

A kognitív fenomenológiai érv a testetlen mesterséges intelligencia tudatossága mellett

Cody Turner

University of Connecticut, USA

A végleges változat megjelent: S. Gouveia (szerk.) *The Age of Artificial Intelligence*: (Vernon Press, 2020)

Ebben a fejezetben két újszerű érvet hozok fel az általam *erős primitivizmusnak* nevezett kognitív fenomenológia mellett, amely szerint létezik a megismerésnek egy olyan fenomenológiája, amely nem redukálható az érzékszervi fenomenológiára, és nem függ attól. Ezután azt állítom, hogy az erős primitivizmus azt jelenti, hogy a fenomenális tudatosság nem igényel érzékszervi feldolgozást. Ez utóbbi állítás a mesterséges intelligencia filozófiájára is hatással van. Ha ugyanis az érzékszervi feldolgozás nem szükséges feltétele a fenomenális tudatosságnak, akkor ebből logikusan következik, hogy a mesterséges intelligencia tudatossága (feltéve, hogy lehetséges) nem igényel megtestesülést. A dolgozat átfogó célja, hogy megmutassa, hogyan lehet az analitikus tudatfilozófia különböző témáit a mesterséges intelligencia filozófiájának egy fontos kérdésére vonatkoztatni.

Az itt vizsgált központi érv a "kognitív fenomenológiai érv a testetlen mesterséges intelligencia-tudat mellett". Ez a következő:

A kognitív fenomenológiai érv a testetlen mesterséges intelligencia tudatossága mellett:

P1. A kognitív fenomenológiára vonatkozó erős primitivizmus igaz.

P2. Ha a kognitív fenomenológiára vonatkozó erős primitivizmus igaz, akkor a fenomenális tudatosság nem igényel érzékszervi feldolgozást.

C1. Ezért a fenomenális tudatosság nem igényel érzékszervi feldolgozást.

P3. Ha a fenomenális tudatosság nem igényel érzékszervi feldolgozást, akkor a mesterséges intelligencia tudatossága (feltételezve, hogy lehetséges) nem igényel megtestesülést.

C2. Ezért a mesterséges intelligencia tudatossága (feltételezve, hogy lehetséges) nem igényel megtestesülést.

Mint jeleztük, először a P1-re összpontosítok, és két érvet mutatok be az erős primitivizmus mellett.

Az első érv egy fenomenológiai érv, amely azon a lehetőségen alapul, amit egyes buddhisták és misztikusok *tiszta tudatnak* neveznek (Smith 2011). A második érv abból merít, amit a *tudat magasabb rendű gondolkodás elméletének* neveznek (Rosenthal 1986, 1993, 2005). Fontos, hogy ezen érvek egyike sem arra irányul (legalábbis ebben a tanulmányban), hogy végérvényesen bizonyítsa a P1 igazságtartalmát. A cél

mindkét érvelés célja, hogy megmutassa, hogy az erős primitivizmus bizonyos elméleti feltételezések mellett hihető. ¹ A céloom itt, más szóval, csupán az alábbi két feltételrendszer megállapítása:

(CON 1) Ha a tiszta tudat lehetséges, akkor a kognitív fenomenológiára vonatkozó erős primitivizmus igaz.

(CON 2) Ha a tudat magasabb rendű gondolkodásemélete igaz, akkor a kognitív fenomenológiára vonatkozó erős primitivizmus igaz.

Fontos, hogy a kognitív fenomenológiára vonatkozó erős primitivizmus akkor is igaz lehet, ha a tiszta tudat lehetetlen, és a tudat magasabb rendű gondolkodás elmélete hamis. A tiszta tudat lehetőségének bizonyítása és a tudat magasabb rendű gondolkodáseméletének valódisága egyszerűen az érvelés két olyan útját jelenti, amely az erős primitivizmushoz vezet. Lehetnek más érvelési utak is.

A tanulmány felépítése a következő. Az I. szakasz bemutatja a kognitív fenomenológiai vitát az elmefilozófiában, és különbséget tesz a kognitív fenomenológia különböző nézetei között (Horgan & Tienson 2002, Strawson 2011, Smithies 2013). A II. és III. szakasz ezután a kognitív fenomenológiával kapcsolatos erős primitivizmus motiválását folytatja a tiszta tudat lehetőségére (II. szakasz) és a tudat magasabb rendű gondolkodáseméletére (III. szakasz) való hivatkozással. A IV. szakasz a P2 érvet motiválja az általam PHEN-nek nevezett elvre való hivatkozással, amely szerint a "fenomenológiai függetlenség" magában foglalja a "metafizikai függetlenséget". A PHEN szerint abból a tényből, hogy a kognitív fenomenológia az érzékszervi fenomenológia hiányában is instanciálható, az a következtetés következik, hogy a fenomenális tudat az agyi érzékszervi feldolgozás hiányában is instanciálható. Az V. szakasz ezután az érvelés P3. pontját motiválja a megtestesült megismerés irodalmában végzett munkára (Varela, Thompson & Rosch 1991) való hivatkozással.

A P3-mal szembeni ellenvetés azt követeli meg, hogy a megtestesülés szükséges a fenomenális tudatossághoz, még akkor is, ha a tudatosság nem függ az érzékszervi feldolgozástól. Azt állítom, hogy az ellenző számára lehetetlen ezt az állítást tenni, ha a megtestesülést (legalábbis részben) a szenzomotoros képességek gyakorlásának képességében határozzuk meg. Ezután megfontolom és végül elvetem azt a felfogást, hogy a tudat autopoietikus enaktivista nézetei (Varela 1997, Di Paolo 2005) a P3 ellenvetéseként felhozhatók. A VI. szakasz zárásként szolgál.

¹ Szigorúan szólva agnosztikus vagyok mind a tekintetben, hogy lehetséges-e a tiszta tudat, mind pedig a tekintetben, hogy igaz-e a tudat magasabb rendű gondolkodás elmélete.

I. A kognitív fenomenológiai vita

Az 1980-as évekig az analitikus elmefilozófiában az volt az általános nézet, hogy a fenomenalitást és a szándékosságot a mentális két különálló aspektusának tekintették. Ez a nézet, amelyet Terrance Horgan és John Tienson (2002) *szeparatizmusnak* nevez, azt vallja, hogy a fenomenális állapotok és a szándékos állapotok metafizikailag függetlenek egymástól, ami azt jelenti, hogy egyik állapotfajta sem a másikból keletkezik, és hogy egy rendszer az egyiket mutathatja, a másikat nem, legalábbis elvileg. A *fenomenálisan tudatos mentális állapot* olyan állapot, amely fenomenális jelleggel rendelkezik, vagyis van valami, amiben olyan, mintha az adott állapotban lenne (Nagel 1974). A *szándékos mentális állapot ezzel szemben* olyan állapot, amely "aboutness"-t mutat, vagy amely reprezentációs jellemzőkkel vagy mentális tartalommal rendelkezik (Fodor 1985). A szeparatizmust hagyományosan az a meggyőződés motiválja, hogy míg az előforduló *érzékszervi* mentális állapotok (pl. az állapot, amelyben akkor vagyunk, amikor aktívan látjuk a piros színt) fenomenális jelleggel rendelkeznek, de nem rendelkeznek intencionális tartalommal, addig az előforduló *kognitív* mentális állapotok (pl. az állapot, amelyben akkor vagyunk, amikor aktívan hisszük, hogy Hartford Connecticut fővárosa) intencionális tartalommal rendelkeznek, de nem rendelkeznek fenomenális jelleggel.

A *kognitív fenomenológia* (röviden CP) az elmefilozófia olyan fogalma, amely a szeparatizmus aláadására szolgál. A CP hívei azt állítják, hogy az előforduló kognitív állapotok mind szándékosan irányítottak, mind (legalábbis bizonyos értelemben) fenomenálisan tudatosak. E támogatók szerint a gondolkodáshoz tapasztalati jelleg társul, ahol a "gondolkodás" alatt a kijelentő magatartások széles skáláját értik, mint például a hit, a kétely, a csodálkozás, a vágyakozás, a szórakozás stb.

A fogalmi térben a kognitív fenomenológia négy különböző nézetét különböztethetjük meg:

- (1) **Eliminativizmus a CP-vel kapcsolatban:** a kognitív fenomenológia nem létezik.
- (2) **Redukcionizmus a CP-ről:** A kognitív fenomenológia redukálható az *érzékszervi fenomenológiára*.
- (3) **Gyenge primitivizmus a CP-ről:** A kognitív fenomenológia nem redukálható az *érzékszervi fenomenológiára*, de *függ az érzékszervi fenomenológiától*.
- (4) **Erős primitivizmus a CP-vel kapcsolatban:** A kognitív fenomenológia nem redukálható az *érzékszervi fenomenológiára*, és *független attól*.

Mielőtt ezt a négy álláspontot megvilágítanánk, fontos kiemelni, hogy a kognitív fenomenológia csak az *előforduló* kognitív állapotokra vonatkozik, vagyis azokra a kognitív állapotokra, amelyeket az alanyok aktívan szórakoztatnak vagy jelképeznek. A *diszpozicionális* kognitív állapotok, mint például az összes olyan meggyőződés, hogy én

jelenleg nem szórakoztató, nem fenomenálisan tudatos. Mert nincs semmi olyan, ami olyan hiedelem, amit jelenleg nem jelképez. Mindezek után forduljunk a CP-vel kapcsolatos négy álláspont felé.

