

Kenéz László

## Kant és egy mesterséges képzelőerő kritikája<sup>1</sup>

### BEVEZETÉS

Kétszáz-egynéhány éve, Königsbergben is bárki tapasztalhatta, hogy eszes lény, ám gyakran azt is tapasztalni volt kénytelen, hogy véges eszes lény. Korántsem volt per-se világos, hogy az emberben az ész miért és mennyiben ütközik korlátokba, s hogy egy efféle lény egyáltalán eszes-e annyira, hogy be is láthassa, miképpen véges. Márpedig ennek híján még a legfelvilágosultabb ember sem tudhatta volna, mibe is keveredik, ha a saját maga értelmére merészel támaszkodni. Ismeretes, hogy akkor és ott Kant orvosolta e problémát s nem is akárhogyan. Nem kevesebbet tett, mint egyszerűen az ember kezébe adta megismerőképessége kézikönyvét, még hozzá pontos utasításokkal, hogy a véges lény csakis a számára rendelt határokon belül maradva forgathassa eszét a saját – s nem kevésbé az emberiség – üdvére. És íme, itt a könyv, benne a spekuláló ember, a véges eszes lény portréja, akinek kitörölhetetlen hajlama, hogy olyan kérdésekbe bonyolódjon, amelyek meghaladják képességeit.

Az idő telt-múlt, s a *Kritikát* még ma is mindenki olvassa, de megismerő lényekként már egyáltalán nem vagyunk egyedül. Az emberiség üdvén intelligens gépek garmadája munkálkodik, még Kalinyingrádban is, ráadásul ki tudja, miféle kitörölhetetlen hajlamokkal tömve. Sőt, nemhogy nem vagyunk egyedül, de sokak szerint az ember megismerés kutatásának rangsorában a gépek még előkelőbb helyet foglalnak el, mint maga az ember. A mesterséges intelligencia kutatása pedig nemcsak ígér, hanem kézzelfogható termékekkel is szolgál. Íme, ott van a gép, emitt meg egy egészen másféle kézikönyv, benne egy halom *flow chart*, amelyekből egy olyan eszes lény portréja bontakozik ki, aki gondolkodik, korábbi ismereteire hagyatkozik, tervez, dönt és egyáltalán, egészében is intelligensen viselkedik. Csak éppen nem véges eszes lény. Pontosabban, képtelen úgy élni, mint mi, s nem azért, mert netán szilíciumból van, hanem mert nem a világban mozog, így szó szerint véve nem is *tapasztal*, határokból sem ütközik, hanem egy olyan elzárt térben leledzik, amelynek gyengéd szoftverközege minden kalandtól megóvjá. Igaz, e gépek olyan gondolati magasságokba is képesek eljutni, ahová mi, emberek csak lépcsőről lépcsőre kapaszkodva tudunk felvándorolni, ám leereszkedni és gyökeret verni képtelenek. Összességében tehát igencsak kétséges, hogy e gépezetek a talán nehézkes, ám mégis meglepően robusztus, s a világgal ténylegesen is megbirkózni képes emberi megismerés modelljének tekinthetők-e egyáltalán.

Másfelől, a megismerő lények palettájára az ember mellé nemcsak az intelligens gépek, de az állatok is felkerültek. Eltelt kétszáz év, s ők már korántsem olyan terem-

<sup>1</sup> Ezúton mondok köszönetet Tózsér Jánosnak értékes észrevételeiért, Blaskó Györgynek pedig a grafikai segítségért.

mények többé, amelyek ottrekedtek volna az embertől őket elválasztó szakadék túoldalán, hanem olyan kognitív lények, amelyek különféle alapképességei – az emberre is átöröklődve – az emberi megismerés kutatási paradigmáinak s így egy *más jellegű* megismerés-modellezésnek is irányt szabnak. Ez utóbbi terepen ugyanis nem az ész jelent kihívást, hanem az élőlények masszív *érzékisége*, így ezek a más elvek szerint épített szerkezetek inkább különféle észlelési feladatok sikeres gépi megoldásait példázzák. Ezek a mesterséges érzékelő rendszerek talán ostobának tűnnek, de legalább hajlékonyak s alapvető vonásaik tekintetében valóban olyanok, mint az érzékelő élőlények. S lásd, a megismerésről így megint csak egészen másféle ábra rajzolható fel. Vegyük most ezt a tervet kézbe, és egy véges lény portréja bontakozik ki előttünk, egy olyan lényé, amely képes egy érzéki világhoz idomulni és pontosan a környezet szabta korlátok mentén járni, akármilyen jellemezze is e korlátokat. Csak nem igazán eszes, s arra sem ösztökélhető, hogy kísértésbe esve, intelligens módon, mint egy saját képességeire hagyatkozva, megpróbálja áthágni a számára rendelt érzéki világ kereteit. Márpedig ez kevés, bármennyire figyelemre méltó is amúgy az a teljesítmény, amit nyújt, hiszen az emberi megismerés nem pusztán alkalmazkodás és környezetbe való beilleszkedés.

Egyfelől ott vannak tehát az okos gépek, amelyek azonban úgyszólván léghajóként lebegnek a világ felett és nem képesek horgonyt vetni a számunkra ismerős és létfontosságú érzéki talajba, másfelől pedig azok a gépek, amelyek ugyan jól túrik a világ szeszélyességét, csak éppen földhözragadtak, mert egy kicsikét sem képesek elemelkedni, hogy a számunkra éppoly ismerős racionalitás világában is eligazodjanak. És a háttérben a két, egymással rivalizáló alkotócsoport, amelyek mindegyike saját gépei számára követeli az emberi megismerés adekvát modelljének titulását.

Feltűnő ez a kétoldali egyoldalúság, amelyet még a kettős – tehát a két típust egy modellbe komponáló s a csorbákat ily módon kiköszörültni próbáló – elgondolások sem szüntetnek meg. Hiába adják fel ugyanis e hibridek tervezői az egységes modell kívánalmát annak érdekében, hogy mindkét típus előnyeit kihasználhassák, úgy tűnik, egyelőre még ez sem elegendő egy valóban hatékony modell létrehozásához. Persze lehet, hogy mindez csak idő kérdése, ám jelen pillanatban nem kell különösképp rosszindulatúnak lenni ahhoz, hogy mégis gyanakodni kezdjünk: az emberi megismerés, a véges eszes lény pszichológiai igényű, egyetemes érvényességre számot tartó, ám egymással rivalizáló, s közös fizikai nevezőre nem hozható gépi modellezései valahol talán zsákutcába tévedtek.

Márpedig egy ilyen gyanú esetén a filozófusok igencsak hajlamosak körülnézni, aztán visszatérést hirdetni valamihez vagy valakihez, annak reményében, hogy onnan új lendületet véve másodjára talán egészen máshová fognak megérkezni. Főként akkor, ha elgondolkodtató analógiák is felmerülnek. Mivel gyanakszunk, és látunk néhány analógiát, visszatérünk Kanthoz, és megvizsgáljuk, vajon *A tiszta ész kritikája* felől közelebbre hogyan értelmezhető a kialakult helyzet. Egyenesen azzal a reménnyel vágunk neki mindennek, hogy Kant megismerésfogalma megoldást nyújthat a kétoldalú gépi hiányosságok problémájára.

Biztató jel, hogy a történeti távolság ellenére az elméleti rokonság jóval közelebbi, mint első pillanatra gondolnánk. Könnyen kezünkbe kaparinthatjuk ugyanis (1) a közös nevezőt, a *kognitív architektúra* fogalmát, amely a mesterséges intelligencia kutatásban explicite, Kantnál pedig, mint alább megmutatjuk, implicite van jelen, s amely így megfelelő kiindulópontként szolgál. Következő lépésként konkrétan is megmutatjuk a

megismerés háromféle szerveződését, nevezetesen (2) Kant, illetve a mesterséges intelligencia kutatás úgynevezett (3) *szimbolikus*, illetve (4) *szubszimbolikus* kognitív architektúráit, s egyúttal körül is határoljuk a kétféle gépi megismerés bizonyos nehézségeit. Ezután térünk rá (5) a megismerésnek mint érzéki és értelmi képzetek összehangolásának kanti problémájára és megoldására, a *sematizmusra*, illetve ehhez kapcsolódóan a *transzcendentális appercepció* funkciójára. Majd levonva bizonyos tanulságokat, ezeken okulva szabjuk ki (6) egy *mesterséges képzelőerő kritikája* során a két gépi architektúra közötti közvetítő tényező lehetőségeinek igencsak szűkre szabott kereteit. Elérve e pontra, már nem visszakoehatunk attól sem, hogy (7) az alapokra rákérdezve az egész problematika metafizikai mélyébe pillantsunk, hogy aztán, egyelőre mit sem kockáztatva, megoldásként a mesterséges megismerés kutatásának más metafizikai alapokon nyugvó alternatíváját ajánljuk megfontolásra.

## (I) FIZIKAI, FUNKCIONÁLIS ÉS KOGNITÍV ARCHITEKTÚRA

Az architektúra szót az építészetből ismerjük, így aztán elég plasztikus ahhoz, hogy más területeken is könnyedén felhasználhassuk. Igencsak hajlunk például arra, hogy az elméleteket – mint különféle fogalmak konstrukcióit – építményként ábrázoljuk, ahol az alapvetőbbek tartják a rájuk épülő újabb fogalmakból kialakított szinteket. Emellett azonban mindennapos az is, hogy az architektúra fogalmát nem csak olyan statikus szerkezetekre használjuk, mint amilyen egy épület és egy elmélet (szerencsés esetben). Példaként, egy üzem működtetése szintúgy egy olyasféle ábrán írható le vagy tervezhető meg a leginkább, amely a termék előállításához szükséges folyamatokat különféle működési láncolatok szerint rendezzi el. Itt azonban már nem a fizikai viszonyokat, hanem azokat az *elvi* kapcsolatokat tesszük láthatóvá, amelyek az üzem funkcionálásának vetületei, vagyis annak a folyamatnak a tagolását adjuk meg, amelynek kezdete a nyersanyag és vége a késztermék, röviden, a gyár működési architektúráját. Egy *fizikai architektúra* így tehát azt adja meg, hogy egy szerkezet miként van fizikailag megépítve, melyik rész melyik más résszel áll fizikai kapcsolatban, milyen statikus alrendszerekre tagolható és így tovább, míg egy *funkcionális architektúra* azt, hogy az adott rendszer hogyan tagolható dinamikusan, azaz milyen struktúrák, kapcsolódások és összefüggések szerint üzemel.

A mesterséges intelligencia konstruálása mind fizikai, mind funkcionális architektúra szerint két fő ágra osztható. A modellek egyik típusa a *fizikai* architektúra tekintetében komplex alegységek bonyolult rendszere, míg a másik egyszerű elemek homogén hálózata. Ismerősebben cseng és többről is árulkodik azonban e rendszerek *elemi funkcionális* architektúráinak neve: szeriális és párhuzamos információfeldolgozás.

