

A nem emberi tudatosság és a specifikusság problémája: Egy szerény elméleti javaslat

Henry Shevlin 

Leverhulme Centre for the Future of Intelligence, University of Cambridge, Cambridge, Egyesült Királyság

Levelezés

Henry Shevlin, Leverhulme Centre for the Future of Intelligence, 16 Mill Lane, Cambridge CB2 1SB, Egyesült Királyság. E-mail: hfs35@cam.ac.uk

Finanszírozási információk

Leverhulme Trust, Grant/Award Number: RC-2015-067

A tudatosság legtöbb tudományos elméletét kihívást jelent az emberi eseteken kívül alkalmazni, mivel a nem emberi rendszerek (biológiai és mesterséges) valószínűleg nem tudják pontosan megvalósítani az emberi architektúrát, amit én *specifitási problémának* nevezek. A tudatelméletekről szóló vita némi háttérnek ismertetése után áttekintem a probléma négy megközelítésének kilátásait. Ezután egy ötödik megoldást, nevezetesen a Jonathan Birch által javasolt *elmélet-közeli megközelítést* vizsgálom. Megvédem ennek egy módosított változatát, amelyet én *szerény elméleti megközelítésnek* nevezek, azzal érvelve, hogy betekintést nyújthat olyan kihívást jelentő esetekben, amelyek egyébként megoldhatatlanok lennének.

KULCSSZAVAK

állati megismerés, tudat, tudatosság, tudatosság mérése, nem emberi tudatosság, más elmék, tudatelméletek, tudatelméletek

1 | BEVEZETÉS

A tudatosság tudománya több fronton is jelentős előrelépést tett, beleértve a különböző elméleti megközelítések kidolgozását és a tudatosság különböző formáinak tesztelésére szolgáló mérőeszközök finomítását. Az egyik legfontosabb kihívás azonban az, hogy hogyan tudjuk felmérni a tudatosságot a nem emberi rendszerekben. Legalábbis elvileg remélhetjük, hogy ugyanazok a tudatelméletek, amelyek célja az emberi tudatos tudatosság előrejelzése, azt is meg tudják mondani, hogy az olyan lények, mint a patkányok, a halak és a rovarok, valamint a mesterséges rendszerek tudatosak-e. Ez az elmélet a tudatosságról szóló elméleteket is megmagyarázza.

Ebben a tanulmányban egy olyan kihívást vizsgálok meg, amellyel a tudatosság számos jelenlegi elmélete szembesül a tudatosság nem-embereknek való tulajdonításával kapcsolatban, és amelyet én *specifitási problémának* nevezek. A oldalon.

Röviden, ez azt érinti, hogy a tudatosság elméleteinek milyen részletességgel kell rendelkezniük, amikor nem emberekre alkalmazzák őket. Ha elméleteinket olyan módon fogalmazzuk meg, hogy részletesen utalunk az emberi kognitív architektúra sajátos aspektusaira, akkor fennáll a hamis negatívumok kockázata, amikor a feldolgozásban mutatkozó kisebb különbségek miatt helytelenül tagadjuk a nem emberi rendszerek tudatosságát. Ha azonban elméleteinket nagyon absztrakt módon fogalmazzuk meg, akkor az ellenkezőjével kell szembenéznünk, nevezetesen azzal, hogy a nem-tudatos rendszereket tévesen tudatos entitásoknak minősítjük.

A tanulmány a következőképpen halad. A 2. szakaszban a tudatelméletekről szóló vita rövid háttérével kezdem, majd a 3. szakaszban összefoglalom a specifikussági problémát, és elmagyarázom, hogy ez hogyan akadályozza az olyan nézetek alkalmazását, mint a *globális munkaterület-elmélet* a nem emberi rendszerekre. A 4. szakaszban a specifikussági problémára adott négy kezdeti választ tárgyalom, amelyeket *konzervativizmusnak*, *minimalizmusnak*, *inkrementalizmusnak* és *elutasításnak* nevezek. Az 5. szakaszban a tudatosság *elmélet-nehezített* és *elmélet-könnyű* megközelítései közötti különbségtételt vizsgálom (Birchnek köszönhetően, hamarosan megjelenik), és azt javaslom, hogy a specifikussági probléma csak akkor jelent jelentős problémát, ha az előbbi módszertan mellett kötelezzük el magunkat. Ugyanakkor azt is megjegyzem, hogy az elméletkönnyű megközelítés nem valószínű, hogy minden olyan választ megad, amelyre a nem emberi tudat tudományától szükségünk van. Végül a 6. szakaszban egy alternatív stratégiát javaslok a specifikussági probléma kezelésére, amelyet én *szerény elméleti megközelítésnek* nevezek. Azt állítom, hogy az elmélet súlyos és az elméletkönnyű megközelítés módszereinek kombinálásával megoldhatjuk a specifikussági problémát, és nehéz esetekben is meghatározhatjuk a tudatosságot, valamint betekintést nyerhetünk abba, hogy a tudatosság mely elméletei a legígéretesebbek.

2 | A TUDATELMÉLETEK VITÁJA

A tudatosság tanulmányozásának egyik központi célja az volt, hogy olyan magyarázatot dolgozzanak ki, amely megmagyarázza, hogy a pszichológiai és idegi feldolgozás egyes formái miért tudatosak, mások pedig miért nem. Az ilyen magyarázatokat kínáló keretműveket jellemzően *tudatelméleteknek* nevezik, az ilyen elméletek értékelését és kritikáját pedig *tudatelméleti vitának*.¹ Bár a tudatelméletek vitája régóta fennálló filozófiai és tudományos elképzelésekből táplálkozik, mint a kognitív tudomány egy olyan speciális területe, amely a filozófusok és a tudósok aktív együttműködésével jár együtt, modern története vitathatatlanul az 1980-as években kezdődött, olyan keretek megjelenésével, mint Rosenthal *magasabb rendű* gondolkodáselmélete (Rosenthal, 1986) és Baars *globális munkaterület-elmélete* (1988). Azóta számos más elmélet is kialakult, amelyek közül kiemelkedik Tononi *integrált információelmélete* (Tononi, 2008), a Ned Block és Victor Lamme által képviselt biológiai elméletek (Block, 2009; Lamme, 2010), valamint Dehaene és Naccache *globális neuronális munkaterülete* (Dehaene & Naccache, 2001).

A terület kutatói által alkalmazott egyik legfontosabb módszer annak vizsgálata, hogy az emberi alanyok milyen körülmények között érzékelik tudatosan vagy tudattalanul a különböző ingereket. Az ilyen jellegű para- digmák közé tartozik a binokuláris rivalizálás (Blake & Logothetis, 2002), a maszkolás (Kouider & Dehaene, 2007) és a dichotikus hallás (Brancucci & Tommasi, 2011), amelyek mindegyike arra épül, hogy a bemeneti vagy feladat körülmények változtatásával szabályozhatjuk, hogy az alanyok képesek-e észlelni és

¹A vita egyik legfontosabb megkülönböztetése a fenomenális tudatosság (amelyet szubjektív élményként értelmezünk) és a hozzáférési tudatosság (az információ önkéntes cselekvésben való felhasználásának képessége, beleértve a jelentéstételt is) közötti különbségtétel (Block, 1995). A könnyebb kifejtés és az elméleti kérdések előítéletének elkerülése érdekében a "tudat" kifejezést a szubjektív tapasztalatra használom, miközben nyitva hagyom annak lehetőségét, hogy a fenomenális tudat és a hozzáférési tudat egymásra utaló kifejezések.

egy inger jelenlétéről szóló jelentés. Az észlelhetőséggel és jelenthetőséggel - a tudatossággal gyakran, bár ellentmondásosan összefüggésbe hozott jelenségekkel - kapcsolatos kognitív és neurális dinamikák azonosításával megkísérélhetjük megalkotni annak a leírását, hogy mi különbözteti meg a feldolgozás tudatos és tudattalan formáit, és ebből kiindulva a tudatosság általános elméletét adhatjuk meg.² A tudatosság tudományában alkalmazott egyéb fontos módszerek közé tartozik a tudatosság szelektív károsodását okozó agykárosodásban szenvedő betegeken végzett kísérleti munka (Weiskrantz, 1986), valamint a tudatos és tudattalan állapotú betegek agyi képalkotása, például általános érzéstelenítésben (Alkire & Miller, 2005).

Továbbra is jelentős viták folynak arról, hogy a tudatosság melyik vezető elméletét kell előnyben részesíteni, ha egyáltalán van ilyen. A terület egésze azonban rendkívül termékenynek bizonyult, rengeteg adatot szolgáltatott az agyi feldolgozás különböző formáiról, ami hatással van a vegetatív állapotú betegek tudatának kimutatására (Casali et al., 2013), és olyan kulcsfontosságú elméleti fogalmak kidolgozásához vezetett, mint például Block megkülönböztetése a fenomenális és a hozzáférési tudatosság között (Block, 1995).

Az emberi tudat alapjainak megértésében elért eredmények ellenére azonban viszonylag kevés előrelépés történt a nem emberi tudattal kapcsolatos vitában.³ Ennek egyik fontos oka az, hogy az ezen a területen dolgozó kutatók túlnyomórészt a *szóbeli beszámolóra* támaszkodnak, hogy egyértelmű bizonyítékot szolgáltatassanak arra, hogy egy adott inger tudatosan feldolgozott. Bár történtek erőfeszítések a verbális jelentésen túllépésre (pl. Tsuchiya, Wilke, Frässle & Lamme, 2015), ezek továbbra is ellentmondásosak. Ennek eredményeképpen számos központi kísérleti paradigma nehezen alkalmazható a nem-emberek tudatosságáról szóló vitákra.⁴ Hasonlóképpen, egyes megközelítések (például Crick & Koch, 1990; Lamme, 2010) kifejezetten az emberben lévő tudat neurális és biológiai mechanizmusainak azonosítására irányulnak, és így nem céljuk, hogy a tudatról olyan beszámolót nyújtsanak, amely általánosítható a nem emberi rendszerekre.

