel, hogy (1) az első kategóriába tartozó szavak, aktusok és állapotok valamilyen beépített („intencionális”) tárggyal rendelkeznek. Az ember valamit akar, valamire gondol, valamilyen hisz stb. Fontosnak tartják azt is, hogy (2) ezeknek a tárgyaknak a megváltoztatása („a vacsoracsillagra gondolok” [szemben azzal], hogy „a hajnalcsillagra gondolok”) nem marad szándékaimra nézve igaz, feltéve, hogy itt mindkét esetben ugyanarról a csillagról van szó, de én ezt nem tudom, s az elsőre gondoltam, de nem a másodikra. Míg szokásos koreferenciális helyettesítésük esetén („minden ember halandó” versus „minden *homo sapiens* halandó”) az igazságérték megmarad. Az a hiedelem, hogy (3) az intencionalitás a mentális egyetlen megkülönböztető jegye, oly erőteljessé vált az utóbbi időben, hogy néhány lelkes hívő szóhasználatában az intencionális szó egyenesen teljesen felváltotta a mentálisat (s eközben úgy gondolták, hogy haladást értünk el!). Végül az intencionalitás valami olyasmi (4), amit egy interpretációként rendelhetünk a dolgokhoz (HAUGELAND, 1978) vagy mint egy „álláspontot” vehetünk fel (DENNETT, 1987): „Ez a személy (kutya, gép) úgy viselkedik, mintha gondolkodna, hinne, akarna és így tovább”. Nem világos még, hogy az (1)-(4) kritériumsor mennyire lesz segítségünkre, vagyis mennyire lesz konstruktív a kognitív elméletet illetően.

Az utolsó meghatározásjelölt azon az explicit kijelentésen alapszik, hogy az kognitív, ami komputációs (PYLYSHYN, 1980, 1984): vagyis a gondolkodás igazából a komputáció, a számítás egy formája (DIETRICH, 1990). A jelen cikkben felvonultatott anyag alapján tévesen még arra a következtetésre is lehetne jutni, hogy az utóbbi felfogás összhangban van a neokonstruktivistákéval. Hiszen a komputacionalizmus is egy módszertani elkülönítésen alapszik, s alig vannak extenzionális elkötelezettségei; s ráadásul még neokonstruktív is!

A komputacionalizmus azonban túl korlátozó, s túlzottan részrehajló a megismerés egy speciális elméleti megközelítése irányában, mégpedig a szimbólum manipuláló felfogás irányába, amely egyszerűen nem az egyetlen létező neokonstruktivista lehetőség.¹

Ennek a lehetséges félreértelmezésnek a tisztázására szögezzük le a következőket: a neokonstruktivizmus azt diktálja, hogy minden kognitív elméletnek *kiszámíthatónak* kell lennie, de nem szükségszerűen komputá-

1 A szimbolikus funkcionalizmus és a nem szimbolikus funkcionalizmus megvitatására lásd HARNAD, 1989, 1993b, d).

ciónak. A kiszámíthatóság csak annyit jelent, hogy az elméletnek eléggé specifikusnak és explicitnek kell lennie ahhoz, hogy egy számítógépes programban formalizálható legyen, s a program képes lesz szimbolikusan ellenőrizni, hogy a modellált jelenségeket jellemző bemenetek mellett, miközben azokat az elveket használja, melyeket az elmélet feltételez, valóban a bejósolt kimeneteket adja-e a program. A komputációs jelleg ehhez (vagyis a komputabilitáshoz) még azt is hozzátenné, hogy csak azokat a sikeres programokat tartaná kognitívoknak, melyek teljesítményüket a szimbólummanipuláció mint kizárólagos elméleti alap segítségével hozná létre. A neokonstruktivizmus viszont minden komputációsan sikeres teljesítménygenerálást kognitívnek tart, akkor is, ha ez transzducerek — effektorok (szimbolikus) szimulációját, más analóg eszközöket, analóg digitális átalakítókat, előre huzalozott speciális célú modulokat, statisztikai szűrőket, párhuzamos feldolgozókat, konnekcionalista hálózatokat, szervomechanizmusokat vagy egyéb trükköket alkalmaz.¹

