

Opponensi bírálat Németh Dezső
„Implicit probabilistic learning: from acquisition to consolidation”
című MTA doktori értekezéséről

Az értekezés a szakma elismert folyóirataiban megjelent tizennégy empirikus és egy áttekintő tanulmányt tartalmaz. A doktori mű ennek megfelelően kiváló angolsággal íródott, jól olvasható, bár a bevezető és módszertani részek erősen repetitívek. A téziszfüzet példamutatóan lényegretörő. A tanulmányok célja az, hogy az Alternáló Szeriális Reakcióidő (ASRT) paradigma használatával feltérképezzék az implicit szekvenciatanulást befolyásoló tényezőket. Az alapvetően exploratív kutatás elsősorban különböző életkori és betegcsoportok bevonásával, esetenként pedig az ASRT paradigma célzott variálásával történt. A tanulmányok jól tükrözik a tudományos vállalkozásban elengedhetetlenül szükséges kutatói elszántságot és kitartást, a kísérletek kivitelezése körültekintő.

A kutatás alapfeltevése, hogy a használt ASRT paradigma képes a repetitív feladatvégzés során lezajló készségtanulást, valamint a feladatsorban rejtetten jelenlévő - spontán módon nem tudatosuló - sorrendi eseménymintázat megtanulását, az implicit szekvenciaspecifikus tanulás jelenségét megragadni. Ezt az alapfeltevést a tanulmányok messzemenően alátámasztják, mindkét jelenség robosztusnak mutatkozott. További alapfeltevés az is, hogy a használt paradigma révén kapott szekvenciaspecifikus tanulást jellemző reakcióidő-különbség alkalmas arra, hogy a különböző helyzetekben megvalósuló implicit tanulási teljesítményt, vagy különböző csoportok implicit tanulási képességét jellemezze.

A bevezető elsősorban a paradigma bemutatására és a „*research landscape*” hivatkozásszerű bemutatására fókuszál. Hiányolom a vizsgált jelenség szisztematikus bemutatását. Arról, pl. hogy mi az implicit tanulás, keveset tudunk meg, nem derül ki, pl., hogy miben, milyen ismérvek alapján különül el más (tanulási) jelenségektől. Hasznos lett volna az implicit tanulás és egyes agyi struktúrák kapcsolatára vonatkozó utalások kapcsán itt kifejteni, hogy milyen működések révén járulnak, járulhatnak hozzá ezek az implicit tanulás megvalósulásához. Annak a leírása is ide kíváncszna, hogy milyen információ-feldolgozás (műveletek, reprezentációk, modellek) szükséges egy sorozatbeli predikciós működés megvalósulásához, esetleg az ASRT paradigmánál szélesebb kontextusban.

A tanulmányok két csoportba oszthatók, az első csoport fókuszja az ASRT paradigma gyakorlási fázisában megvalósuló szekvenciatanulással, a második csoport a gyakorlást követő, feladatvégzés nélküli eltelt időnek (*offline* időszaknak), valamint az ez alatti alvásnak a szerepét vizsgálja a szekvenciával kapcsolatos tudás konszolidációjában.

Az első két tanulmányban (2.1, 2.2) az ASRT paradigmát különböző korcsoportokban alkalmazták. Az eredményeket úgy értelmezték, hogy az implicit tanulási teljesítmény 12-13 évnél fiatalabb korban nagyobb mértékű, mint idősebb korban. A 2.2 tanulmányban azt is vizsgálták, hogy a rejtett sorozatszabályosságra vonatkozó utalás és annak felfedezésére buzdító instrukció milyen hatással van a teljesítményre, valamint a sorozatok egyes