Mindazonáltal a tudatosság számos vezető elmélete ambiciózusabb, és a tudatosság általános elemzését valamilyen pszichológiai vagy kognitív mechanizmus, például a metakogníció vagy a globális adás szempontjából próbálja megadni. Következésképpen az ilyen elméletek azt remélik, hogy képesek lesznek megválaszolni azt a kérdést, hogy mely nem emberi rendszerek tudatosak, azáltal, hogy meghatározzák, melyek azok, amelyek a vonatkozó folyamatokat végrehajtják. Ennek gyakorlati megvalósítása egy adott nem emberi rendszer esetében természetesen számos kihívást jelentene, és aprólékos vizsgálatot igényelne a rendszer kognitív architektúrájának részleteinek meghatározása érdekében, de legalábbis az átfogó módszertanunk meglehetősen egyszerű lenne.

Az az elképzelés, hogy az emberi tudatról alkotott elméleteinket felhasználhatnánk arra, hogy a nem emberi tudattal kapcsolatos kérdéseket ilyen közvetlen módon rendezzük, ellentmondásos (lásd az 5. szakaszt ennek az "elmélet-hangsúlyos" megközelítésnek a megvitatására). Én azonban úgy vélem, hogy a vezető elméletek közül sok - köztük a globális munkaterület-elmélet (a továbbiakban GWT; Dehaene & Naccache, 2001), a munkamemória-elméletek (Baddeley, 1992), a magasabb rendű gondolkodás elméletei (Rosenthal, 2005), Prinz attended intermediate representations (AIR) nézete (Prinz, 2012) és a figyelem sémaelmélete (Graziano, 2013) - alkalmasak erre a stratégiára, és ezekre együttesen a tudatosság *kognitív* elméleteiként fogok hivatkozni. Az ilyen teoretikusok, most azzal fogok érvelni, hogy egy olyan kihívással szembesülnek, amelyet a specifikusság problémájának nevezek.

²Ahogy a szövegben is szerepel, a tudatosság és a jelentési kötelezettség közötti kapcsolatáról jelenleg is jelentős viták folynak. Lásd Block (2007) hasznos összefoglalóját.

³A nem emberi tudatosságról természetesen a kognitív tudomány más ágaiban, nevezetesen a kognitív etológiában és az állati elmék filozófiájában is kiterjedtebb viták folytak (lásd pl. Bekoff, 2010; Godfrey-Smith, 2016; Griffin, 2001). További tárgyalásokat lásd az 5. szakaszban.

⁴Néhány figyelemre méltó kivétel a majmokon végzett binokuláris rivalizációs kísérletek (Stoerig, Zontanou & Cowey, 2002) és a késleltetett nyomkövetéses szemlinkelési kísérletek nyulakon (Thompson, Moyer & Disterhoft, 1996).

3 | A SPECIFIKUSSÁGI PROBLÉMA

Röviden, a specifikusság problémája az a kihívás, hogy miként lehet a tudatosságot alkotó kognitív mechanizmusokat az általunk preferált elmélet szerint úgy megfogalmazni, hogy azok az emberi eseten túl is alkalmazhatóak legyenek. Az olyan folyamatok, mint a globális aktiváció, a munkamemória, a figyelem és a metakogníció valószínűleg mind olyan módon nyilvánulnak meg a nem emberi rendszerekben, amelyek többé-kevésbé drámaian eltérnek az emberi esettől. A nehézséget az jelenti, hogy hogyan tudjuk azonosítani, hogy ezek közül a különbségek közül melyek azok, amelyek a tudatossággal kapcsolatos kérdések szempontjából fontosak, és melyek azok, amelyeket félretehetünk. Előjáróban meg kell jegyezmem, hogy a fent vázolt specifikussági probléma nem újszerű. A "liberalizmus" és a "sovinizmus" közötti helyes egyensúly megtalálása a mentalitás és a tudatosság nem emberi rendszereknek való tulajdonításában például régóta kísérti az elme funkcionális megközelítéseit (Block, 1978), és számos filozófus tárgyalt már hasonló kihívásokat a tudat specifikus tudatelméletek számára. Így Prinz (2018) az AIR elméletének a hullőkre és kétéltűekre való alkalmazhatóságát vizsgálva megjegyzi, hogy "információfeldolgozó rendszereik bizonyos tekintetben hasonlítanak a miénkhez, de olyan szempontból is különböznek, ami csökkenti a bizalmat a tudatosság tulajdonításakor" (189. o.). Hasonlóképpen, amikor a GWT-t kerti csigákra próbálja alkalmazni, Schwitzgebel (hamarosan megjelenik) megjegyzi, hogy "az információ széles körben terjed a csigák agyában, lehetővé téve az összehangolt cselekvést", és felteszi a kérdést, hogy ez "elég globális munkaterület-e" ahhoz, hogy a tudatosság bizonyítékának számíthasson (26. o.). A legjelentősebb talán Carruthers (2018a, 2018b) legújabb munkájában felveti az emberi tudatosság általa preferált beszámolója, nevezetesen a GWT által felvetett specifikussági probléma egy változatát, és arra a következtetésre jut, hogy ez értelmetlenné teszi az állati tudatosság kérdéseit (lásd a 4.4. szakaszt).⁵

Amit azonban újszerűnek és fontosnak tartok a jelen vitában, az a specifikussági probléma mint *általános* probléma jelentősége a tudatosság kognitív elméletei számára a nem emberi rendszerekre való alkalmazásuk során. Első illusztrációként tekintsük Dehaene és Naccache globális munkaterület-elméletét. Ez az elmélet időnként nagyon absztrakt kifejezésekkel fogalmazódik meg, például abban az állításban, hogy "a tudatosság nem más, mint az egész agyra kiterjedő információmegosztás" (Dehaene, 2014, 165. o.). Máshol azonban úgy tűnik, hogy Dehaene sokkal szűkebb látásmódot tart szem előtt, például azt állítva, hogy "a jelentéskészség a tudatos állapot kulcsfontosságú jellemzője" (165. o.). Ez felveti azt a kérdést, hogy melyek azok a képességek, amelyek szigorúan *kritikusak* a tudatosság szempontjából releváns értelemben vett globális közvetítéshez, és hogy egy rendszer felépítése mennyire térhet el az emberi esettől, miközben mégis tudatos marad. Vegyük például, hogy az emberi globális munkaterület a legkülönbözőbb kognitív funkciók, például a jelentés, az önismeret és a hosszú távú tervezés számára tesz elérhetővé információkat. Az, hogy ezek közül melyik szükséges ahhoz, hogy egy rendszer a szoros értelemben vett globális munkaterülettel rendelkezzen, egyáltalán nem nyilvánvaló. Ahogy Carruthers (2018a) fogalmaz: "[a]z a globális közvetítés egyes aspektusai fontosabbak, vagy relevánsabbak a tudatosság kérdése szempontjából, mint mások?" (p. 54).

E probléma egy változata vitathatatlanul felmerül a tudatosság minden kognitív elmélete esetében, amennyiben olyan széles körben meghatározott kognitív funkciókra hivatkoznak, amelyek számos különböző módon valósulhatnak meg különböző biológiai vagy mesterséges rendszerekben. Vegyük például a tudat magasabb rendű megközelítését, amelynek minden változata nagyjából a tudatot a saját mentális állapotaink egyfajta tényleges vagy lehetséges tudatosságával azonosítja. Ez a tudatállapotok formáját öltheti

⁵A tudatosságkutatásban a specifikussági probléma ezen esetein kívül természetesen az összehasonlító filozófiában és a kognitív tudományokban is számos olyan esetet találunk, ahol a pszichológiai állapotok nem embereknek való tulajdonításával kapcsolatos módszertani kérdések a fajok közötti kognitív különbségek miatt merülnek fel. Lásd például Buckner (2015) a megismerés esetét, Carruthers (2013) a munkamemóriát, valamint Clayton és társai (2007) az epizodikus és "epizódszerű" memóriát.

saját mentális állapotaink belső, kvázi-perceptuális tudatosságának formája (Lycan, 1996) vagy magasabb rendű gondolatok (Rosenthal, 2005). Ahhoz azonban, hogy az ilyen elméleteket nem emberi rendszerekre is alkalmazhassuk, meg kell határoznunk, *hogyan mi számít* érzékelésnek vagy gondolatnak a kívánt értelemben, és itt összetett fogalmi kérdések merülnek fel. A metakogníció - tágabb értelemben a saját mentális állapotok inter-nális reprezentálásának képessége - számos különböző formában jelenik meg, és különböző mértékben megtalálható a különböző nem emberi rendszerekben. A legegyszerűbb esetben ez magában foglalhatja az érzékelési inputok belső és külső eredetű változásainak egyszerű megkülönböztetését (Merker, 2005), de a fogalom olyan kifinomultabb képességeket is magában foglal, mint például annak felismerése, hogy másoknak a sajáttól teljesen eltérő meggyőződések lehetnek (Kaminski, Call & Tomasello, 2008), és hogy képesek vagyunk reprezentálni a különböző meggyőződésekbe vetett bizalmunk mértékét (Terrace & Son, 2009). Valószínűleg minden esetben a nem-emberek számára a metakognitív reprezentációk fogalmának megfogalmazásához és strukturálásához rendelkezésre álló kognitív erőforrások nagymértékben eltérnek az emberek által birtokoltaktól. Például nyelvi képességeink és gazdag fogalmi repertoárunk révén egy ember úgy reprezentálhatja magát, mint aki komplex logikai struktúrájú gondolatokkal rendelkezik. Az, hogy az önreprezentáció ilyen összetett formái szükségesek-e ahhoz, hogy a tudatosság szempontjából releváns értelemben vett magasabb rendű gondolatokkal rendelkezünk, nem triviális válasz.

Nem szeretném túlhangsúlyozni a sajátossági probléma leküzdésének nehézségeit. Joggal remélhetjük, hogy a tudatosságról szóló elméleteinkben van *valamilyen* megfelelő szintű specifikusság, amely lehetővé teszi számunkra, hogy a tudatosságot pontosan tulajdonítsuk a nem emberi rendszereknek. A kihívás az, hogy hogyan tudjuk ezt azonosítani. Ebben a tekintetben fontos megjegyezni, hogy a tudatelméletek kidolgozásához használt kísérleti módszerek nem feltétlenül alkalmasak a feladatra. Az olyan technikák segítségével, mint a maszkolás és a binokuláris rivalizálás, meghatározhatjuk, hogy például az embernél a globális munkaterület aktiválása elegendő a tudatossághoz, de az a kérdés, hogy mit *tekintsünk* globális munkaterületnek nem emberi esetekben, másfajta elméletalkotást és vizsgálatot igényel.