Az első esetben egyetlen, ma már elképesztő sebességgel dolgozó processzor, vagyis adatkiolvasó működik. A *szeriális* adatfeldolgozás működési architektúrája tehát, mint neve is sugallja, egy szeriában dolgozó szerkezet sajátja. Az előkészítő részleg feladata az, hogy az információt a processzor számára olvasható formára hozza és az adatkiolvasóhoz juttassa, a funkcionálás veleje pedig az egymás utáni, *szukcesszív* adatfeldolgozás. Ezért is olyannyira csábító a gondolat, hogy ezt az egyedi proceszszort, munkamódszere miatt, a *figyelem* analogonjának tekintsük. Ezt a típust mindenki ismeri, hiszen ilyenek a személyi számítógépek is, s mindannyiunk számára kétség-

telen, hogy ezek a gépek meggyőző intelligens viselkedést produkálnak, pontosan úgy, mintha ott benn egy „nagykoponya” üldögélne.

A másik, talán kevésbé ismert építési mód a *párhuzamos* feldolgozás elve szerint konstruált szerkezet. Ez esetben nem egyetlen, hanem számos, bár jóval egyszerűbb adatkiolvasó kapcsolódik össze hálózattá, oly módon, hogy az egyes processzorok által kiolvasott adatok szerint működésbe lépő egység kimenete egy vagy több másik processzor bemeneteként szolgál. A teljes rendszer végső kimenetéhez tehát minden egyes processzor adatkiolvasása hozzájárul, s mivel ezek az adatkiolvasó és továbbító folyamatok *szimultán* zajlanak, az információfeldolgozás alapvetően párhuzamos. Ezek a rendszerek egyforma elemekből, nevezetesen rengeteg egyszerű, sőt, akár butának is nevezhető processzorból állnak össze, amely szerkesztés bevallott ihletője az élőlények *agya*. E konstrukció háttérében tehát az a közismert gondolat áll, hogy az idegsejtek vélhetően szintén egyfajta hálózatba kapcsolódva munkálkodnak lelkünkön, egyenként alig csak töredékkel hozzájárulva az egész rendszer működéséhez.

A fizikai architektúrával a továbbiakban nemigen foglalkozunk, és alapvető szemléleti váltással csakis a magasabb szintű működések elvei felé fordulunk. Ez pedig annál is inkább fontos fejlemény, mert egy ilyen megközelítés már korántsem áll olyan messze Kanttól, mint a technikai részletek. Még nyilvánvalóbb ez akkor, ha a mesterséges intelligencia kutatásának fentebb vázolt két funkcionális architektúráját, a szeriális és a párhuzamos adatfeldolgozást – vagyis úgyszólván az üzemek termelési folyamatát – egyenesen *megismerésként* nevesítjük, vagyis pszichológiai értelemben releváns *kognitív architektúráként* vesszük szemügyre. A kognitív architektúrát a továbbiakban úgy tekintjük tehát, mint az általában vett funkcionális architektúra azon speciális esetét, amely a megismerésnek nevezett működés feltételeit adja meg részleteiben, s amelynek funkcionális elvi kiinduló nyersanyaga a világ, végterméke pedig a világról nyert ismeret. Még egyszerűbben, valamely kognitív architektúrát megadni tulajdonképpen nem más, mint azon kijelentések halmazára hivatkozni, amely kijelentésekkel a megismerés *elvi* hogyanjainak kérdését válaszoljuk meg.

S e pontról már látható, hogy *A tiszta ész kritikája* is éppen ezt teszi, nevezetesen arra a kérdésre ad választ, hogyan működik a világmegismerés, s pontosan azokat a kereteket adja meg, más szóval éppen azt a kognitív architektúrát vázolja fel, amelynek terminusai egyfelől a megismerendő dolog, másfelől pedig a létrejövő ismeret (*cognitio*). És pontosan ilyen válaszokat adnak a mesterséges intelligencia kutatásának azon pszichológiai igényű megfogalmazásai is, amelyek az emberi megismerés tulajdonképpeni kognitív architektúrájának meghatározását tűzik célul maguk elé, vagyis, a *klasszikus kognitívizmus* (NEWELL–SIMON 1976) és a *radikális konnekcionizmus* RUMELHART–MCCLELLAND 1988) is. Megvan tehát az optika, amelynek birtokában a közös vonások, már ha vannak, jól kivehetőek lesznek. Lássuk tehát, hogyan ad számot Kant és a mesterséges intelligencia kutatásának kétféle paradigmája a megismerés elvi mikéntjéről. Előbb azonban még néhány bevezető szó.

A kognitív architektúra mint speciális funkcionális architektúra leírásához, mint fentebb már láthattuk, laza, ám egyelőre megfelelő keretet biztosíthat egy *indusztriális metafora*. A megismerés így a kognitív ágens megismerési szekciójának üzemelése, úgyszólván az a gyártási menet, amelyben a nyersanyagok előírt feldolgozásának eredményeként végül elébünk pottyan a késztermék, jelen esetben az ismeret. Az architektúra *kimerítő* leírása így nem más, mint végigkövetni a termék elkészül-

tének útját, minden egyes ponton kiemelve azokat az elvi tényezőket, amelyek betartásával az előállítás zajlik. Ha azonban kevesebb is beérjük – s jelen esetben beérjük –, arra a kérdésre, hogyan működik a megismerés, egy elemibb választ is nyerhetünk akkor, ha csak azt nézzük meg, hol akadhat meg valamiképp az ismeret előállítása. Más szóval, azokat a szükséges feltételeket igyekszünk meghatározni, amelyek nélkül a megismerés *nem* funkcionál. A következőkben tehát sorra vesszük azt a három kognitív architektúrát, három megismerésüzemet, amely azzal az univerzális igénnyel lép fel, hogy meghatározza az emberi megismerés működésének elvi hogyanjait. Először Kanttal foglalkozunk, majd a klasszikus kognitívizmus modelljével, végül pedig, kevésbé közismert volta miatt némileg bővebben, a radikális konnektionizmus megismerési architektúrájával, mindhárom esetben kifejezetten olyan diszfunkcionális mód elvi okai után kutatva az adott rendszerben, amely a megismerés nélkülözhetetlen feltételeire világíthat rá.

## (2) TRANZSCENDENTÁLIS ÜZEM

A megismerés legalapvetőbb szükséges feltételeit Kant transzcendentális – vagyis nem a megismert tárgyakkal, hanem a tárgyak megismerésének *mikéntjével* foglalkozó – vizsgálódásaiból emelhetjük ki. A világmegismerésnek, még pontosabban a természet megismerésének így, mint tudjuk, *A tiszta ész kritikája* szerint két elemi gyökere van, a *fogalom (conceptus)* és a *szemlélet (intuitio)*. Mivel azonban „tartalom nélkül üres a gondolat, fogalom nélkül vak a szemlélet” (KANT 2004, B 75), sem a fogalom, sem a szemlélet nem elegendő önmagában az objektív ismeret létrejöttéhez. Röviden, nem működik a megismerés, ha a fogalom nélkülözi a szemléleti tartalmat, vagy ha a szemléletbe nem játszanak bele a fogalmak. Határozzuk meg azonban pontosabban is ezt az összefüggést.

Kant a megismerés e két összetevőjének jellemzésére a következő jelzőket használja: az értelmet *spontaneitás*, az érzékiséget pedig *receptivitás* jellemzi. Ez módot ad arra, hogy a megismerés folyamatát két irányból is felvázolhassuk, a megismerés menetét természetesen nem időbeli lefolyásként, hanem logikai lépések sorozataként értve.

*Egyrészt* tehát, az objektív ismeretet célzó megismerés elvi kiindulópontjaként a spontán értelem szolgálhat, amely működéséhez nem szorul rá az érzékiségre, mivel az elme *képzeltalkotó* képessége folytán módunk van arra, hogy tárgyakat gondoljunk el. Az elme spontaneitásának termékei, vagyis az *értelmi képzetek*, így lehetővé teszik, hogy a megismerés folyamatát elsőként, modern episztemológia kifejezéssel élve, egy *top-down* vektorral, az elmétől a világ felé haladva vegyük szemügyre. Mivel az elme értelmi képzetei *fogalmak*, ily módon tehát fogalmakkal közelítünk a megismerendő dologhoz. Ám így csak félútig juthatunk, mert fogalma által elgondolva a tárgyat, az még – ahogy a *Kritikában* áll – nem lesz adva. Sőt, Kant egyenesen arra int bennünket, hogy az elmétől a világ felé haladva e ponton meg is kell állnunk, mert ha ebből az irányból erőltetnénk a továbbhaladást, tekintet nélkül az érzékiségre, tévútra jutnánk. A fogalmak általi elgondolás ugyan komoly távlatokat nyit meg, ám az érzékiség nélkül e lehetőség nem gyümölcsöző.

„[...] hiába terjesztjük ki a fogalmakat a mi érzéki szemléletünkön túlra, ez a kiterjesztés mit sem segít. Mert ez esetben üres fogalmakkal van dolgunk [...], objektív realitás nélküli, pusztá gondolati formákkal [...] (KANT 1787/2004, B 148.)

Egy efféle függetlenítési próbálkozás ráadásul metafizikai dogmatizmushoz vezethet, hiszen ha feltesszük, hogy a fogalmak önmagukban is hozzáférnek a világhoz, és úgy jeleníthetik meg a dolgokat – miként Kant a kritikai fordulat előtt megfogalmazza –, „ahogyan azok vannak” (KANT 1770/1988a, 187), akkor nincs is szükség ismeretkritikára. Egy ilyen feltételezés kísértése ugyanis az, hogy a világ felőli tudakozódást pusztá fogalmi konstrukciókkal intézhetjük el. Összegezve, *A tiszta ész kritikája* nem ad módot arra, hogy csakis fogalmak segítségével jussunk objektív ismeretek birtokába.

*Másrészt*, okulva az előbbiekből, kíséreljük meg most a másik irányból, az érzéki-ségtől indulva, úgyszólván tehát a világtól az elme felé haladva, megrajzolni az objektív ismeret elvi gyártási útját. A dolog magában (Ding-an-sich) mint a megismerési folyamat kiindulópontja afficiálja az elmét, amely ennek következtében képzeteket „fogad magába”, mint Kant mondja (KANT 1787/2004, B 75). E képzetek, amelyek minden gondolkodás előtt adva lehetnek, a *szemléletek* (KANT 1787/2004, B 131). Ebből az irányból, a világ felől, az elme receptivitására hivatkozva – vagyis, egy másik kurrens kifejezéssel élve, *bottom-up* vektorral – azonban ugyancsak félútig juthatunk. E képzetek által adva lesz ugyan valami, amit Kant *jelenségnek* nevez (KANT 1781/2004, A 124), ám ez önmagában megint csak nem elégséges az objektív ismerethez. Hogy miért, az jól látható, ha feltesszük, hogy ezzel a vektorral nyomulunk tovább. Ekkor ugyanis egyenesen a tudomány lehetőségét tennénk kockára, vagy egyszerűbben, elménk azon képességéről kellene lemondanunk, hogy fogalmak által elgondolva a *priori* módon ismerjük meg valamit. Ha ugyanis az emberi megismerést teljes mértékben a tárgy befogadása határozná meg, akkor le kellene mondanunk a világmegismerésben az olyan fogalmakról (okság, szükségszerűség), amelyeknek kizárólagos a *posteriori* módon való használata végzetesen legyengítené a tudást.

