

Doktori (PhD) értekezés

A ROT-modell: a rejtett oksági tényezők szerepének modellezése a kovariancia-alapú  
okszági tanulás során

Kurucz Gyózzó

Debreceni Egyetem  
BTK  
2021

**A ROT-modell: a rejtett oksági tényezők szerepének modellezése a kovariancia-  
alapú oksági tanulás során**

Értekezés a doktori (Ph.D.) fokozat megszerzése érdekében  
a pszichológia tudományágban

Írta: Kurucz Győző okleveles pszichológus

Készült a Debreceni Egyetem Humán Tudományok doktori iskolája  
(Pszichológia programja) keretében

Témavezető: Dr. ....

(olvasható aláírás)

A doktori szigorlat bizottság

elnök: Dr. ....

tagok: Dr. ....

Dr. ....

A doktori szigorlat időpontja: 20 .....

Az értekezés bírálói:

Dr. ....

Dr. ....

Dr. ....

A bírálóbizottság:

elnök: Dr. ....

tagok: Dr. ....

Dr. ....

Dr. ....

Dr. ....

A nyilvános vita időpontja: 20 .....

## Nyilatkozat

„Én, Kurucz Győző teljes felelősségem tudatában kijelentem, hogy a benyújtott értekezés önálló munka, a szerzői jogi nemzetközi normáinak tiszteletben tartásával készült, a benne található irodalmi hivatkozások egyértelműek és teljeseek. Nem állok doktori fokozat visszavonására irányuló eljárás alatt, illetve 5 éven belül nem vontak vissza tőlem odaítélt doktori fokozatot. Jelen értekezést korábban más intézményben nem nyújtottam be és azt nem utasították el.”

.....  
Kurucz Győző

## **Köszönetnyilvánítás**

Szeretném köszönetemet kifejezni a családomnak, kifejezetten feleségemnek és gyermekeimnek, a kitartó türelmükért, témavezetőmnek az elmúlt évek során nyújtott szakmai támogatásáért és tanácsaiért. Ugyancsak hálás vagyok kollégáimnak, akik akár tudtukon kívül is hozzájárultak azon szakmai hitvallás és hivatástudat kialakulásához, amely – reményeim szerint – a dolgozatban is megjelenik.

# Tartalomjegyzék

Ábrák jegyzéke.....	6
Táblázatok jegyzéke.....	8
Bevezetés.....	9
Matematikai jelölések magyarázata.....	13
<b>1. Szakirodalmi áttekintés.....</b>	<b>15</b>
1.1. Az oksági indukció kutatásának áttekintése.....	15
1.1.1. Az intervenció szerepe az oksági tanulásban.....	18
1.1.2. Események időbeli eloszlásának szerepe az oksági tanulásban.....	20
1.1.3. Előzetes ismeretek szerepe az oksági tanulásban.....	21
1.1.4. Megközelítések és magyarázatok.....	23
1.2. A kognitív modellezés Bayes-i irányzata.....	25
1.3. Az oksági Bayes-hálók formalizmus áttekintése.....	28
<b>2. A ROT-modell bemutatása.....</b>	<b>33</b>
2.1. A ROT-modell központi alapfeltevései.....	35
2.2. A mérlegelt hipotézisek formális leírása.....	38
2.3. Az oksági kapcsolat meglétére vonatkozó becslés lépéseinek részletezése.....	42
2.3.1. Magyarázatok generálása a megfigyelt esetekre.....	42
2.3.2. A rejtett oksági tényezők jelenlétére vonatkozó bizonytalanság felülvizsgálata.....	47
2.3.3. Az oksági kapcsolatra vonatkozó hipotézisek felülvizsgálata (az oksági support kiszámítása).....	48
<b>3. A ROT-modell illeszkedésének vizsgálata.....</b>	<b>55</b>
3.1. Alternatív modellek.....	56
3.2. A modellek illeszkedésének vizsgálata összegzett kovariancia-információ alapján való tanulás esetén.....	58
3.2.1. A vizsgált modellek predikciói.....	58
3.2.2. Mások vizsgálatai.....	61
3.2.3. Első kísérlet: $p(K=1 E=1)$ és $p(K=1 E=0)$ hatásának vizsgálata, kis virtuális mintanagyság esetén.....	70
3.2.4. Második kísérlet: $p(K=1 E=0)$ és virtuális mintanagyság ( N ) hatásának vizsgálata.....	74

3.2.5. Harmadik kísérlet: $p(K=1 E=1)$ és $p(K=1 E=0)$ hatásának vizsgálata nagy virtuális mintanagyság esetén.....	79
3.3. Illeszkedés vizsgálata online tanulás esetében.....	87
3.3.1. Negyedik kísérlet: valószínűségi kontraszt ( $\Delta P$ ) hatásának vizsgálata.....	87
<b>4. Diskusszió.....</b>	<b>95</b>
<b>Felhasznált irodalom.....</b>	<b>100</b>
<b>Mellékletek.....</b>	<b>113</b>
A. A ROT-modell forrása.....	114
B. SS-modell forrása.....	117
C. Support-modell forrása.....	118

## Ábrák jegyzéke

1. ábra Egy egyszerű, két változó együttes gyakorisági eloszlását ábrázoló kontingencia tábla.....	14
2. ábra Két esemény (előzmény és következmény) együttes gyakorisági eloszlását ábrázoló kontingencia tábla.....	15
3. ábra Oksági Bayes-háló, egy három változóból álló oksági rendszer ábrázolására.....	29
4. ábra Egy egyszerű oksági Bayes-háló, egy diák vizsgán nyújtott teljesítményének modellezésére.....	31
5. ábra A ROT-modell által vizsgált, az előzmény és következmény közötti oksági kapcsolat meglétére, illetve hiányára vonatkozó hipotézisek.....	38
6. ábra Példa a béta-eloszlás alakjának (sűrűségfüggvényének) változására a két paraméterének ( $\alpha$ és $\beta$ ) függvényében.....	40
7. ábra A support-modell által vizsgált, az előzmény és következmény közötti oksági kapcsolat meglétére, illetve hiányára vonatkozó hipotézisek.....	57
8. ábra Az oksági súlyok ( $\omega_0, \omega_1$ ) bizonytalanságát ábrázoló együttes valószínűségi eloszlás alakulása az paraméter különböző értékei mellett.....	58
9. ábra A vizsgált modellek predikcióinak alakulása (függőleges tengely) a következmény előzmény jelenlétében ( $p(K=1 E=1)$ ), illetve hiányában ( $p(K=1 E=0)$ ) tapasztalt valószínűségének, illetve a virtuális mintanagyságnak ( $N$ ) a különböző szintjei mellett.....	60
10. ábra Lu és mtsai. (2008) 3. kísérletének eredményei (átlagos értékelések) a generatív feltételben, és a vizsgált modellek lineáris regresszióanalízis segítségével előkészített becslései (kör: ROT-modell, négyzet: SS-modell, háromszög: support-modell). A hibasávok a sztenderd hiba alapján számolt 95%-os konfidencia-intervallumok.....	65
11. ábra Az első kísérlet egyes feltételeiben adott ítéletek alakulása (az oszlopok az ítéletek mediánját, a hibasávok az átlagot, és az átlag körüli $\pm 1$ sztenderd hibát jelölik; a négyzet az SS-modell, a kör a ROT-modell, a háromszög a support-modell lineáris regresszióanalízis segítségével előkészített becslése).....	73
12. ábra A második kísérlet egyes feltételeiben adott ítéletek alakulása (az oszlopok az ítéletek mediánját, a hibasávok az átlagot, és az átlag körüli $\pm 1$ sztenderd hibát jelölik; a négyzet az SS-modell, a kör a ROT-modell, a háromszög a support-modell lineáris regresszióanalízis segítségével előkészített becslése).....	77
13. ábra A harmadik kísérlet egyes feltételeiben adott ítéletek alakulása (az oszlopok az ítéletek mediánját, a hibasávok az átlagot, és az átlag körüli $\pm 1$ sztenderd hibát jelölik; a négyzet az SS-modell, a kör a ROT-modell, a háromszög a support-modell lineáris regresszióanalízis segítségével előkészített becslése).....	82