HARNAD (1989) tárgyalja a formális-szimbolikus szimuláció és a fizikai-oki *implementáció* különbségeit. Egyetértek PYLYSHYNnel (1978) abban, hogy a megismerés számítógépes szimulációja (mely úgy akarja modellálni teljesítményünket, ahogy „igazából végezzük azt”) és a mesterséges intelligencia (mely számítógépes *l'art pour l'art*) közötti különbség meglehetősen mesterkélt, olyan időleges és önkényes tényezőkön alapul, mint az, hogy jelenlegi modelljeink nem elég általánosak (játékproblémák) és hogy nem vagyunk elég szigorúak azokban a viselkedési vagy teljesítménykorlátokban, melyeket alkalmazunk a modell megítélésére. Jelenleg (főként azért, mert nincsen koherens riválisa a komputabilitásnak mint a kognitív elmélet ellenőrizhetőségének, s a fizikai okságnak mint implementációjának) nincs okunk kételkedni abban, hogy végső soron mindegyik megközelítés *konvergálni* fog, ahogy közelítünk a teljes szervezet teljesítményének átfogó modelljéhez (egy olyan modellhez, mely képes átmenni a „Teljes Turing Teszten”). Mindenesetre a kognitív tudományban ugyanolyan jellemző az elméleti életre, mint a fizikában az, hogy az elméletet az adatok aluldeterminálják.

A számítógépes vagy komputációs modellálás rosszul megalapozott bírálatai javarészt arra mennek vissza, hogy 1) félreértik az alulmeghatározottságot és az általánosságot — számos módon ki lehet számítani egy faktoriálisan, van azonban több azonos paraméterekkel dolgozó mód egy vi-

1 Nemigen lehet „trükknek” nevezni valamit, ha az eljut a teljes mértékű teljesítmény utánzásáig (HARNAD, 1992b).

lágemyetem kialakítására? vagy akár egy egész szervezet kialakítására (lásd DENNETT, 1978)? Továbbá arra mennek vissza, 2) hogy homályosak fogalmaink a mechanizmusokról és a magyarázatról. Még a kifinomult kritikusok is (például SEARLE, 1980) önkéntelenül is a mai tudományos helyzet átmeneti primitívségére alapozták az ellenpéldákat ezen az új területen, s örökkévalónak tekintették a mai játékp problémák korlátozott távlatait. A neokonstruktivizmus a releváns általánosítás e kritikákkal szemben, s nem a szimuláció, a játékp problémák és a szimbólum manipuláló modulok kétértelműségei. Felvettem már (HARNAD, 1989), hogy talán a neokonstruktivizmus az, amit SEARLE (1980) értett azon vagy kellett volna értsen azon, hogy mi is a gyenge mesterséges intelligencia-pozíció.

Mindazoknak a trükköknek, melyek tényleg működnek, igazi fizikai funkciójuk van, mely szimbolikusan szimulálható, szimbolikusan nem implementálható azonban. Vagyis ezek komputálhatóak, de nem kizárólagosan komputációsok.

A komputacionalistákat rendkívüli mértékben befolyásolták a természetes nyelv olyan vonásai, mint hogy minden „tényállás” potenciálisan leírható (bármilyen kívánt megközelítési szintig — lásd STEKLIS és HARNAD, 1976; HARNAD, 1993c); valamint a komputáció olyan tulajdonságai, mint hogy minden „folyamat” komputációsán szimulálható (egy megközelítésig); ez is a Church-Turing tétel egyik változata, mely szerint minden komputációsán leírható.