A CP-vel kapcsolatos eliminativisták elismerik, hogy a megtörtént kognitív tevékenységek *hozzáférési tudatosságot* jelentenek, de tagadják, hogy az ilyen tevékenységek *fenomenálisan tudatosak* lennének. Egy mentális állapot akkor és csak akkor tudatos, ha elérhető a verbális jelentés, a racionális következtetés és/vagy a viselkedés szándékos irányítása számára (Block 1995). Miközben az eliminativisták elismerik, hogy az előforduló vágyak, hiedelmek, csodálkozások stb. rendelkezésre állnak ezekhez a felhasználásokhoz, elutasítják azt az elképzelést, hogy van valami, amiben ezekben a kognitív állapotokban lenni olyan, mintha az ember lenne. ²Ezzel szemben a redukcionista elismerik, hogy a megismerés fenomenológiát foglal magában, de fenntartják, hogy a kognitív fenomenológia a gondolkodást jellemzően kísérő érzékszervi fenomenológiával magyarázható, mint például a mentális képzetek és a belső beszéd.

Az eliminativizmus és a redukcionizmus közötti vita a CP-ről végső soron talán csak verbális vita. Ugyanez nem mondható el a redukcionizmus és a primitivizmus közötti vitáról. A primitivisták egyetértenek a redukcionistaakkal abban, hogy az érzékszervi fenomenológia különböző formái, például az auditív képzetek (azaz a belső beszéd) és a vizuális képzetek (azaz a mentális képek) gyakran kísérik gondolatainkat. A redukcionistaiktól eltérően azonban a primitivisták lemondanak arról az állításról, hogy ez a kísérő érzékszervi fenomenológia *kimeríti a megismerés tapasztalati jellegét*.

Gondoljunk a <a buszok sárgák> gondolatra. Ennek a gondolatnak minden egyes jelzője kapcsolódhat, és gyakran kapcsolódik is, egy külön érzékszervi jelenségkarakterhez, amely azt képviseli. Például az egyik esetben a gondolatot egy sárga busz mentális képe reprezentálhatja, míg egy másik esetben egy belső beszédbeli mondat, amely azt mondja: "A BUSZOK SÁRGAAK." A primitivisták azt állítják, hogy bár ugyanannak a gondolatnak ez a két tokeningje különbözik érzékszervi jellegük tekintetében, mégis van egy közös "mi-az-az-szerűség", amely végigvonul rajtuk. Ez a közös tapasztalati jelleg a primitivisták szerint egy *sui generis* kognitív fenomenológiát jelöl, amely nem redukálható az auditív és/vagy vizuális képek bármilyen kapcsolatára.

A gyenge primitivizmus és az erős primitivizmus közötti különbségtétel a CP-vel kapcsolatban ritkán jelenik meg a kognitív fenomenológiával foglalkozó szakirodalomban (Levine 2011 kivétel). Ezt a megkülönböztetést az a belátás motiválja, hogy a fenomenális irredukálhatóság nem vonja maga után a fenomenális függetlenséget. X fenomenálisan akkor és csak akkor függ Y-től, ha X nem instanciálható Y hiányában. A gyenge primitivizmus azt állítja, hogy míg a kognitív fenomenológia irreducibilis az érzékszervi fenomenológiára,

²Lásd Carruthers és Veillet (2011) érvelését az eliminativizmus mellett.

a kognitív fenomenológia nem tud megvalósulni az érzékszervi fenomenológia hiányában. Ahogy Levine megállapítja, a gyenge primitivizmus esetében "továbbra is az a helyzet, hogy egy kognitív tartalom fenomenálisan csak úgy tapasztalható meg, ha az valamilyen érzéki prezentációra gyakorolt hatásán keresztül valósul meg" (Levine 2011: 112). Az erős primitivizmus ezzel szemben azt állítja, hogy a kognitív fenomenológia mind fenomenálisan irreducálható, mind fenomenálisan független az érzékszervi fenomenológiától.

II. A "tisza tudatosság" érve az erős primitivizmus mellett

A CP-vel kapcsolatos primitivizmus mellett szóló legnépszerűbb érvek némelyike fenomenológiai jellegű. Ezek a fenomenológiai érvek gyakran *fenomenális kontrasztesetek* formájában jelennek meg (Chudnoff 2015). Például egy angolul beszélő fenomenológiáját szembeállítják egy olyan személy fenomenológiájával, aki nem beszél angolul. Arra kérnek bennünket, hogy az érvelés kedvéért képzeljük el, hogy mind az angolul beszélő, mind a nem angolul beszélő ugyanazt az angol mondatot hallja, amelyet ugyanaz a személy mond ki pontosan ugyanazon a helyen. Ebben a képzeletben, így szól az érvelés, a két beszélő között általános fenomenális különbség van, annak ellenére, hogy az érzékszervi fenomenológiájuk azonos. Ez a fenomenológiai felismerés azt sugallja, hogy a nyelv megértésével kapcsolatban kell lennie egy redukálhatatlan kognitív fenomenológiának.

Levine (2011) helyesen jegyzi meg, hogy a CP-vel kapcsolatos primitivizmus mellett szóló legtöbb létező fenomenológiai érv legfeljebb a gyenge primitivizmust támogatja (a gyenge primitivizmust "tisztátalan CP-nek", az erős primitivizmust pedig "tisza CP-nek" nevezi):³

"Amikor a tiszta és a tisztátalan CP közötti különbségtételről van szó, nyilvánvalónak tűnik, hogy ez az érv [a primitivizmus fenomenológiai kontraszt érve] nem támogathat többet, mint a tisztátalan CP-t. Hiszen az összes példa a vizuális megjelenések, hangáramok és hasonló megtapasztalásának módjait érinti. Az érv olyasvalakit szólít meg, aki a tudatos életében nem talál mást, mint érzéki fenomenológiát, és arra készíti, hogy észrevegye az ezen érzéki megjelenések közötti különbségeket, amelyek csak a kognitív behatolásra való hivatkozással magyarázhatók. Elég tisztességes. Mégis, ettől még nem jutunk el a tiszta CP-hez. Még mindig lehetséges, mindazok ellenére, amit ez az érv bizonyít, hogy egy kognitív tartalom csak úgy jelenhet meg egy tudatos szubjektum számára, hogy befolyásolja valamilyen érzékszervi sokféleség megjelenési módját. Ez a tisztátalan CP" (Levine 2011: 115- 116).

Ebben a fejezetben egy fenomenológiai érvet kínálok az erős primitivizmus mellett, amely a "tisza tudat" lehetőségén alapul. A tiszta tudat egy olyan fenomenális állapot, amely állítólag meditációval elérhető. Azt állítják, hogy a tiszta tudatállapotok során a szándékos

³ Figyelemre méltó kivétel Chudnoff (2015). Ott egy olyan fenomenális kontrasztos esetet mutat be, amely valójában a CP-vel kapcsolatos erős primitivizmust támasztja alá.

a tapasztalat struktúrája eltűnik, és nem marad más, mint egy nyitott fenomenális mező. David Woodruff Smith (2011) így írja le a tiszta tudat állapotát: "A kitartó meditációban eltűnik a tárgyérzet, azaz a szándékos tartalom vagy a tárgyat meghatározó érzék eltűnik, és eltűnik az én vagy a szubjektum érzete. Vagyis a szubjektum-objektum - vagy szubjektum-akt-tartalom-objektum - szokásos intencionális struktúrája feloldódik egy ilyen meditatív állapotban. Ekkor csak a tudatosságnak ez az áramló fázisa van" (Smith 2011: 481). A "tiszta tudat" érve az erős primitivizmus mellett a következőképpen ábrázolható:

A tiszta tudatosság érve az erős primitivizmus mellett:

PP1. A tiszta tudatosság lehetséges.

PP2. Ha a tiszta tudat lehetséges, akkor a kognitív fenomenológiára vonatkozó erős primitivizmus igaz.

C1. Ezért igaz az erős primitivizmus a kognitív fenomenológiáról.

A PP1-et a mély meditatív élmény fenomenológiája támasztja alá. Visszavághatunk a PP1 ellen, például azzal az érveléssel, hogy a meditáció megbízható ismeretelméleti útmutató a saját fenomenális állapotaink természetére vonatkozóan. Elutasíthatnánk a PP1-et azzal is, hogy támogatjuk Franz Brentano (1874) tézisét, miszerint az intencionalitás a mentális jelzője. Ha Brentanónak igaza van abban, hogy a mentális állapotok szükségszerűen intencionálisak, akkor ebből az következik, hogy a tisztán tudatos mentális állapotok lehetetlenek, mivel az ilyen állapotok állítólag nélkülözik a szándékosságot. Steven Katz (1978) ezen a vonalon érvel a tiszta tudatosság lehetősége ellen. Katz szerint a tiszta tudatállapotok azért elérhetetlenek, mert minden tudatos élményt (beleértve a misztikus élményeket is) valamilyen fogalmi séma jelenléte közvetít: "Magát az élményt, valamint azt a formát, amelyben beszámol róla, olyan fogalmak alakítják, amelyeket a misztikus hoz magával, és amelyek alakítják élményét" (Katz 1978). Ebben a tanulmányban a PP1 plauzibilitását magától értetődőnek veszem. Ismétlem, fő célom itt a PP2, vagyis annak a feltételes állításnak az igazságtartalmának a megállapítása, amely szerint a kognitív fenomenológiára vonatkozó erős primitivizmus igaz, ha a tiszta tudat lehetséges.