14. ábra A rejtett oksági tényezők jelenlétének valószínűségére vonatkozó ítéletek alakulása a harmadik kísérletben (az oszlopok az ítéletek mediánját, a hibasávok az átlagot, és az átlag körüli $\pm 1$ sztenderd hibát jelölik).....	84
15. ábra A negyedik kísérletben alkalmazott, az egyes eseménytípusokat ábrázoló „kártyák” .....	89
16. ábra A negyedik kísérlet egyes feltételeiben adott értékelések átlagai az egyes kísérleti feltételekben (a hibasávok az átlag körüli $\pm 1$ sztenderd hiba intervallumot jelölik).....	91
17. ábra A ROT-modell két változata által becsült tanulási görbék a negyedik kísérlet egyes feltételeire (háromszög: , kör: , négyzet: ).....	92
18. ábra A negyedik kísérlet résztvevőinek válaszai, és a számukra egyedileg készített, becsült tanulási görbék közötti kereszt-korrelációk értékeinek eloszlása a ROT-modell két változatánál.....	93

## Táblázatok jegyzéke

1. táblázat A következmény feltételes valószínűségi eloszlásának alakulása a vizsgált hipotézisek mellett.....	41
2. táblázat A kovariancia-információ egyes eseménytípusaira adott lehetséges magyarázatok.....	43
3. táblázat Néhány példa a Bayes-faktor értelmezésének szemléltetésére.....	49
4. táblázat A kovariancia-információk résztvevők általi rangsorolása Griffiths és Tenenbaum (2005) 2. és 3. kísérletében, valamint a vizsgált modellek predikciói a kísérletekben alkalmazott kovariancia-információkra.....	64
5. táblázat Saito és Shimazaki (2013) kísérletében alkalmazott kovariancia-információk, a résztvevők által adott leggyakoribb válaszok, valamint a vizsgált modellek által adott becslések.....	68
6. táblázat A résztvevők által adott válaszok átlaga, valamint a vizsgált modellek becslései az első kísérlet egyes feltételeiben.....	74
7. táblázat A résztvevők által adott válaszok átlaga, valamint a vizsgált modellek becslései a második kísérlet egyes feltételeiben.....	78
8. táblázat A résztvevők által adott válaszok átlaga, valamint a vizsgált modellek becslései a harmadik kísérlet egyes feltételeiben.....	83
9. táblázat A generatív-, és preventív oksági tényezők jelenlétének valószínűségére vonatkozó ítéletek, és a ROT-modell ez irányú becslései a harmadik kísérlet egyes feltételeiben.....	85

## Bevezetés

Világunkban az események folyását az azok között működő oksági kapcsolatok bonyolult hálózata határozza meg. Egy ilyen világban rendkívül hasznos képesség ezen oksági kapcsolatok megismerésének, valamint ezen ismereteknek a viselkedés megtervezésében való felhasználása. Az események folyását meghatározó oksági összefüggések ismerete lehetővé teszi hogy bejósoljunk bizonyos eseményeket, hogy elvárásokat fogalmazzunk meg a viselkedésünk következményeit illetően, vagy - a viselkedéstervezés szempontjából hasznos - oksági magyarázatokat találjunk bizonyos jelenségekre. Sloman (2005) az oksági gondolkodást, azaz az oksági kapcsolatokra vonatkozó ismeretek megszerzésének és alkalmazásának képességét ezért kognitív rendszerünk egyik központi képességének tekinti.

Az oksági kapcsolatokról való gondolkodás számos alkalmazott területen központi jelentőségű, de legalábbis nagyon fontos szerepet tölt be. Így például az időjárás-előrejelzésben, a gazdasági folyamatokkal kapcsolatos előrejelzésekben (Einhorn és Hogarth, 1985), vagy a betegségek kialakulásának és terjedésének megértésében (Galea és mtsai., 2010) szükséges a jelenségek háttérében álló oksági összefüggésrendszerek megértése, és azok alkalmazása a jelenségről való gondolkodás során – a komplex rendszerek működésének megértésében az oksági gondolkodás szerepe elhanyagolhatatlan (Kurucz, 2014). Az események oksági jellegének felismerése ugyancsak segíthet a múltbeli események alaposabb megértésében, és ezáltal egy – a történelmi összefüggéseket is magyarázni képes – történelem felvázolásában (Kurucz, 2010a).

Érdekes, hogy bár a filozófiában az oksági kapcsolatokra vonatkozó kérdések – mind azok metafizikáját (áttekintésért ld. Schaffer, 2016), mind episztemiológiáját (pl. Eberhardt, 2009; Williamson, 2004, Hume és Kant ez irányú munkásságának áttekintéséért ld. De Pierris és Friedman, 2013) tekintve – régóta széles körű érdeklődésnek örvendenek, azok megismerésére és alkalmazására való képességünk a pszichológia tudományának csak igen korlátozott érdeklődését vonzotta egészen a XX. század végéig. A XX. század első felében Jean Piaget-t említhetjük, aki elsősorban a fizikai szinten is megragadható oksági jelenségek (például a felhők és égitestek mozgása, a szél keletkezése, a bicikli és gőzgép működése stb.) gyermekek általi értelmezését vizsgálta, a kognitív fejlődés jellegének és folyamatának alaposabb feltárása érdekében (Piaget, 1930). A szociálpszichológia elsősorban a társas

kapcsolatok keretén belül foglalkozott az oksági attribúcióval – azzal a problémakörrel, hogy milyen módon azonosítjuk azokat az ágenseket, akik felelőssé tehetőek bizonyos következményekért (Hewstone, 1989; Hilton és Slugoski, 1986; Kelley, 1973). Megemlíthetjük még továbbá Albert Michotte-ot, az oksági gondolkodás egyik korai kutatóját, aki egy jellegzetes perceptuális jelenség – egy tárgy mozgásának másik tárgyra való „áttérjedésének” – mélyre hatoló vizsgálatával (Michotte, 1963) indította meg az oksági kapcsolatok alapvető észlelési folyamatokban betöltött szerepének széles körű kutatását.

A XX. század végére a kognitív pszichológia és általában a pszichológia érdeklődése is felélénkült az oksági gondolkodás iránt, azonban ez az érdeklődés sokkal általánosabb érvényű volt, mint a korábbi megközelítések, nem korlátozódott a perceptuális jelenségekre, vagy a szociális- illetve fizikai területekre, a középpontjában pedig az oksági kapcsolatokra vonatkozó ismeretek megszerzése, az oksági tanulás állt. Az oksági tanulás vizsgálatában újdonságot jelentett a korábbiakhoz képest, különböző modern eszközök és eljárások alkalmazása a matematika és számítástudomány területéről, amelyek egyben egzakt nyelvezetet is biztosítottak az oksági kapcsolatokról való gondolkodás számára. Ugyancsak jellemző volt az oksági kapcsolatok megismerésére vonatkozó, korábban domináns Hume-i filozófia háttérbe szorulása és a Kanti elképzelések előtérbe kerülése (ami megnyilvánult például abban, hogy a top-down folyamatokra nagyobb hangsúly helyeződött), valamint modern statisztikai eszközök – egyebek között a Bayes-i statisztika – alkalmazása. A három terület (számítástudomány, filozófia, statisztika) integrációja az oksági tanulás – és tágabban véve az oksági gondolkodás – vizsgálatának egy, a korábbiaknál sokkal szélesebb körű és hatékonyabb vizsgálatát tette lehetővé (Glymour, 2003; Holyoak és Cheng, 2011).