4 | NÉGY MEGOLDÁS A SPECIFIKUSSÁGRA

Az 5. és 6. szakaszban felvázolom azt az ígéretes módszertant, amelyet a tudatelméletünk sajátosságainak kalibrálására használunk, hogy lehetővé tegyük a nem emberekre való pontos alkalmazásukat. Először azonban hasznos lesz feltérképezni néhány tágabb elméleti terepet. Ezt szem előtt tartva most röviden áttekintek négy lehetséges megközelítést, és meghatározom, hogy szerintem milyen kezdeti nehézségek merülnek fel mindegyiknél. Hangsúlyoznám azonban, hogy a következő rövid megjegyzések nem a kiütéses ellenvetéseknek szólnak. Sőt, az alább tárgyalt megközelítéseket nem tekintem a specifikussági probléma megoldásainak kimerítőnek, és még csak nem is feltétlenül a legjobb vagy legkifinomultabb válaszoknak, amelyeket elfogadhatnánk. Inkább azért választottam ki őket, mert viszonylag világos és egyszerű megoldásokat próbálnak adni a kérdésre, és így segítenek körvonalazni a lehetséges válaszok szélesebb körét.

4.1 | Konzervativizmus

A tudatosság elméleteink nem emberi rendszerekre való alkalmazásának egyik egyszerű stratégiája az lenne, ha a konzervativizmus oldalán tévednénk, és - egyértelmű bizonyítékok

hiányában - azt feltételeznénk, hogy az emberben a tudatosságra vonatkozó jelölt mechanizmusunkhoz kapcsolódó képességek teljes skálája alapvető fontosságú. A GWT konzervatív olvasata például úgy kezdődhetne, hogy

azzal a megfigyeléssel, hogy a globálisan sugárzott információ az emberekben felhasználható a reflektív gondolkodásban, a mérlegelésben, az önismeretben és a szóbeli jelentésben, és ebből arra következtetnek, hogy mindezek a képességek szükségesek a tudatossághoz. Ezzel szemben egy konzervatív magasabb rendű gondolkodás teoretikusa megkövetelheti, hogy minden tudatos lény rendelkezzen az összetett, koncepcionálisan strukturált önreflexív gondolkodás képességével. Ez egyértelmű választ adna a specifikussági problémára, amennyiben a tudatosság csak azokra a rendszerekre korlátozódna, amelyek az emberben a tudatosság szempontjából kritikusnak ítélt kognitív mechanizmusok nagyon közeli analógiáival rendelkeznek.

Természetesen egy konzervatív nézetből valószínűleg az következik, hogy kevés, vagy egyáltalán nem emberi állatnak van tudata, azon az egyszerű alapon, hogy az ember számos, a természetben egyedülállónak tűnő kognitív adottsággal rendelkezik, mint például kifinomult nyelvi és fogalmi képességeink. Ez persze nem vezethet bennünket arra, hogy ezt eleve elutasítsuk. Az az elképzelés, hogy a tudatosság és a magasabb kognitív funkciók szorosan összefüggnek, és hogy a nem emberi állatokból következésképpen teljesen hiányzik a tudatosság, számos védelmezőt talált mind az átmeneti, mind a történelmi időkben (lásd pl. Carruthers, 1999; Davidson, 1975).

Ennek ellenére kevés modern olvasó érzi magát kényelmesen egy olyan nézet mellett, amely a tudatot kizárólag az embereknek tulajdonítja. Bár vitatkozhatunk arról, hogy mekkora bizonyítási súlyt kell adni ezeknek az intuitív megfontolásoknak, ha van egyáltalán, a konzervativizmus fent vázolt egyszerű formája szempontjából talán az a fontosabb aggodalom, hogy egy ilyen álláspont indokolatlanul antropocentrikus lenne, nem annyiban, hogy önmagában tagadná a nem emberi rendszerek tudatosságát, hanem inkább abban, hogy motiválatlanak tűnik kizárni a tudatot a miénktől eltérő kognitív architektúrájú rendszerekből, *pusztán* e különbségek alapján. Természetesen teljesen lehetséges, hogy a kognitív architektúra *bizonyos* különbségei döntőek lesznek annak szempontjából, hogy egy rendszer tudatos-e vagy sem, de nem szabad eleve azt feltételeznünk, hogy *minden* ilyen különbség releváns lesz.

Hogy illusztráljam a dolgot, azt javaslom, hogy nem túlzás azt feltételezni, hogy létezhet egy olyan lény, amelynek intelligenciája és viselkedési képességei nagyjából megegyeznek az emberekével, de kognitív felépítésében jelentősen különbözik tőlünk, esetleg inkább térképszerű, mint fogalmi belső reprezentációs rendszerrel rendelkezik (Boyle, 2019; Camp, 2007), vagy a miénktől jelentősen eltérő tulajdonságú rövid távú memóriarendszerrel rendelkezik. Bár lehet, hogy ezek a különbségek olyan mértékűek, hogy a rendszer nem tudatos, a tudatosság lehetőségének teljes elvetése egy ilyen rendszerben csak azért, mert *a* releváns kognitív mechanizmusaiiban jelentős architekturális különbségeket kapunk, kissé alulmotiválnak tűnik, nem is beszélve a rejtélyességről.

Következésképpen, bár elismerem, hogy lehetne még többet is mondani, hogy a konzervatív nézet vonzóvá váljon, azt javaslom, hogy legalábbis első pillantásra nem tűnik ígéretes válasznak a sajátosság problémájára.

4.2 | Liberalizmus

Egy másik egyszerű megközelítés a tudatot a szóban forgó elmélet által javasolt kognitív képességekkel és mechanizmusokkal azonosítaná a lehető legnagyobb mértékű általánosságban. A GWT esetében ez az egész rendszerre kiterjedő információmegosztás bármely formáját jelentheti, míg a tudatosság nagyon liberális metakognitív megközelítése a tudatot bármely olyan rendszerben jelenlévőként azonosíthatja, amely képes saját mentális állapotainak tulajdonságait más kognitív folyamatok bemeneteként használni (vö. Shea, 2012b).

Miközben ez a liberális megközelítés elkerüli a konzervativizmus bűnei alól, az ellenkező váddal szemben is sebezhető, nevezetesen, hogy túlságosan *nagyvonalú* a tudatossághoz való viszonyulásban. Jelenségek

mint a globális információmegosztás és a metakogníció a legszabadabb értelmezésükben mind a természetben, mind a létező mesterséges rendszerekben mindenütt jelen vannak, és valószínűleg a rendszer tágabb kognitív képességeitől, például a szenzomotoros képességektől vagy az általános intelligenciától függetlenül is megvalósíthatók. Egy játékpéldával élve, tekintsük a hálózati időprotokollt (NTP), amely a belső órák szinkronizálására szolgáló eszköz a hálózaton belüli eszközök között. Legalábbis látszólag ez a globális információmegosztás egy formájának minősülhet (legalábbis a GWT néhány radikálisan liberális értelmezése szerint), mégis nehéz komolyan venni annak lehetőségét, hogy egy ilyen viszonylag egyszerű folyamat tudatosságot eredményezne.

Ezt az intuíciót bevallottan nem mindenki osztja; vannak, akik üdvözlik a tudatosságnak a természetes és mesterséges rendszerekben mindenütt jelenlévőnek tekintését. Az integrált információelmélet prominens védelmezői például úgy vélik, hogy a tudatosság "valószínűleg széles körben elterjedt az ani- málisok között, és ... kis mennyiségben még bizonyos egyszerű rendszerekben is megtalálható" (Tononi & Koch, 2015, 11. o.). Bár az ilyen nézeteket nem lehet a karosszékéből kizárni, érdemes elismerni, hogy jelentős mértékben erőszakot tesznek a tudat elterjedéséről alkotott előelméleti nézeteinken. Schwitzgebel jól fogalmazza meg a lényegét, amikor megjegyzi, hogy "[a]z elméletalkotásunk előtt tudjuk, gondoljuk, hogy a tudatos entitások körébe nem tartoznak bele a protonok vagy az egyszerű logikai kapuk" (24. o.).

Az ilyen elmélet előtti közhelyek jelentősége önmagában is vitatott, de legalábbis egy befolyásos nézet szerint (Lewis, 1972) az elmélet keresése során az kell, hogy legyen az egyik irányadó korlát, hogy az elmélet a lehető legnagyobb mértékben rendszerezze az elmélet előtti ítéleteinket arról, hogy mely állapotok és rendszerek tudatosak. Még ha nem is kívánjuk ezt a konkrét módszertant elfogadni, valószínűleg akkor is kívánunk majd *némi* súlyt adni a tudatossággal kapcsolatos előelméleti ítéleteknek, nem utolsósorban annak elismerése miatt, hogy ezek az ítéletek szerepet játszanak a feltárni kívánt kezdeti jelenség azonosításában.⁶ A legtöbben közülünk nyugodtan elvetnénk például egy olyan tudatelméletet, amely azt állítja, hogy a gyerekek tizenéves korukig nem tudatosak, vagy hogy a fonálférgek tudatosak, a kutyák pedig nem. Anélkül tehát, hogy szentnek tekintenénk a preteoretikus előítéleteinket, jó okunk lehet arra, hogy szerepet szánjunk nekik a tudatosság tudományos elméleteinek értékelésében. Következésképpen, amennyiben egy liberális megközelítés durván ellentétes az előítéleteinkkel - például ha azt állítjuk, hogy egy nagyon egyszerű robot is lehet tudatos -, akkor *ceteris paribus* kevésbé lesz vonzó, mint az, amelyik a tudatosságot úgy mutatja be, hogy nem tesz erőszakot a preteoretikus attitűdjeinken.