A PP2 első látásra tartható, mivel a tiszta tudatállapotok definíció szerint érzékszervi fenomenológia hiányában is léteznek. Más szóval, ha meditációval lehetséges olyan tudatállapotot elérni, amely mentes minden mentális tartalomtól (beleértve az érzékszervi tartalmakat is), akkor ebből az következik, hogy a kognitív fenomenológia nem redukálható az érzékszervi fenomenológiára, és független attól (azaz igaz az erős primitivizmus a kognitív fenomenológiára vonatkozóan). Én két utat látok, amit járhatnánk

a PP2 elleni tiltakozásról. Az első az, hogy *tagadjuk, hogy a tiszta tudat magában foglalja a fenomenális tudatot*. Talán a tiszta tudatállapotok állítólagos "üressége" annak köszönhető, hogy az ilyen állapotok nem rendelkeznek fenomenális jelleggel. Ha a tiszta tudati állapotok nélkülözik a fenomenális jelleget, akkor a tiszta tudat nem egyfajta kognitív fenomenológia, ellentétben azzal, amit a PP2 sugall. A tiszta tudatosság hívei azonban nagyon határozottan kijelentik, hogy a tiszta tudatállapotban való tartózkodáshoz társul egy *mi-mi-mi-szerűség*. Ahogy Smith hangsúlyozza: "Ez a meditatív élmény *fenomenális*: "fenomenálisan" jelenik meg a tudatban, és a mibenlét egyszerűen a tudatban való létezés, mintegy várva, hogy mi más jelenhet meg vagy nem jelenhet meg a tudatban" (Smith 2011: 481). Ez a Smith-idézet arra utal, hogy a fenomenalitás már a tiszta tudat fogalmába is bele van építve. Ha ez így van, akkor a jelenlegi ellenvetés (miszerint a tiszta tudat nem fenomenális) a lényege szerint inkább a PP1 elleni ellenvetésnek tűnik, mint a PP2 elleni ellenvetésnek, ahogyan azt korábban feltételeztük.

A PP2-vel szembeni ellenvetés egy másik módja az, ha *elismerjük, hogy a tiszta tudat fenomenális, de tagadjuk, hogy a tiszta tudat kognitív lenne*. Ezen ellenvetés szerint a tiszta tudat fenomenológiája nem tartalmaz sem érzékszervi fenomenológiát, sem kognitív fenomenológiát. Az ellenvetés szerint egy fenomenálisan tudatos X mentális állapot akkor és csak akkor minősül kognitívnek, ha X-et fenomenálisan tudatos kognitív attitűd és/vagy fenomenálisan tudatos kognitív tartalom jellemzi. A probléma az, hogy a tiszta tudatállapotok sem kognitív attitűdöket, sem kognitív tartalmakat nem foglalnak magukban, mert az ilyen állapotok definíciójuk szerint a tapasztalat intencionális struktúrájának eltűnésével szinonimák. A tiszta tudat fenomenológiája tehát nem egyfajta kognitív fenomenológia, vagyis a tiszta tudat lehetősége nem vonja maga után a kognitív fenomenológiára vonatkozó erős primitívizmust. Vagy legalábbis így szól az ellenvetés.

A PP2-vel szembeni ezen ellenvetésre az egyik válaszlehetőség az, hogy a látszat ellenére a tiszta tudatállapotokat valójában valamilyen kognitív attitűd és/vagy kognitív tartalom jellemzi, és ezért kognitív fenomenológiai állapotokat jelölnek. Más szóval, a PP2 híve azt állíthatja, hogy az az állítás, miszerint a tiszta tudat állapotai kognitív természetűek, összeegyeztethető a misztikusok azon állításával, miszerint a tiszta tudat állapotai nem mutatnak felismerhető intencionális struktúrát. E két állítás összeegyeztetésének egyik módja, ha azt állítjuk, hogy a tiszta tudat egy intencionális mentális állapot, de olyan intencionális állapot, amelynek tárgya a semmi fogalma, vagy talán maga a nem-intentionalitás fogalma. Másképpen fogalmazva, az elképzelés szerint a tiszta tudat olyan intencionális állapot lehet, amely a nem-reprezentáció állapotát képviseli. A tiszta tudat ilyen értelmezése lehetővé teszi a PP2 támogatója számára, hogy azt állítsa, hogy a tiszta tudat egyfajta

a kognitív fenomenológiának (mivel kognitív tartalmakat foglal magában), miközben elismeri, hogy a tiszta tudatállapotok első személyű, fenomenológiai szempontból mentesnek *tűnnek az* intencionalitástól.

A tiszta tudat e javasolt elemzése sok tekintetben hasonlít ahhoz, ahogyan az elmefilozófusok elemzik a *hangulatok* fenomenológiáját. A hangulatokról, például az elragadtatásról és a szorongásról is azt gondolják néha, hogy nem szándékos, irányítatlan mentális állapotok. Brentano tézisének (lásd fentebb) hívei azonban elutasítják a hangulatoknak ezt a nem-intencionális felfogását, és ehelyett azt állítják, hogy "miközben a hangulatok *látszólag* nem irányulnak semmire, közelebbről megvizsgálva kiderül, hogy valójában különleges típusú tárgyakra irányulnak: testi állapotokra vagy szokatlan külső tárgyakra, például a világ egészére, határozatlan szándékú tárgyakra vagy gyakran változó tárgyakra (lásd Goldie 2000, 2002, Seager 2002, Seager és Bourget 2007, Tye 2008)" (Mendelovici 2013: 8-9). Ahogyan egyes filozófusok megpróbálják megmagyarázni a hangulatok látszólagos nem-intentionálisát, itt arra utalok, hogy a PP2 hívei megpróbálhatják megmagyarázni a tiszta tudati állapotok látszólagos nem-intentionálisát. Pontosabban, ugyanúgy, ahogyan a szorongás vagy az elragadtatás állapotainak szándékos tárgya lehet "a világ egésze" vagy "minden", a PP2 híve azt sugallhatja, hogy a tiszta tudatállapotok tárgya "a szándékos világ hiánya" vagy "a semmi". Ha ez igaz, akkor nem meglepő, hogy a tiszta tudatállapotokról (tévesen) úgy számolnak be, hogy nem tartalmaznak szándékosságot.

A PP2 ellenzője a dialektika ezen pontján azzal vádolhat, hogy témát váltok, vagy legalábbis azzal, hogy félreértem a tiszta tudat fogalmát. A misztikusok állítása - ragaszkodhat az ellenző - nem pusztán az, hogy a tiszta tudat látszólag nélkülözi az intencionalitást; ez inkább az a szilárdabb állítás, hogy a tiszta tudat *valójában* nélkülözi az intencionalitást. Erre válaszul a PP2 támogatója talán csak megkettőzhetné a szavakat, és kijelenthetné, hogy a tiszta tudat éppen azért "kognitív", mert a tiszta tudati állapotok a fenomenológiai látszat ellenére valójában intencionálisak.

Lehet, hogy a tiszta tudat definíció szerint nem szándékos, és a fenti elképzelés, miszerint a tiszta tudat állapotai szándékosan képviselik a nem-reprezentáció állapotát, téves. Elismerem, hogy a tiszta tudat állapotai nem nevezhetők megfelelően "kognitív" fenomenális állapotoknak, ha ez a helyzet. Ugyanis ismétlem, a fenomenális állapotok plauzibilisen csak akkor minősülnek "kognitívnek", ha valamiféle szándékos kognitív attitűd és/vagy kognitív tartalom jellemzi őket. A tény azonban az, hogy az erős primitivizmus mellett fenomenológiai érv adható anélkül is, hogy a tiszta tudat fogalmára hivatkoznánk. *Még ha a tiszta tudat nem is egyfajta kognitív*

fenomenológia, nem nehez elképzelni egy kissé módosított fenomenológiai érvet az erős primitivizmus mellett, amely a meditáció fenomenológiáján alapul. Érvelhetnénk például azzal, hogy a "mindfulness" meditáció (vagy az érzékszervi deprivációs tartályban való elmerülés) során a kijelentéses tartalmak elszigetelt tudatosítása fenomenológiai bizonyítékot jelent az erős primitivizmus alátámasztására.

Smith szerint "a figyelmes önismeret a célja annak a gyakorlatnak, amelyet az egyik buddhista hagyományban "mindfulness"-nek vagy *vipassana-nak* neveznek. Ebben a gyakorlatban a meditatív tapasztalat előkészíti a tudatosság önmagában való megtapasztalását - úgymond a tiszta tudatosságot -, és ezután lehetővé teszi, hogy észrevegyünk bizonyos gondolatokat, érzéseket vagy érzeteket, amelyek a tudatfolyamban esetleg megjelennek" (Smith 2011: 481). A tételes tartalmak elszigetelt tudatossága nem tiszta tudatosság, de vitathatatlanul olyan fenomenális tapasztalat, amely hitelt ad az erős primitivizmusnak. Ha ugyanis azt mondjuk, hogy az ember fenomenális mezeje kimerül a propozicionális tartalmak tudatosításában, akkor azt mondjuk, hogy olyan tudatállapotot élünk át, amely kognitív fenomenológiát foglal magában, minden kísérő érzékszervi fenomenológia hiányában.

III. A "magasabb rendű gondolkodás" érve az erős primitivizmus mellett

Most rátérek az erős primitivizmus melletti második érvre, amelyet a magasabb rendű gondolkodás (HOT) tudatelmélete motivál. A tudat magasabb rendű gondolkodás elmélete, amelynek legfőbb képviselője David Rosenthal (1986, 1993, 2005), azt állítja, hogy egy X mentális állapot akkor válik tudatossá, ha X-et egy magasabb rendű gondolat tárgyának tekintjük. Rosenthal a HOT és az elsőrendű mentális állapot közötti kapcsolatot reprezentációs kapcsolatnak tekinti (szemben mondjuk az ismeretséggel). A HOT-elméletből az erős primitivizmusra való érvelés a következő alapszerkezetű:

A magasabb rendű gondolkodás érve az erős primitivizmus mellett:

QP1. A HOT elmélet igaz.

QP2. A HOT elmélet erős primitivizmust von maga után a CP-vel kapcsolatban.

C1. Ezért a CP-re vonatkozó erős primitivizmus igaz.