Egy újabb áttekintés szerint (Danks, 2009) az oksági gondolkodás kutatásának két nagyobb területét különböztethetjük meg, amelyek között meglehetősen hiányos az integráció. A kutatások egyik nagy csoportja az oksági ismeretek felhasználását helyezi a középpontba, ilyen az oksági (prediktív és diagnosztikus) következtetések (Cummins, 1995; Kurucz, 2010b; Sloman és Hagmayer, 2006), vagy az oksági ismereteken alapuló döntések (Joyce, 1999) vizsgálata. A kutatások másik csoportja az oksági ismeretek megszerzésével foglalkozik, és magában foglalja az oksági kapcsolatok perceptuális szintű tanulását (pl. White, 1995), valamint az oksági indukciót, amely nem közvetlen percepción, hanem inkább következtetésen alapuló tanulás, amely során a rendelkezésünkre álló adatok alapján

következtetünk az oksági kapcsolat meglétére, vagy hiányára. Jelenleg úgy tűnik, hogy a tanulás ezen két formája megkülönböztethető viselkedéses szinten, és eltérő idegrendszeri alapokkal is rendelkeznek (ld. például Roser, Fugelsang, Dunbar, Corballis és Gazzaniga, 2005; Schlottmann és Shanks, 1992).

A dolgozat témája az oksági kapcsolatok megismerése, az oksági tanulás, ezen belül is az oksági indukció. A téma modern megközelítésének az egyik központi eleme, hogy az oksági tanulást a kognitív rendszerünk által megoldandó induktív problémának tekinti. Eszerint az oksági tanulás jellegzetessége, hogy a tanulás tárgyai – az oksági kapcsolatok – közvetlenül nem figyelhetők meg, csupán közvetett bizonyítékok alapján tudunk következtetni azokra, amelyek azonban önmagukban nem oksági jellegűek. Ebben a felfogásban a legfontosabb kérdés hogy milyen ismeretekkel és milyen kognitív folyamatokkal, valamint ezen folyamatokba kódolt feltevésekkel kell rendelkeznie egy intelligens ágensnek ahhoz, hogy megismerhesse ezeket az oksági kapcsolatokat, és ezen ismeretek alapján hatékonyan manipulálhassa az események folyását (Holyoak és Cheng, 2011). A dolgozatban bemutatjuk az oksági kapcsolatok kovariancia-információn, azaz az események együttjárására vonatkozó információ alapuló tanulásának egy Bayes-i modelljét, a ROT-modellt. A modell központi szerepet szán a tanulási helyzetben nem megfigyelhető – rejtett – oksági tényezőknek a tanulás folyamatában; a modell rövid elnevezése is innen származik (ROT = Rejtett Oksági Tényezők). Saját kísérleteink, valamint mások eredményei alapján arra törekszünk, hogy megvizsgáljuk a ROT-modell illeszkedését a kovariancia-alapú oksági tanulás humán adataira, és ezen keresztül megítéljük alkalmasságát arra, hogy képet kaphassunk az oksági magyarázatok és rejtett oksági tényezők oksági tanulásban betöltött szerepéről. A ROT-modellt számítógépes program formájában is implementáltuk, ami lehetővé teszi egzakt és egyértelmű predikciók levezetését – ilyen módon a modell predikcióinak feltárása is hatékonyabb lehet, valamint a modell illeszkedése kvantitatív módon is megítélhető (Gupta, 2008; Sun, 2008, 2009).

A dolgozat felépítése a következő. Elsőként áttekintjük az oksági tanulás kutatásának legjellemzőbb kérdéseit és megközelítéseit a kognitív pszichológián belül, a kognitív modellezés Bayes-i irányzatának legmarkánsabb jellemzőit, valamint az oksági Bayes-hálók formalizmust, amely utóbbi fontos alapját képezi a ROT-modell működésének. Ezután ismertetjük a modell részletes formális leírását, amely során bemutatjuk az annak

középpontjában lévő alapfeltevéseket, valamint olyan részletességgel leírjuk a modell működését, ami lehetővé teszi annak számítógépes implementációját is. Végül megvizsgáljuk a ROT-modell becsléseit olyan jellegzetes feltételek mellett, amelyeket gyakran alkalmaznak az oksági indukció kutatása során, majd ellenőrizzük a modell becsléseinek illeszkedését empirikus adatokra. Az empirikus adatok egyik része saját kísérleteinkből, másik része pedig egyéb szakirodalmi forrásokból származik. Hogy pontosabb képet kapjunk a modell megfelelőségéről, összevetjük annak illeszkedését egyéb, a szakirodalomból ismert modellek illeszkedésével is. A diszkusszióban mérlegeljük a ROT-modell alkalmasságát az emberi oksági tanulás magyarázatára, valamint felvetjük azokat a lehetőségeket, amelyek a modell továbbfejlesztése érdekében megfontolandók lehetnek.

## Matematikai jelölések magyarázata

A dolgozatban, elsősorban a ROT-modell ismertetése során, törekszünk a modell matematikai részletekbe menő leírására, annak érdekében, hogy annak működése könnyen megérthető, és maga a modell könnyen implementálható legyen számítógépes formában. Itt röviden tisztázzuk azt, hogy milyen jelöléseket alkalmaztunk, amelyek esetleg eltérhetnek a megszokottól.

A dolgozatban a változókat – ide értve a különböző eseményekre utaló változókat is – dőlt római nagybetűvel jelöltük (pl.  $X$ ). Az eseményekre utaló változók esetében gyakori, hogy a változó értékét is feltüntettük, amely utalhat arra, hogy az esemény bekövetkezett (pl.  $X=1$ ), vagy arra, hogy nem következett be (pl.  $X=0$ ). A különböző paramétereket, mint például a valószínűségi eloszlások paramétereit görög kisbetűkkel jelöltük (pl.  $\alpha, \delta$ ). Bizonyos változók, illetve paraméterek megkülönböztetésére alsó-, illetve felső indexeket használtunk (pl.  $X^+, X^-, \chi_a, \chi_b$ ).

Az egyes egyszerű-, vagy összetett események valószínűségét  $p()$ -vel jelöltük. Amikor egy konkrét esemény valószínűségét szeretnénk volna ismertetni, akkor a zárójelen belül az eseményt jelző változót annak értékével együtt tüntettük fel (pl.  $p(X=1)$ ). Ha egy változó konkrét értékének a valószínűsége helyett a változó valószínűségi eloszlását szeretnénk volna megadni, akkor a zárójelben egyszerűen a változót jelölő szimbólumot tüntettük fel (pl.  $p(X)$ ). Több változó együttes valószínűségi eloszlását  $p(X, Y)$  módon adtuk meg. Egy változó függvényyszerűen megadható valószínűségi eloszlását a szokásos módon adtuk meg, feltüntetve az eloszlás típusát, és paramétereit is (pl.  $p(X) \sim Normal(\mu, \sigma)$ , azaz az  $X$  változó normál eloszlást követ  $\mu$  átlaggal és  $\sigma$  szórással). Egy egyszerű esemény feltételes valószínűségét  $p(X=1|Y=1)$  módon, a feltételes valószínűségi eloszlását pedig  $p(X|Y=1)$  módon jelöltük.