Ennek következményeként amikor felmerül, hogy egy analóg vagy más nem szimbolikus folyamat zajlik, a komputacionalisták erre azt reagálják, hogy de hát akkor nem is kognitív (vagy még árulkodóbban, azt, hogy még nem, vagy nem teljesen kognitív).¹

Ez a korlátozás meglehetősen önkényesnek tűnik, hiszen a komputacionalitás olyan fogalom, mely később jött létre, mint az olyan elméletelőtti gondolatok, mint az „észlelés” és a „gondolkodás”. Miért kellene az

1 Korábban (HARNAD, 1987) felvázoltam egy háromszintű reprezentációs rendszert a kategorizációs képesség magyarázatára. Ott a szimbolikus rendszernek nevezett rész nyilván komputációsabb, míg az ikonikus rendszer analógabb. A kategoriális rendszer közvetít e kettő között, lehorgonyozva a szimbólumrendszert a nem szimbolikus rendszerbe (HARNAD, 1992, 1993c). Nem látok okot arra azonban, miért ne nevezhetnénk az egész háromszintű rendszert kognitívnak. A HARNAD - (1987) modell nem formalizált vagy számítógépesen ellenőrzött modell, s így nem igazán neokonstruktív még; világos motivációja azonban az esetleges számítógépes tesztelhetőség, és játékos teszteléseket már kezdtünk is végezni (HARNAD, HANSON és LUBIN, 1991, 1994).

legyen a kérdés, hogy ezek a tevékenységek „világosan” kognitívek-e, s miért kellene ez attól függhön, hogy az agy úgy döntött-e, hogy megfelelő eszközökkel implementálja őket, megfelelőkkel a komputacionalisták szemében?

Végző soron a neokonstruktivisták és a komputacionalisták közötti vita, akárcsak a percepció kapcsán a konstruktivisták és a gibsoniánusok közötti vita, csak neokonstruktíven dönthető el.

De vajon ez a dolgozat maga neokonstruktív-e? Nem. Pusztán módszertani és megalapozó jellegű. Bár a pszichológiának jelenleg a jelentős adatokon kívül egy neokonstruktív elméletre van szüksége, ez nem azt jelenti, hogy hagyjuk abba a töprengést arról, hogy egyáltalán mit is csinálunk. Még számos lehetőség van a megalapozó és elméletelőtti fogalomalakításra. Így például feltételezve, hogy mindannyian egyetértünk, hogy a szervezet működését végző soron okilag kell magyarázni (vagyis fizikailag implementálható mechanizmusokkal), van-e bármilyen koherens alternatíva az elméletalkotás neokonstruktív korlátaival szemben? Vajon koherens-e az analóg-digitális megkülönböztetés? És mik a komputációs implementáció határai a megismerés terén, szembeállítva a komputációs szimulációval? Vagyis a megismerés milyen mértékig szimbólummanipuláció ténylegesen? Feltehetően a megismeréskutatás következő néhány évtizede erről fog szólni.

A kézirat elfogadva: 1993. november

IRODALOM

- CATANIA, A. C. and HARNAD, S. (eds.), 1988, *The Selection of Behavior, The Operant Behaviorism of BF Skinner: Comments and Consequences*, New York: Cambridge University Press.
- CHAITIN, G., 1975, Randomness and mathematical proof, *Scientific American* 232, 47—52.
- DAVIS, M., 1958, *Computability and unsolvability*, Manchester: McGraw-Hill.
- DAVIS, M., 1965, *The undecidable*, New York: Raven.
- DENNETT, D. C., 1978, Why not the whole iguana? *Behavioral and Brain Sciences*, 1, 103—104.
- DENNETT, D. C., 1982, The myth of the computer: An exchange, *N. Y., Review Books*, XXIX (11), 56.
- DENNETT, D. C., 1978, *The Intentional Stance*, Cambridge MA: MIT Press.
- DENNETT, D. C., 1993, Discussion, In: BOCK, G. R. and MARSH, J. (Eds.), *Experimental and Theoretical Studies of Consciousness*, CIBA Foundation Symposium 174. Chichester: Wiley.
- DIETRICH, E., 1990, Computationalism, *Social Epistemology*, 4, 135—154.