Két fontos, egymással összefüggő tisztázó megjegyzésre van szükség az érvelés hatályát illetően. Először is, az érvelés csak a HOT-elmélet azon változatait érinti, amelyek a *fenomenális tudatosság*, vagyis a tapasztalat "mi-az-az-szerűségének" magyarázatát célozzák. Erre a kikötésre azért van szükség, hogy kapcsolatot tudjunk teremteni a HOT-elmélet és az erős primitivizmus igazságtartalma között. A HOT-elmélet azonban mindenekelőtt az *állapottudat* elmélete, nem pedig a fenomenális tudaté. A

A "tudatállapot" kifejezés, ahogyan itt használom, azt az elképzelést jelöli, hogy a tudatos mentális állapot olyan állapot, amelyben az ember tudatában van annak, hogy van. Tehát S alany X mentális állapota akkor és csak akkor rendelkezik állapottudattal, ha S tudatában van X-nek. Az állapottudat és a fenomenális tudat nem feltétlenül ugyanaz a dolog, ami azt jelenti, hogy a HOT-elméletek nem feltétlenül a fenomenális tudat elméletei.

Valójában jelentős viták folynak arról, hogy mi a kapcsolat az állapottudat és a fenomenális tudatosság között (ha van egyáltalán). Egyes filozófusok, mint például Lycan (1996), azt állítják, hogy az állapottudat a fenomenális tudat szükséges feltétele. Mások, mint például Block (2011), úgy vélik, hogy a fenomenális tudatosság felülmúlja az állapottudatot abban az értelemben, hogy valaki annak ellenére lehet fenomenálisan tudatos X-ről, hogy nincs tudatában X-nek. Én semleges maradok az állapottudat és a fenomenális tudatosság viszonyát illetően. Az egyetlen pont, ami e vita szempontjából számít, az az, hogy az erős primitivizmus melletti HOT-érv csak a HOT-elmélet azon változataira vonatkozik, amelyek az állapot- és a fenomenális tudatot is meg akarják magyarázni.

Most aggódhatunk amiatt, hogy a HOT-elméletek csak arra alkalmasak (és csak arra, hogy az állapot-tudatot (nem pedig a fenomenális tudatot) magyarázzák), és hogy következésképpen a HOT-érv az erős primitivizmus mellett nem állja meg a helyét. Ezt az aggodalmat cáfolja az a tény, hogy Rosenthal HOT-elmélete, amely a nézet legismertebb változata, mind a fenomenális tudat, mind az állapottudat magyarázatára alkalmas. Maga Rosenthal mondja: "Ahogy sokan, köztük én is, ezt a kifejezést használják, az, hogy van valami, ami az ember számára olyan, mintha egy állapotban lenne, egyszerűen annak a szubjektív látszata, hogy az ember abban az állapotban van. Valóban, Block (2011: 424) idéz engem ebben az értelemben: "Az, hogy milyen az, amikor valakinek fájdalma van, ennek a kifejezésnek a vonatkozó értelmében egyszerűen az, hogy milyen az, amikor valaki tudatában van annak, hogy fájdalma van"" (Rosenthal 1997: 433)" (Rosenthal 1997: 433).

A második tisztázó megjegyzés az érvelés hatókörével kapcsolatban az, hogy az érvelés csak a HOT-elmélet *aktualista* változataira vonatkozik. Az aktualista HOT-elméletek szerint a magasabb rendű gondolatokat kölcsönző tudatosság a *bekövetkező* mentális állapotok, vagyis olyan mentális állapotok, amelyeket egy szubjektum aktívan fenntart. Rosenthal HOT-elmélete az aktualista HOT-elmélet egyik változata. Az aktualista HOT-elmélettel szemben áll a *diszpozicionális* HOT-elmélet, amelynek vezető képviselője Peter Carruthers (2005). A diszpozicionális HOT-elmélet azt állítja, hogy a releváns magasabb rendű gondolatok diszpozicionális mentális állapotok, szemben a bekövetkezettekkel. Gennaro elmagyarázza, hogy "az alapgondolat [a diszpozicionális HOT-elmélet mögött] az, hogy egy élmény tudatos státusza annak köszönhető, hogy a magasabb rendű gondolkodás számára *elérhető*... Így nem történik tényleges HOT" (Gennaro

2018). A diszpozicionális HOT-elmélet definíció szerint nem tartozik az erős primitivizmusra vonatkozó HOT-érv hatálya alá. Ennek oka, hogy a diszpozicionális HOT-elméletek nem képesek a fenomenális tudatosságot magyarázni. Ahogy az I. fejezetben említettük, a fenomenálisan tudatos állapotok természetüknél fogva occurrent mentális állapotok, nem pedig diszpozíciók. Ugyanis lehetetlen fenomenálisan tudatosnak lenni egy olyan mentális állapotról, amelyet az ember éppen nem instanciál. Rosenthal egyetért ezzel a vélekedéssel, hiszen azt mondja, hogy "az, hogy valaki diszpozicionáltan gondol valamire, nem teszi tudatossá az adott dolgot, hanem csak potenciálisan tudatossá" (Rosenthal 2004: 28). Mivel a diszpozicionális HOT-elméletek nem tudnak számot adni a fenomenális tudatosságról, nem tartoznak az erős primitivizmus HOT-érvének hatálya alá.

Ebben a tanulmányban az érvelés kedvéért feltételezem, hogy Rosenthal HOT elmélete hihető.

Bár a HOT-tudatelmélettel szemben számos ellenvetés van (lásd pl. Sebastián 2019), a nézet plauzibilitását az a tény támasztja alá, hogy a nézet a szakirodalomban befolyásos és népszerű tudatelméletnek számít. Mint említettük, e fejezet célja nem Rosenthal HOT-elméletének védelme (azaz a QP1), hanem annak a feltételes állításnak a megállapítása, amely szerint a HOT-elmélet igazságából (azaz a QP2) következik az erős primitivizmus. Ezért a kérdés: Rosenthal HOT-elmélete magával vonja-e az erős primitivizmust a kognitív fenomenológiáról?

Kétkelkedhetünk a QP2-ben azon az alapon, hogy a HOT elmélet szerint az elsőrendű érzékszervi állapotok szükségesek a fenomenális tudatossághoz. Ha a HOT elmélet szerint a fenomenális tudatossághoz elsőrendű érzékszervi állapotok jelenléte szükséges, akkor feltehetően hamis, hogy a HOT elmélet erős primitivizmust von maga után a kognitív fenomenológiával kapcsolatban. Valójában a HOT elmélet vitathatóan összeegyeztethetetlen az erős primitivizmussal, ha azt állítja, hogy az elsőrendű érzékszervi állapotok szükségesek a tudatossághoz. Ugyanis hihetően nem lehet egyszerre az, hogy (a) a kognitív fenomenológia nem redukálható az érzékszervi fenomenológiára és független attól, és (b) a fenomenális tudat szükségessé teszi az elsőrendű érzékszervi állapotok jelenlétét. Más szóval, ha egy mentális állapotnak érzékszervi inputra van szüksége ahhoz, hogy fenomenálisan tudatos legyen, akkor lehetetlen, hogy fenomenálisan tudatos mentális állapot létezzen érzékszervi fenomenológia hiányában.

Az erős primitivizmus mellett érvelő HOT-érvelő szerencséjére a HOT-elmélet valójában nem kötelezi el magát a (b) pont mellett. Az elmélet ugyanis azt vallja, hogy a magasabb rendű gondolatok önmagukban is elegendőek a tudatossághoz. Rosenthal pontosan ezt mondja: "És a 'milyen az, ami' ezen értelmezése alapján az elmélet valóban azt állítja, hogy a HOT elégséges ahhoz, hogy legyen valami, ami olyan, mintha valaki abban az állapotban lenne, amelyet a HOT leír, még akkor is, ha ez az állapot nem következik be" (Rosenthal, 1997: 433-434). Nevezzük ezt "szufficienciafeltételnek":

Az elégséges állapot: Magányos magasabb rendű gondolatok elegendőek a fenomenális tudatossághoz.

Rosenthal elfogadja az elégségségi feltételt, hogy figyelembe vegye a megtévesztés eseteit. Ahol van képviselő, ott a megtévesztés lehetősége is fennáll. Ez a modális igazság a következő kérdést veti fel, amelyre a HOT teoretikusnak választ kell adnia. Nevezetesen, mi történik, ha egy magasabb rendű gondolat félreprezentálja egy előforduló elsőrendű észlelési állapot tartalmát? A keletkező tudati állapotot az elsőrendű állapot valóságos tartalma, vagy a magasabb rendű állapot tévesen reprezentált tartalma jellemzi? Az előbbi lehetőség Rosenthal szerint nem megfelelő, mert úgy tűnik, hogy a magasabb rendű állapotot teljesen irrelevánssá teszi a fenomenális tudat létrehozását illetően. Annak érdekében, hogy a magasabb rendű gondolatok lényeges szerepet játsszanak a tudat létrehozásában, Rosenthal tagadja, hogy a fenomenalitáshoz első rendű érzékszervi állapotokra lenne szükség. Így Rosenthal szerint a tudatosság a céltalanság eseteiben instanciálódik, vagyis olyan esetekben, amikor egy magasabb rendű gondolat olyan elsőrendű állapotot reprezentál, amely valójában nem létezik. Sam Coleman így foglalja össze Rosenthal álláspontját:

Milyen érzés úgy tudatosítani magunkban, hogy egy piros londoni busz percepciót testesítünk meg, amikor nincs ilyen percepció? A tudatos mentális élet a látszatokból áll, hangsúlyozza Rosenthal: így, mivel az alany számára a látszat ugyanaz, függetlenül attól, hogy egy piros londoni busz-percepciót tokenizál-e vagy sem (feltéve, hogy ugyanaz a HOT van a helyén), az élménye mindkét esetben megkülönböztethetetlen - ahogy a hallucinációk is megkülönböztethetetlenek lehetnek a valóságos társaiktól. Ennek megfelelő megjegyzések vonatkoznak a félreprezentációra is: A HOT-ok szabályozzák a szubjektum tudatfolyamát alkotó szubjektív mentális megjelenéseket (Coleman, 2015: 2705).