4.3 | Inkrementalizmus

Megpróbálhatunk a konzervativizmus és a liberalizmus közötti irányt választani egy harmadik megközelítés elfogadásával, amelyet *inkrementalizmusnak* nevezek. Röviden, ezt a kifejezést olyan nézetekre fogom használni, amelyek a tudatosságot fokozatosan érkezőnek tekintik. Itt nem csupán arról van szó, hogy a különböző rendszerek különböző tudatossággal rendelkeznek, amelyek intenzitásukban, komplexitásukban vagy gazdagságukban különböznek egymástól (ebben valószínűleg minden teoretikus egyetértene), hanem arról, hogy maga a tudat alapvető tulajdonsága nem egy mindent vagy semmit kérdés. Ahelyett, hogy például azt kérdeznénk, hogy *van-e* vagy *nincs valami, ami olyan*, mint egy rák vagy egy számítógép, inkább azt kérdezhetnénk, hogy milyen *mértékben van* valami, ami olyan, mint egy ilyen rendszer.

A fent vázolt inkrementalizmus egy olyan álláspont, amellyel szimpatizálhatunk, vagy amelyet a jelen vitától teljesen független okokból elutasíthatunk. Remélhetjük azonban, hogy

⁶Lásd még Boyd (1999) a "programszerű" és "magyarázó" meghatározások szerepéről a természetes fajták esetében.

ha ilyen álláspontot képviselünk, a sajátosság problémája könnyen kezelhető. Például kombinálhatnánk egy inkrementalista nézetet a globális munkatér-elmélet melletti elkötelezettséggel, és javasolhatnánk, hogy a "teljes tudatossághoz" az emberi globális munkatér teljes kognitív architektúrája szükséges, miközben megengedhetnénk, hogy más rendszerek bizonyos *fokú* tudatossággal rendelkezzenek, amennyiben a rendszer megközelíti ezt az architektúrát. Alternatív megoldásként a teljes tudatosságot az emberi lények gazdag metakognitív képességeivel azonosíthatnánk, miközben megengedhetnénk, hogy a metakogníció kezdetlegesebb formáival rendelkező lények is rendelkezzenek valamilyen kisebb mértékű tudatossággal.

Az inkrementalista megközelítések nyilvánvalóan vonzóak, mivel lehetővé teszik számunkra, hogy széles hálót vessünk ki azon rendszerek körére, amelyek tudatosnak minősülhetnek, miközben fenntarthatjuk, hogy bizonyos fontos értelemben egy olyan élőlény, mint egy homár vagy egy egyszerű robot, sokkal *kevésbé* tudatos, mint mi. A tudat evolúciós megközelítéséhez is jól illeszkednek, mivel azt sugallják, hogy a tudat fokozatosan alakulhatott ki olyan formákban, amelyek nagyon különböznek attól, amit mi az emberi tapasztalatokhoz társítunk.

Sokak számára azonban alapvető akadályt jelent majd egy ilyen nézet elfogadásában maga a tudatossági fokozatok fogalma. Bár kétségtelenül igaz, hogy olyan tudatos élményeket élünk át, amelyek tisztasága, élénksége és gazdagsága tekintetében különböznek egymástól, mint például ébredéskor vagy elalváskor, vagy kábítószer vagy alkohol hatása alatt, mindezekben az esetekben mégis arról van szó, hogy van *valami, amihez hasonló* (Nagel, 1974) a vonatkozó élmény; vagyis még mindig egyértelműen tudatos állapotban vagyunk. Ahogy Carruthers (2018a) fogalmaz: "[o]lyan lehetünk félig éberek, vagy csak részben éberek. De néhány olyan állapot, amelyben az ember félig öntudatos állapotban van, mégis teljesen (egyértelműen) fenomenálisan tudatos" (53. o.; lásd még Simon, 2017). Az ilyen familiáris példák tehát nem segítenek abban, hogy értelmet nyerjen az elképzelés, hogy mit jelentene az, ha egy élőlény a tudatosság bizonyos (nem teljes) fokával rendelkezne. Nem kívánok ezen a ponton túlságosan megpihenni, már csak azért sem, mert a filozófusok között mélyen vitatott kérdés, hogy a tudatosság fokozatokban jöhet-e létre vagy homályosságot mutat-e (Papineau, 2003; Rosenthal, 2019). Mindazonáltal, amennyiben az olvasó a tudatosság fokozatait rejtélyes fogalomnak tartja, ez csökkenteni fogja a specifikussági probléma inkrementális megoldásának vonzerejét.

Az inkrementalizmus másik aggálya az, hogy miként tudunk elvi kritériumokat felállítani a különböző nem emberi rendszerek közötti hasonlóság értékelésére, amelyek közül néhánynak a miénktől gyökeresen eltérő kognitív architektúrája lehet. Ha egy rendszer például rendelkezik tételesen strukturált gondolkodással, de nem rendelkezik szenzomotoros képességekkel, akkor azt kell-e mondanunk, hogy több vagy kevesebb tudattal rendelkezik, mint egy ellentétes adottságokkal rendelkező rendszer? A lényeg nem az, hogy az ilyen összehasonlítások értelmetlenek vagy lehetetlenek, hanem inkább az, hogy csak a különböző kognitív képességek relatív fontosságára vonatkozó további teo- retikai perspektíva alapján lehet őket elvégezni, és ezt a perspektívát valószínűleg nem az általunk támogatni kívánt konkrét tudatelmélet biztosítja. A GWT alapvető kerete például önmagában nem határozza meg, hogy a globálisan sugárzott információk myr- iad fogyasztói rendszerei közül melyek különösen fontosak a tudatosság szempontjából.

Annak a kérdésnek a megválaszolásához, hogy mely rendszerek számítanak a tudatosság szempontjából, megpróbálhatunk olyan léziós vizsgálatokból származó bizonyítékokra hivatkozni, amelyekben bizonyos kognitív képességek hiányoznak vagy drámai módon károsodnak, miközben a tudatosság nagyjából épnek tűnik, mint például az epizodikus emlékezet elvesztése a hippocampusz sérülésében szenvedő betegeknél (lásd pl. Burgess, Maguire & O'Keefe, 2002). Az ilyen esetek azonban nem döntenek el azt a kérdést, hogy milyen

mértékben lehet jelen a tudatosság. Dönthetünk például úgy, hogy a verbális beszámolót helyezzük előtérbe, és azt állíthatjuk, hogy a betegek, akik még mindig képesek beszámolni egyidejű élményeikről, teljes mértékben tudatosak. De azt is állíthatnánk, hogy az emlékek néhány másodpercnél hosszabb ideig való megtartásának képességétől megfosztva a páciensek, akiknek

súlyos anterográd amnéziával küzdő betegeknél alacsonyabb volt a tudatossági szint, mint a nem verbális betegeknél, akiknek teljesen ép volt az epizodikus emlékezetük. Az ilyen kérdések mélyreható és nagyrészt elméleti jellegűek, és úgy tűnik, hogy nem valószínű, hogy csak empirikus adatokra való hivatkozással megválaszolhatók, és bár ezt nem tekintem úgy, hogy ez az inkrementalizmus mint a tudat tudományának szélesebb perspektívája ellen szól, de legalábbis azt sugallja, hogy talán nem nyújt egyszerű megoldást a specifikusság problémájára.

4.4 | Elutasítás

Az utolsó megközelítés, amelyet meg kívánok vizsgálni, a tudatosság kognitív elméleteinek a nem emberekre való alkalmazásával kapcsolatos *elutasítás*. Ez alatt azt a nézetet értem, hogy a nem emberi tudatossággal kapcsolatos kérdések nem azért rosszul felvetettek, mert a nem emberi rendszerek nem egyszerűen tudatosak, hanem azért, mert a tudatos és tudattalan kategóriák nem alkalmazhatók rájuk érthetően. Ez Carruthers (2018a, 2018b) megfontolt álláspontja, aki a fenti 3. szakaszban felvetettekhez hasonló megfontolások alapján azt állítja, hogy a nem emberi tudattal kapcsolatos kérdések rosszul vannak felfogva, és nincs rájuk semmilyen informatív vagy hasznos válasz.

Bár jelen összefüggésben nem lehetséges Carruthers érvelésének részletes összefoglalása, a főbb pontok a következőkben foglalhatók össze. Először is, Carruthers azt állítja, hogy a globális műsorszórás az emberben a tudatosság egyik erős jelölt mechanizmusa. Másodsor, azt állítja - kontra inkrementalizmus -, hogy az emberi tudatosság "minden vagy semmi", és hogy az emberi állapotaink vagy fenomenálisan tudatosak, vagy nem.⁷ Harmadsor, azt sugallja, hogy "a globális műsorszórás fokozatokat ismer el a fajok között" (2018a, 48. o.), és (ahogy én is érveltem fentebb), hogy nincsenek elvi módszerek annak meghatározására, hogy az emberekben a globális munkaterület mely aspektusai kritikusak és melyek mellékesek a tudatos tapasztalatainkhoz.

E megfontolások fényében azt javasolja, hogy az állati tudatossággal kapcsolatban két álláspontot lehet képviselni, attól függően, hogy a globális közvetítést pusztán a tudatossághoz kapcsolódónak tekintjük-e, vagy a jelenség teljes mértékben reduktív leírását alkotjuk. Ha a for- mer "kvalia-realista" álláspontot képviseljük, akkor "az állati tudatosság kérdése megoldhatatlanná válik ... [f]agy nem lehet tudni, hogy a munkaterület mely összetevői elegendőek a kvaliák jelenlétéhez" (2018b, 198. o.). Ezzel szemben, ha az utóbbi nézetet fogadjuk el, ahogy Carruthers sürgeti, akkor el kell ismernünk, hogy a "tudat" kifejezés az emberi kognitív dinamikának egy olyan sajátos tulajdonságát ragadja ki, amelyet kevés, ha egyáltalán más állat fog pontosan instanciálni, éppen annyiban, amennyiben számtalan különböző módon hasonlítanak és különböznek tőlünk a fajok között. Következésképpen, állítja Carruthers, "az állati tudattal kapcsolatban nincs ténszerűség", sőt, "az állati tudatosság kérdése már nem is érdekel jelentős mértékben" (2018b, 200. o.).⁸

Carruthers érvei kifinomultak és több tanulmányban és egy hamarosan megjelenő könyvben (Carruthers, 2020) is kifejtésre kerülnek. Mindazonáltal következtetései tagadhatatlanul radikálisak - talán elég radikálisak ahhoz, hogy sokan azt feltételezzék, hogy az ezeket megalapozó következtetéseknek hibásaknak kell lenniük. Ahogy Schwitzgebel (hamarosan megjelenő) fogalmaz: "[a]zzal a meggyőződéssel, hogy van valami, ami ez

⁷Bár Carruthers úgy tűnik, hogy ezt az állítást nagyrészt fogalmi alapon védi (pl. azt állítja, hogy "nehéz elképzelni egy olyan érzékelési állapotot, amely részben fenomenálisan tudatos, részben pedig nem"), azt is megjegyzi, hogy a globális adás neurális dinamikája is ezt a "minden vagy semmi" jellemzőt mutatja: A neurális populációk aktivációi vagy elérik a "globális gyűjtáshoz" szükséges küszöbértéket, vagy nem (Dehaene, 2014).