Szögezzük le tehát: ahhoz, hogy az emberi objektív ismeret létrejöjjön, a két ág külön-külön szükséges, de nem elégséges. Nem csak elgondolnunk kell a tárgyat (csak értelmi képzettel rendelkezni róla), s nem csak adva kell hogy legyen valamely jelenség egy szemléletben (csak érzéki képzet birtokába jutni, valamely afficiáló dolog hatása folytán), hanem adva kell lennie a tárgynak. Ehhez pedig elengedhetetlen az értelem és érzékiség összjátéka.

S ennyi egyelőre elegendő is számunkra, így tehát, mielőtt áttérnénk a gépi architektúrák taglalására, húzzuk alá még egyszer a tézist: ha a megismerés a „két szélső pólus” (KANT 2004, A 124), azaz érzékiség és értelem együttműködésének gyümölcse, akkor nincs megismerés *érzéki tartalommal bíró fogalmak*, illetve *fogalmak szerint rendezett szemléletek* nélkül. Ugyanaz a kívánalom tehát két perspektívából is ábrázolható: az értelem felől tekintve, a fogalomhoz *hozzá kell adni* azt, ami a szemléletben adott, az érzékiség felől tekintve pedig, a szemléletet kell értelmessé tenni, vagyis mint valamely egyedít megjelenítő képzetet, fogalmak *alá rendelni* (KANT 2004, B 75).

A következőkben a gépek kapcsán szintén logikai képet festünk a mesterséges megismerés szerkezetéről. Fentebb már azt is említettük, hogy a kognitív architektúra az első esetben egy szeriális, a másodikban pedig egy párhuzamosan megosztott információfeldolgozás specifikált alakja, s csak azt kell nevesíteni, hogy magasabb szinten, megismerésként miként is történik mindez. A két szerveződést Smolensky nyomán *szimbolikus* és *szubszimbolikus* kognitív architektúráként fogjuk megnevezni (SMO-

LENSKY 1996), és a legegyszerűbb eljárást követve, a kitűzött elemi cél érdekében, a gépek esetén is azt vizsgáljuk meg, mikor *nem* működik az egyik, illetve a másik architektúra szerint kialakított gép. Még pontosabban, most azt vesszük szemügyre, milyen elvi okai lehetnek annak, ha képtelenné válik valamely kognitív teljesítményre az első, illetve a második típusú szerkezet.

### (3) SZIMBOLIKUS ÜZEM

Mint mindannyian tudjuk, a *komputer* számításokat végző gép. A számítógép valami módon képes arra, hogy összeadjon, szorozzon, függvényműveleteket végezzen, sőt, olyan bonyolult operációkat hajtson végre, mint amilyen például a sakkozás. Nem világos azonban, hogy mitől lesznek a számítógépek pszichológiai értelemben is relevánsak, azaz miért lehetnek az emberi megismerés adekvát modelljei. Nos, ha egy géptől megkérdezzük, mennyi öt meg hét, és jó választ ad, aztán helyesen szoroz, rajzol grafikont vagy éppen ügyesen ad sakkot, akkor korántsem tűnik túlzásnak az emberi intelligenciának is ugyanolyan kognitív architektúrát tulajdonítani, mint ami lehetővé teszi az ilyesféle meggyőző viselkedést produkáló gépek építését. Ha tehát az intelligens számítógép számításokat végez, és így tudja, hogy öt meg hét az tizenkettő, márpedig ez már *megismerés*, akkor feltehető, hogy végső soron mi is ezt tesszük valamilyen módon, ha szintén tudjuk az eredményt. Kérdés persze, hogy mit is jelent az, hogy *számítás*?

A klasszikus kognitívizmus szerint e számítás *komputáció*, azaz szabály vezérelte *szimbólummanipuláció*. Egy szimpla példát hozva arra, hogy mi is ez tulajdonképpen, gondoljunk az abakuszra. Amikor a golyós számológépen összeadást végzünk, akkor szabályt követünk (így és így kell összeadni) és szimbólumokkal (az egyes kövek különféle számértékeket reprezentálnak) manipulálunk (a szabályokat követve csoportosítjuk őket, választjuk szét stb.). Ám ugyancsak szabályokat követve szorzunk és végzünk függvényműveleteket, sőt, sakkozunk, még hozzá mindannyiszor szimbólumokkal dolgozva. Ha tehát a megismerést úgy modellezzük, hogy meghatározzuk a szabályokat, és pontosan leképezzük a különféle szimbólumokat, majd ezeket egy fizikai hordozón is megvalósítjuk, oly módon, hogy fizikailag biztosítjuk a manipuláció lehetőségét, akkor pontosan egy *szimbolikus architektúrát* implementáltunk.

A kognitív architektúra szintjén maradvá az itt szükséges feltételeket firtatva könnyen felismerhető tehát, hogy ha nincsenek szabályok, ha a manipulált elemeknek nincs jelentése, vagyis nem szimbólumok, vagy ha a manipuláció nem követi a szabályokat, akkor nem is beszélhetünk megismerésről. E hármastételrendszerből itt egyet emelünk ki, a *legalapvetőbbet*. A szimbolikus kognitív architektúra szerinti megismerő működés gyártószalagja be sem indul, ha a szimbólumoknak nincs jelentése, míg szabályok nélkül, vagy összevissza manipulálva a gép még elvegetálhat (akár helyesen is működvé). A műveletek viszont, ha az egyes elemek nem bírnak szemantikai értékkel, teljességgel meddő, pusztán fizikai működések.

Ám hogyan tehet szert jelentésre egy fizikai rendszer? Nyilván úgy, hogy valaki (a programozó) szemantikai értékekkel ruházza fel az egységeket, ezek kombinációit, illetve az egységek szabályalapú manipulációit. Ez az eljárás pedig a megismerés esetén nem más, mint vonatkozásokat képezni a megismerendő világ bizonyos egységei, aspektusai és eseményei, illetve a megismerő rendszer részei és működései

között. Ha a megismerés tehát ilyen vonatkozások szabályokat követő kombinálása, esetleg új vonatkozások szabályokat követő kiképezése, akkor a megismerés terméke, az ismeret nem más, mint a világ különféle aspektusaival egy az egyben összefüggésben álló struktúrák, azaz jelentéssel bíró fizikai egységek, nevezetesen fizikai szimbólumok rendszere. Ilyen vonatkozás nélkül nincs kognitív teljesítmény, tehát, összefoglalva, ha a megismerés szimbólummanipuláció, akkor a *szimbólumok lehorgonyzása* (HERNAD 1996) nélkül nincs megismerés.

Ám milyen feladatot ró ez a szimbolikus üzem megbízott vezetésére? Nehezet. A gyártórészleget hozzá kell kapcsolni a világhoz, ez rendben. Igen ám, csak hogy ellentétben az összeadással, a szorzással és a függvényműveletekkel, de még a sakkal is, a világ, a *mi emberi világunk* jóval szövevényesebb valami annál, hogy egyszerűen csak néhány, sőt, jó néhány alapvető szimbólummal és szabállyal elintézhetnénk. Nyilván nem ezek szaporításához kell folyamodni (vagyis mintegy az üzemvezetés szájába rágni, mi a teendő minden egyes esetben), mivel ez irdatlan, sőt végtelen feladatot róna a programozókra (DREYFUS 1972). Ha viszont ragaszkodunk ahhoz, hogy *egyetlen* adekvát kognitív architektúra alapján interpretáljuk az emberi megismerést, akkor az *univerzális* igényvel épített rendszernek minden emberi kognitív teljesítményt produkálnia kell tudni. Ez pedig annyit tesz, hogy a szimbolikus elvű gépezeteknek is el kell tudniuk igazodni az érzékiség világában, mivel mi emberek, eligazodunk. Összefoglalva tehát: ha a megismerés komputáció, akkor a szimbólumoknak magával az aktuális világgal kell valamilyen módon egy folytonosan frissített vonatkozásban állniuk, mert e nélkül egy szerkezet nem lehet az emberi megismerés adekvát modellje.

Mielőtt lendületből kapcsolatot teremtenénk Kant tartalom nélküli fogalmai és a lebegő szimbólumok között, nézzük meg közelebről a másik mesterséges kognitív architektúra sajátosságait is. Mint fentebb már megneveztük, ezen álláspont szerint a párhuzamos feldolgozásban működő gépek kognitív teljesítményét egy *szubszimbolikus* architektúrának köszönhetjük. Az imént láthattuk, hogy milyen lenne a megismerés, ha a szimbólumok felől ragadnánk meg, most pedig úgy tűnik, hogy valamiképpen e szimbólumok *alá* kell mennünk.

#### (4) SZUBSZIMBOLIKUS ÜZEM

Az a gyár, amelynek működését most megvizsgáljuk, fizikai tekintetben igencsak egyszerű: egyszerű elemek és az ezeket összekötő huzalok sokasága, semmi egyéb. S bár e szerkezet előzetesen tehát nem tagolt, egy tanulási folyamatot követően, amelynek elvi alapja a bemenetek és kimenetek összefüggésének rögzítése, a szerkezet képessé válik bizonyos kognitív teljesítményekre. A *párhuzamosan megosztott feldolgozás* (Parallel Distributed Processing, PDP) elvét követő modellek ilyen teljesítményeit úgynevezett *emergens* jegyek segítségével tudjuk kellőképp értékelni. Az emergencia ez esetben nem más, mint kiemelkedés, amelynek „ahonnan”-ja az alapvető fizikai szint működési leírása. Ezáltal lehetővé válik a rendszer viselkedésének beillesztése egy olyan leírási közegbe, amely magasabb szintű struktúrák szerint tagolt, mint az alapszint. Egyszóval, azt tudjuk mondani, hogy a gép *felismeri* az ‘a’ betűt, holott végső soron semmi egyéb nem történt, mint hogy a rendszer fizikai szerkezete egy bizonyos aktivitásmintázatot – s ennek következtében megfelelő kimenetet – generált. Hozzuk azonban mindezt közelebb egy egyszerű példa segítségével.