Ez a szufficiencia feltétel magyarázza, hogy a HOT elmélet miért feltételez erős primitivizmust. Ugyanis azt mondani, hogy a HOT-ok képesek a fenomenális tudatosságot első rendű érzékszervi állapotok hiányában is instanciálni, pontosan azt jelenti, hogy a kognitív fenomenológia nem függ az érzékszervi fenomenológiától. Bevallottan nem én vagyok az első, aki ezt az érvet felhossa. Richard Brown és Pete Mandik (2012) kifejezetten kapcsolatot teremt a HOT szufficienciafeltétele és a kognitív fenomenológia között:

Ez a HOT-okkal kapcsolatos elégséges állítás a HOT-elmélet és a kognitív fenomenológia képviselői által képviselt egyik központi tézis közötti legfontosabb közös pont: Egy gondolat adhat okot arra, hogy legyen valami, amihez hasonlít, és nem teszi ezt annak erejével, hogy van valamilyen együttesen előforduló nem kognitív állapot (Brown & Mandik, 2012: 4).

Érdeemes kiemelni, hogy Brown és Mandik nem tesznek különbséget a kognitív fenomenológia különböző nézetei között a tanulmányukban. Megállapítható, hogy Brown & Mandik a kognitív fenomenológiáról szóló primitivizmussal foglalkozik, de nem tesz különbséget a gyenge

primitivizmus és erős primitivizmus. Különbséget teszek e két változat között, és azt a konkrét állítást teszem, hogy a HOT-elmélet az erős primitivizmust támogatja.

Az erős primitivizmus melletti HOT-érv ellenfele a dialektika ezen szakaszában elismerheti, hogy a szufficienciafeltétel beépült Rosenthal HOT-elméletébe, de azt állítja, hogy a HOT-elméletnek vannak más aktualista változatai is, amelyek nem támogatják a szufficienciafeltételt. ⁴Sam Coleman (2015) például az aktualista HOT-elmélet egy olyan változatát védi, amely állítása szerint nem kötelezi el magát a szufficienciafeltétel mellett. Coleman az ő nézetét a tudatosság idézőjeles magasabb rendű gondolkodásának (QHOT) elméletének nevezi. A QHOT-elmélet szerint a magasabb rendű gondolat és az elsőrendű állapot közötti kapcsolat nem mentális reprezentáció (ahogy Rosenthal gondolná), hanem "mentális idézés". Coleman azt állítja, hogy a mentális idézés a mentális reprezentációval ellentétben "valódi intimitást tesz lehetővé az érzéki állapotainkkal, az érzéki tudatban" (Coleman, 2015: 2711). ⁵Ennek az állításnak a következménye Coleman számára az, hogy a QHOT-elméletben az elsőrendű érzéki állapotok szükségesek a fenomenális tudatossághoz. Coleman kifejezetten elutasítja azt az elképzelést, hogy az izolált QHOT-ok elegendőek a tudatossághoz:

Mi a helyzet a céltalansággal? Egy magányos QHOT nem lesz elegendő a tudatos állapothoz. Itt van, hogy miért. A QHOT-ok feltételezhetően szubjektív tudatosságot biztosítanak. Ha azonban egy QHOT-nak nincs beágyazható érzékszervi célpontja, akkor legfeljebb a semmi (állapotának) szubjektív tudatosságát keltheti, mivel teljes egészében hiányzik belőle az elsőrendű tartalom. De a semmi állapotának szubjektív tudatossága egyáltalán nem is szubjektív tudatosság, mivel a szubjektív tudatosságnak (szándékosan) valamiről kell szólnia. A szó szerint semmiről szóló tapasztalat egyszerűen nem tapasztalat. Ezért egy üres QHOT nem hoz létre szubjektív tudatosságot - vagy tudatos állapotot. A QHOT-ok csak akkor képesek tudatossá tenni a szubjektumot valamiről, ha van valami (érzékszervi tartalom), amiről tudatában lehet (Coleman 2015: 2722).

Ha a QHOT elmélet valóban azt állítja, hogy az elsőrendű érzékszervi állapotok szükségesek a tudatossághoz, akkor hamis, hogy az aktualista HOT elmélet erős primitivizmust feltételez. Az erős primitivizmus melletti HOT-érvelés hívei itt azzal válaszolhatnak, hogy azt állítják (contra Coleman), hogy a QHOT-elmélet elkötelezett a szufficienciafeltétel mellett. A válasz, amelyre én gondolok, a *tiszta tudat lehetőségére való apellálást* foglalja magában. A tiszta tudat fogalma, ahogyan azt tárgyaltuk, azt sugallja, hogy a fenomenális tapasztalat létezhet mindenféle szándékos struktúra nélkül is. Ha a tiszta tudat lehetséges, akkor Coleman téved, amikor azt állítja, hogy "a szó szerint semmiről szóló tapasztalat egyszerűen nem tapasztalat" (Coleman, 2015: 2722). Pontosabban: okkal feltételezhető, hogy a cél nélküli

⁴Carruthers (2008) pontosan erre a pontra mutat rá, amikor a szufficiencia feltételét tárgyalja.

⁵A "mentális idézet" fogalma nem azonos a "mentális ismeretség" fogalmával, ahol az ismeretség a ruszelli értelemben értendő. Ehhez kapcsolódóan Coleman QHOT-elmélete nem szinonimája a Hellic (2007) által védett magasabb rendű ismeretelméletnek.

A QHOT-ok elegendőek a fenomenális tapasztaláshoz a tiszta tudatosság lehetőségét figyelembe véve. És ha ez így van, akkor a QHOT-elmélet elkötelezi magát a szufficiencia feltétel mellett.

Az érvelés kedvéért azonban tegyük fel, hogy Colemannek igaza van abban, hogy a QHOT-elmélet elutasítja a szufficienciafeltételt. A QP2 híve talán elfogadja ezt a feltételezést, de ennek ellenére ragaszkodik ahhoz, hogy a QHOT-elmélet erős primitivizmust von maga után a kognitív fenomenológiával kapcsolatban.

Konkrétabban, a QP2 híve egyetérthet Colemannel abban, hogy a fenomenális tapasztalatok szükségszerűen *valamiről szólnak*, de elutasítja Coleman feltételezését, hogy ennek a "valaminek" *érzékszervi* állapotnak kell lennie. Feltehetjük a kérdést: mi történik, ha egy QHOT egy *elsőrendű gondolatot* (szemben egy elsőrendű érzékszervi állapottal) tekint tárgyának? Coleman feltehetően azt állítja, hogy a tudat ezekben az esetekben nem instanciálódik, mert - hogy megismételjük - úgy véli, hogy a fenomenális tapasztalat megköveteli a QHOT-elméletben az elsőrendű érzékszervi állapotok jelenlétét. Egy beszélgetőpartner azonban azzal a kérdéssel nyomást gyakorolhat Colemanre, hogy miért kell az elsőrendű mentális állapotoknak érzékszervi jellegűeknek lenniük. Ha a QHOT-ok úgy működnek, hogy az elsőrendű érzékszervi állapotokat tudatossággal ruházzák fel, akkor miért ne működnének úgy is, hogy az elsőrendű kognitív állapotokat is tudatossággal ruházzák fel? A következetesség vitathatatlanul megköveteli, hogy az elsőrendű kognitív állapotok is fenomenálisan tudatossá váljanak az elmélet szerint, ha egy QHOT-ba ágyazódnak. Brown & Mandik ezt a pontot a HOT-elmélet általános megvitatásával összefüggésben vetik fel. ⁶

Ha az elsőrendű gondolatok QHOT-okba ágyazva válnak fenomenálisan tudatossá, akkor a QHOT-elmélet erős primitivizmust feltételez a kognitív fenomenológiával kapcsolatban. Ez a feltétel akkor is igaz, ha elfogadjuk, hogy a QHOT-elmélet elutasítja a szufficienciafeltételt, és azt mondja, hogy az elsőrendű mentális állapotok szükségesek a fenomenális tudatosság adományozásához. Semlegesen állok ahhoz, hogy a QHOT-elmélet magában foglalja-e az erős primitivizmust. Lehet, hogy az erős primitivizmus melletti HOT-érv csak Rosenthal HOT-elméletének vonatkozásában tartható, és ezért szűkíteni kellene a hatókörét. Mindazonáltal a HOT-érv nyilvánvalóan továbbra is érdekes, mivel Rosenthal HOT-elmélete meglehetősen kiemelkedő a tudatossággal foglalkozó szakirodalomban.

IV. 2. premissza: Az erős primitivizmustól az érzékszervi feldolgozás szükségtelenségéig

⁶ Brown & Mandik (2012) a HOT-elmélet általános tárgyalásakor szintén arra hívja fel a figyelmet, hogy a konzisztencia

megköveteli az elsőrendű kognitív állapotok fenomenálisan tudatossá tételét.

Eddig bemutattam a kognitív fenomenológiai vitát az elmefilozófiában, és két érvet mutattam be az erős primitivizmus igaza mellett a kognitív fenomenológiával kapcsolatban: az általam tiszta tudatból származó érvnek és a HOT-elméletből származó érvnek nevezett érvet. Bizonyára még sok mindent lehetne mondani mindezekről az érvekről és általában az erős primitivizmus kilátásairól. E dolgozat célja egyszerűen az erős primitivizmus mint plauzibilis nézet motiválása, majd annak bemutatása, hogy az erős primitivizmus igazsága vitathatóan magával vonja a testetlen mesterséges intelligencia tudatosságának lehetőségét.