⁸Természetes ellenvetés lehet, hogy az állati tudattal kapcsolatos kérdéseknek redukálhatatlan *erkölcsi* érdeke van,

amennyiben különleges kötelezettségeink vannak a fájdalomérzésre képes érző lényekkel szemben (lásd pl. Regan, 2004; Singer, 1989).

Carruthers azonban máshol (1999, 2005) amellett érvelt, hogy az állatjólét erkölcsi kérdései nem a tudatosság kérdéseitől függenek, hanem a vágy frusztrációjának fogalmaival fogalmazhatók meg (lásd még Dawkins, 2012).

szeretünk kutyának lenni, mint amilyenek valaha is lehetnénk, hogy egy okos filozófiai érv az ellenkezőjére valójában megállja a helyét" (24. o.).⁹ A magam részéről, bár elutasítom Carruthers nézetét, a jelen dolgozat hátralévő részében nem tudok megfelelő választ adni rá. Következésképpen, és ennek a mulasztásnak a fényében elismerem, hogy a kérdést, amely a következő értekezést keretbe foglalja, feltételes módban kell érteni: Feltételezve, hogy az elutasítás téves, és az ani- mal tudatosság kérdése valóban lényegi kérdés, milyen lehetséges választ adhatunk a speci- ficitás problémájára?

5 | A TUDATOSSÁG ELMÉLETEKKEL TELI ÉS ELMÉLETEKKEL KÖNNYÍTETT MEGKÖZELÍTÉSEI

Eddig azzal érveltem, hogy a specifikusság problémája komoly kihívást jelent a tudatosság kognitív elméleteinek nem emberi rendszerekre való alkalmazása szempontjából. A következőkben azzal kívánok érvelni, hogy bár a probléma megkövetelheti, hogy felülvizsgáljuk az állati tudat megközelítésére szolgáló módszereink egy részét, mégsem megoldhatatlan. Ahhoz, hogy felvázolhassuk a megoldást, először is meg kell vitatnunk a nem emberi tudatosság problémájának két átfogó megközelítése közötti kulcsfontosságú különbséget, nevezetesen az "elméletnehéz" és az "elméletkönnyű" megközelítést (Birch, hamarosan megjelenik).

Röviden, az elmélet-központú megközelítés az, amit a fenti vitában hallgatólagosan feltételeztünk, nevezetesen, hogy az állati tudattal kapcsolatos kérdéseket úgy kell megközelítenünk, hogy megvizsgáljuk, rendelkeznek-e az emberi tudatról szóló legjobb elméletünk által azonosított releváns mechanizmusokkal. E megközelítés szerint annak meghatározásakor, hogy mely nem emberi rendszerek tudatosak, a következő kihívásokkal kell szembenéznünk: először is azonosítani azokat a mechanizmusokat, amelyek az egyes mentális állapotokat tudatossá teszik az emberekben, másodsor pedig azonosítani, hogy mely nem emberi rendszerek valósítják meg a vonatkozó mechanizmusokat, és így rendelkeznek legalább néhány tudatos állapottal. Ez a megközelítés, ahogyan fentebb kifejtettük, a specifikussági probléma formájában erőteljes kihívásba ütközik.

Ezzel a kihívással szembesülve inkább egy egészen más stratégiát kereshetnénk a tudat jelenlétének értékelésére a nem emberi rendszerekben. Az egyik ilyen módszer az, amit Birch "theory-light" megközelítésnek nevez (Birch, hamarosan megjelenik). Röviden, az elmélet-könnyű megközelítés szerint nem kell előre komoly elméleti kötelezettségeket vállalnunk a tudat mögöttes mechanizmusait illetően. Ehelyett azt javasolja, hogy a tudatosság plauzibilis viselkedéses markereinek emberi esetekre való hivatkozással történő azonosítására, majd ezek állatokra való alkalmazására támaszkodjunk.

Egy egyszerű példával élve, tegyük fel, hogy a viselkedéses bizonyítékok alapján arra a következtetésre jutunk, hogy az emberek nem képesek elvégezni egy T feladatot, amely S ingerrel kapcsolatos, és amelyben S-t szubliminálisan mutatják be. Ennek alapján, ha felfedeznénk, hogy valamilyen nem emberi rendszer, az NH képes a T feladat elvégzésére, akkor ezt joggal tekinthetnénk bizonyítéknak arra, hogy a rendszer tudatos, azon az alapon, hogy a T feladatot egy ember nem tudná elvégezni tudatos feldolgozás nélkül. Ezt a bizonyítékot tovább erősítené, ha azt találnánk, hogy az NH több, a T-hez hasonló, különálló feladat elvégzésére is képes, amelyek mindegyike az embernél tudatosságot igényel.

⁹Figyeljük meg a híres (és ellentmondásos) *Cambridge-i nyilatkozatot a tudatosságról*, amelyet 2012-ben számos vezető idegtudós írt alá, és amely azt állította, hogy "a bizonyítékok súlya azt jelzi, hogy az ember nem egyedülálló a tudatosságot létrehozó neurológiai szubsztrátumok birtokában" (2. o.).

Először is meg kell jegyezni, hogy az elmélet-light megközelítés vitathatatlanul megragadja a kognitív etológia és az állati elmék filozófiájának nagy részében rejlő implicit módszertant. Az e területen dolgozó kutatók általában nem fogadnak el kezdetben explicit tudatelméletet, hanem ehelyett neurobiológiai és viselkedési adatok széles körére támaszkodnak, hogy előzetes ítéleteket hozzanak arról, hogy mely állatok valószínűleg tudatosak (Barron & Klein, 2016; Bekoff, 2010; Birch, 2017; Griffin, 2001). Vegyük például a Griffin és Speck (2004) következő módszertani megállapítását:

Bár egyetlen bizonyíték sem szolgáltat abszolút bizonyítékot a tudatosságra, [az] erősen utaló bizonyítékok halmozódása jelentősen növeli annak valószínűségét, hogy egyes állatok legalább egyszerű tudatos gondolatokat és érzéseket tapasztalnak (Griffin & Speck, 2004, 5. o.).

Természetesen az elmélet-light megközelítés által támogatott következtetés hibás és bizonytalan lenne; előfordulhat például, hogy a T feladatot az embereknél kiszolgáló mechanizmusok eltérnek a nem emberi NH rendszer által használt mechanizmusoktól, így a T feladat NH általi végrehajtása nem jár együtt a tudattal. Hasonlóképpen az is előfordulhat, hogy a tudatosság valamilyen globálisan szükséges feltétele (pl. az öntudat képessége), amely az embernél mindig jelen van, az NH-nál mindig hiányzik. Ebben az esetben a T feladat elvégzésének képessége nagyon jó bizonyíték lehet a tudatosságra az emberben és más olyan rendszerekben, amelyekben a tudatos tapasztalás háttérfeltételei megvalósulnak, míg az NH-hez hasonló rendszerekben nagyon kevés bizonyítékot szolgáltat a tudatosságról.

Az ilyen megfontolásokat természetesen nem lehet félvállról venni. Mindazonáltal az elmélet-light megközelítés reménye az, hogy az ilyen markerek nagy számának vizsgálatán keresztül *ésszerű következtetéseket tudunk* levonni a tudatosság valószínűségére vonatkozóan a különböző rendszerekben. Birch például megjegyzi, hogy legalább három olyan feladatot, amely az embernél tudatosságot igényel - nevezetesen a nyomkövetéses kondicionálást, a fordított tanulást és a multiszenzoros tanulást - a jelek szerint a méhek is el tudnak végezni. A tudatosság markereinek e csoportosulása fényében azt sugallja, hogy a méhek kiváló tudatosság-jelöltek, mivel megfelelnek az emberi tudatosság három kulcsfontosságú markerének (Birch, hamarosan megjelenik; lásd még Shea, 2012a).

Úgy tűnhet, hogy az elmélet-könnyű megközelítés egyszerű javulást kínál az elmélet-hangsúlyos megközelítéshez képest, lehetővé téve a nem-emberek tudatosságának ésszerű értékelését, miközben ügyesen megkerüli a specifikusság problémáját. Bár egyetértek Birch-csel abban, hogy az elmélet-light megközelítés értékes módszer a nem-emberek tudatosságának értékelésére, számos fontos korlátja van.

Úgy vélem, hogy három olyan eset van, amikor ez a legjobb esetben is kevésbé informatív, a legrosszabb esetben pedig félrevezető lehet, ha túl szigorúan alkalmazzák. Először is, vegyük észre, hogy az elmélet-light megközelítés némileg híján van azoknak az erőforrásoknak, amelyekkel meghatározható, hogy egy rendszer mely mentális állapotai tudatosak. Bizonyos esetekben ésszerű lehet arra következtetni, hogy egy adott nem-emberi *érzékelési* állapot tudatos, ha például olyan tanulási vagy viselkedési formák (például a nyomkövetéses kondicionálás) bemeneteként szolgálhat, amelyekről okunk van feltételezni, hogy az embernél tudatosságot igényelnek. Létezik azonban egy sor más állapot - érzelmi állapotok, vágyak és állandó reprezentációs állapotok, mint például a hiedelmek -, amelyek nem értékelhetők ilyen módon.

Bizonyos esetekben ennek több mint elméleti jelentősége lehet. Képzeljük el, hogy okunk van feltételezni (esetleg fiziológiai mutatók alapján), hogy egy szervezet nagyfokú stresszben van, és szeretnénk meghatározni, hogy etikai alapon van-e okunk nyugtatót adni. Még akkor is,

ha okunk lenne azt hinni, hogy a szóban forgó élőlény képes lenne

bizonyos tudatos állapotok esetében a tudatosság különböző markereivel való rendelkezése alapján ez nem engedné meg, hogy automatikusan arra következtessünk, hogy a stressz állapota önmagában egy tudatos negatív érzellemmel társul. Az sem világos azonnal, hogy egy elmélet-centrikus megközelítés hogyan oldhatná meg ezt a kérdést, amennyiben az ilyen érzelmek esetleg nem szolgálhatnak inputként az általunk a tudatosság markereiként azonosított intelligens viselkedésfajtákhoz.