elemeinek célszerű szétválogatása révén sikeresen különböztettek meg a sorozat egyes szabályos jellemzőinek leképeződésében szerepet játszó folyamatokat. A 2.3 tanulmányban meggyőzően demonstrálták, hogy hipnózis hatására megnő az ASRT paradigmában a szekvenciaspecifikus tanulást jellemző reakcióidő-különbség. A 2.4 tanulmányban alkohol-függő és normál csoportban vetették össze az ASRT paradigmában mutatott szekvenciatanulási teljesítményt. A csoportokat végrehajtó funkcióbeli teljesítmény szerint szétválasztva kapott eredményeik azt sugallják, hogy az ASRT teljesítmény a végrehajtó funkciók erősségével csökkenhet. A 2.5 tanulmányban a szerzők egy ötletes kísérleti elrendezést vezettek be abból a célból, hogy elkülönítsék a motoros események és az ingeresemények idői mintájának szerepét az ASRT paradigmában. Az eredmények szerint az effektusban mindkettőnek szerepe van. A paradigmát a 3.4 tanulmányban a szekvenciatanulás *offline* konszolidációjának vizsgálatában is használták, az eredmények szerint a motoros komponens a konszolidációban nagyobb szerepet játszik, mint a perceptuális. A 2.6 tanulmányban az ASRT paradigmát kettős feladat helyzetekbe helyezték. A feladatok közül a párhuzamos mondatmegértési feladat esetében találták a legkisebb szekvenciaspecifikus tanulási hatást. A 2.7 tanulmányban áttekintették a szekvenciatanulást munkamemória terhelés mellett vizsgáló szakirodalmi tanulmányokat, és azt találták, hogy míg a munkamemória-terhelés az explicit szekvenciatanulást, esetleg a készség tanulást befolyásolja, addig az implicit szekvenciatanulásra vélhetően nincs számottevő hatással. A 2.8 tanulmányban autizmussal élő gyerekek ASRT-ben mutatott teljesítményét hasonlították össze életkorban illetve IQ-ban illesztett csoporttal; a teljesítménymintázat a három csoportban nem tért el. A 2.9 tanulmányban idős, enyhe kognitív deficitben szenvedők csoportjának ASRT-ben mutatkozó teljesítményét hasonlították össze egészséges kontroll csoporttal, és azt találták, hogy a szekvenciatanulási teljesítmény az enyhe kognitív deficitben szenvedő csoportban kisebb volt. További elemzések azt mutatták, hogy a csoportközi különbséget a kísérleti blokkok elején lehetett megfigyelni, a blokkok második felére a különbség lecsökkent. A 2.10 tanulmányban obstruktív alvási apnoében szenvedő és egészséges kontroll csoport ASRT paradigmában mutatott szekvenciatanulási teljesítményét hasonlították össze, de nem találtak különbséget.

A 3.1 tanulmányban az ASRT paradigmában megmutatkozó szekvenciatanulási hatás *offline* időszaka alatti alvás konszolidációra gyakorolt hatását vizsgálták idős és fiatal felnőtt csoportban, de nem találtak a szekvenciatanulással kapcsolatos hatásokat. A 3.2 tanulmányban az alvással töltött *offline* szakasz konszolidációra gyakorolt hatását vizsgálták obstruktív alvási apnoében szenvedő és egészséges csoportban. Míg a készség tanulásban a konszolidációs hatás eltért, a szekvenciatanulás konszolidációjában nem mutatkozott különbség. A 3.3 tanulmányban az *offline* szakasz hosszának (fél nap – egy hét intervallum) konszolidációra gyakorolt hatását vizsgálták idős és fiatal felnőtt csoportban. Az idős csoportban a szekvenciatanulás konszolidációja kisebb mértékű volt, mint a fiataloknál, bár az *offline* szakasz hosszának nem volt kimutatható hatása. A 3.5 tanulmányban egyéves *offline* szakasz esetében meggyőzően demonstrálták az ASRT-paradigmában mérhető szekvenciaspecifikus tudás megtartását.

Átfogó kritikai megjegyzések

1.

A legtöbb pszichológiai kísérletben alapvető követelmény, hogy az összehasonlítandó feltételekben ugyanazt a mérőeszközt alkalmazzák. Az ASRT paradigmában mérhető szekvenciaspecifikus tanulást jellemző reakcióidő-különbség ennek a feltételnek véleményem szerint nem felel meg. A kísérletileg manipulált változók az ASRT paradigmában a konstans válasz-inger idő miatt nem függetlenek a mérőeszköztől, a paradigma ugyanis különböző helyzetekben, különböző csoportokban eltérő idői jellemzőkkel rendelkező eseménysorozatokat hoz létre. Azaz pl. a disszertációban használt 120 ms válasz-inger időkülönbség esetén egy 800 ms reakcióidejű (pl. idős felnőtt) személynek 920 ms inger- (vagy válasz-) követési idővel jellemezhető sorozatot, egy 400 ms reakcióidővel jellemezhető (pl. fiatal felnőtt) személy 520 ms inger- (vagy válasz-) követési idővel jellemezhető sorozatot mutatnak be. Mivel rendkívül valószínűtlennek tűnik, hogy az idői jellemzők eltérése ne befolyásolná valamilyen módon a szekvenciatanulást (a disszertációban - pl. 118-119. oldalon - a szerző is utal erre), ezért úgy gondolom, hogy a feltételek közötti összehasonlításokon alapuló következtetéseket - különösen eltérő reakcióidők esetén - jelentős fenntartással kell kezelni.