Tegyük fel tehát az érvelés kedvéért, hogy a kognitív fenomenológiára vonatkozó erős primitivizmus igaz. Miért gondoljuk, hogy ennek a fenomenológiai igazságnak metafizikai implikációi vannak? Más szóval, miért gondoljuk, hogy a kognitív fenomenológiai P2 érv a testetlen mesterséges intelligencia tudatossága mellett igaz? Miért gondoljuk, hogy abból a tényből, hogy a kognitív fenomenológia nem függ az érzékszervi fenomenológiától, következik az a következtetés, hogy a fenomenális tudat nem függ az agyban zajló érzékszervi feldolgozástól? A rövid válasz az, hogy a P2 a *legjobb magyarázatra való következtetés* eredménye. A legjobb magyarázat arra, hogy a kognitív fenomenológia miért nem függ az érzékszervi fenomenológiától, az, hogy az érzékszervi feldolgozás nem szükséges feltétele a fenomenális tudatnak.

A P2 védelmezője arra is rámutathat, hogy az elmefilozófiában bevett szokás a tudat természetére vonatkozó metafizikai következtetéseket levonni fenomenológiai premisszák alapján. Két példa ilyen évrre a "Transzparencia érv az elsőrendű reprezentacionalizmus mellett" (Tye, 2002) és a "Perifériás belső tudatosság érve az önreprezentacionalizmus mellett" (Kriegel, 2009). Mindkét érv azt feltételezi, hogy a tudatos tapasztalat fenomenológiája tájékoztathat bennünket a tapasztalat mögöttes metafizikai természetéről. Az "Átláthatóság" érvvel kapcsolatban például az az elképzelés, hogy az elsőrendű reprezentacionalizmus (egy metafizikai tézis) levezethető abból a tényből, hogy a tapasztalat egyetlen introspektíve észlelhető aspektusa a világra irányított reprezentációs tartalma (egy fenomenológiai tézis). A kognitív fenomenológia P2 érvelése a testetlen mesterséges intelligencia tudatossága mellett hasonló érvelést alkalmaz. Pontosabban, a premissza a következő elvre támaszkodik (nevezzük "PHEN"-nek)

PHEN: A fenomenológiai függetlenség magában foglalja a metafizikai függetlenséget.

A PHEN szerint abból a tényből, hogy X fenomenológiailag független Y-től, az a következtetés következik, hogy X az Y metafizikai alátámasztása nélkül is instanciálható.

a kognitív fenomenológiával kapcsolatban az az elképzelés, hogy ha a kognitív fenomenológia független az érzékszervi fenomenológiától, akkor a kognitív fenomenológia az érzékszervi feldolgozás hiányában is instanciálható, ami azt jelenti, hogy az érzékszervi feldolgozás nem előfeltétele a fenomenális tudatosságnak.

A PHEN ellen tiltakozhatunk azon az alapon, hogy az elvnek nyilvánvaló ellenpéldái vannak. Például a <fájdalom> fenomenológiailag független a <C-rostok tüzelésétől>, tekintve, hogy az utóbbi egyáltalán nem fenomenológiai. Ebből a tényből azonban nem következik, hogy a <fájdalom> metafizikailag független a <C-rostok tüzelésétől>, vagy hogy a fájdalomérzet nem egyenértékű a C-rostok tüzelésével. ^{7A} <fájdalom> metafizikailag függhet a <C-rostok tüzelésétől> (azaz a többszörös realizálhatóság tézise hamis lehet), annak ellenére, hogy az előbbi fenomenológiailag független az utóbbitól.

Ez a PHEN-nel szembeni ellenvetés a "fenomenológiai függőség/függetlenség" fogalmának (érthető) félreértésén alapul. A fenomenológiai függőség/függetlenség viszonya a feltételezés szerint olyan viszony, amely két *fenomenális* tulajdonság között áll fenn, ami azt jelenti, hogy a relációknak mindkettőnek tapasztalati úton megnyilvánulónak kell lenniük. Mivel a C-rostok tüzelése nem tapasztalati értelemben manifeszt, nincs értelme azt mondani, hogy a <fájdalom> fenomenológiailag független a <C-rostok tüzelésétől>. Van értelme azonban azt mondani, hogy az agenciális tapasztalat fenomenológiailag független az erkölcsi tapasztalattól, vagy hogy a konatív tapasztalat fenomenológiailag független a kognitív tapasztalattól, vagy hogy a kognitív tapasztalat fenomenológiailag független az érzékszervi tapasztalattól. Ugyanis ezekben az esetekben minden reláta tapasztalati értelemben megnyilvánuló tulajdonság.

Természetesen a PHEN még ezzel a pontosítással is ellentmondásos marad. Lehet például, hogy bár az érzékszervi feldolgozás (pl. a Koch által "érzékszervi forró zónának" nevezett (Koch, Massimini, Boly & Tononi 2016)) jelenléte a tudatosság fejlődési előfeltétele, mégis igaz, hogy idősebb korunkban egyfajta *sui generis* kognitív fenomenológiát fejlesztünk ki, amely az érzékszervi fenomenológia hiányában is instanciálható. A P2-vel szembeni ezen ellenvetésben első látásra nincs semmi konceptuálisan problematikus. A bizonyítás terhe azonban egyenesen az ellenvetőt terheli, hogy megmagyarázza, hogyan lehetséges, hogy (a) a fenomenális tudat metafizikailag függ az érzékszervi feldolgozástól, de (b) a kognitív tapasztalat fenomenológiailag nem függ az érzékszervi tapasztalattól. Nem tudok olyan magyarázatról, amely jelenleg létezik erre vonatkozóan a szakirodalomban. A P2 nagyon is hihető, amíg nincs ilyen magyarázat.

⁷ Köszönöm Bill Lycan-nak, hogy felhívta a figyelmemet erre a kifogásra.

Mielőtt rátérnénk a kognitív fenomenológiai érvelés P3-ára, érdemes felhívni a figyelmet arra, hogy a kognitív fenomenológiával kapcsolatos gyenge primitivizmus egyértelműen nem azt jelzi, hogy az érzékszervi feldolgozás nem szükséges a tudatossághoz. Ez pontosan azért van így, mert a gyenge primitivizmus azt állítja, hogy a kognitív fenomenológia nem instanciálható az érzékszervi fenomenológia hiányában (annak ellenére, hogy az érzékszervi fenomenológiára nem redukálható).

V. 3. premissza: Az érzékszervi feldolgozás szükségtelenségétől a testetlen mesterséges intelligencia tudatosságáig

Tegyük fel, hogy mind a P1, mind a P2 igaz, és hogy következésképpen a fenomenális tudatosság nem igényel érzékszervi feldolgozást. Hogyan érinti ez a következtetés a mesterséges intelligencia filozófiáját? Különösen, milyen érvelés áll a P3 mögött? *Lényeges, hogy a P3 azon az ellentmondásos feltételezésen alapul, hogy a mesterséges intelligencia tudata elvileg lehetséges.* Jelenleg a kognitív tudomány és az elmefilozófia nyitott kérdése, hogy a mesterséges intelligencia rendszerek képesek-e a fenomenális tudatosságra. Egyes tudósok (pl. Searle 1980) azzal érvelnek, hogy a tudat csak biológiai organizmusoknál lehetséges, míg mások (pl. Chalmers 2011) az elme komputationalista megközelítését támogatják, amely szerint a gépi tudat lehetséges. Itt nem hozok érveket a mesterséges intelligencia tudatosságának lehetősége mellett vagy ellen. Inkább csak feltételezem, hogy a mesterséges intelligencia tudata lehetséges, és a következő kérdéssel foglalkozom (nevezzük "a megtestesülés kérdésének" (EQ)):

EQ: A mesterséges intelligencia tudatossága, ha létezne, megtestesülést igényelne?

Az EQ a mesterséges intelligencia filozófiájának egyik sürgető kérdése. Tudni akarjuk, hogy a testetlen AI-rendszerek (mint például a neurotrofikus szuperszámítógépek vagy a kifinomult személyi asszisztensek - talán az Apple "Siri" vagy a Google "Alexa" jövőbeli változatai) képesek-e a mi-mi-mi hasonlóságra. ⁸ Végül is valószínű, hogy az ilyen rendszerek a nem túl távoli jövőben képesek lesznek átmenni a Turing-teszten. Bostrom és Muller (2016) felmérése szerint "a szakértők becslése szerint 2040-2050-re 50% fölé, 2075-re pedig 90% fölé emelkedik a 'magas szintű gépi intelligencia' elérésének valószínűsége" (Muller 2016: 4).

Az egyik lényeges aggodalom itt az, hogy *elkerülhetetlenül* tudatot fogunk *tulajdonítani* a jövőbeli testetlen mesterséges intelligenciáknak. Talán pszichológiailag kivitelezhetetlen megtagadni a fenomenális tapasztalatot egy mesterséges intelligenciától.