Az elmélet-könnyű megközelítés második korlátja az, hogy olyan rendszerekről is képes ítéletet mondani, amelyek nem mutatják a tudatosság releváns markereit, és fennáll a "hamis negatív eredmények" kockázata. Míg az olyan képességek, mint a nyomkövető kondicionálás, ha jelen vannak, bizonyítékot szolgáltathatnak a tudat jelenlétére, nem világos, hogy miért kellene a hiányukat a tudat hiányának bizonyítékaként kezelnünk, különösen, ha az embertől teljesen eltérő kognitív architektúrájú rendszerekkel, például mesterséges rendszerekkel vagy egyszerűbb állatokkal van dolgunk. Gondoljunk például arra, hogy egy olyan élőlénynek, amely erősen támaszkodik egyetlen érzékszervi modalitásra, például a szaglásra, kevés oka lehet arra, hogy elsajátítsa vagy megőrizze a multiszenzoros tanulás képességét. Hasonlóképpen, egy olyan organizmus, amely nagyon stabil környezeti jelzésekkel rendelkező környezetben él, talán kevés előnnyel járna a fordított tanulás képességéből, annak ellenére, hogy más okból (pl. kifinomult szociális intelligenciával rendelkezik a saját környezetével való kapcsolatok kezelésében) intuitívan hihető tudatosság-jelölt. Tágabb értelemben felhívnam a figyelmet arra a tényre, hogy az emberi lények számos tudatos élménye, például a sok érzékelési élményben és testi érzetben szerepet játszó egyszerű kvaliák legalábbis első látásra nem kapcsolódnak a kifinomult tanulási képességekhez, ami legalábbis azt a lehetőséget sugallja, hogy a tudatossághoz talán eleve nincs is szükség ilyen képességekre. Az ilyen esetek kezelésében, ahol a tudatosság jelzői hiányoznak, az elmélet- könnyű megközelítés híveinek tehát vagy dogmatikusan ki kell zárniuk a tudatot, vagy pedig agnosztikusnak kell maradniuk, és ezzel el kell ismerniük, hogy a megközelítés alkalmazása korlátozott.

Az elmélet-light megközelítés utolsó korlátja, hogy előfordulhatnak olyan esetek, amikor a markerek félrevezetőek lehetnek, ami a tudatosság téves attribúciójához és hamis pozitív eredményekhez vezethet. Különösen az olyan mesterséges rendszerek lehetőségére gondolok, amelyeket kifejezetten úgy alakítottak ki, hogy a tudatosság általunk javasolt markereit mutassák, miközben egyébként meglehetősen leegyszerűsítettek. Így legalábbis *prima facie* lehetségesnek tűnik, hogy egy mérnök olyan robotot építhet, amely képes a nyomkövetésre, a fordított tanulásra és a multiszenzoros tanulásra, miközben egyébként nagyon gyenge tudatosság-jelölt (például nem képes a célvezérelt viselkedés más formáira). Bár nem zárhatjuk ki, hogy egy ilyen rendszer tudatos lenne, ez bizonyára egészen másfajta eset lenne, mint egy biológiai organizmus, amely ugyanezeket a jeleket mutatja, és ezért különleges kezelést igényel. Ideális esetben azt kívánnánk, hogy ilyen esetben - és erre az elmélet-light megközelítés nem tűnik alkalmasnak -, hogy legyen valamilyen elvi módszer annak értékelésére, hogy a tudatosság a megfelelő markerek jelenléte *ellenére* is hiányozhat-e.

6 | SZERÉNY ELMÉLETI MEGKÖZELÍTÉS

Azt javasolnám tehát, hogy bár az elmélet-light megközelítés elkerülheti a specifikussági problémát, alkalmazhatóságának is vannak bizonyos korlátai, különösen a filogenetikailag távolabbi szervezetek vagy egzotikusabb rendszerek esetében. A dolgozatnak ebben az utolsó részében egy stratégiát szeretnék javasolni az elmélet-nehezített és az elmélet-könnyű megközelítés konstruktív kombinálására oly módon, amely lehetővé teszi számunkra, hogy leküzdjük a két megközelítés korlátait. Kezdetben lazán megfogalmazva a dolgot, azt

javaslom, hogy az elmélet-súlyos és az elmélet-könnyű megközelítések egyfajta *dinamikusan működő*

egyensúlyt, ahol az egyik felismerése tájékoztatja és korlátozza a másikat. Ezt nevezem én *szerény elméleti megközelítésnek*.

Ennek a stratégiának számos előnye van. Először is, és ez a legfontosabb, ez adhat nekünk némi kezdeti vásárlást a specifikussági problémára. Nagyjából az az elképzelés, hogy a tudatosság markerei megközelítéssel azonosítanánk az erősebb és gyengébb tudatossági jelölteket jelentő fajokat (vö. Birch, 2018). Innen kiindulva megvizsgálhatjuk, hogy kognitív architektúrájuk miben hasonlít és miben különbözik az emberétől. Ezeket a bizonyítékokat aztán kombinálhatjuk a tudatosság általunk preferált elméletével, hogy segítsen meghatározni az elméletünk megfelelő specifikussági szintjét.

A GWT-re alkalmazva ezt például úgy kezdhethetjük, hogy először azonosítjuk a tudatosság elmélettől független markereit, és ezek segítségével azonosítjuk a tudatosságra erősen esélyes nem emberi rendszereket. Tegyük fel, hogy a példa kedvéért azt találjuk, hogy ezek a fajok (a) rendelkeznek az egész agyra kiterjedő információmegosztás valamilyen formájával; (b) képesek szelektíven figyelni az érzékszervi ingerekre; és (c) a globális aktivációval kapcsolatos "győztes mindent visz" információs dinamikát mutatnak. Ugyanakkor azt is megállapítjuk, hogy e lények némelyikének hiányzik az epizodikus memória és az általános munkamemória területe. Ezek az eredmények legalábbis némi okot adnak arra, hogy úgy gondoljuk, hogy a GWT megfelelő specifikussági szintje, amelyen a GWT-t az összehasonlító pszichológia céljaira ki kell írni, olyan, amely magában foglalja a figyelmet és a "mindent visz a győztes" információs dinamikát, de nem tesz lényeges utalást az epizodikus memóriára vagy a terület általános munkamemóriájára. Miután elméletünket ilyen módon meghatároztuk, alkalmazhatjuk azt a nehezebb esetekre, például olyan rendszerekre, amelyek a tudatosság néhány, de nem minden már azonosított jelzőjét mutatják. Ha a kérdéses rendszerek mégis rendelkeznek az elméletünk által javasolt kognitív mechanizmusokkal (a vizsgálat előző szakaszában azonosított specifikussági szinten), akkor lehet, hogy van okunk feltételezni, hogy tudatosak.

Konkrét példaként, amikor ez a módszer alkalmazható, tekintsük a tengeri kardszárnyú lándzsás úszókat (1. ábra). Ezek az élőlények két különböző szakaszból álló életciklussal rendelkeznek. Lárvális stádiumukban a lándzsás úszók szabadon úsznak, de felnőttkorukban beássák magukat a tengerfenékbe, és nagyrészt helyhez kötöttek maradnak. Bár a lanceletek rendelkeznek érzékszervekkel (egyetlen előre néző szemmel) és több ezer neuronból álló aggyal, neuroanatómiájuk lényegesen kevésbé összetett, mint a legegyszerűbb gerinceseké. Viselkedésük köre is ennek megfelelően korlátozott, és



1. ÁBRÁZAT Egy kifejlett lándzsás úszómadár. Attribúció: © Hans Hillewaert. Creative Commons Attribution- Share Alike 4.0 International

licenc alapján.

a lanceletek még a mozgékonyabb lárvastádiumban is végeznek néhány meglehetősen egyszerű viselkedési "rutint", például felfelé úsznak a vízoszlopban, hogy táplálkozzanak, vagy gyorsan elúsznak a veszély elől.

Mondanom sem kell, hogy kevés okunk van azt várni, hogy a lanceletek kifinomult vagy sokoldalú viselkedést mutassanak, amely a tudatosság bizonyítékául szolgálhatna; nem kell például táplálékra vadászniuk, és így kevés tanulási tevékenységet kell végezniük. Ha tehát kizárólag egy olyan elmélet-centrikus megközelítésre támaszkodnánk, amely a tudat jelenlétének megállapításához olyan viselkedéses markereket használ, mint a nyomkondicionálás, akkor valószínűleg agnosztikusnak kellene maradnunk a lándzsás úszók tudatosságát illetően. A fent vázolt szerény elméleti megközelítéssel azonban talán valamivel jobban járunk. Tegyük fel például, hogy gondos tesztelés után úgy találjuk, hogy a lanceletek csak néhány tudati jelzőt mutatnak. Ennek ellenére megcáfolhatnánk, hogy a Lanceletek rendelkeznek a globális munkaterület valamilyen analógiájával, különösen azokkal a jellemzőkkel, amelyeket a vizsgálat előző szakaszában a tudatosság szempontjából kritikusnak azonosítottunk. Ez arra adhat okot, hogy a Lanceleteket potenciális vagy akár ésszerű tudatosság-jelölteknek tekintsük.¹⁰ Másik lehetőség, hogy a Lanceletek neurobiológiája túl egyszerű vagy túlságosan eltérő ahhoz, hogy megfeleljen a tudatosságról alkotott elképzelésünknek, és így nemleges választ kapunk. Bármelyik esetben azonban olyan meghatározásra jutottunk volna, amelyet - bár hibásan és bizonytalanul - nehéz, sőt talán lehetetlen lenne pusztán az elmélet-centrikus megközelítéssel elérni.¹¹

A lanceletek esete remélhetőleg arra szolgál, hogy bemutassa, hogyan segíthet egy szerény elméleti megközelítés a specifikussági probléma megoldásában és a nehéz esetek megítélésében. Hasonló módszerrel lehetne megállapításokat tenni a tudat jelenlétéről a mesterséges rendszerekben. Így azt találhatnánk, hogy egy egyébként intelligens mesterséges rendszer, amely nem mutatja a tudatosság egyik viselkedéses jelzőjét sem, mégis olyan belső felépítéssel rendelkezik, amely nagymértékben hasonlít az emberi globális munkaterülethez, és így valószínűsíthetően tudatossági jelölt. Ezt a megállapítást azonban csak akkor lehetne megtenni, ha lenne némi alapunk arra, hogy a GWT-t megfelelő szintű specifikussággal írjuk le.