2.

A tanulmányokban visszatérő probléma, hogy hogyan lehet reakcióidő-különbségeket összehasonlítani olyan csoportok között, amelyek között a reakcióidő alapvetően eltér. Ez nem triviális probléma, több gondolatmenet, több modell is felállítható, amelyek különböző eredményekre is vezethetnek, itt a „kutatói szabadságfokok” száma meglehetősen nagy, az alkalmazott korrekció lehetséges hatásainak mérlegelése elengedhetetlen, mert a legtöbb esetben az ismert előjelű, de ismeretlen mértékű korrekciós hatások miatt csak bizonyos eredménymintázatok esetében lehet komolyan vehető következtetést levonni.

Pl. ha az egyik csoportban az alap reakcióidő 300 ms, ami a kezelés hatására 330 ms-ra nő, a másik csoportban pedig az 500 ms alap reakcióidő nő 550 ms-ra, akkor egy varianciaanalízis azt mutathatja, hogy a kezelés hatása a két csoportban eltér, hiszen az egyikben 30, a másikban 50 ms változást okoz. Érvelhetünk viszont úgy is, hogy a hatás nem additív, hanem arányos, így mindkét esetben +10% reakcióidő növekedést mérhetünk, ami alapján a kezelés hatása nem tér el a két csoportban. Ez a gondolatmenet pl. az időskori változások általános lassulási megközelítésében (ld. pl. Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103(3), 403–427.) jól értelmezhető, és legtöbbször a reakcióidők normalizálása (egy jól választott kontroll-feltételben mért értékhez való viszonyítás) révén vehető figyelembe. Ebben a példában erős állítás pl. akkor lenne tehető, ha a lassabb csoportban még normalizálás mellett is kisebb hatást találnánk (azaz p. 500 ms-ról 505 ms-ra nőne a kezelés hatására a reakcióidő, szemben a másik csoport 300-330 ms eredményével).

A bemutatott tanulmányokban nem került kialakításra egy olyan módszer, ami ezt a problémát konzisztensen kezeli, egyes esetekben alkalmaznak valamilyen korrekciót, máskor nem.

A 2.1 tanulmányban a szerzők z-transzformációt alkalmaztak. Ha jól értelmezem, személyenként „összeöntötték” a magas és alacsony frekvenciájú próbákra vonatkozó egyes reakcióidőket, és ezen személyi eloszlások mért átlagát és szórását alapul véve személyenként z-transzformálták az adatokat. Nem világos, hogy az így kapott adatokkal aztán mi történt, hogyan alakították ki a személyt jellemző két értéket. Ez az eljárás nehezen értelmezhető annak a fényében, hogy a disszertáció tanulmányaiban a résztvevőket a próbákban kapott reakcióidők mediánjaival jellemzik (amit egyébként helyes döntésnek tartok az egyéni reakcióidő-eloszlások ferdesége miatt). A 23. oldalon a szerzők is hasonlóan bírálják ezt a módszert, de nem kínálnak más megoldást, és végül a nyers reakcióidő adatok elemzésére alapozzák következtetéseiket, ami nyilvánvalóan a hatások eltúlzásával járhat a magasabb reakcióidejű csoportok esetében (ugyanaz igaz a 2.2. tanulmány esetén is). Érdekes lett volna itt pl. a z-transzformáció csoportszinten történő elvégzését, vagy a fent említett, arányokon alapuló logikát is megvizsgálni, és az egyes eljárások (beleértve a nyers reakcióidő adatok használatát is) torzító hatásait számba venni.

Feltételezem, hogy csoportszintű z-transzformációt alkalmaztak a 2.9 tanulmányban, de ez a szövegből nem derül ki. A 3.2 tanulmányban a reakcióidő adatokat az első epochban mért reakcióidővel normalizálják, az ebből a döntésből adódó hatások mérlegelése viszont hiányzik.

Érdekes kérdés, hogy mikor kell korrekciót alkalmazni. Amikor nem véletlenszerűen beosztott csoportok közötti összehasonlításról van szó, a csoportok közötti alapvető reakcióidő-különbségeket minden esetben szükséges figyelembe venni; a szignifikáns csoport-hatás hiányából nem vonható le következtetés a csoportok közötti egyenlőségre, így a helyes eljárás az, hogy minden esetben korrekciót alkalmazunk, ahogy az a 2.9 tanulmányban is történt.

3.