Ha a kezünk alá adnak egy tánckart, ahol minden egyes táncosnak csak néhány mozdulatot engedünk meg, továbbá meghatározzuk azt is, hogy az egyes táncosok szomszédai mely mozdulataira mit csináljanak, akkor a nézőtérrel a tánckar egészének mozgása, mondjuk bizonyos kirajzolódó térbeli alakzatok, mint a tánckar „kimenetei”, emergens jegyek lesznek. Ezeknek a jegyeknek a létrehozásában az egész tánckar részt vesz, így ezek mintegy *megosztlanak* a résztvevők között, s amennyiben *ilyenként* utalnak valamire (jelentenek valamit, képeznek vonatkozást valami mással, a megismerés esetén a világgal) akkor már belátható, hogy miért nevezhetjük őket *megosztott reprezentáció*nak. Mivel pedig a megosztott reprezentáció és a szimbólum szerkezete egybevágó, úgy tűnik, megvan a szimbólum, úgy, ahogyan a párhuzamos feldolgozásban megjelenik, következésképp, az ez *alatti* szint leírása a megismerés tulajdonképpeni adekvát megragadása. Ez pedig, mint már mondtuk, nem más, mint az egységeket és ezek kapcsolatait leíró elemi fizikai szint.

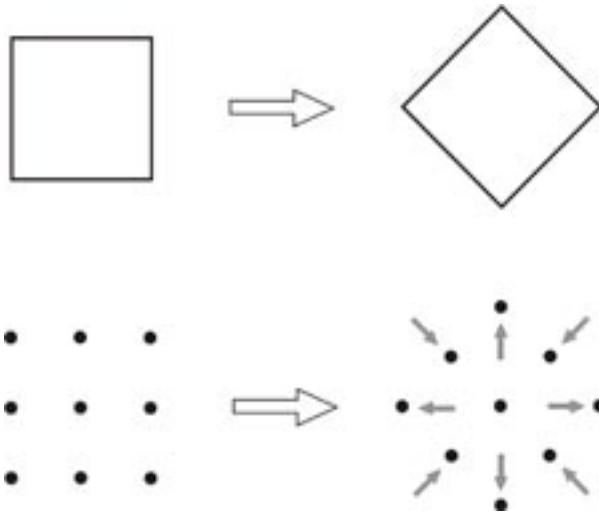
Mi jellemzi tehát azt a bázist, amelyből bizonyos jegyek emelkednek ki a tanulás során? Térjünk vissza a tánckarhoz. Mint észrevehetjük, előzetesen meg kellett szabnunk, hogy melyik táncos mit csináljon, vagyis *szomszédai* mozgásaira miképpen reagáljon. Ez pedig valamely *szabály* megadása. Csak nem mindegy, milyen nyelvezetben, mert a lényeg itt rejlik. A párhuzamos feldolgozás is szabályra épít, csak éppen ezeket az elemi szintű, egyszerű szabályokat tekinti elsődlegesnek az emergens zóna egységei között fennálló vagy felállítható, magasabb szintű szabályokkal szemben. Ellenkéntben tehát a szimbolikus architektúrában működő modellel, amelynek szerkezeti felépítése hű tükre lesz a lefedett – azaz a rendszer elemeinek vonatkozásait képező – terület szemantikai struktúrájának, a szubszimbolikus szint szemantikailag *intranszparens* réteg lesz (SMOLENSKY 1996).

A példánál maradva, az egyes táncosokat instruáló utasítások között így *nem* fogjuk fellelni a ‘négyzet’ kifejezést, és akárhogy is vizsgáljuk az egyes táncosok mozgását, illetve az ezt szabályozó összefüggéseket, nem fogunk tudni arra következtetni, hogy az egységek *együttese* milyen szemantikai értékkel rendelkeznek. A rendszer működésének alapszintű leírása tehát az egyes egységek működési szabályait adja meg, s az ismeret létrejötte így nem más, mint az egységek e szabályok szerint kialakuló aktivitásmintázata, vagyis röviden, megfelelő *kapcsolatok* (*connections*) kialakítása. Ha pedig még azt is felidézünk, hogy az idegsejtek analóg módon szerveződnek rendszerré, akkor nem kell külön kommentálni, hogy miképpen válik pszichológiailag relevánssá ez az elképzelés egy materialista metafizika keretében.

Azonban (és itt kezdődnek a bajok, mint ahogy, mellesleg, itt kezdhethetjük el építeni a Kanthoz vezető hidat) a kiképzett megosztott reprezentációk, bár rendkívüli módon „hajlékonyak”, vagyis akár hiányos *inputok* is *outputra* bírhatják őket, így igen kemény terepen is megfelelő működést tesznek lehetővé, s ezenkívül meglepően ellenállóak a rendszer tényleges sérüléseivel szemben is, azaz „kíméletesen hanyatlóak” (vö. CLARK 1996, 131 skk.), az általuk nyert kognitív teljesítmények szintje nem éri el az emberi megismerés teljesítményeinek színvonalát. Az egyik legfontosabb kritikai pont, amelyet most szükségünk szerint felidézünk, éppen e gyenge teljesítmény miéértjére mutat rá, nevezetesen arra, hogy ezek a megosztott reprezentációk – a szimbólumokkal való formális izomorfiájuk ellenére – *mégsem* egyenértékűek a szimbolikus architektúra pontos fizikai vetülettel is rendelkező egységeivel, mivel nem képesek olyan kognitív teljesítmények alapjául szolgálni, amelyekhez viszont a rivális szimbolikus architektúra egységei szilárd bázist nyújtanak (FODOR–PYLYSHYN 1988). A trenírozott – tehát kikép-

zett és így emergens, megosztott reprezentációk segítségével leírható kognitív teljesítményt mutató – hálózatok bizonyos feladatokban ugyanis igencsak rosszul teljesítenek, nevezetesen a megosztott reprezentációk nagyon nehezen szolgálnak olyan elemi, ám az emergens zóna egységeire épített *újabb szintű* műveletek kiindulópontjával, mint amilyenek lehetnek például a propozicionális logikai operációk.

Ehhez járul továbbá az a párhuzamosan megosztott funkcionális architektúrából következő másik hátrány, hogy – ellentétben az úgynevezett szimbolikus architektúrában épített, tehát szemantikailag transzparens fizikai építménnyel rendelkező szakértő rendszerekkel – egy ilyen gép nem lesz képes *számot adni* arról, hogy miként jutott el a kiadott eredményhez. Röviden, nem lesz képes indokolni és beláttatni, hogy az általa adott eredmény miért helyes. S ez nem is csoda, mivel a kiképzett emergens jegyet az elemi szintű egységcsoportok aktivitása mindig *másként* alakítja ki (vö. CLARK 1996, 163). Ellentétben a szimbolikus architektúra magasabb egységeivel, amelyek mereven kompozicionálisak, a megosztott reprezentáció *nem* az. Tánckari példánknál maradva, bár úgy tűnik, hogy az emergens alakzat, esetünkben a négyzet, az öt kiképző tánckar tagjaiból állna össze, sőt, bizonyos értelemben ez így is van (amennyiben a táncosok fizikai testén szupervenial), ám az alakzat rotációjával a négyzet ugyanazon részét a látszólagos elfordulás következtében más táncosok fogják reprezentálni (lásd az ábrát).



**I. ábra.** A felső két alakzatot balról jobb felé nézve minden további nélkül vehetjük úgy, mint egyazon négyzet rotációját, ám ha elemekből építjük fel az alakzatot (alsó sor), akkor az egyes elemek elmozdításával létrejött második, új négyzet részeit más elemek reprezentálják az új pozícióban, mint a kiinduló helyzetben, nevezetesen, ami csúcs volt, oldal lett és fordítva. Szubszimbolikus szinten az elemek – és nem az egész – elfordulásának szabályait adjuk meg: ezt jelzik a szürke nyilak.

Ez azonban azzal jár, hogy nem lesz *közvetlen* átmenet a két alakzat között az emergens szinten, vagyis emberibben is megfogalmazva, a rendszerben nem egyazon négyzet fordul el (nem egyetlen reprezentáció aktivizálódik kétszer), hanem két különböző aktivitásmintázat követi egymást. Ám ahhoz, hogy kiképzett alakzatok segítsé-

gével tovább léphessünk egy még magasabb szintre, az emergens alakzatot *ugyanazon* tárgy különféle megjelenéseiként kellene vennünk. Ehhez viszont, és az ezekhez kapcsolódó kognitív teljesítményekhez, például ahhoz, hogy egy logikai következtetés (mint megismerési teljesítmény) ne legyen szétfolyó, s így maga a gép is számot adhasson menetéről, az egységet biztosító megoldásra, nevezetesen *fogalmakra* lenne szükség.

Világos persze, hogy a logikai feldolgozás mint megismerés az általában vett megismerés erősen szűkített fogalma, de tény, hogy az ilyen teljesítmények éppannyira jellemzik az emberi kogníciót. Ha tehát fenn akarjuk tartani, hogy *egyetlen* adekvát kognitív architektúra szerint kell megragadnunk a megismerést, akkor egy ilyen *univerzális* igényű számadásnak a magasabb kognitív teljesítményekkel is boldogulnia kell. Márpedig akár szubszimbolikus a megismerés, akár nem, nincs emberi megismerés *fogalmak* képviselte általános érvény nélkül. A hálózatok esetén ez pedig azt jelenti, hogy ha önálló fogalmi műveleteket akarunk modellezni, akkor eleve implementálnunk kell azokat a szemantikai struktúrákat, amelyeket modellezni kívánunk, pontosan úgy, ahogyan a szimbolikus modell esetében történik. Más szóval, a magasabb fogalmi szint struktúrái *aprioriak* a rendszer tapasztalataihoz képest, mert ezekből őket kiképezni (hatékonyan) nem lehetséges. Ez pedig ellentétben áll a hálózati modellezés kiindulópontjával, mert előzetes struktúrákat feltételez. Összefoglalva tehát, ha a megismerés megfelelően súlyozott kapcsolatok rendszerének kialakítása, az ismeret pedig nem más, mint kapcsolatok konfigurációja, akkor nincs megismerés, ha a rendszer nem képez ki olyan emergens egységeket is, amelyek minden szempontból képesek fogalmakként viselkedni. Mivel e ponton megint csak szinte kitapintható az összefüggés Kanttal, érdemes megállni és némi reflexióval összekötve, egybefoglalni azt, ami-re a két rivális modell típus kapcsán eddig jutottunk.