A szerény elméleti megközelítésnek más előnyei is vannak, többek között az, hogy képes egyértelmű ítéleteket hozni a tudatosságról olyan kilátástalan esetekben, mint a fent tárgyalt gerrymandered robot. Konkrétan, ha egyszer már alkalmaztuk az elmélet-könnyű megközelítés által szorgalmazott tudatossági markerek módszerét, hogy a tudatosságról szóló elméletünket a specifikusság egy adott szintjén kiteljesítsük, bizonyos esetekben elvi okunk lehet arra, hogy a markerek jelenlétét mint megbízható jelzőt *figyelman kívül hagyjuk*. Képzeljük el tehát, hogy minden olyan biológiai organizmus, amely $M_1 \dots M_n$ viselkedési markereket mutat, rendelkezik egy $C_1 \dots C_n$ jellemzőkkel rendelkező globális munkaterülettel. Ha ezután találkozunk egy olyan mesterséges rendszerrel, amelyet kifejezetten az $M_1 \dots M_n$ markerek megjelenítésére építettek, de amelyből hiányzik a $C_1 \dots C_n$ jellemzők egésze vagy nagy része, akkor jó okunk lehet kételkedni abban, hogy a rendszer tudatos. Ez potenciálisan lehetővé tenné számunkra, hogy kizárjuk az olyan rendszereket, mint Searle Kínai szobája (Searle, 1980), amelyek intuitíve gyenge tudatosság-jelöltek, annak ellenére, hogy viselkedésbeli kifinomultságuk szuperfinom. Végül is, míg a kínai szoba számos markerrel rendelkezik.

¹⁰Az elképzelés, hogy a lánctalpak a globális munkaterület analógiájával rendelkezhetnek, nagyon valószínűtlennek tűnhet (és talán az is), de érdemes megjegyezni, hogy maga Dehaene is azt állítja, hogy "a tudat egy hasznos eszköz, amely valószínűleg már régen kialakult az evolúció során" (Dehaene, 2014, 244. o.).

¹¹Ez csupán egy példa, és nem áll szándékomban érdemi állításokat tenni a Lancelet-tudatról. A kérdéssel foglalkozó kutatók közül a legtöbben úgy vélik, hogy a Lanceletek valószínűleg nem tudatosak; Feinberg és Mallatt például azt állítják, hogy "a [Lancelet] érzékelési útvonalainak többsége csak három vagy négy neuronszintet tartalmaz a külső inger és a motoros válasz

között ... [és] nem az érzékszervi tudatossághoz szükséges hosszú feldolgozási hierarchia" (Feinberg & Mallatt, 2016, 48-49. o.). Megjegyzendő azonban, hogy ezt az állítást csak annak a nézetüknek a fényében teszik, hogy a tudatossághoz hierarchikus érzékszervi feldolgozásra van szükség.

(nevezetesen a szóbeli beszámolóra való képesség), akkor valószínűsíthetően hiányozni fog a szerény elméleti megközelítés által a tudatosság szempontjából relevánsnak ítélt mechanizmusok - legyen szó akár a munkamemóriáról, a figyelemről vagy az egész rendszerre kiterjedő információmegosztásról - bármilyen ana- logiája. Természetesen az is előfordulhat, hogy lehetetlen olyan rendszert építeni, amely képes az $M_1 \dots M_n$ tulajdonságok megjelenítésére anélkül, hogy ne rendelkezne a $C_1 \dots C_n$ tulajdonságokkal is. Ez önmagában is rendkívül informatív eredmény lenne, ami talán arra utalna, hogy a szóban forgó markerek mélyen kapcsolódnak a tudatossághoz.

A szerény elméleti megközelítéssel kapcsolatos eddigi vitám arra összpontosított, hogyan használhatnánk a tudatosság jelölőit a tudatosság adott elméletének csiszolására és pontosítására, hogy az alkalmazhatóvá váljon a trükkös nem emberi esetekre is. Befejezésül azonban két további hasznos alkalmazást szeretnék megemlíteni, amelyek nem állnak rendelkezésre az elmélet-közeli megközelítés számára önmagában. Először is, miután elméletünket a megfelelő fokú specifikussággal meghatároztuk, megvizsgálhatjuk, hogy egy élőlény mely állapotai tudatosak, még azok is, amelyeket kezdetben nem mutattak ki a markerek. Elvileg ez egy egyszerű dolog lenne: Miután például arra a következtetésre jutottunk, hogy egy élőlény idegrendszere a globális munkaterület megfelelő analógiáját instanciálja, meghatározhatnánk, hogy egy adott mentális állapotot a munkaterületen sugároznak-e, vagy ehelyett nem központosított útvonalakon keresztül szabályozzák a viselkedést. Az ilyen kérdések gyakorlati megválaszolása természetesen sokkal nehezebb lenne, de ha az elmélet bizonyos mértékét integráljuk egy tágabb megközelítésbe, akkor legalább láthatunk egy lehetséges stratégiát a kezelésükre.

A szerény elméleti megközelítés másik alkalmazási területe az lenne, hogy a tudatosságra vonatkozó jelölőinket arra használjuk, hogy segítsünk meghatározni a tudatosság viszonylag több és kevésbé hihető elméleteit. Így, miután a markerek segítségével azonosítottuk az erősebb és gyengébb tudatosság-kandi- datokat, felmérhetjük, hogy egy adott elméletet koherens módon úgy lehet-e megfogalmazni, hogy az erősebb jelölteket tartalmazza, míg a gyengébbeket kizárja. Ha felfedeznénk, hogy az elmélet kidolgozásának nincs olyan elvi módja, amely akár csak megközelítőleg is megfelelne a tudatossági markerek módszerével végzett meghatározásoknak, akkor viszont kétségek merülnének fel az elmélet plauzibilitásával kapcsolatban. Durván fogalmazva, ha egy elméletet nem lehet úgy megfogalmazni, hogy a csimpánzokat is a tudatos rendszerek közé soroljuk anélkül, hogy a zsebszámológépeket is bevonjuk, akkor ez egy megdönthető ok lesz a kérdéses elmélet megkérdőjelezésére. Egy ilyen meghatározás természetesen már intuíció alapján is megtehető. A tudatosság markereinek felhasználásával azonban gazdagabb és elvi alapokon nyugvó módunk lesz arra, hogy erős és gyenge tudatossági jelölteket állapítsunk meg anélkül, hogy pusztán az intuícióna hagyatkoznánk, és így árnyaltabb tesztet kapunk, amelynek alávethetünk egy adott elméletet.

7 | KÖVETKEZTETÉS

Ebben a tanulmányban három fő állítást fogalmaztam meg. Az első állításom az volt, hogy a tudatosság kognitív elméletei fontos kihívással szembesülnek a nem emberi rendszerekre való alkalmazásuk során, nevezetesen a specifikusság problémájával. Másodsor, a nem emberi rendszerek tudatosságának meghatározására szolgáló alternatív módszert vizsgáltam, nevezetesen Birch elméletkönnyű megközelítését, és azt javasoltam, hogy számos erénye ellenére van néhány fontos korlátja. Végül azt állítottam, hogy a legjobb stratégia a specifikussági probléma megoldására és a nem-emberek tudatosságának elvi és pontos meghatározására egy olyan szerény elméleti megközelítés, amely a tudatelméleti viták hagyományosabb technikáit integrálja az elmélet-könnyű megközelítés által javasolt tudatos

módszer jelölőivel. Azt javasoltam, hogy a tudatosság markereinek módszerét alkalmazva meghatározhatjuk a tudatelméletek alkalmazásakor alkalmazandó megfelelő specifikussági szintet. Miután ezt megtettük, kiterjeszhetjük elméleteinket a lehetséges esetek feltárására.

a nem emberi tudat, mind a biológiai, mind a mesterséges tudat, amelyet nehéz lenne pusztán az elmélet-centrikus megközelítéssel kezelni.