⁸ Nem egyértelmű, hogy a mélytanuló rendszerek ugyanúgy számítanak-e "érzékszervnek", mint az agy.

rendszer, amely verbálisan és viselkedésileg megkülönböztethetetlen egy normális felnőtt beszélgetőpartnertől. Ha ez a helyzet, és a testetlen mesterséges intelligenciák *képtelenek* a tapasztalatszerzésre, akkor nagyon is fennáll az az aggodalom, hogy az emberiséget pszichológiailag becsapják, és azt hiszik, hogy az érzéketlen számítógépes programok fenomenálisan tudatosak. Ezt nevezem én a *becsapott disztópia forgatókönyvének*. Az EQ-ra választ kell adni annak érdekében, hogy biztosítsuk a duped-dystopia forgatókönyv elkerülését. A P3 azt állítja, hogy a mesterséges intelligencia tudata nem igényel megtestesülést, feltételezve, hogy a fenomenális tudatosság nem igényel érzékszervi feldolgozást. A P3-at motiváló alaptétel az, hogy *a megtestesülés csak akkor szükséges a tudatossághoz, ha az érzékszervi feldolgozás szükséges a tudatossághoz*. Ahhoz, hogy ellenálljon ennek a tételnek, a P3 ellenzőjének valahogyan be kell bizonyítania, hogy a tudathoz akkor is szükség van a megtestesülésre, ha a tudat nem teszi szükségessé az érzékszervi feldolgozást. Első látásra nehéz felismerni, hogyan lehetne ezt a teljesítményt elérni.

Itt először is meg kell jegyeznünk, hogy a "megtestesülés" gyakran (legalábbis részben) a szenzomotoros képességek gyakorlásának képességeként értendő. Vegyük például a kognitív tudományokban a *megtestesült megismerés megközelítését* (Varela, Thompson & Rosch 1991). Amint azt a neve is jelzi, a megtestesült megismerés megközelítése azt állítja, hogy a kognitív folyamatok megtestesülése megtestesülést igényel.

A megközelítés ezáltal elveti az elme hagyományos számítási szemléletét, amely szerint a megismerés alapvetően belső információfeldolgozás kérdése. Lényeges, hogy a megtestesült megismerés megközelítésének képviselői a megtestesülés fogalmának meghatározásakor jellemzően a szenzomotoros képességekre hivatkoznak. Monica Cowart például azt állítja, hogy a megtestesült megismerésről szóló vita keretében "a megtestesülés alapvető fogalmát tágan értelmezve úgy értelmezzük, mint egy organizmus szenzomotoros képességeinek egyedi módját, amely lehetővé teszi számára, hogy sikeresen lépjen kölcsönhatásba a környezeti résszel" (Cowart 2004). Hasonlóan Chris Letheby is úgy definiálja a megtestesült megismerés megközelítését, mint azt a tézist, amely "azt vallja, hogy a fogalmi és nyelvi gondolkodás az agy perceptuális és szenzomotoros rendszereiben gyökerezik" (Letheby 2012: 403). Ha a megtestesülés definíció szerint szenzomotoros képességek birtoklását (és ezáltal érzékszervi feldolgozást) követeli meg, akkor a P3 szükségszerűen igaz. Ugyanis lehetetlen, hogy a megtestesülés a tudat előfeltétele, ha a megtestesüléshez szenzoros feldolgozás szükséges, és kikötöttük, hogy a szenzoros feldolgozás *nem* előfeltétele a tudatnak.

Az érvelés kedvéért feltételezem, hogy a megtestesülés definíció szerint nem követeli meg a szenzomotoros képességek birtoklását. Még ha el is fogadjuk ezt a feltételezést, a bizonyítás terhe továbbra is a P3 ellenzőjét terheli, hogy megmagyarázza, hogyan lehet a megtestesülés szükséges a fenomenális tudatossághoz, tekintve, hogy a tudatosság nem függ az érzékszervi feldolgozástól.

Vannak

a kifogást benyújtó különböző módokon járhat el az ilyen magyarázat megadásában. Az én szemszögemből nézve a P3 ellen csak úgy lehet kifogást emelni, ha a következő háromféle állítást tesszük:

1. **állítás:** Az *életben maradás* szükséges feltétele annak, hogy fenomenálisan tudatosak legyünk.
2. **állítás:** A testet öltés *az élet* szükségszerű feltétele.
3. **állítás:** A fenomenális tudatosság nem igényli a szenzomotoros képességek meglétét.

E hármás állítás sikeres előadása egyet jelent annak bemutatásával, hogy a tudatossághoz testet öltés szükséges, de nem érzékszervi feldolgozás, és így egyet jelent a P3 hamisságának bemutatásával. Természetesen felmerül a kérdés, hogy ez a három állítás *összhangban van-e egymással*. A tudatfilozófiában az *autopoiesis enaktivizmus* néven ismert álláspont (Varela 1997, Di Paolo 2005) mind az 1., mind a 2. állítás mellett elkötelezett. Nem világos azonban, hogy az autopoiesis enaktivizmus összeegyeztethető-e a 3. állítással. Az autopoietikus enaktivizmus mellett vagy ellen szóló érveket itt nem veszem figyelembe. Inkább feltételezem, hogy az autopoietikus enaktivizmus plauzibilis, és továbblépek, hogy megkérdézzem, vajon a nézet fegyverként felhasználható-e a P3 elleni ellenvetésként. E kérdés értelmezéséhez szükség van az enaktivizmus megértésére, és különösen az *autopoietikus* enaktivizmus és a *szenzomotoros* enaktivizmus közötti különbségtétel megértésére.

Az enaktivista tudatelméletek a tudatot interaktív természetűnek tekintik, nem pedig passzívnak, ami azt jelenti, hogy az enaktivista szerint a fenomenális tapasztalat olyasvalami, amit mi csinálunk, nem pedig olyasvalami, ami bennünk zajlik le. Ahogy Jan Degenaar és J. Kevin O' Regan megjegyzik: "A tudat enaktív megközelítései a környezet állítólagos belső reprezentációi helyett az interakciós mintákat és az aktív elköteleződést hangsúlyozzák" (Degenaar & Regan 2017: 393). Az enaktivizmus a külső világgal való testi interakciók hangsúlyozása miatt a tudat megtestesült elméletét képviseli. 9 Minden enaktív elmélet számára égető kérdés: pontosan milyen interaktív képességek szükségesek és elégségesek a tudatossághoz? Ez a kérdésfeltevés felveti az *autopoietikus* enaktivizmus és a *szenzomotoros* enaktivizmus közötti döntő különbséget. A nézet utóbbi változata azt állítja, hogy a releváns interaktív képességek a szenzomotoros képességek. Pontosabban, a szenzomotoros enaktivizmus azt állítja, hogy a szenzomotoros képességek birtoklása szükséges és elégséges is a tudatossághoz (O'Regan és Noë 2001; O'Regan 2011; Degenaar és O'Regan

⁹ Az "enaction" fogalma a *The Embodied Mind* (Varela, Thompson és Rosch 1991) című könyvből származik, amelyben a szerzők a kognitív tudomány forradalmisítására tesznek kísérletet azáltal, hogy a kognitív tudományt összekapcsolják Merleau-Ponty (1945) fenomenológiai wokjával a megtestesülésről.

2015). Ezzel szemben az autopoietikus enaktivizmus lehangolja a szenzomotoros képességek szerepét, ehelyett a tudatosság és az étellel kapcsolatos interaktív képességek közötti elválaszthatatlan kapcsolat vélelmzésére összpontosít. Az "életet" ezen elmélet kontextusában az "autopoiesis" fogalmával határozzák meg. Ahogy Degenaar és Regan tisztázza: "Az enaktív irodalomban az életet jellemzően autopoézisként értelmezik (Maturana & Varela 1980). Az élő szervezetek spontán aktív, önfenntartó és öntermelő rendszerek, és azt javasolták, hogy fenomenológiánkat ennek az önteremtő vagy "autopoietikus" szerveződésnek a fényében kell megértenünk" (Degenaar & Regan 2015: 396). Degenaar és Regan továbbá megvilágítja, hogy az "autopoiesis" szükségszerűen magában foglalja a környezettel való megtestesült interakciót: "Egy autopoietikus rendszernek metabolizálnia kell és folyamatosan alkalmazkodnia kell a környezetéhez, hogy folytatni tudja tevékenységét. Az autopoiesis definíciója szerint nem vonhatja ki magát az interaktív kontextusból" (Degenaar & Regan 2015: 397-398).

Az autopoietikus enaktivizmus tehát azt mondja, hogy az élet szükséges a tudatossághoz, ahol az életet alapvetően interaktív folyamatnak tekintjük, amely a megtestesülést követeli meg. Éppen ezért az autopoietikus enaktivizmus elkötelezett a fenti 1. és 2. állítás mellett. A P3 ellenzőjének lényeges kérdése, mint jeleztük, hogy az autopoietikus enaktivizmus összeegyeztethető-e a 3. állítással, amely szerint a szenzomotoros képességek birtoklása nem szükséges a tudatossághoz. Van okunk kételkedni abban, hogy a kettő összeegyeztethető. Az autopoietikus enaktivizmust ugyanis az a meggyőződés motiválja, hogy "a perceptuális tudatosság a perceptuális képességeken túl *további* korlátokat is magában foglal" (Degenaar & Regan 2015: 397-398). Más szóval, míg az autopoietikus enaktivisták tagadják, hogy a szenzomotoros képességek birtoklása *elégleges feltétele* a tudatosságnak, addig hajlamosak egyetérteni a szenzomotoros enaktivistákkal abban, hogy az említett képességek birtoklása *szükséges feltétele* a tudatosságnak. Lehetséges-e, hogy az autopoietikus enaktivisták fenntartják a 3. állítást, annak ellenére, hogy az autopoietikus enaktivisták többsége kifejezetten elfogadja azt a tételt, hogy a tudatossághoz szenzomotoros képességek szükségesek? Azt állítom, hogy a válasz erre a kérdésre nemleges. Nekem úgy tűnik, hogy az autopoietikus enaktivizmus és a 3. állítás kölcsönösen ellentmondásosak, mert az autopoietikus folyamatok vitathatatlanul szenzomotoros készségektől függenek.