A szerzők értelmezése a legtöbb esetben az, hogy az ASRT paradigmában megfigyelhető gyakori és ritka tripletek utolsó elemére vonatkozó reakcióidő-különbség a gyakori triplet „gyorsító” hatását tükrözi. Ez egy attraktív értelmezés, aminek azonban nem látom a kísérleti alátámasztását. Ugyanennyire jogos értelmezés az is, hogy a reakcióidő-különbség a ritka tripletekre adott válasz lelassulásának a következménye. Pl. az 59. oldal 2.3.1 ábráján látható Triplet x Feltétel interakció oka ugyanúgy lehet az is, hogy a hipnózis feltételben az alacsony gyakoriságú tripleteknél a lassulás nagyobb mértékű (sőt, az ábrán – „szemre” az alacsony gyakoriságú triplet reakcióideje „lóg ki” a másik három reakcióidő közül - ami persze ebben a kísérleti tervben valójában lényegtelen). Az sem zárható ki, hogy a különbség két, ellenkező előjelű hatás eredője. Tömören szólva, egyértelmű, hogy a sorozatok mintázottsága révén a reakcióidők jellegzetesen megváltoznak, azt az értelmezést viszont, hogy ez „gyorsulás” eredménye, nem látom alátámasztottnak.

4.

A felvett nagyszámú kísérlet és a paradigma ugyanolyan paraméterekkel történő használata jó lehetőséget adott volna arra, hogy az egyes kísérleteknél előre megbecsüljék a szükséges mintanagyságot, és így elkerüljék az esetlegesen alacsony statisztikai erőből származó problémákat. Egy ilyen számítás arra is választ adhatna, hogy mekkora különbség kimutatására volt reális esély pl. a 2.8 tanulmányban.

5.

A statisztikai eljárásokkal kapcsolatos megjegyzések

A bemutatott kutatás alapvetően exploratív jellege miatt konzervatív statisztikai módszerek alkalmazására lett volna szükség, a kutatás során alkalmazott eljárások ezzel szemben számos ponton növelik a nullhipotézis téves elutasításának valószínűségét.

5.1

A tanulmányok túlnyomó részében az ANOVA számításokban olyan post-hoc elemzést (Fisher LSD) alkalmaztak, amelyről jól ismert, hogy a nullhipotézis téves elutasításának valószínűségét (egy speciális eset kivételével) az elfogadott szint fölé tolja (ld. pl. Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: a model comparison perspective* (2nd ed). Mahwah, N.J: Lawrence Erlbaum Associates., 229-230. o.). Ennek pl. a 2.1, 2.2. tanulmány következtetéseit tekintve konkrét jelentősége lehet.

5.2

A statisztikai elemzések során az általánosan elfogadott 5%-os alfa szintet a szerzők több esetben (pl. 2.2., 2.3, 2.9 tanulmány) figyelmen kívül hagyták. Különösen aggályos ez azokban az esetekben, ahol pl. egy 9.5%-os p-értékkel jellemzett többszörös interakciót (2.2 tanulmány) a fentebb említett Fisher LSD teszttel követik, amely alkalmazásának ráadásul feltétele, hogy a megfelelő ANOVA hatás szignifikáns legyen.

5.3

A szignifikanciaszintek összehasonlítása révén nem lehet érvényes statisztikai következtetést levonni (ld. pl. Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). *Erroneous analyses of interactions in neuroscience: a problem of significance*. *Nature Neuroscience*, 14(9), 1105–1107. <https://doi.org/10.1038/nn.2886>). A 2.6 tanulmányban (94. o.) a Feltétel x Triplet interakció feloldása a szignifikanciaszintek összehasonlításán alapszik, ami a levont következtetés alátámasztására alkalmatlan. A Feltétel x Triplet interakció feloldására a tripletek közötti reakcióidő-különbségek statisztikai összehasonlítása lenne alkalmas.

A fentiek fényében különösen aggályos a 2.3 kísérletben (60. o.) bemutatott számítás, amelyben egy nem szignifikáns ($p = 21\%$) interakciós hatás feloldásaként Fisher LSD páros

összehasonlításokban a szignifikanciaszintek összehasonlítása alapján vonnak le következtetést.

5.4.