## (ÁTMENETI SZÁMADÁS)

Az alapproblémát így interpretáljuk tehát: a klasszikus kognitivisták architektúra szimbolikai fogalmak ugyan (enyhébben: kezelhetők fogalmakként), ám nincsenek közvetlenül a világban lehorgonyozva, a megosztott reprezentációk pedig szemantikailag ugyan lehorgonyoztak, ám nem fogalmak. Ha ez áll, akkor módunk van arra, hogy a két gépi architektúra problémáit a kanti megismerés két feltétele, a fogalom és a szemlélet felől közelítve kísérjük meg tárgyalni. A fogalomnak szemléleti tartalom kell (különben az elme saját gondolataiba záródik és egy láthatatlan világról kezd el mesélni), s ezt nyújtja az érzékiség egyedi szemlélete, a szemléletnek pedig fogalom kell, különben az elme hiába fogad magába képzeteket a tárgy által afficiálva (hiába alakítunk ki tehát reprezentációkat a környezet hatására), a megismerés folyamata megreked, a szemlélet vak lesz, hiszen ilyenkor ugyan nézünk, csak éppen nem látunk. Márpedig a szimbolikus architektúra gépeinek fogalmi üresek, míg a hálózatok kiképzett, alulról épülő megosztott reprezentációinak kialakításába nem játszanak bele a fogalmak. Megint másként: a szimbólumoknak „érzéki jelleget” kell kölcsönözni (KANT 1787/2004, B 75), hogy az aktuális világban lehorgonyozhassuk őket, a megosztott reprezentációknak pedig értelmi színezetet kell adni, hiszen a hálózat maga nem képes olyan magasabb egységek kialakítására, amely alá aztán újabb egyedi tapasztalatait (a mindig aktuális környezet további behatásait) közvetlenül szubszumálni tudná. Következés-

képp, ha e sejtelem nem csalóka, és a fenti, a három megismerési architektúra közötti analógia megáll a lábán, akkor ahhoz, hogy a kétféle mesterséges architektúrát elviekben összehétkétsük, tulajdonképpen csak Kant nyomába kell erednünk.

Egyelőre tekintsünk el attól, hogy mi az ára annak, ha meg akarjuk menteni az *a priori* megismerést úgy, hogy a tapasztalatról se kelljen lemondanunk. A későbbiekben erről lesz még szó. Annyit azonban mindenképpen érdemes előre megjegyezni, hogy a világmegismerés kérdése Kantnál ebből az igényből fakadóan az érzéki és az értelmi *képzetek összehangolásának* kérdésévé alakul át. Igaz, megmarad maga a dolog, mint afficiátor, amely a megismerési folyamat egyik terminusaként szemléletet, azaz közvetlenül az egyedi tárgyra vonatkozó képzetet (*repraesentatio*) gerjeszt az elmében, ám a fogalom hatóköre, mint tudjuk, *csakis* a jelenségekre lesz alkalmazható. Márpedig a jelenség empirikus úton keletkezett képzet, s a képzet végső soron – *elmemódosulás* (KANT 1781/2004, A 97). A mesterséges intelligencia kontextusában ez azonban azt jelenti, hogy a megismerés egységéhez a hibrid – tehát mindkét típus felhasználására építő – modellezésében *csakis* a szimbólumok és megosztott reprezentációk összekapcsolását kell megoldani, s a tulajdonképpeni probléma közege *csakis* a szűkebb értelemben vett rendszer *internális* kontextusa lesz. Ez kognitív szinten annyit tesz, hogy a megismerés problémáját nem szükséges egy rendszerfüggetlen világ és a megismerő rendszer összefüggésében, hanem elégséges a világ kialakított (érzéki) *reprezentációi* és a megismerő rendszer eleve adott (*a priori*) *reprezentációinak* összefüggésében tárgyalni. Később látni fogjuk, hogy ezzel a különféle kognitív teljesítményeket összehangoló tényezők jelöltjeinek sorából csak egyvalami esik ki: *maga a megismerendő világ*. Ehelyett viszont egy másféle problémabog keletkezik, méghozzá már magánál Kantnál: az érzéki és az értelmi képzetek, nevezetesen a tiszta értelmi fogalmak *különneműségének* problémája (KANT 1787/2004, B 176). Természetesen mindehhez majd ellenőriznünk kell annak jogosultságát is, hogy Kant képzetfogalmát egy lapon emlegethetjük-e a mesterséges intelligencia kutatás reprezentációfogalmával. A következőkben azt azonban elsőként minden további nélkül megvizsgálhatjuk, mit kezd az érzéki tapasztalás és fogalmi elgondolás heterogenitásával a könnigsbergi mester.

## (5) SEMATIZMUS ÉS TRANSCENDENTÁLIS APPERCEPCIÓ

Mint korábban felidéztük, Kant alaptézisében a következőt szögezi le: egyfelől, az értelem nélkül az érzékiség ad ugyan jelenséget, de nem adja egy empirikus megismerés tárgyát (KANT 1787/2004, A 124), röviden, az, amit ad, *nem a tárgy*, és másfelől, az érzékiség nélkül az értelem, bár fogalmakat bocsáthat a megismerés rendelkezésére, mégsem képes tárgyat adni, röviden a tárgyat ugyan elgondolhatóvá teszi, de *nem adja*. Ebből pozitíve: a megismeréshez tárgynak kell adódnia. Ám hogyan lehetséges egy olyan együttműködés, amely pontosan ezzel szolgál, ha a fogalmak és a szemléletek ily módon különeműek?

Mint ugyancsak ismeretes, Kant az érzékiség és az értelem összehangolásához egy közvetítő képességhez folyamodik, illetve e képesség produktumát illeszti közvetítőként a szemlélet és a fogalom közé, amely így „egyfelől a kategóriával, másfelől a jelenséggel [áll] az egyneműség viszonyában” (KANT 1787/2004, B 177). E közvetítő képesség a *képzelőerő*, terméke a *séma*, maga az összeillesztés művelete pedig a

*sematizmus*. Mivel azonban az egyneműség biztosítója egy olyan momentum, amely mind az érzéki, mind pedig a tiszta értelmi képzetek tekintetében meghatározó, nevezetesen az *idő* – mint a képzetek bármiféle *összekapcsolásának* formai feltétele –, először gondoljuk meg, mit von maga után egy ilyen problémamegoldó stratégia.

Vegyük rögtön észre, hogy az idő fogalmával az a logikai tér, amelyben érzékiség és értelem egybekapcsolódásának feltétele megfogalmazódott, egy újabb, *temporális* dimenzióval bővül, hiszen a megismerés két ágának összeboronálásához elsőként az érzéki képzetek és az értelmi fogalmak *egyenkénti* idői szintéziseit kell biztosítani. Másképpen, az időre való hivatkozás szükségessé teszi, hogy a különmemű képzetek összehangolásának mikéntjét egy mélyebb feltételrendszer közegébe ereszkedve vegyük fontolóra, és itt emeljük ki azokat a tényezőket, amelyek az egyáltalában vett képzet létrejöttéhez szükségesek.

Mindez világosan kitűnik a tiszta értelmi fogalmak dedukciójából *A tiszta ész kritikájának* első kiadásában. Itt Kant a megismerés szubjektív forrásaként három tényezőt nevez meg, az *érzék*et, a *képzelőtehetséget* és az *appercepciót*, mindhárom képesség funkcionálását idői (szukcesszív) szerkezetben felvázolva. Így egy érzéki képzethez az apprehenzió szintézise, míg egy asszociatív kapcsolat létrejöttéhez a reprodukció szintézise szükséges (KANT 1781/2004, A 98–102), ám a számunkra igazán izgalmas momentum a harmadik tényezővel bukkan fel.

Az *appercepció* feltételrendszerében ugyanis, tehát a fogalmak, vagyis az értelem ágán egy olyan mozzanat tűnik elő, amely mellett a gépi intelligencia kapcsán nem mehetünk el szó nélkül, nevezetesen a *tudat*. Még pontosabban, a fogalmak mint értelmi *egységek* szintézisének azon feltétele, hogy minden elgondoltról tudjuk, ugyanaz, mint amit az előző pillanatban gondoltunk el – egyszerűen, a *rekonstrukció* szintézise.

„[...] ez az egy tudat az, mely a sokféleséget, az egymás után megnézett és azután reprodukált dolgot is egy képzetben egyesíti. [...] [Így] mindig találni kell egy tudatot, még akkor is, ha hiányzik belőle a szembetűnő világosság, mert e tudat nélkül teljességgel lehetetlenek a fogalmak, s így lehetetlen a tárgyak megismerése.” (KANT 2004, A 103–104.)

Látható tehát, hogy e gondolatmenetet követve, az ismeret létrejötte felé tájékozódva, az apprehenzió (az érzékiség ága) és a reprodukció (a képzelőerő ága) szintéziséen túl fel kell tennünk egy átfogó egyesítő tényezőt az értelem ágán. Ahogyan Kant megfogalmazza:

„Egyetlen ismeret sem keletkezhet bennünk, s az ismereteknek semmilyen egymás közti összekapcsolása és egysége sem jöhet létre a tudat azon egysége nélkül, mely a szemléletek minden adatát megelőzi, és amelyre vonatkoztatva a tárgyakról alkotott minden képzet egyáltalán lehetségessé válik. E tiszta, eredendő változtathatatlan tudatot nevezem *transzcendentális appercepciónak*.” (KANT 2004, A 107.)

Amennyiben a tudat valamilyen formája nélkül nincs megismerés, akkor a transzcendentális appercepció nem más, mint annak mindennek előtt feltett lehetősége, hogy bármely képzetemet magamhoz tartozóként foghassam fel. Ez a felismerés pedig igen csak megrendítő. Ha ugyanis az volt az alapötletünk, hogy a transzcendentális és a mesterséges kognitív architektúrák között megpillantott analógián felbuzdulva, a gépek megismerési teljesítményeiben mutatkozó hiányosságokat a két architektúra kanti modellt követő egyesítésével orvosoljuk, akkor be kell látnunk, hogy ez az út egy komprehenzív tudat feltételezése nélkül járhatatlan. Erre a kérdésre később még visszatér-

rünk, de most nem erőltetjük a továbbhaladást, különösen azért nem, mert ha Kantnál a fogalom mint az értelmi képzet szintézise nem valósulhat meg valamely tudat nélkül, akkor ezt a képzetfogalmat nem vehetjük minden további nélkül ekvivalensnek a mesterségesintelligencia-kutatás reprezentációfogalmával. Ha tehát közös nevezőnek tekintjük a képzetet és a gépi intelligenciák reprezentációját, ezt csakis a legtágabb értelemben tehetjük. Kétségtelen persze, hogy a képzet (*Vorstellung*) szerkezetileg izomorf mind a szimbólummal, mind pedig a megosztott reprezentációval annyiban, hogy vonatkozást képez valamivel, ami tőle különbözik, nevezetesen a reprezentált tényállással. Ha tehát megelégszünk ennyivel, akkor nem lehet baj.