		Y	
		1	0
X	1	$N_a$	$N_b$
	0	$N_c$	$N_d$

1. ábra

Egy egyszerű, két változó együttes gyakorisági eloszlását ábrázoló kontingencia tábla

A dolgozatban többször is közöltünk kovariancia-információt, mint két változó együttes gyakorisági eloszlását, amelyet hagyományosan kontingenciatábla segítségével ábrázolhatunk (ld. 1. ábra). Folyószövegben, illetve táblázatokban, ábrákon a kovariancia-információt halmazelméleti jelölést (pl.  $\{N_a, N_b, N_c, N_d\}$ ) használva, vagy mátrixként (1) tüntettük fel.

$$\begin{matrix} N_a & N_b \\ N_c & N_d \end{matrix} \quad (1)$$

A nagy görög  $N$  (nú) a virtuális mintanagyságot – azaz a megfigyelt adatok összességét – jelöli,  $N_a, N_b, N_c, N_d$  pedig az egyes cellagyakoriságokat. Magát a kovariancia-információt, mint a megfigyelt adatok összességét  $\mathcal{D}$  -vel (kalligrafikus D) jelöltük.

Hipotézisek jelölésére a  $\mathcal{H}$  -t (kalligrafikus H) használtuk, a különböző hipotéziseket alsó indexszel különböztettük meg egymástól (pl.  $\mathcal{H}_0, \mathcal{H}_1$ ).

Több matematikai művelet jelét egyszerűsített formában alkalmaztuk. A halmozott összeg, illetve halmozott szorzat műveleteket például kifejezetten arra a célra használtuk, hogy bizonyos változók értékei mentén összegezzük, illetve összeszorozzuk az operandusokat – ezekben az esetekben csak a változó nevét tüntettük fel a műveleti jel alatt.

A halmozott összeget például így jelöltük:  $\sum_X X$ , amely tehát  $X$  változó minden egyes értéke mentén való összegzést jelent. A halmozott szorzat esetében hasonlóan jártunk el, tehát  $\prod_X X$  az  $X$  változó minden értéke mentén való szorzást jelent.

Integrálszámításnál a  $\int_i^j f(X)d(X)$  jelölést használtuk, amely  $f(X)$  függvény alatti területet jelenti  $[i, j]$  zárt tartományon belül.

# 1. Szakirodalmi áttekintés

## 1.1. Az oksági indukció kutatásának áttekintése

Az események közötti oksági kapcsolatok kovariancia-információ alapján való tanulásának vizsgálata során arra törekszünk, hogy feltárjuk, hogyan vagyunk képesek az események együttjárásának megfigyelése révén az oksági kapcsolatra vonatkozó ismeretekre szert tenni. A kérdés vizsgálata során feltételezzük, hogy a rendelkezésre álló információ szerkezete jellegzetes: egy esemény (következmény) jelenlétének, illetve hiányának gyakorisága egy másik esemény (előzmény) jelenléte, illetve hiánya esetén. Az ilyen jellegű információt hatékonyan ábrázolhatjuk egy kontingenciatáblában (2. ábra).

		<i>következmény</i>	
		<i>jelen van</i>	<i>nincs jelen</i>
<i>előzmény</i>	<i>jelen van</i>	a	b
	<i>nincs jelen</i>	c	d

2. ábra

Két esemény (előzmény és következmény) együttes gyakorisági eloszlását ábrázoló kontingencia tábla

A kovariancia-információ ennek megfelelően négy részből tevődik össze, amelyekre, mint eseménytípusokra hivatkozunk. Az a-típusú esemény akkor következik be, ha mind az előzmény, mind a következmény jelen van, a b-típusú esemény akkor, ha az előzmény jelen van, de a következmény elmarad. A c-típusú esemény akkor következik be, ha az előzmény nincs jelen, de a következmény igen, a d-típusú esemény pedig akkor, ha mind az előzmény, mind a következmény elmarad.

Az oksági indukció kutatása jellemzően kísérletes módszerekre támaszkodik, amelyen belül több jellegzetes vizsgálati paradigmával is találkozhatunk. Közös ezekben a paradigmákban, hogy az állati- és emberi asszociatív tanulás vizsgálati paradigmáinak számos jellemzőjét megőrizték (Glymour, 2003). A kutatások során alkalmazott különböző eljárásokat itt a kovariancia-információ bemutatásának módja, valamint a tanulás eredményének, azaz az oksági ismeretek mérésének a módja mentén tekintjük át.

A kovariancia-információ prezentációjának módját illetően találkozunk olyan eljárással, amely során a kísérleti személy a kovariancia-információt összegzett formában látja, az egyes eseménytípusok gyakorisága formájában (pl. Griffiths és Tenenbaum, 2005; Lu, Yuille, Liljeholm, Cheng és Holyoak, 2008), vagy egy listaként (például Waldmann és Hagmayer, 2005). Ezek előnye lehet, hogy kiküszöbölik az emlékezeti folyamatok hatását, ugyanakkor némileg életidegen vizsgálati helyzetet teremtenek, a hétköznapi világban ugyanis ritkán állnak rendelkezésünkre ilyen formában összegzett adatok, sokkal jellemzőbb, hogy az események együttjárására vonatkozó információt a körülöttünk zajló események folyamából kell megszereznünk (Buehner, 2005). A valós élettel nagyobb összhangban van a kovariancia-információ online prezentációja, amikor az egyes eseménytípusokat egyesével, egymás után mutatják be (pl. Waldmann, 2001; Waldmann és Holyoak, 1992; Sobel és Kushnir, 2006), amely egyébként bizonyos esetekben – mint például a beavatkozások oksági tanulásra gyakorolt hatásának vizsgálata esetén – kizárólagos.

A tanulás során elsajátított ismeretek, hiedelmek mérésének tipikus módja azok közvetlen mérése egyszerű ítéleteken keresztül, amelyek az oksági kapcsolat erősségére (pl. Sobel és Kushnir, 2006; Waldmann és Hagmayer, 2005), ritkább esetben az oksági kapcsolat pusztán meglétére, azaz az oksági struktúrára (pl. Griffiths és Tenenbaum, 2005; Lu és Mtsai, 2008; Saito és Shimazaki, 2013) vonatkoznak. Bizonyos esetekben azzal is találkozunk, hogy a kísérleti személyeknek a megfigyelt kovariancia-információ alapján több, az érintett változók közötti kapcsolatrendszerre vonatkozó hipotézis közül kell választaniuk (pl. Lagnado és Sloman, 2002; Steyvers, Tenenbaum, Wagenmakers és Blum, 2003). Bár találkozhatunk vele, a kutatási gyakorlatban kevésbé elterjedt a megszerzett ismeretek viselkedéses mérése (pl. Waismeyer, Meltzoff és Gopnik, 2015; Young, 2014).

Nehezíti a kovariancia-alapú tanulás univerzális magyarázatának megtalálását, hogy a kovariancia-információ prezentálásának, illetve az oksági ismeretek mérésének a módja befolyásolja a tanulás eredményét, valamint, ha normatív szempontból tekintjük, akkor a tanulás hatékonyságát is (Collins és Shanks, 2006; Perales és Shanks, 2008; Shou és Smithson, 2015b; Vallée-Tourangeau, Payton és Murphy, 2008; White, 2003). Ennek egyik következménye, hogy a számos magyarázat, elképzelés között nincs olyan, ami az eddig összegyűlt tapasztalatok nagyon széles körét képes lenne magyarázni (Perales és Shanks, 2007).