Összefoglalva: P3 azt állítja, hogy a mesterséges intelligencia tudatossága (feltételezve, hogy lehetséges) nem igényel megtestesülést, ha a fenomenális tudatosság nem igényel érzékszervi feldolgozást. A P3 elleni tiltakozáshoz meg kell magyarázni, hogyan lehet a megtestesülés szükséges a fenomenális tudatossághoz, ha a tudatosság nem függ az érzékszervi feldolgozástól. Először is rámutattam, hogy az ellenzőnek lehetetlen ezt a teljesítményt elérni, ha a megtestesülést (legalábbis részben) úgy definiáljuk, hogy a képességet

szenzomotoros készségek gyakorlása. Ezután megfontoltam és végül elvettem azt az elképzelést, hogy az autopoietikus enaktivizmus a P3 ellenvetéseként felhozható.

VI. Következtetés

A kognitív fenomenológiai érv P1 a testetlen mesterséges intelligencia tudatossága mellett azt állítja, hogy a kognitív fenomenológiára vonatkozó erős primitivizmus igaz. Itt nem tettem kísérletet az erős primitivizmus igazságtartalmának megállapítására. Ehelyett csupán azt állítottam, hogy az erős primitivizmus hihető bizonyos elméleti feltételezések mellett; különösen azt a feltételezést, hogy a tiszta tudat lehetséges és/vagy azt a feltételezést, hogy Rosenthal HOT-elmélete igaz. Az érvelés P2 része azt állítja, hogy az érzékszervi feldolgozás nem szükséges feltétele a fenomenális tudatnak, ha az erős primitivizmus igaz. Azzal érveltem, hogy a P2-t erősen alátámasztja az, amit én PHEN-nek nevezek, de elismertem, hogy a PHEN-nek ésszerűen ellen lehet állni. Végül bemutattam, hogy bár a P3-at lehet kifogásolni, joggal mondhatjuk, hogy a P3 az érvelés legkevésbé vitatható előfeltevése. Az a következtetés, hogy a mesterséges intelligencia tudata nem teszi szükségessé a megtestesülést, korántsem triviális. Egy ilyen következtetés kiszélesíti a tudatos mesterséges intelligenciák lehetséges típusainak körét, és vitathatóan még azt is sugallja, hogy a tudat két alapvetően különböző fajtája létezik: az érzékszervi alapú tudat és a test nélküli tudat.

Idézett
művek

- Arango-Muñoz, Santiago (megjelenés előtt). 'Kognitív fenomenológia és metakognitív érzések'. *Mind and Language*.
- Block, Ned (1995). "A tudat egyik funkciójával kapcsolatos zűrzavarról". *Brain and Behavioral Sciences* 18 (2):227--247.
- Block, Ned (2011). "Az érzékelési tudatosság túlsordul a kognitív hozzáférése". *Trends in Cognitive Sciences* 15 (12):567-575.
- Block, Ned (2011). "A tudat magasabb rendű megközelítése nem létezik". *Analysis* 71 (3):419 - 431.
- Brentano, Franz (1874). *Pszichológia empirikus szempontból*. Routledge.
- Brown, Richard & Mandik, Pete (2012). 'On whether the Higher-Order Thought Theory of Consciousness Entails Cognitive Phenomenology, or: Milyen az, amikor valaki azt gondolja, hogy P?'. *Philosophical Topics* 40 (2):1-12.
- Carruthers, Peter (2005). *Tudatosság: Essays From a Higher-Order Perspective*. Oxford University Press UK.
- Carruthers, Peter (2008). "A tudatosság magasabb rendű elméletei". *Stanford Encyclopedia of Philosophy*.
- Carruthers, Peter & Veillet, Bénédicte (2011). 'The Case Against Cognitive Phenomenology'. In Tim Bayne & Michelle Montague (szerk.), *Kognitív fenomenológia*. Oxford University Press. pp. 35.
- Chalmers, David (2011). "A megismerés tanulmányozásának számítási alapja". *Journal of Cognitive Science* 12 (4):323-357.
- Chudnoff, Elijah (2015). 'Fenomenális kontraszt érvek a kognitív fenomenológiához'. *Filozófia és fenomenológiai kutatás* 90 (2):82-104.
- Coleman, Sam (2015). 'Idézőjeles magasabb rendű gondolkodáselmélet'. *Philosophical Studies* 172 (10): 2705-2733.
- Cowart, Monica (2004). Megtestesült megismerés. *Internet Encyclopedia of Philosophy*.
- Degenaar, Jan & O'Regan, J. Kevin (2017). 'A szenzomotoros elmélet és az enaktivizmus'. *Topoi* 36 (3): 393-407.
- De Jesus, Paulo (2016). 'Autopoietikus enaktivizmus, fenomenológia és az élet és az elme közötti mély kontinuitás'. *Fenomenológia és a kognitív tudományok* 15 (2):265-289.
- Di Paolo, Ezequiel A. (2005). "Autopoiesis, adaptivitas, teleológia, ügynökség". *Fenomenológia és a kognitív tudományok* 4 (4):429-452.

- Fanaya, Patrícia Fonseca (2020). "Autopoietikus enaktivizmus: cselekvés és reprezentáció Peirce fényében újrazvizsgálva". *Synthese* 198 (Suppl 1):461-483.
- Fodor, Jerry A. (1985). "Fodor útmutatója a mentális reprezentációhoz: Az intelligens nénikék vándorútja - Mecum". *Mind* 94 (373):76-100.
- Gennaro, Rocco J. (2018). 'A tudatosság magasabb rendű elméletei'. *Internet Encyclopedia of Philosophy*.
- Hellie, Benj (2007). 'A magasabb rendű intencionalizmus és a magasabb rendű ismeretelmélet'. *Philosophical Studies* 134 (3):289--324.
- Horgan, Terence & Tienson, John (2002). 'A fenomenológia intencionalitása és az intencionalitás fenomenológiája'. In David J. Chalmers (szerk.), *Philosophy of Mind: Classical and Contemporary Readings*. Oup Usa. pp. 520-533.
- Katz, Steven T. (1978). "Nyelv, ismeretelmélet és miszticizmus". In *Mysticism and Philosophical Analysis*. Oxford University Press. pp. 22--74.
- Kirchhoff, Michael (2018). 'Autopoiesis, szabad energia és az élet-elme folytonossági tézis'. *Synthese* 195 (6):2519-2540.
- Koch, Christof & Massimini, Marcello & Boly, Melanie & Tononi, Giulio. (2016). 'A tudatosság neurális korrelátumai: Progress and problems'. *Nature Reviews Neuroscience*. 17. 307-321. 10.1038/nrn.2016.22.
- Kriegel, Uriah (2009). 'Önreprezentacionalizmus és fenomenológia'. *Philosophical Studies* 143 (3): 357-381.
- Kriegel, Uriah (2015). *A tudatosság változatai*. Oxford University Press.
- Krueger, Joel (2006). "A tiszta tapasztalat változatai: William James és Kitaro Nishida a tudatosságról és a megtestesülésről". *William James Studies* 1
- Letheby, Christopher (2012). A megtestesült megismerés védelmében: válasz Fred Adamsnek. *Fenomenológia és a kognitív tudományok* 11 (3):403-414.
- Levine, Joseph (2011). "A gondolkodás fenomenológiájáról". In Tim Bayne és Michelle Montague (szerk.), *Kognitív fenomenológia*. Oxford University Press. pp. 103.'
- Lycan, William G. (1996). *Tudat és tapasztalat*. MIT Press.
- Mendelovici, Angela (2013). Intencionalizmus a hangulatokról. *Gondolat: A Journal of Philosophy* 2 (1): 126-136.
- Merleau-Ponty, Maurice (1962). *Az érzékelés fenomenológiája*. Routledge.

- Nagel, Thomas (1974). "Milyen érzés denevérek lenni? *Philosophical Review* 83 (október):435-50.
- Rosenthal, David (2005). *Tudat és elme*. Oxford University Press UK.
- Rosenthal, David M. (1993). "A magasabb rendű gondolatok és a függelékelmélet a tudatosság". *Filozófiai pszichológia* 6 (2):155-66.
- Rosenthal, David M. (1997). A tudat perceptuális és kognitív modelljei. *Journal of the American Psychoanalytic Association* 45.
- Rosenthal, David M. (1986). "A tudat két fogalma". *Philosophical Studies* 49 (May):329-59. Rosenthal, David (2004). 'A magasabb rendű elmélet fajtái'. In Rocco J. Gennaro (szerk.), *Higher-Order. A tudatosság elméletei: An Anthology*. John Benjamins.
- Searle, John R. (1980). "Elmék, agyak és programok". *Behavioral and Brain Sciences* 3 (3):417-57.
- Sebastián, Miguel (2019). 'Drop it like it's HOT: A magasabb rendű gondolkodáselméletek ördögi regressziója'. *Philosophical Studies* 176 (6):1563-1572.
- Smith, David (2011). "Nibbanikus (vagy tiszta) tudatosság és azon túl". *Philosophia* 39 (3):475-491.
- Smithies, Declan (2013). 'A kognitív fenomenológia természete'. *Philosophy Compass* 8 (8): 744-754.
- Strawson, Galen (2011). "Kognitív fenomenológia: a való élet". In Tim Bayne & Michelle Montague (szerk.), *Kognitív fenomenológia*. Oxford University Press. pp. 285--325.
- Tye, Michael (2002). "A reprezentacionalizmus és a tapasztalat átláthatósága". *Noûs* 36 (1):137-51.
- Varela FJ (1997) "Az élet mintái: az identitás és a megismerés összefonódása". *Brain and Cognition* 34(1):72-87. Varela, Francisco; Thompson, Evan & Rosch, Eleanor (1991). *A megtestesült elme: Cognitive Science és az emberi tapasztalat*. MIT Press.
- Ziff, Paul (1958). "A robotok érzései". *Erkenntnis* 19:64.