Tekintsünk el tehát az iménti kudarctól, és jusson eszünkbe, hogy Kant nemcsak reprodukív, hanem egy *másféle* képzelőerőről is beszél. Ez pedig talán módot ad arra, hogy ne az idő felől közelítsünk a sematizmushoz. Vizsgáljuk meg hát a *produktív képzelőerőt* is, amely képesség éppen azáltal különböztethető meg az érzékiségtől és rokonítható az értelemmel, hogy működéséhez nem szorul rá az afficiált érzékiségre, ám mégsem azonos az értelemmel, mert a sokféleséget *nem* intellektuálisan kapcsolja egybe (KANT 1787/2004, B 151). A produktív képzelőerő azon képességünk tehát, amely által „akkor is megjeleníthetjük egy tárgy képzetét, ha az szemléletünkben nincs jelen”, s így annyiban rokona az értelemnek, hogy *spontán* és ugyancsak „az érzékelés a *priori* meghatározásának képessége” (KANT 1787/2004, B 152). A kérdés persze éppen az, *hogyan* határozza meg az érzékelést. Ha a képzelőerő *spontán* is annyiban, hogy nem egyszerűen meghatározott, mint az érzékiség, *meghatározó* tevékenysége során mi módon lenne tekintettel a fogalmakra abban, ahogyan az egyedi szemléleteket egyesíti? És ezzel a kérdéssel visszaértünk kiindulópontunkhoz, a sematizmushoz.

Mint már említettük, az érzékiség és az értelem közötti közvetítőt Kant *sémának* nevezi, amely séma „önmagában soha nem egyéb a képzelőerő termékénél” (KANT 1787/2004, B 179). Emellett az is biztosan állítható róla, hogy nem *képmás*, vagyis nem olyan megjelenítés, amely valamit annak konkrét meghatározottságában ábrázol. A képmás érzéki adottság, s ha a séma ilyen lenne, akkor nem lehetne megoldás, mivel tovább is fennállna érzéki és értelmi képzet heterogenitásának problémája. Ha viszont éppen úgy jelenítené meg tárgyát, mint azt a fogalomban általánosságban elgondoljuk, akkor megint csak nem jutnánk előbbre.

Mi történik tehát, amikor egy fogalom kapcsán csak elképzeljük a neki megfelelő érzéki adottságot? Idéztük fentebb: a képzelőerő folytán akkor is megjeleníthetjük egy tárgy képzetét, ha az szemléletünkben nincs jelen, márpedig ha a fogalom útmutatása nyomán megjelenítünk valamit, de az érzékileg mégsem lesz adott, akkor az nem is korlátozódik egy konkrétumra, mint ahogy egy képmás. Mégsem semmi, sőt, éppen elegendő ahhoz, hogy egy adott szemlélet számára meghatározóvá váljon. Lássuk, mi teszi ezt lehetővé.

„[Az empirikus fogalom] mindenkor közvetlenül a képzelőtehetség sémájára vonatkozik, mint olyan *szabályra*, amelynek nyomán egy bizonyos általános fogalomnak megfelelően határozzuk meg szemléletünket.” (KANT 1787/2004, B 180. Kiemelés tőlem – K. L.)

Továbbra is kérdés persze, hogy miképp, ám egy lépéssel mégis előbbre jutottunk. Ha ugyanis a képzetet a legtágabb értelemben véve, pusztán valamire való vonatkozásában vesszük, és a sémát csakis formális szempontból, a *szabály* felől közelítjük meg, akkor a vizsgálódás kiterjeszhető. Így feltehető és megvizsgálható egy *mester-*

séges képzelőerő, mint olyan eljárás lehetősége, amely valamilyen módon, valamely szabály segítségével biztosítja a hálózat aktivitásmintázatainak alapuló megosztott reprezentáció egyesítését a szimbólumokkal.

## (6) EGY MESTERSÉGES KÉPZELŐERŐ KRITIKÁJA

Ismeretes, hogy a *kritika* kifejezés, ha Kanttal kapcsolatosan használjuk, vizsgálatot jelent. Egy mesterséges képzelőerő kritikája ennyiben tehát nem több, mint annak feltérképezése, milyen elvi keretek szabnak határt egy efféle képesség feltételezésének. A kritika szó másik, a mi korunkhoz közelebb álló jelentése, amikor a bírálat értelmében használjuk, így egyelőre nem is merül fel, ám hamarosan látni fogjuk, végül mégiscsak előkerül.

Kezdeként eleveintsük fel az alapproblémát: a gépek egyik típusa erős a fogalmak, a másik típus pedig a szemléletek terén. Az összehangoláshoz tehát olyasvalami kellene, amely bilaterális s mint ilyen, mindkét kognitív architektúrában gyökeret ereszt. Első interpretációnk szerint Kantnál a szemléleti és a fogalmi ág összehoronzását egy alapvető formális mozzanat, nevezetesen a képzet mint olyan temporalitása alapján végezhetjük el. Ily módon viszont legvégül a tudathoz kell folyamodnunk, mint a képzetszintézis végső biztosítékához. Elsőként tehát afelől kell bizonyosságot nyernünk, hogy jogosan állunk meg e ponton, mert a mesterséges intelligencia kontextusában valóban nem lehetséges olyan tényező, amely kiváltani lenne képes a transzcendentális appercepció „gondolkodom”-ját. Másodsor: a produktív képzelőerő kapcsán felvettük annak lehetőségét, vajon pusztán formális alapon, a sémát csakis a szabály fogalma alapján meghatározva, kapcsolat teremthető-e érzékiesség és értelem között. Haladjunk sorjában.

Vegyük tehát az értelmi képzetet úgy, ahogyan Kant veszi, vagyis mint olyan szintetikus terméket, amelynek egységét végső soron a transzcendentális appercepció funkciója garantálja. Kant számára persze értelmetlen feltenni azt, hogy létezhet az elmében olyan képzet, amely nem tartozhat valamely „gondolkodom”-hoz, mégsem teljesen reménytelen a kérdés, vajon más, *nem fenomenológiai* tényező is betöltheti-e a különféle képzetek összekapcsolásának funkcióját, ha az elmét mint a megismerés lókusztát csakis harmadik személyű perspektívából közelítjük meg, mint ahogyan tesszük ezt a mesterséges intelligencia kutatásában. Ha ugyanis fel tudnánk mutatni valamit, ami alkalmas a *transzcendentális appercepció kiváltására*, tehát éppen olyan módon egyesíti egy összetett rendszer különféle reprezentációit, mint a „gondolkodom” teszi Kantnál, akkor meg lenne a keresett tényező, amely kantiánus módon házasítja össze a kétféle géptípus reprezentációit. Ez pedig annyit tesz, hogy az előző fejezet zsákutcája talán mégis járható útnak bizonyulna.

Lássuk csak: ha pusztán arról van szó, hogy egyazon rendszerhez tartozóként kell elismerni egy reprezentációt, vajon nem alkalmas-e a programozó arra, hogy ezt megtegye? Sőt, még tovább menve, maga a gép fizikai felépítése és viselkedése – hiszen harmadik személyű perspektívából máshoz, még ha lenne is, nem férünk hozzá – nem szolgáltat-e elégséges alapot annak jogosultságára, hogy úgy ítéljünk, maga a gép valamely reprezentációt relevánsnak tekint és így magához tartozóként ismer el?

Vegyük először ez utóbbi lehetőséget: vajon a gép fizikai architektúrája elégséges-e ahhoz, hogy a kétféle géptípusból egyetlen rendszer keletkezzék? A szubszimbo-

likus kognitív architektúra tárgyalása során említettük, hogy a legalapvetőbb fizikai szint szemantikailag intranszparens réteg. Míg tehát a szimbolikus architektúra fizikai strukturálásából elvileg akár ki is olvasható, hogy egy adott rész a világ mely aspektusát reprezentálja, hiszen a rendszer egységei és kapcsolatai egy az egyben fennálló megfeleltetésben állnak a reprezentált területtel, addig a megosztott reprezentáció a rendszer fizikai architektúráján nem pillantható meg. Hiába is illesztenénk össze a két gépet, pusztán fizikai alapon, a PDP architektúrája miatt nem dönthető el, hogy egy bizonyos reprezentáció a rendszerhez tartozik-e vagy sem: az egybekapcsolt gépnek először valamit produkálnia kell, vagyis megismerőként kell *viselkednie*.

Am egy ilyen gép viselkedése szolgáltatathat-e alapot arra, hogy különféle reprezentációkat egységbe fogva megszervezze önnön kognitív egységét s így módon valósítsa meg a kétféle reprezentáció egyesítését? Nem arról van-e inkább szó, hogy a megfelelő viselkedés *cél*, nem pedig elvi kiindulópont? Nézzük csak, ha egyszer egy hibrid modell úgy viselkedik, hogy tanúságát adja: mind az érzékiség, mind pedig a racionalitás területén elboldogul, továbbá arra is képes, hogy e két területet összekapcsolja, méghozzá legalább úgy, mint egy ember, akkor már *megoldottuk*, méghozzá a gyakorlatban, a két gépi architektúra egyesítésének a problémáját, s így a mesterséges képzelőerő egész eddigi elméleti firtatása túlhaladottnak tekinthető.

Vegyük hát a másik esetet, hiszen a programozó nyilván alkalmas fóruma lehet annak, hogy döntsön, valamely reprezentációt a megismerő rendszerhez sorol-e vagy sem. Am ily módon, mint egyesítő tényező, rendszeren kívüli, s amennyiben rá bízunk a kétféle típus egyesítéséhez szükséges, eleve adotként feltételezendő szintetizáló funkciót, úgy mindig a rendszer felett kell bábáskodnia. Itt persze ellentétként fel lehetne hozni, hogy Kantnál sem arról van szó, hogy minden egyes alkalommal, minden egyes képzet kapcsán bölintanom kell: igen, ez a tudatomhoz tartozik, hanem arról az elvi lehetőségről, hogy ezt megtehessem. A tényleges azonosítás amúgy is kivitelezhetetlen lenne, mivel az „empirikus tudat önmagában szétszórt valami, és nem függ össze a szubjektum azonosságával” (KANT 1787/2004, B 133). Am, ha a programozó ez az egyesítő tényező, és éppen az emberi megismerés modellezéséről van szó, akkor az ő esetében is fel kell tennünk egy, nem egyszerűen az ő „empirikus tudatán” túli tényezőt, hanem egy, az adott megismerési apparátuson teljes mértékben kívül álló tényezőt. Vagyis csak arrébb toltuk a problémát, márpedig ez jelen esetben regresszust indít be.

Ha tehát ily módon nem boldogulunk, akkor a transzcendentális appercepció útján valóban nem haladhatunk tovább, mert nem tudjuk mással kiváltani, így viszont végső soron csakis a tudat garantálhatja érzékiség és értelem megismerésként való egyesítését. Következésképp, marad a képzelőerő. Tartsuk azonban szem előtt azt is, hogy ha lemondunk a tudatról, akkor le kell mondanunk a kanti képzetfogalomról is. A gépi reprezentációk tehát nem valamely tudat által szintetizált egységek, hanem pusztán valamire vonatkozó entitások. Lássuk, milyen esélyünk van a megosztott reprezentáció és a szimbólumok házasítására a produktív képzelőerő felől közelítve.