Vallée-Tourangeau és mtsai. (2008) egyik kísérletében például a kísérleti személyek három csoportjának oksági ítéleteket kellett alkotnia vírusok és betegségek közötti kapcsolatokról, a rendelkezésükre bocsájtott kovariancia-információ alapján. A kísérleti csoportok különböztek a kovariancia-információ prezentálásának a módjában: az első csoport szöveges állításokba ágyazva ismerte meg az egyes eseménytípusok gyakoriságát, a második csoport számára kontingenciatáblákban mutatták be az információkat, ahol a gyakoriságokat nem számokkal, hanem megfelelő mennyiségű, az események típusára egyértelműen utaló képecskével ábrázolták, a harmadik csoport pedig a kovariancia-információ hierarchikus elrendezésével találkozott, ahol a legfelső szinten a következmény gyakorisági eloszlását, a következő szinten pedig az előzmény gyakorisági eloszlását ábrázolták. Jelentős különbségeket találtak a csoportok között az adott oksági ítéletek tekintetében – az első csoportban (egyszerű szöveges prezentáció) ítélték a leggyengébbnek az oksági kapcsolatot, a harmadik csoportban (hierarchikus prezentáció) pedig a legerősebbnek.

Shou és Smithson (2015b) több kísérletet mutat be az oksági ismeretek mérésére szolgáló kérdések hatásával kapcsolatban, amelyekben azt találták, hogy a válaszok eloszlását befolyásolja az, hogy a kérdés az oksági kapcsolat erősségére, az oksági kapcsolat meglétére, vagy jövőbeli események bejósolására vonatkozik. Azt is találták továbbá, hogy a becslések egyrészt kevésbé magabiztosak, másrészt változatosabbak voltak amikor az előzmény ismeretében kellett a következményre becslést tenni (prediktív következtetés), mint amikor a következmény ismeretében kellett az előzményre vonatkozó becslést tenni (diagnosztikus következtetés).

Az oksági kapcsolatok online tanulása esetében azzal is számolnunk kell, hogy az oksági ismeretek mérésének gyakorisága a tanulás során, jelentősen befolyásolhatja a tanulás végére kialakult hiedelmeket – amint azt Catena, Maldonado, Megías, és Frese (2002), Catena, Perales, és Maldonado (2004), illetve Collins és Shanks (2002) is demonstrálták.

Bár a vizsgálatok sokszínűek a területen, az eredmények – úgy tűnik egybehangzóan – azt mutatják, hogy jók vagyunk az oksági kapcsolatok tanulásában akkor is ha pusztán kovariancia-információ áll rendelkezésünkre. Ezt már gyermekek esetében is határozottan jelzi, hogy pusztán az események megfigyelése révén képesek a kívánt hatást létrehozni a megfelelő manipuláció kiválasztásával (Waismeyer és mtsai., 2015), ami még látványosabb

akkor, amikor van lehetőségük a cselekedeteik következményeiből tanulni (McCormack, Bramley, Frosch, Patrick és Lagnado, 2016; Schulz, Kushnir és Gopnik, 2007). Pusztán kovariancia-információ alapján meglehetősen hatékonyan tudunk egyszerűbb oksági rendszereket (több esemény közötti oksági kapcsolatok rendszerét) megismerni (Ahn és Dennis, 2000; Baetu és Baker, 2009; Kushnir, Gopnik, Lucas és Schulz, 2010). Az oksági kapcsolatok tanulásában ugyanakkor számos más jelzőingerre is támaszkodunk, amelyek befolyásolhatják a kovariancia-információ felhasználásának módját a tanulás során; ilyen például a saját cselekedeteinek (beavatkozásaink) hatásának megfigyelése, vagy az események időbeli eloszlásának jellegzetességei. A meglévő ismereteink, hiedelmeink és elvárásaink ugyancsak jelentősen befolyásolhatják az oksági tanulást (Lagnado, Waldmann, Hagmayer és Sloman, 2007).

### **1.1.1. Az intervenció szerepe az oksági tanulásban**

Ha beavatkozunk az események folyásába, és ezzel valamilyen módon befolyásoljuk a történéseket, az nagyon hasznos információ lehet az események közötti oksági kapcsolatok megismerésében. Az ilyen beavatkozás, intervenció nagyon erős bizonyítékot szolgáltat az oksági kapcsolatok megismeréséhez, ezáltal az oksági tanulás rendkívül hatékony eszköze lehet (Campbell, 2007; Hagmayer, Sloman, Lagnado és Waldmann, 2007; Schulz, Gopnik és Glymour, 2007). Az intervenció hatékonyságát igazolják azok a vizsgálatok is, amelyek a kovariancia-információ puszta megfigyelése alapján történő tanulással vetik azt össze (pl. Lagnado & Sloman, 2004; Saito & Shimazaki, 2014; Steyvers és mtsai., 2003; Waldmann & Hagmayer, 2005).

Steyvers és mtsai. (2003) például egy kísérletükben arra kérték vizsgálati személyeiket, hogy képzeljenek el három földönkívülit, akik közül az egyik képes olvasni egy másik földönkívüli elméjében (azzal a kitéttel, hogy ez a képessége nem minden esetben működik). A földönkívüliek kommunikációs mintázatainak egyszerű megfigyelését követően két lehetséges hipotézis közül kellett választani a vizsgálati személyeknek, amelyek arra vonatkoztak, hogy ténylegesen mely földönkívüli mely másik földönkívüli gondolatait olvassa. Az adatok zajos jellege (a gondolatolvasás nem sikerül mindig) ellenére a vizsgálati személyek a puszta véletlennél jobb találati arányt értek el, tehát viszonylag hatékonyan használták fel a kovariancia-információt az oksági kapcsolatok megismerésére. Ami

érdekesebb, hogy egy másik kísérletükben, ahol hasonló volt a feladat, de a vizsgálati személyek itt tetszőlegesen beavatkozhattak (megváltoztathatták egy földönkívüli gondolatait), a megfelelő hipotézis kiválasztása sokkal egyszerűbbnek bizonyult, a kísérleti személyek találati aránya nagymértékben javult az előző kísérletben tapasztaltnál képest.

Érdekes kérdés lehet, hogy milyen módon választanak az emberek a lehetséges beavatkozások közül, tehát ha van lehetőségük választani, hogy miként avatkozzanak be egy rendszer működésébe, akkor mi határozza meg a döntésüket? Bramley, Lagnado, és Speekenbirk (2015) kísérletükben azt találták, hogy a vizsgálati személyek jelentős része kifinomult intervenciós stratégiát alkalmazott: kezdetben egyszerű beavatkozásokat alkalmazva nagy mértékben leszűkítették a lehetséges hipotézisek körét azokra a hipotézisekre, amelyek megfelelően magyarázták az adatokat, majd egyre specifikusabb beavatkozásokat választottak a fennmaradt hipotézisek közötti döntés érdekében. Coenen, Rehder, és Gureckis (2015) két intervenciós stratégia megjelenését vizsgálták. Az egyik stratégia a lehetséges hipotézisek közötti döntést (diszkriminációs stratégia), a másik bizonyos hipotézisek alátámasztását (konfirmatív stratégia) célozta. A vizsgálati személyek tényleges beavatkozásaira a két stratégia egyfajta keveréke volt jellemző, ugyanakkor képesek voltak stratégiát váltani, ha a feladat azt igényelte.