- FRAENKEL, A. A., BAR-HILLEL, Y. and LEVY, A., 1973, *Foundations of set theory*, New York: Elsevier.
- GIBSON, E. J., 1969, *Principles of perceptual learning and development*, Engelwood Cliffs NJ: Prentice Hall.
- GIBSON, J. J., 1979, *An ecological approach to visual perception*, Boston: Houghton Mifflin
- GYR, J., WILLEY, R., and HENRY, A., 1979, Motor-sensory feedback and geometry of visual space: a replication, *Behavioral and Brain Sciences*, 2, 59—94.
- HARNAD, S., 1976, Induction, evolution and accountability, In: HARNAD et al., 1976, 58—60.
- HARNAD, S., 1982, Neoconstructivism: A unifying theme for the cognitive sciences, In: SIMON, T. and SCHOLLES, R. (eds.), *Language, mind and brain*, Hillsdale NJ: Erlbaum, 1—11.
- HARNAD, S., 1984, What are the scope and limits of radical behaviorist theory? *The Behavioral and Brain Sciences*, 7, 720—721.
- HARNAD, S., 1987, The induction and representation of categories, In: HARNAD, S. (ed.), 1987, *Categorical Perception*, The Groundwork of Cognition, New York: Cambridge University Press.
- HARNAD, S., 1989, Minds, Machines and Searle, *Journal of Theoretical and Experimental Artificial Intelligence*, 1, 5—25.
- HARNAD, S., 1990a, The Symbol Grounding Problem, *Physica D* 42, 335—346.
- HARNAD, S., 1990b, Lost in the hermeneutic hall of mirrors Invited Commentary on: Michael Dyer: Minds, Machines, Searle and Harnad, *Journal of Experimental and Theoretical Artificial Intelligence*, 2, 321—327.
- HARNAD, S., 1991, Other bodies, Other minds: A machine incarnation of an old philosophical problem, *Minds and Machines*, 1, 43—54.
- HARNAD, S., 1992a, Connecting Object to Symbol in Modeling Cognition, In: CLARKE, A. and LUTZ, R. (Eds.), *Connectionism in Context*, Springer Verlag.
- HARNAD, S., 1992b, The Turing Test Is Not A Trick: Turing Indistinguishability Is A Scientific Criterion, *SIGART Bulletin*, 3, (4), (October), 9—10.
- HARNAD, S., 1993a, Discussion, In: BOCK, G. R. and MARSH, J. (Eds.), *Experimental and Theoretical Studies of Consciousness*, CIBA Foundation Symposium 174., Chichester: Wiley.
- HARNAD, S., 1993b, *Grounding Symbols in the Analog World with Neural Nets*, Think (Special Issue on machine Learning) (in press).
- HARNAD, S., 1993c, The Origin of Words: A Psychophysical Hypothesis, In: DURHAM, W. and VELICHKOVSKY, B. (Eds.), *Muenster*, Nodus Pub.
- HARNAD, S., 1994, Does the Mind Piggy-Back on Robotic and Symbolic Capacity? To appear, In: MOROWITZ, H. (ed.), *The Mind, the Brain, and Complex Adaptive Systems*.