A *tiszta ész kritikájában* az objektív percepció ismeret, amely szemléletként (*intuitio*) közvetlenül vonatkozik a tárgyra, ám ha ez a vonatkozás közvetve történik, akkor fogalom (*conceptus*). De mi ez a közvetítő? Miként Kant mondja, a fogalom valamely *ismérv* által vonatkozik tárgyára, amely sok dologban közös (KANT 1787/2004, B 377). A szemlélet viszont, mivel közvetlenül vonatkozik a tárgyra, egyedi. Következésképp, a fogalom szemével nézve csakis általánosságot látnánk (ha Kant nem tagadná az

intellektuális szemlélet lehetőségét, ám tagadja), a szemlélet fejével pedig egyedieket gondolhatnánk el (ha nem lenne értelmetlen feltenni, hogy minden egyes szemléletben adott individualitásra külön bejáratú fogalmat tartunk fenn). Ha tehát valamilyen szabály fogalma felől kell összehangolnunk az egyedít nyújtó szemléletet (valamely individuumra vonatkozó szenzorikus reprezentációt) és az általánost megragadó fogalmat (általános érvényű szimbólumot), akkor *egyedi és általános* között kell kapcsolódást találnunk. Vegyünk tehát Kant nyomán egy empirikus fogalmat, a *kutya* fogalmát és nézzük meg, hogyan határozható ez meg.

„A kutya fogalma valamilyen szabályt jelent, amelynek nyomán képzelőtehetségem általánosságban meg tudja rajzolni egy négylábú állat alakját anélkül, hogy [...] a tapasztalat kínálta egy-egy különös alakra kellene korlátozódnia.” (KANT 1787/2004, B 180.)

Ezzel tehát körülírtuk egy érzéki fogalom (jelen esetben a kutya) sémájának létrehozását, úgy, ahogyan azt a képzelőerő az értelemre való tekintettel előállítja. Igen ám, csak hogy bennünket éppen az érdekel, hogyan eszközölhető egy ilyen művelet valamely szabály alapján akkor, ha csakis a harmadik személyű perspektíva adottságaira hagyatkozhatunk. Mivel e perspektívából a képzelet produktumaihoz egyébként sem férhetünk hozzá (a tudatra, a fentebb kifejtett okok folytán amúgy sem apellálhatnánk), kénytelenek vagyunk a szabályt, amelyre hagyatkozva mi, emberek, megrajzoljuk egy négylábú állat képzeletbeli alakját, a két szélső pólus felől meghatározni. És most jön a meglepetés: akár a fogalom, akár a szemlélet felől kiindulva próbálkozunk is ezzel, *végtelenségbe* ütközünk.

Vegyük kiindulópontként az értelmet. Tudjuk, a fogalmakhoz hozzá kell adni a tárgyat. Csakis a fogalmakra tekintettel így ugyan megadhatunk formális szabályokat, amelyek alapján adott individuumokról eldönthető, hogy ez alá a fogalom alá tartoznak-e vagy sem. Ám a mesterséges intelligencia kontextusában, ha a szimbólumokat vesszük kiindulópontul, és a világ egyedi dolgait előzetesen megadott szabályok alapján reméljük osztályozni, akkor meglehetősen rigid rendszert kapunk, amelynek kielégítő működéséhez árnyalásra, s tulajdonképpen a szabályok túlbujánzó szaporítására lenne szükség. *Képzeljük csak el* (ha már egyszer mi, emberek, ezt minden további nélkül megtehetjük), hányféleképpen jelenhet meg csakis vizuálisan egy kutya, ha beleértjük ebbe az általunk jól identifikálható, de hiányos megjelenéseket is, s utána gondoljunk bele, hogy milyen kihívást jelenthet mindezt csakis formális szabályok segítségével körülhatárolni.

Ám ha a másik irányból kísérjük meg szabályok megfogalmazását, akkor se jutunk messzebb. Láttuk, a szemléleteket fogalmak alá kell rendelni. De az egyedi felől indulva éppen arra nincs szabály, hogy melyik általános alá szubszumáljuk az adott tapasztalatot. Ez az egyedi ide is tartozhat és oda is, ám hogy melyik szabályt tekintjük mervadónak, ahhoz újabb szabályt kellene bevezetni és így tovább. Így, ha a fogalom felől indulunk, és a világ érzékek közvetítette eseményeit kell kategorizálnunk, akkor nem probléma a szubszumpció, csak éppen a végtelen változatosság okoz gondot a szabályok megfogalmazásában, ha pedig az egyedi esemény felől közelítjük meg a fogalmak alá rendelést, akkor az ítélőerő végtelen regresszusába ütközünk (KANT 2004, B 172).

Ha kiolvashatjuk is tehát a *Kritikából*, hogy egy fogalom sémája szabályt követ, ezek után már tudhatjuk, hogy ez a szabály nem lehet formális eljárás. Csakhogy, ha nem ilyen, akkor pozitíve mi jellemzi? Hogyan is rajzoljuk meg azt a képzeletbeli alakot?

Sajnos Kant e ponton nem árul el többet, sőt, le is zárja a továbbhaladást a sokat idézett passzussal.

„Értelmünknek a jelenségekre és pusztá formájukra vonatkoztatott sematizmusa az emberi lélek mélyén rejtőző művészet, melynek valódi fogásait aligha olvashatjuk ki a természetből: soha nem tárulnak fel a szemünk előtt.” (KANT 1787/2004, B 180–181)

Következésképp, Kant nyomában haladva, ezen az úton sem jutottunk tovább. Ismeretes azonban e pont *metafizikai* jelentősége, mind a klasszikus német idealizmus, mind pedig Heidegger számára (vö. TENGYELI 1988, 98–99; HEIDEGGER 2000, 135), nem hagyhatjuk tehát szó nélkül. Ha azonban ahhoz, ami a mélyben rejtőzik, sem introspektíve, sem pedig pusztán formális szabályok segítségével nem férünk hozzá, ám továbbra is kitarunk Kant és a mesterségesintelligencia-kutatás analógiája mellett, akkor be kell végeznünk a kanti értelmében vett kritikát, és át kell térnünk a mai értelemben vett kritikára. E ponton ugyanis az egész vizsgálódásnak véget lehetne vetni azzal a felvetéssel, hogy a mesterséges képzelőtehetség lehetetlensége semmit sem jelent, mert az analógia *egyszerűen nem áll fenn*. Ha tehát ezt a párhuzamot továbbra is fenn akarjuk tartani, akkor hangnemet kell váltanunk, egyúttal tudatosítva a kritika közegét is, nevezetesen azt, hogy immár metafizikai dimenzióban járunk.

Nézzünk tehát körül, elsőként visszafelé: ha sem a tudatot, sem pedig a képzelőerő természet adta művészetét nem tudjuk kiváltani mesterségesen, vagyis harmadik személyű perspektívából nincs egyesítő fórum, továbbá a szabályra mint formális egyesítő elvre nem hivatkozhatunk, akkor mi marad még egyáltalán? E helyzetben, a fenomenológiát kényszerűen nélkülözve, a szabály felől pedig mindkét irányból a végtelenségbe sodródva, van még egyáltalán olyan közös elvi momentum, amelyhez folyamodva netán kapcsolatba hozhatjuk egymással a kétféle gépi architektúrát?

Szeretnénk azt mondani, hogy igen: *maga a világ*. Csakhogy, ha ebben az irányban kívánunk keresgélni, akkor el kell hagynunk azt a keretet is, amelyben Kant és a mesterséges intelligencia kutatása mozog, s amelyben az elmefüggetlen, illetve rendszerfüggetlen világ csak a megismerés egyik ágában, nevezetesen az empirista vektor kiindulópontjaként kap szerepet, s csakis érzéki *képzet*ként (a Ding-an-sich által afficiált elme képzeteként), illetve *reprezentáció*ként (a környezeti hatást reprezentáló fizikai mintázatként). Ne feledjük, Kantnál a fogalom kérlelhetetlenül rá van utalva az érzékiségre, és nem férhet közvetlenül hozzá a dolgokhoz úgy, amint azok vannak. S ugyanígy, a kognitívizmus szimbólumai sem képesek közvetlenül a világban gyökeret verni. Ez pedig annyit tesz, mint korábban már utaltunk rá, hogy a megismerés ilyen megközelítésében a világot a megismerő rendszerben a világ érzéki reprezentációja *helyettesíti*. Kantnál szubjektíve, vagyis csakis a megismerő rendszerre való tekintettel ez a helyettesítés nem okoz ugyan gondot, mert a transzcendentális appercepció mint feltétel elégséges alapot nyújt a megismeréshez, vagyis a világ afficiált érzéki képzetek és csakis a jelenségekre érvényesíthető fogalmak temporális összehangoltságához. Ám a gépek esetén ilyen tényezőt nem leltünk fel, formális szabályokkal pedig hiába próbálkoztunk. Fel kell tehát vetnünk annak lehetőségét, hogy egyáltalán alkalmas metafizikai keretben mozog-e a mesterséges megismerés kutatása.

## (7) METAFIZIKAI KULCSOK

Ha valóban felállítható egy összefüggés, amely szerint a világ helyettesítése a világ reprezentációjával azzal jár, hogy fenomenológiai tényező (első személyű perspektíva) bevonása nélkül elveszik az érzékiség és értelem, az egyedi tapasztalat és az általános fogalom összehangolásának lehetősége, a megismerés mesterséges modellezésében pedig a szimbolikus és a szubszimbolikus architektúra egyesítése, akkor ebből a helyzetből kiutat a *világ* csak akkor jelenthet, ha egy olyan metafizika felé mozdulunk el, amely ezt a helyettesítést nem engedi meg. Mivel azonban ekkor fenomenológia híján leszünk, nem a *realizmus* felé kell tájékozódni, hanem egy olyan metafizika felé, amely a világot és a megismerő rendszert nem a reprezentáció fogalmára hagyatkozva rendeli egymáshoz. Csakhogy lehetséges-e egyáltalán olyan elmélet, amely megfelel e kívánalomnak? A válasz érdekében nézzünk magunk mögé ismét egy pillanatra.