Felmerülhet a kérdés, hogy mi teszi ennyire hatékonyá az oksági kapcsolatok tanulását a beavatkozások eredményének megfigyelése révén, az események pusztán megfigyelésével szemben? Egy plauzibilis magyarázata lehet az intervenció által nyújtott előnynek, hogy cselekedeteinket az azokat követő események legvalószínűbb okának tekintjük, és ezért hajlamosabbak leszünk azzal párosítani őket (Lagnado és Sloman, 2004; Sobel és Kushnir, 2006). A magyarázat azt feltételezi, hogy gondolkodásunk alkalmazkodik a környezet jellemzőihez, tehát ez esetben ahhoz, hogy – az események természetes időbeli eloszlásának megfelelően – az okok időben mindig megelőzik a következményeiket.

Az intervenciók hatása az oksági tanulásra ugyanakkor nem egyoldalú: az adekvát intervenciós stratégia támogathatja az oksági tanulást, ugyanakkor az oksági kapcsolatokra vonatkozó erős hiedelmeink miatt időnként hajlamosak lehetünk olyan intervenciós stratégiát választani, amely ahelyett, hogy segítene reális képet alkotni az oksági kapcsolatról, tovább mélyítheti esetlegesen téves elképzeléseinket. Az intervenciók ilyen alkalmazása megfigyelhető nem csak felnőtteknél (Goedert, Ellefson és Rehder, 2014), de már

gyerekeknél is (Legare, 2012). A folyamat hozzájárulhat az úgynevezett oksági illúziók, azaz az oksági kapcsolatokra vonatkozó téves hiedelmek kialakulásához, megerősödéséhez és fennmaradásához (Yarritu és Matute, 2015; Yarritu, Matute és Luque, 2015).

### **1.1.2. Események időbeli eloszlásának szerepe az oksági tanulásban**

Az okok időben megelőzik a következményeiket. Ezt, a megfigyelt eseményekben rejlő, az időben egymásra következésre vonatkozó információt már néhány éves gyerekek is hatékonyan alkalmazzák (Rottman, Kominsky és Keil, 2014), felnőtteknél ugyancsak fontos szerepe lehet annak megállapításában, hogy a megfigyelt események közül melyik az ok (előzmény), és melyik a következmény (Burns és McCormack, 2009; Rottman és Keil, 2012).

Az előzmény és következmény időbeli távolsága ugyancsak fontos információ lehet az oksági kapcsolat megismerése során. Úgy tűnik, arra számítunk, hogy a következmények rövid időn belül követik az azokat kiváltó okokat – legalábbis erre utalnak azok az eredmények, amelyek szerint kevésbé vagyunk hajlamosak megállapítani az oksági kapcsolatot két esemény között, ha nagyobb a köztük lévő idői távolság, annak ellenére, hogy adott az események közötti statisztikai összefüggés (Greville és Buehner, 2007; Lagnado és Sloman, 2006; Lagnado és Speekenbirk, 2010; Shanks, Pearson és Dickinson, 1989). Az időbeli távolság hatását feltehetően elvárásaink közvetítik; amennyiben a kérdéses események közötti időbeli távolság megfelel elvárásainknak, akkor annak káros hatása nem jelentkezik (Buehner és May, 2002, 2004), vagy bizonyos esetekben a hatás megfordítható (Buehner és McGregor, 2006).

Buehner és May (2002) például egy kísérletükben arra kérték a résztvevőket, hogy egy számítógépes játékban állapítsák meg, hogy egy bizonyos cselekedetük (lövés gránátvetővel) hatékony-e (előidéz-e robbanást a célterületen)? A kísérleti személyek többször is próbálkozhattak; a gránát az esetek kb. 75%-ban idézett elő robbanást. Az instrukcióban arról is tájékoztatták a kísérleti személyeket, hogy rajtuk kívül mások is lőhetnek ki gránátokat, ezért bizonyos robbanások ennek is betudhatók. A kísérleti személyek egyik csoportja továbbá arról értesült, hogy a gránátok késleltetés nélkül robbannak, míg a kísérleti személyek másik csoportja úgy tudta, hogy a gránátok késleltetve robbannak. Azok a kísérleti személyek, akik úgy tudták, hogy a detonáció azonnali, kevésbé hatékonynak ítélték a cselekedetüket, ha a gránátok ténylegesen késleltetve robbantak, mintha azonnal robbantak

volna. Azoknál viszont, akik úgy tudták, hogy a gránátok késleltetve robbannak, nem jelentkezett ilyen markáns különbség; hasonlóan hatékonyak ítélték a cselekedetüket, akár azonnal akár késleltetve robbantak a gránátok – a késleltetés hatása tehát eltűnt, amikor a kísérleti személyek – az erre vonatkozó instrukció nyomán – elvárták azt.

A következmény időbeli bejósolhatósága egy további fontos jelzőinger, amely az oksági kapcsolatok tanulásában fontos szerepet tölthet be. A bejósolhatóságot nagyban befolyásolja az előzmény és a következmény bekövetkezése között eltelt idő változékonysága; minél nagyobb ez a változékonyság, annál nagyobb a bizonytalanság a következmény bekövetkezésének pontos idejét illetően. A következmény időbeli bejósolhatóságának fokozása valóban serkentően hat az oksági tanulásra (Greville és Buehner, 2016; Lagnado és Speekenbirk, 2010).

A kérdéses események időbeli eloszlása fontos jelzés lehet az oksági kapcsolat meglétére, hiányára, vagy éppen az oksági szerepekre vonatkozóan; ezen információ oksági tanulásban betöltött központi szerepe mellett érvel például Buehner (2005) és Lagnado és mtsai. (2007). Az oksági kapcsolatok megismerése során valóban gyakran előnyben részesítjük az idői eloszlásra vonatkozó információt a kovariancia-információval szemben (Cobos, López, Cano, Almaraz és Shanks, 2002; Lagnado és Sloman, 2006; McCormack, Frosch, Patrick és Lagnado, 2015), ugyanakkor a kovariancia-információ szerepe korántsem elhanyagolható, az idői eloszlásra vonatkozó információ és a kovariancia-információ integrálásának módja meglehetősen kifinomult formát is ölthet (Buehner és McGregor, 2009; Greville és Buehner, 2007; Mendelson és Shultz, 1976; Saito és Shimazaki, 2012).

### **1.1.3. Előzetes ismeretek szerepe az oksági tanulásban**

Természetesen a beavatkozások, intervenciók működésére, valamint az ok és okozat idői eloszlására vonatkozó ismereteken túl egyéb ismeretek is szerepet játszhatnak az oksági tanulás során, befolyásolhatják annak folyamatát és eredményét. A hatások egy része a kérdéses oksági kapcsolatra vonatkozó specifikus ismereteknek tudható be, más részük absztrakt ismeretekre vezethetőek vissza, amelyek az oksági jelenségekről való gondolkodást széles körűen meghatározhatják.

Bár a kovariancia-alapú oksági tanulás modelljeinek nagy része azt feltételezi, hogy a kovariancia-információt alkotó eseménytípusok értelmezése normatív – azaz az a-, és d-

típusú eseményeket (ahol mind az előzmény, mind a következmény jelen van, vagy hiányzik) az oksági kapcsolat mellett szóló pozitív bizonyítékként, míg a b-, illetve c-típusú eseményeket (ahol az előzmény jelen van, a következmény nem, vagy vice versa) az oksági kapcsolat ellen szóló, negatív bizonyítékként kezelik – az emberek esetében az eseménytípusok értelmezése gyakran eltérhet a normától, és hajlamosak lehetnek a pozitív bizonyítékokat negatívként értelmezni, a negatív bizonyítékokat pozitívként, sőt azok bizonyíték-értékét is megkérdőjelezhetik (White, 1998, 2000). Dennis és Ahn (2001) szerint a kovariancia-információ ilyen dinamikus értelmezése lehet felelős a primácia-hatásért is, amely szerint az oksági kapcsolatok online tanulása során a kezdeti próbák hatása maradandó, azok jelentős mértékben meghatározhatják a tanulás végén kialakított véleményt. A primácia-hatás egy lehetséges magyarázata, hogy a kezdeti próbák során kialakított – az oksági kapcsolatra vonatkozó – hipotézis meghatározza a későbbi próbák során prezentált információk értelmezését, mégpedig olyan módon, hogy azok a kezdeti hipotézist támasszák alá (Luhmann és Ahn, 2011; Mandel és Vartanian, 2009).