A szerzők a módszertani részekben több helyen jelzik, hogy Greenhouse – Geisser korrekciót alkalmaztak, ami alapvető módszertani követelmény, mert bizonyos ANOVA struktúrák esetén korrekció nélkül a nullhipotézis téves elutasításának valószínűsége növekedne (megjegyzem, hogy a korrekciót nem érdemes Mauchly-teszt szignifikanciájára alapozni, ld. pl. Maxwell, és Delaney. 2004, 542.o., Keselman, H. J., Rogan, J. C., Mendoza, J. L., & Breen, L. J. (1980). Testing the validity conditions of repeated measures F tests. Psychological Bulletin, 87(3), 479.). Egyes tanulmányokban azonban nem teljesen világos, hogy történt-e korrekció vagy sem. Van, ahol a módszertannál jelzik, de nincs olyan ANOVA ahol alkalmazható lenne (2.1 tanulmány); van, ahol nem jelzik, de úgy tűnik, hogy alkalmazták (a 2.3. tanulmányban a blokk főhatás alapján - 58. oldal alja), és több helyen elképzelhető, hogy nem történt korrekció (a 2.2 tanulmányban, a 43. oldal tetején leírt hármas interakció alapján; a 2.6 tanulmányban a 94. oldalon bemutatott számítás alapján; a 2.8. tanulmányban a 115. oldal középső szakaszában közölt Epoch főhatás számítás alapján).

5.5

A 2.3 tanulmányban a szabadsági fokok alapján úgy tűnik, hogy a 14 fős kísérleti csoport végrehajtói funkciót jellemző értékeinek mediánja mentén történő kettéosztása két hatfős csoportot eredményezett, azaz két főt - gondolom a rangsorban a 7-8.-at - kihagyták az elemzésből. Ezt érdemes lett volna a szövegben is megemlíteni.

A 2.4 tanulmánynál az EPOCH szempontot tartalmazó hatásoknál a szabadsági fokok tévesek, (4,25) szerepel (4,112) helyett (71. oldal).

A 2.5 tanulmánynál az ANOVA-ban (82. o.) a szabadsági fokok tévesek. $F(1,23)$ helyett $F(1,32)$; illetve $F(4,20)$ helyett $F(3,96)$ lenne helyes.

A 2.6 tanulmánynál az ANOVA-ban a szabadsági fok (94. o.) téves: $F(2,24)$ helyett $F(2,50)$ lenne helyes. A feltüntetett p-érték egyébként a téves $F(2,24)$ szabadsági fokot tükrözi.

6.

A legtöbb tanulmányban a résztvevők tökéleteshez közeli pontosságot értek el, ami miatt a pontosságadatokat sokszor nem is elemzik. Ez azt mutatja, hogy a résztvevők a „gyorsan, de pontosan” instrukcióból alapvetően a pontosságot érezték fontosabbnak, ami a reakcióidő, mint mérőszám hasznosságát csökkentheti (mert a résztvevő nem a döntési folyamat végén, hanem saját szempontjai szerint, kissé kivárva adja meg a választ). Az állandó válasz-inger idői távolságot a résztvevők a paradigma sebességének kényelmi optimalizálására használhatják. Ezért érdemes lenne megfontolni a gyorsaság hangsúlyozását, különösen azért, mert a téves találatokból adódó reakcióidő-torzítás egy négyalternatívás diszkriminációs feladatban viszonylag alacsony.

Apróbb megjegyzések

7.

A 2.2 tanulmányban az ingerek eltértek, ezért a kétfajta ingerre a reakcióidők különbözhetnek, ezek szerepét célszerű lett volna a csoportokon belül és a csoportok között is kiegyenlíteni. Érdekes lenne azt is megfontolni, hogy az explicit feltétel az implicittel szemben egy kettős feladat helyzet.

8.

A 2.8.1. táblázatban mit jelentenek az „ASRT learning” oszlopban feltüntetett számok?

9.

Azokat a színes ábrákat, ahol a színnek megkülönböztető szerepe van, érdemes lett volna színesben nyomtatni (pl. 34. o). A szövegben helyenként a félkövér jelölések elvesztek (37. o). Azokban a tanulmányokban, amelyekben a szerzők célja az életkori különbségek feltérképezése volt, érdemes lett volna az ábrákon az életkorokat (pl. a csoportátlag alapján) intervallumskálán ábrázolni.

10.

Az összefoglaló táblázatokból kimaradt a 2.10 tanulmány, valamint a 3.2 és 3.3 tanulmányok felcserélődtek.

Összefoglalásként elmondható, hogy Németh Dezső MTA doktori értekezésben bemutatott tudományos tevékenysége, az implicit statisztikus tanulás ASRT-paradigmában megnyilvánuló hatásával kapcsolatos eredményei az MTA doktora cím megszerzéséhez elegendőek, javasolom a nyilvános védés kitűzését.

Dr. Horváth János

Budapest, 2017-08-07