Mint tudjuk, értelem és érzékiség eddig taglalt problémájának tulajdonképpen közege a *racionalizmus* és az *empirizmus* azon vitája, amelyet éppen Kant simított el a maga sajátlagos megoldásával. Ismeretes, ha azt valljuk – nagyon egyszerűsítve az álláspontokat –, hogy a megismerésben nincsenek velünk született ideák, akkor *mindent*, amit az elme csak tartalmaz, az egyedi, faktikus tapasztalásból kell valamiképpen kinyerni. Márpedig így baj lesz, mert az olyan fogalmakról, amelyek érvényességét nem lehet egyedi esetek alapján behatárolni, le kell mondanunk a tudományban (a világmegismerésben), lévén, hogy ha fel is bukkannak az elmében, azon jogukat, hogy a világra vonatkoztatva használhassuk őket, nem leszünk képesek csak a tapasztalásra való hivatkozással dedukálni. Röviden, ha következetesek vagyunk, *szkeptikusnak* kell lennünk, ezzel pedig végzetes sebet ejtünk a természettudományon. Másrészt, ha a fogalmakból indulunk ki (hogy már eleve immúnissá tegyük őket a kétellyel szemben), és felteszünk egy innát eszköztárat, amelyet éppen az olyan kínos problémákra való tekintettel állítunk össze, mint a szükségszerűség, az okság és így tovább, akkor az emberi megismerést a *dogmatizmus* karjába taszítjuk, hiszen a tudománynak elegendő lesz *csakis* ezekre a képzetekre hagyatkoznia ahhoz, hogy a világot feltérképezhesse, és adott esetben a tapasztaláshoz egyáltalán nem is szükséges folyamodnia.

Mindez ismételtén arra int, hogy sokkal mélyebb régiókban rejlik az az alapvető probléma, amelybe a mesterséges intelligencia kutatása – a kognitív tudomány térhódításától vérszemet kapva és pszichológiailag releváns megismeréseméletek megfogalmazását célozva készakarva, a technikai területek felől érkezve pedig talán akaratlanul, de végzetesen – belebonyolódott. Ha mindaz, amit eddig állítottunk, áll, akkor a gépi architektúrák gondolkörének mélyén maga Kant rejtőzik, egészen pontosan az a *metafizikai diszkusszió*, amelyben az empirizmus vagy racionalizmus dilemmája értelmes, megoldást sürgető probléma. A szimbolikus kognitív architektúra felmenője így a racionalizmus, míg a szubszimbolikusé az empirizmus (vö. PLÉH 1996, 299), méghozzá oly módon, hogy egy bizonyos, alapvető metafizikai vonást e kései utódok fizimiskáján is felfedezhetünk.

A gépi architektúrák fentebb felvázolt szerkezetének ismeretében talán megengedhető azt állítani, hogy a szimbolikus architektúra alapján működtetett gép egy olyan dogmatikus, aki *a priori* módon ugyan sok mindenről képes beszámolni, ám ha az *ehhez* a világhoz való konformitást várjuk el tőle, akkor folytonos *a posteriori* igazítgatásra szorul; a szubszimbolikus architektúra szerint működő hálózat pedig egy olyan empi-

rista, aki egyedi tapasztalatokhoz jut a világról, ám csakis ezekből építkezve a tulajdonképpeni logikai feldolgozáshoz mégsem képes felemelkedni, s teljesítménye így végső soron egy *a priori* struktúra beültetésével javítható.

Kant szerint azonban az *ember* az érzékiségre is rászoruló, ám tiszta logikai feldolgozásra is képes lény. Amennyiben tehát a kétféle géptípus példányainak viselkedését úgy értelmezzük, mint a *két komponensből* szerveződő emberi megismerés ekként vagy akként hiányos változatait, akkor ezek összedolgozásához értelemszerűen egy közvetítő szükséges. Ezt hiposztazáltuk fentebb mesterséges képzelőerőként, majd amellet érveltünk, hogy ez nem működőképes. Csakhogy *feltétlenül* el kell-e kötelezni magunkat amellet, hogy az emberi megismerés érzékiségre szoruló, *a priori* elveket követő logikai feldolgozás? Kikerülhetetlen netán, hogy a megismerést csakis érzéki és értelmi reprezentációk összehangolásának tekintsük?

Tudjuk, Kant számára az addig rejtőző metafizika kulcsát a tárgy és képzet kapcsolatának problémája adta meg (KANT 1772/1988b, 241). Ám az, amit Dilthey üzen az alábbi, jól ismert idézetben, nem ad-e kulcsot egy másféle metafizikához?

„[...] az eddigi ismeretelmélet, mind az empirista, mind Kanté, valamely – a pusztta megjelenítéshez [Vorstellen] tartozó – tényállásból magyarázta a tapasztalatot és a megismerést. A Locke, Hume és Kant konstituálta megismerő szubjektum ereiben nem valódi vér folyik, hanem a pusztta gondolkodási tevékenységnek tekintett ész híg leve.” (DILTHEY 1974, 66.)

Ha azonban Kant szemére vethető, hogy az a metafizika, amelynek alapkérdése képzet és tárgy kapcsolata, a megismerést a *pusztta megjelenítéshez* rögzíti, ily módon pedig reprezentációk relációira redukálja, akkor ugyanebben a mesterséges intelligencia kutatása is elmarasztható. Egy *agens* – tehát egy *tényleges* környezetben *ténylegesen* cselekvő kognitív lény – esetén ugyanis *maga a világ* jut szóhoz, hiszen a világot semmilyen képviselőt sem helyettesítheti, ha egy lény arra kényszerül, hogy benne létezzen. Egy világba ágyazott lény fogalma így nem ismeretelméleti, hanem sokkal inkább *ontológiai* (metafizikai) kifejtést kíván. S itt adódik mód arra, hogy e fejezet kiinduló kérdését megválaszoljuk. Ha ugyanis a megismeréshez ekként nem a mesterséges intelligencia episztemológiai fogalomtárával, hanem egy *mesterséges egzisztencia* lehetőségfeltételeit firtatva, a kognitív létezés ontológiai fogalma felől kell közelítenünk, akkor pontosan egy olyan metafizikának nyitunk teret, amelyben a világ és a megismerő *agens* összefüggését *nem* a reprezentáció fogalmára hagyatkozva interpretálhatjuk.

Más kérdés persze, hogyan fejthető ki egy fenomenológiát (első személyű perspektívát) nélkülöző, de nem a reprezentációs paradigmához kapcsolódó kutatás metafizikai kerete. Egy olyan megismerésemélet ugyanis, amely a megismerést a megismerő létezése felől közelíti meg, éppen a kontinentális hagyomány fenomenológiai vonulatában fogalmazódott meg a legpregnansabban (HEIDEGGER 1989, 165 skk.), arról nem is szólva, hogy a kortárs kognitív tudomány felől a megismerés bármiféle antirepresentációs elgondolása, első személyű perspektíva nélkül, egy gesztussal behaviorizmusnak minősíthető. Válasz híján egyelőre tehát megelégszünk annyival, hogy összefoglaljuk, hogyan jutottunk idáig, hogy egy kérdéssel hagyassuk nyitva a problémát.

Elsőként megvizsgáltuk az analógiát, amelyet egyfelől a klasszikus kognitívizmus modelljeinek esetén jelentkező szimbólum-lehorgonyzási probléma és Kant szemlélet nélküli fogalmai között, másfelől pedig a PDP-hálózatok megosztott reprezentációi és a fogalmak alá nem szubszumált szemléletek között állapítottunk meg, majd a

konfirmációból erőt merítve megkíséreltük a két gépi modell kantiánus összehangolását. E célból foglalkoztunk a sematizmussal és a reproductív képzelőerővel, ám ily módon a gépi kontextustól idegen appercepciófogalom felbukkanása kapcsán viszakozni kényszerültünk. Ezután más úton indultunk tovább, de a produktív képzelőerőt a szabály és az ítélőerő fogalmi felől megközelítve sem jutottunk tovább. Mindez ahhoz a gondolathoz vezetett, hogy a két gépi architektúra közötti utolsó közvetítő jelölt, a világ, csak azon metafizikai keret megváltoztatásával juthat szóhoz, amelyben a megismerés mind Kant, mind a mesterségesintelligencia-kutatás számára reprezentációk összehangolására korlátozódik. Következzék végül tehát a kérdés: ha az emberi megismerés mesterséges modellezése a megismerés fogalmát eddig egy olyan metafizikai felől közelítette meg, amelyet alapvetően a reprezentáció fogalma határoz meg, ám a megismerés mint a világban való létezés módja egy másféle metafizika felől is tárgyalható, akkor lehetőség nyílik az emberi megismerés egy egészen *másféle* gépi modellezésére, s végső soron a megismerés *másféle* természettudományos megközelítésére is?

## IRODALOM

- CLARK, A. 1996. *A megismerés építőkövei. Filozófia, megismeréstudomány és a párhuzamos megosztott feldolgozás*. Ford.: Pléh Csaba. Budapest: Osiris.
- DILTHEY, W. 1974. Bevezetés a szellemtudományokba. Kísérlet a társadalom és a történelem tanulmányozásának alapvetésére. Előszó. In *A történelmi világ felépítése a szellemtudományokban. Tanulmányok*. Ford.: Erdélyi Ágnes. Budapest: Gondolat.
- DREYFUS, H. 1972. *What Computers Can't Do*. New York: Harper & Row.
- FODOR, J. – PLYSHYN, Z. 1988. Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition*, 28.
- HEIDEGGER, M. 1989. *Lét és idő*. Ford.: Vajda Mihály et al. Budapest: Gondolat.
- HEIDEGGER, M. 2000. *Kant és a metafizika problémája*. Ford.: Ábrahám Zoltán, Menyes Csaba. Budapest: Osiris.
- HERNAD, I. 1996. A szimbólum-lehorgonyzás problémája. In Pléh Cs. (szerk.): *Kognitív tudomány*. Budapest: Osiris–Láthatatlan Kollégium.
- KANT, I. 2004. *A tiszta ész kritikája*. Ford.: Kis János. Budapest: Atlantisz.
- KANT, I. 1988a. Az érzékelhető és az értelemmel felfogható világ formájáról és elveiről. Ford.: Tengelyi László. In Tengelyi L.: *Kant*. Budapest: Kossuth. 179–228.
- KANT, I. 1988b. Levelek Marcus Herzhez. Második levél, 1772. február 21. Ford.: Vidrányi Katalin. In Tengelyi L.: *Kant*. Budapest: Kossuth. 240–247.
- NEWELL, A. – SIMON, H. 1981. Computer Science as Empirical Inquiry. In Haugeland, J. (ed.): *Mind Design*. Cambridge: MIT Press.
- PLÉH Cs. 1996. A gondolat elveszett rendjének nyomában. In Clark, A.: *A megismerés építőkövei. Filozófia, megismeréstudomány és a párhuzamos megosztott feldolgozás*. Budapest: Osiris. 295–307.
- RUMELHART, D. – MCCLELLAND, J. 1988. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition 1–2* Cambridge: MIT Press.
- SMOLENSKY, P. 1996. A konnekcionizmus helyes kezeléséről. In Pléh Cs. (szerk.): *Kognitív tudomány*. Budapest: Osiris.
- TENGYELI L. 1988. *Kant*. Budapest: Kossuth.