ORCID

Henry Shevlin  <https://orcid.org/0000-0002-7753-3281>

HIVATKOZÁSOK

- Alkire, M. T. & Miller, J. (2005). Általános érzéstelenítés és a tudat neurális korrelátumai. *Progress in Brain Research*, 150, 229-244. [https://doi.org/10.1016/S0079-6123\(05\)50017-7](https://doi.org/10.1016/S0079-6123(05)50017-7).
- Baars, B. J. (1988). *A tudatosság kognitív elmélete*. New York, NY: Cambridge University Press. Baddeley, A. (1992). Tudatosság és munkamemória. *Consciousness and Cognition*, 1(1), 3-6.
- Barron, A. B. & Klein, C. (2016). Mit mondhatnak a rovarok a tudat eredetéről. *Proceedings of the National Academy of Sciences*, 113, 4900-4908.
- Bekoff, M. (2010). *Az állatok érzelmi élete*. Novato, CA: New World Library.
- Birch, J. (2017). Az állati érzékenység és az elővigyázatosság elve. *Animal Sentience*, 2(16), 1-16.
- Birch, J. (2018). *Az állati érzékenység dimenziói: Szemcsék, valencia, egység és áramlás*. (Publikálatlan kézirat).
- Birch, J. (megjelenés előtt). A gerinctelenek tudatának kutatása: Three approaches. *Nouūs*. <https://doi.org/10.1111/nous.12351>
- Blake, R. & Logothetis, N. K. (2002). Vizuális verseny. *Nature Reviews. Neuroscience*, 3(1), 13-21. <https://doi.org/10.1038/nrn701>.
- Block, N. (1978). Bajok a funkcionalizmussal. *Minnesota Studies in the Philosophy of Science*, 9, 261-325.
- Block, N. (1995). A tudat funkciójával kapcsolatos zűrzavarról. *Behavioral and Brain Sciences*, 18, 227-287.
- Block, N. (2007). Tudatosság, hozzáférhetőség és a pszichológia és az idegtudományok közötti háló. *Behavioral and Brain Sciences*, 30(5), 481-548.
- Block, N. (2009). A tudatosság főbb elméleteinek összehasonlítása. In M. Gazzaniga (Ed.), *The cognitive neurosciences IV* (pp. 1111-1123). Cambridge, MA: MIT Press.
- Boyd, R. N. (1999). Fajok, komplexitás és többszörös megvalósulás: Megjegyzések Millikan *Történelmi fajok és a speciális tudományok* című művéhez. *Filozófiai tanulmányok: An International Journal for Philosophy in the Analytic Tradition*, 95 (1/2), 67-98.
- Boyle, A. (2019). Mások elméjének feltérképezése. *Review of Philosophy and Psychology*, 10(4), 747-767. <https://doi.org/10.17863/CAM.37430>.
- Brancucci, A. & Tommasi, L. (2011). "Binaurális rivalizálás": Dichotikus hallás mint a tudatosság neurális korrelátumának vizsgálatának eszköze. *Brain and Cognition*, 76(2), 218-224. <https://doi.org/10.1016/j.bandc.2011.02.007>
- Buckner, C. (2015). A megismerés tulajdonságklaszter-elmélete. *Philosophical Psychology*, 28(3), 307-336.
- Burgess, N., Maguire, E. A. & O'Keefe, J. (2002). Az emberi hippocampusz és a térbeli és epizodikus emlékezet. *Neuron*, 35(4), 625-641. [https://doi.org/10.1016/S0896-6273\(02\)00830-9](https://doi.org/10.1016/S0896-6273(02)00830-9)
- Camp, E. (2007). Gondolkodás térképekkel. *Philosophical Perspectives*, 21, 1-145.
- Carruthers, P. (1999). Szimpátia és szubjektivitás. *Australasian Journal of Philosophy*, 77(4), 465-482.
- Carruthers, P. (2005). *Tudatosság: Essays from a higher-order perspective*. New York, NY: Oxford University Press.
- Carruthers, P. (2013). A munkamemória evolúciója. *Proceedings of the National Academy of Sciences*, 110(Supplement 2), 10371-10378.
- Carruthers, P. (2018a). Összehasonlító pszichológia tudat nélkül. *Consciousness and Cognition*, 63, 47-60. Carruthers, P. (2018b). Az állati tudatosság problémája. *Proceedings and Addresses of the American Philosophical Association*, 92, 179-205.
- Carruthers, P. (2020). *Emberi és állati elmék: A tudatosság kérdéseinek lezárása*. New York, NY: Oxford University Press.
- Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., ... Massimini, M. (2013). Az érzékszervi feldolgozástól és a viselkedéstől független, elméleti alapú tudatossági index. *Science Translational Medicine*, 5(198), 198ra105. <https://doi.org/10.1126/scitranslmed.3006294>
- Clayton, N. S., Salwiczek, L. H., & Dickinson, A. (2007). Epizodikus emlékezet. *Current Biology*, 17(6), R189-R191.

- Crick, F. & Koch, C. (1990). A tudat neurobiológiai elmélete felé. *Seminars in the Neurosciences*, 2, 263-275.
- Davidson, D. (1975). Gondolkodás és beszéd. In S. Guttenplan (Ed.), *Elme és nyelv* (pp. 7-23). Oxford: Oxford University Press.
- Dawkins, M. S. (2012). *Miért számítanak az állatok: Állati tudatosság, állatjólét és emberi jólét*. Oxford: Oxford University Press.
- Dehaene, S. (2014). *A tudat és az agy: Az agy kódolja gondolatainkat*. New York, NY: Viking Press.
- Dehaene, S. & Naccache, L. (2001). A tudatosság kognitív idegtudománya felé: Alapvető bizonyítékok és egy munkaterületi keretrendszer. *Cognition*, 79, 1-37.
- Feinberg, T. E. & Mallatt, J. M. (2016). *A tudat ősi eredete: Hogyan hozta létre az agy a tapasztalatot*. Cambridge, MA: MIT Press.
- Godfrey-Smith, P. (2016). *Más elmék: A polip és az intelligens élet evolúciója*, London: William Collins. Graziano, M. S. A. (2013). *A tudatosság és a szociális agy*. New York, NY: Oxford University Press.
- Griffin, D. R. (2001). *Állati elmék: Beyond cognition to consciousness*. Chicago, IL: University of Chicago Press.
- Griffin, D. R. & Speck, G. B. (2004). Új bizonyítékok az állati tudatosságról. *Animal Cognition*, 7(1), 5-18. <https://doi.org/10.1007/s10071-003-0203-x>
- Kaminski, J., Call, J. & Tomasello, M. (2008). A csimpánzok tudják, amit mások tudnak, de azt nem, amit ők hisznek. *Cognition*, 109(2), 224-234. <https://doi.org/10.1016/j.cognition.2008.08.010>. <https://doi.org/10.1016/j.cognition.2008.08.010>
- Kouider, S. & Dehaene, S. (2007). A feldolgozás szintjei a nem-tudatos észlelés során: A vizuális maszkolás kritikai áttekintése. *Philosophical Transactions of the Royal Society B*, 362, 857-875.
- Lamme, V. A. (2010). Hogyan változtatja meg az idegtudomány a tudatról alkotott képünket. *Cognitive Neuroscience*, 13, 204-220.
- Lewis, D. (1972). Pszichofizikai és elméleti azonosítások. *Australasian Journal of Philosophy*, 50(3), 249-258.
- Lycan, W. G. (1996). *Tudat és tapasztalat*. Cambridge, MA: MIT Press.
- Merker, B. (2005). A mobilitás kötelezettségei: A tudatosságra való áttérés szelekciós nyomása az állati evolúcióban. *Consciousness and Cognition*, 14(1), 89-114.
- Nagel, T. (1974). Milyen érzés denevérek lenni? *Philosophical Review*, 83, 435-456.
- Papineau, D. (2003). Létezhet-e a tudat tudománya? *Filozófiai kérdések*, 13(1), 205-220. <https://doi.org/10.1111/1533-6077.00012>.
- Prinz, J. (2012). *A tudatos agy: Hogyan generál a figyelem élményt*. New York, NY: Oxford University Press.
- Prinz, J. (2018). Figyelem, munkamemória és állati tudatosság. In K. Andrews & J. Beck (Eds.), *The Routledge handbook of philosophy of animal minds (The Routledge handbook of philosophy of animal minds)* (pp. 185-195). Abingdon: Routledge/Taylor & Francis Group.
- Regan, T. (2004). *Az állatok jogai*. Berkeley & Los Angeles: University of California Press. Rosenthal, D. (2005). *Tudat és elme*. New York, NY: Oxford University Press.
- Rosenthal, D. (2019). Tudatosság és bizalom. *Neuropsychologia*, 128, 255-265. <https://doi.org/10.1016/j.neuropsychologia.2018.01.018>.
- Rosenthal, D. M. (1986). A tudat két fogalma. *Philosophical Studies*, 49(May), 329-359. Schwitzgebel, E. (megjelenés előtt). Van-e valami, amiben kerti csigának lenni? Filozófiai témák. Searle, J. R. (1980). Elmék, agyak és programok. *Behavioral and Brain Sciences*, 3(3), 417-457.
- Shea, N. (2012a). Módszertani találkozások a fenomenális fajtával. *Filozófia és fenomenológiai kutatás*, 84(2), 307-344. <https://doi.org/10.1111/j.1933-1592.2010.00483.x>. <https://doi.org/10.1111/j.1933-1592.2010.00483.x>
- Shea, N. (2012b). A jutalom előrejelzési hibajelek metareprezentációsak. *Nous*, 48(2), 314-341. <https://doi.org/10.1111/j.1468-0068.2012.00863.x>
- Simon, J. A. (2017). Homályosság és zombik: Miért nincs határeset a "fenomenálisan tudatos". *Philosophical Studies*, 174(8), 2105-2123. <https://doi.org/10.1007/s11098-016-0790-4>. <https://doi.org/10.1007/s11098-016-0790-4>.
- Singer, P. (1989). Minden állat egyenlő. In T. Regan & P. Singer (szerk.), *Állati jogok és emberi kötelezettségek* (pp. 215-226). Upper Saddle River, NJ: Pearson.
- Stoerig, P., Zontanou, A. & Cowey, A. (2002). Tudatos vagy nem tudatos: Az agykérgi vakság értékelése négy ember és egy majom esetében. *Cerebral Cortex*, 12(6), 565-574. <https://doi.org/10.1093/cercor/12.6.565>

- Terrace, H. S. & Son, L. K. (2009). Összehasonlító metakogníció. *Current Opinion in Neurobiology*, 19(1), 67-74. <https://doi.org/10.1016/j.conb.2009.06.004>. <https://doi.org/10.1016/j.conb.2009.06.004>
- Thompson, L. T., Moyer, J. R. & Disterhoft, J. F. (1996). Nyulaknál a nyomkövetéses szemlinkelés a tanulási képességek heterogénitását mutatja mind a korcsoportok között, mind az életkoron belül. *Neurobiology of Aging*, 17(4), 619-629. [https://doi.org/10.1016/0197-4580\(96\)00026-7](https://doi.org/10.1016/0197-4580(96)00026-7).
- Tononi, G. (2008). A tudat mint integrált információ: Egy ideiglenes kiáltvány. *The Biological Bulletin*, 215(3), 216-242.
- Tononi, G. & Koch, C. (2015). Tudatosság: Itt, ott és mindenhol? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1668), 1-18.
- Tsuchiya, N., Wilke, M., Frässle, S. & Lamme, V. A. (2015). Nem jelentési paradigmák: A tudatosság valódi neurális korrelációinak kivonása. *Trends in Cognitive Sciences*, 19, 757-770.
- Weiskrantz, L. (1986). *Vakság: Egy esettanulmány és következményei*. New York, NY: Oxford University Press.

Hogyan kell idézni ezt a cikket: Shevlin H. Non-human consciousness and the specificity problem: A modest theoretical proposal. *Mind & Language*. 2021;36:297-314. <https://doi.org/10.1111/mila.12338>.