- HARNAD, S., HANSON, S. J., and LUBIN, J., 1991, Categorical Perception and the Evolution of Supervised Learning in Neural nets, In: POWERS, D. W. and REEKER, L. (Eds.), *Working Papers of the AAAI Spring Symposium on Machine Learning of Natural Language and Ontology*, pp. 65—74, Presented at Symposium on Symbol Grounding: Problems and Practice, Stanford University, March 1991.
- HARNAD, S., HANSON, S. J. and LUBIN, J., 1994, Learned Categorical Perception in Neural Nets: Implications for Symbol Grounding, In: HONAVAR, V. and UHR, L. (eds.), *Symbol Processing and Connectionist Network Models in Artificial Intelligence and Cognitive Modelling: Steps Toward Principled Integration*, (in press).
- HARNAD, S., STEKLIS, H. D. and LANCASTER, J. B. (eds.), 1976, Origins and Evolution of Language and Speech, *Annals of the New York Academy of Sciences* 280.
- HAUGELAND, J., 1978, The nature and plausibility of cognitivism, *Behavioral and Brain Sciences*, 1, 215—260.
- HAUGELAND, J., 1985, *Artificial intelligence*, The very idea, Cambridge MA: MIT/Bradford.
- HAYES, P., HARNAD, S., PERLIS, D. and BLOCK, N., 1992, Virtual Symposium on Virtual Mind, *Minds and Machines*, 2, 217—238.
- HEYTING, A., 1971, *Intuitionism*, An introduction, New Jersey: Humanities.
- JOHNSON-LAIRD, P. M., 1983, *Mental models*, Cambridge MA: Harvard University Press.
- KLEENE, S. C., 1969, *Formalized recursive functionals and formalized realizability*, Providence, R.: American Mathematical Society.
- PYLYSHYN, Z. W., 1978, Computational models and empirical constraints, *Behavioral and Brain Sciences*, 1, 93—127.
- PYLYSHYN, Z. W., 1973, What the mind's eye tells the mind's brain: A critique of mental imagery, *Psychological Bulletin*, 80, 1—24.
- PYLYSHYN, Z. W., 1980, Computation and cognition: Issues in the foundations of cognitive science, *Behavioral and Brain Sciences*, 3, 111—169.
- PYLYSHYN, Z. W., 1981, The imagery debate: Analogue media versus tacit knowledge, *Psychological Review*, 88, 16—45.
- PYLYSHYN, Z. W., 1984, *Computation and cognition*, Cambridge MA: MIT/Bradford.
- RABIN, M. O., 1977, Complexity of computations, *Communications of the Association of Computer Machinery*, 20, 625—633.
- SEARLE, J. R., 1980, Minds, brains and programs, *Behavioral and Brain Sciences*, 3, 417—424.
- SEARLE, J. R., 1983, *Intentionality*, N. Y: Cambridge University Press.
- SHANNON, L. E. and WEAVER, W., 1949, *The mathematical theory of communication*, Urbana: University of Illinois Press.
- SKINNER, B. F., 1984a, Methods and theories in the experimental analysis of behavior, *Behavioral and Brain Sciences*, 7, 511—546.

- SKINNER, B. F., 1984b, Reply to Harnad, *Behavioral and Brain Sciences*, 7, 721—724.
- STEKLIS, H. D. and HARNAD, S. R., 1976, From hand to mouth: Some critical stages in the evolution of language, *Annals of the New York Academy of Sciences*, 280, 445—455.
- TURING, A. M., 1964, *Computing machinery and intelligence*, In: ANDERSON, A. (ed.), *Minds and Machines*, Englewood Cliffs NJ: Prentice Hall.
- ULLMAN, S., 1980, Against direct perception, *Behavioral and Brain Sciences*, 3, 373—415.
- WITTGENSTEIN, L., 1953, *Philosophical investigations*, New York: Macmillan.
- WITTGENSTEIN, L., 1967, *Remarks on the Foundations of Mathematics*, Cambridge, Mass.: M. T. Press.

ISTVÁN HERNÁD

NEOCONSTRUCTIVISM: A UNIIFYING THEME FOR THE COGNITIVE SCIENCES

Behavioral scientists studied behavior; cognitive scientists study what generates behavior. Cognitive science is hence theoretical behaviorism (or behaviorism is experimental cognitivism). Behavior is data for a cognitive theorist. What counts as a theory of behavior? In this paper, a methodological constraint on theory construction — „neoconstructivism” — will be proposed (by analogy with constructivism in mathematics): Cognitive theory must be computable; given an encoding of the input to a behaving system, a theory must be able to compute (an encoding of) its outputs. It is a mistake to conclude, however, that this constraint requires cognitive theory to be computational, or that it follows from this that cognition is computation.