A mindennapok során gyakran találkozunk olyan információval, ami többértelmű, vagy hiányos, például nem utal egyértelműen a kérdéses események jelenlétére, vagy hiányára. Az ilyen bizonytalan információkat jellemzően nem ignorálják az emberek, hajlamosak felhasználni azokat az oksági tanulás során (Marsh és Ahn, 2003, 2009; Shou és Smithson, 2015a). Ezen információk értelmezését meghatározza a vizsgált oksági kapcsolatról alkotott vélemény, és az értelmezésük olyan lesz, ami konzisztens ezzel a véleménnyel (Marsh és Ahn, 2003, 2009).

Amikor egyes események közötti oksági kapcsolatról tanulunk, nem csak a kérdéses kapcsolatra vonatkozó specifikus ismeretekre teszünk szert, hanem absztrakt ismeretekhez is jutunk, amelyek például események jellegzetes kategóriájának oksági szerepére, esetleg maguknak az oksági kapcsolatok működésének a jellegzetességeire vonatkoznak (Griffiths és Tenenbaum, 2009; Kemp, Goodman és Tenenbaum, 2010). Ilyen absztrakt ismeret lehet például, amikor általában úgy hisszük, hogy a gyógyszerek szedése segít meggyógyulnunk. Ezek az ismeretek – mivel általánosabbak, mint egy konkrét kapcsolatra vonatkozó ismeretek – más oksági kapcsolatok tanulását is befolyásolhatják. Pontosan erre utal az absztrakt transzfer jelensége: a korábban elsajátított ismeretek befolyásolják más tanulási helyzetekben a tanulás folyamatát, vagy eredményét (Glautier, 2004; Hagmayer, Meder, von

sydow és Waldmann, 2011; Waldmann és Hagmayer, 2001). Ilyen absztrakt ismeretek lehetnek még például az oksági kapcsolatok erősségére vonatkozó elvárások is, mint például az, hogy az oksági kapcsolatok inkább erősek, tehát az okok nagy bizonyossággal létrehozzák a következményt (Yeung és Griffiths, 2015).

#### **1.1.4. Megközelítések és magyarázatok**

Az oksági kapcsolatok tanulására vonatkozó magyarázatokat, modelleket jellemzően három fő csoportba szokták sorolni, amelyek között találunk például eltéréseket abban, hogy milyen szerepet tulajdonítanak a kovariancia-információnak a tanulás során, valamint abban is, hogy mennyire hangsúlyosnak tartják a bottom-up, illetve top-down folyamatokat a tanulás folyamatában. A megközelítéseket jellemzően asszociatív-, szabályközpontú-, illetve modellközpontú megközelítésként említi a szakirodalom.

Az asszociatív megközelítés az oksági tanulást a klasszikus kondicionálás egy jellegzetes esetének tekinti, amely során egy bizonyos előzmény és egy bizonyos következmény közötti asszociáció alakul ki azok következetes párosítása (vagy éppen annak elmaradása) nyomán (Dickinson, 2001; Escobar és Miller, 2012; Shanks és Dickinson, 1988). Ennek a feltevésnek megfelelően a vizsgálatok jellemzően az asszociatív tanulás vizsgálatából ismerős ingerversengéses paradigmákat alkalmazták az oksági tanulás kutatásában is (pl. Blanco, Baeyens és Beckers, 2014; Van Hamme és Wasserman, 1994; Wasserman és Berglan, 1998). Az irányzat a tanulás eredményének jellemzően az egyes események közötti oksági kapcsolatok súlyát tekinti. Az asszociatív megközelítés egyik korai nagy sikerű, és több jelenséget is magyarázni képes modellje a Rescorla-Wagner modell (Rescorla és Wagner, 1972), amely eredetileg a klasszikus kondicionálásban tapasztalható blokkolás jelenségének magyarázatára született, ugyanakkor több-kevesebb sikerrel (és néhol kisebb módosításokkal, pl. Van Hamme és Wasserman, 1994) alkalmazták már az oksági tanulás során tapasztalt jelenségek magyarázatára is.

A szabályközpontú megközelítés az oksági tanulást tisztán induktív folyamatnak tekinti, amely során a rendelkezésre álló kovariancia-információból valamilyen egyszerű szabály segítségével következtethető ki a kérdéses oksági kapcsolat erőssége. Az egyik legkorábbi és legjelentősebb elképzelés a  $\Delta P$  szabály (Jenkins és Ward, 1965; Ward és Jenkins, 1965), amely lényegében a következmény előzmény jelenlétében, illetve előzmény hiányában

tapasztalt valószínűsége közötti kontrasztot (azaz a köztük lévő különbséget) ábrázolja:  $\Delta P = p(K=1|E=1) - p(K=1|E=0)$ . Változatos elképzelésekkel találkozhatunk itt (áttekintésért ld. Perales és Shanks, 2007), amelyek különböznek például abban, hogy a kovariancia-információt alkotó egyes eseménytípusok közül melyiket tekintik az oksági kapcsolatra nézve pozitív, illetve negatív bizonyítéknak, vagy hogy milyen súlyokat tulajdonítanak az egyes eseménytípusoknak, tehát milyen erős bizonyítéknak tekintik azokat az oksági kapcsolatra vonatkozóan. Míg az asszociatív elképzelések jellemzően az online tanulásra alkalmazhatók, a szabályközpontú elképzelések jelentős része az összesített kovariancia-információra támaszkodik, bár találunk olyan elképzelést is, amely az oksági kapcsolatra vonatkozó következtetés lépésről lépésre való felülvizsgálatát feltételezi (pl. Catena és mtsai., 2002).

Mind az asszociatív-, mind a szabályközpontú megközelítés a bottom-up folyamatokra helyezi a fő hangsúlyt a tanulás során – ezekkel szemben a modell-központú megközelítés a top-down folyamatok központi szerepét hangsúlyozza az oksági kapcsolatok tanulásában. Eszerint az oksági kapcsolatokra vonatkozó elképzeléseink, vagy hipotéziseink felülvizsgálata zajlik a rendelkezésre álló bizonyítékok – a kovariancia-információ – alapján (Sloman, 2005; Waldmann, 1996). Ezek az elképzelések és hipotézisek irányíthatják a tanulás folyamatát, például befolyásolhatják, hogy milyen jelentőséget tulajdonítunk az egyes információknak, vagy hogy milyen intervenciós stratégiát választunk egy bonyolult kapcsolatrendszer megismerése során (Coenen és mtsai., 2015). A modellközpontú megközelítések egy korai konkrét modellje Cheng power PC modellje (Cheng és Novick, 1992; Cheng, 1997), amelynek egyik központi gondolata, hogy a rendelkezésre álló információ értelmezése során abból az egyszerű feltevésből indulunk ki, hogy ha egy következménynek van két alternatív oka, amelyek egyaránt képesek azt létrehozni, akkor azok között nincs interakció (tehát nem befolyásolják egymás hatását). Ez a feltevés segíthet például az egyszerűbb oksági kapcsolatok megismerésében, mivel így egy kiválasztott előzmény hatása a kérdéses következményre megítélhető anélkül, hogy az összes többi alternatív oksági tényezőt tekintetbe vennénk.

A modellközpontú megközelítés legjelentősebb magyarázatai a kognitív modellezés Bayes-i irányzatához sorolhatók, amely – úgy tűnik – nagyon sikeres integratív szerepet tölthet be az oksági gondolkodással kapcsolatos jelenségek megértésében (Holyoak és

Cheng, 2011), mivel képes lehet a jelenségek széles körét magyarázni, és a vizsgálható kérdések körét is képes kiterjeszteni, például az intervenciók hatására, a bonyolult oksági kapcsolatrendszerek tanulására, illetve a konkrét- és absztrakt ismeretek hatására vonatkozóan.

## **1.2. A kognitív modellezés Bayes-i irányzata**

Alapvető kihívás a környezetünk megismerése során, hogy a rendelkezésünkre álló információ sokszor rendkívül zajos, vagy többértelmű; mi mégis könnyedén megbirkózunk ezzel a problémával. A rendelkezésünkre álló kétdimenziós képek alapján például egyértelmű határokkal rendelkező objektumokat észlelünk magunk körül, egy nagyobb társaságban, jelentős zajban is képesek vagyunk követni egy beszélgetést, kisgyermekként könnyedén tanulunk újabb szavakat, kategóriákat, és rövid idő alatt elsajátítjuk azok helyes alkalmazását a beszéd során. Tenenbaum, Kemp, Griffiths, és Goodman (2011) szerint a megismeréskutatás egyik legizgalmasabb kérdése, hogy hogyan vagyunk képesek annyi mindent, olyan rövid idő alatt megtanulni, kevés és meglehetősen zajos információ alapján? Ez az a kérdés, illetve probléma, amely a megismeréskutatás Bayes-i megközelítésének középpontjában áll.

A kognitív modellezés Bayes-i irányzata egy modern, néhány évtizedes múlttal rendelkező interdiszciplináris megközelítés, amely a megismerés problémájának vizsgálata érdekében merít a pszichológia, statisztika, számítástudomány és filozófia legmodernebb vívmányaiból. A megközelítés hatékonyságát igazolja, hogy azt a megismeréskutatás számos területén sikeresen alkalmazták, mint például a szövegértés, kategóriák tanulása, oksági tanulás, vagy a szenzomotoros koordináció (áttekintésért ld. Chater és Oaksford, 2008; Rahman és Ogilvie, 2006). A megközelítés a felnőtt kognitív teljesítmény magyarázata mellett a kognitív fejlődés kutatása számára is érdekes lehetőségeket tartogat, újszerű kérdések megfogalmazását teszi lehetővé (Gopnik és Bonawitz, 2014; Schultz, 2007).

A célunk itt, hogy röviden áttekintsük a megismeréskutatás Bayes-i megközelítésének legmarkánsabb jellemzőit – a hiedelmek valószínűségként való ábrázolását, a strukturált reprezentációk alkalmazását, valamint a Bayes-tétel központi szerepét –, és felvázoljuk a megközelítéssel szemben megfogalmazott legjellemzőbb kritikákat.

A Bayes-i megközelítés a környezetünkre vonatkozó hiedelmeinket valószínűségként ábrázolja, és ehhez a valószínűségek szubjektív értelmezését alkalmazza. Ilyen szubjektív valószínűségek segítségével fejezi ki például azt, hogy mennyire tartunk hihetőnek egy hipotézist, amely a megfigyelt adatok keletkezését magyarázhatja, vagy mennyire tartjuk esélyesnek azt, hogy egy bizonyos esemény be fog következni. A szubjektív valószínűségek alkalmazása megengedi azt, hogy különböző személyek ugyanarra a hipotézisre, vagy predikcióra vonatkozóan eltérő hiedelmekkel rendelkezzenek – erre utal a szubjektív jelző. Williamson (2004) szerint, ha úgy tekintjük, hogy az aktuális hiedelmeink az aktuális ismereteinken alapulnak, akkor az eltérő hiedelmek problémáját episztemiológiai alapon objektivizálhatjuk; ha ugyanis a hiedelmekben tapasztalható egyének közötti eltérések magyarázhatók az aktuális ismeretekben való eltérésekkel, akkor az ismeretek gyarapodása során a hiedelmek között tapasztalható eltéréseknek csökkenniük kell, míg az összes lényeges ismeret birtokában ezek az eltérések teljesen eltűnnek.

A környezetünkre vonatkozó ismeretek ábrázolására a Bayes-i megközelítés változatos, az ábrázolni kívánt ismeretkör jellegzetességeihez igazodó, gazdagon strukturált reprezentációs eszközöket alkalmaz, amelyek az adott ismeretkörön belül komplex valószínűségi következtetéseket tesznek lehetővé. Ilyen eszköz például az oksági Bayes-háló formálizmus, amely az oksági ismeretek ábrázolására, valamint valószínűségi oksági következtetések egzakt levezetésére alkalmas, ennek megfelelően az oksági gondolkodás kutatásában nagy mértékben támaszkodnak rá. Egyéb területeken olyan specifikus reprezentációs eszközöket alkalmaznak, mint például a probabilisztikus kontextusfüggetlen nyelvtanok a beszéd-feldolgozásban, a kétdimenziós rácsok a téri tájékozódásban. Hasznos felismernünk ugyanakkor, hogy az ismeretek ilyen módon való ábrázolása szigorú normatív megszorítások alkalmazását jelent a megismerési- és tanulási folyamatokra nézve, ezért a megközelítés inkább egy racionális nézőpontot képvisel.

A Bayes-i megközelítés további központi eleme, hogy a megismerési folyamatokat mint egyszerű valószínűségi következtetéseket képzelel el, amelyeket az ismereteket ábrázoló strukturált reprezentációk alapján vezethetünk le. Ennek a valószínűségi következtetésnek az általános formáját ragadja meg a Bayes-tétel (2).

$$p(\mathcal{H}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathcal{H})p(\mathcal{H})}{p(\mathcal{D})} \quad (2)$$

A Bayes-tétel – lényegét tekintve – egy intuitív elképzelést ragad meg, amely szerint a rendelkezésre álló adatok ( $\mathcal{D}$ ) alapján felülvizsgáljuk a hiedelmeinket, feltevéseinket ( $\mathcal{H}$ ) – valójában a Bayes-tétel a hiedelmek felülvizsgálatának matematikailag egzakt, racionális módja. A tételben  $p(\mathcal{H})$  az előzetes hiedelmeinket, bizonyos feltevéseink, elméleteink plauzibilitását, valószínűségét ábrázolja (ezek az adott ismeretkörre vonatkozó strukturált reprezentációk alapján levezethetők), valamint bizonyos kognitív folyamatok jellegzetes beállítódását is reprezentálhatja, amelyek a tanulás folyamatát alapvetően meghatározhatják. A  $p(\mathcal{D}|\mathcal{H})$  kifejezi, hogy mennyire tartjuk elképzelhetőnek a megfigyelt adatokat, azaz tapasztalatainkat, egy bizonyos feltevés/hipotézis mellett. Ez ugyancsak levezethető az alkalmazott strukturált reprezentációk segítségével, amelyek ez esetben generatív modellekként működnek – azt írják le, hogy hogyan keletkeznek az adatok az azokat létrehozó folyamatokra vonatkozó feltevések és megszorítások, valamint a valóságra vonatkozó bizonyos elképzelések, hipotézisek mellett.

A megismeréskutatás Bayes-i irányzata a megismerési folyamatok Marr-i értelemben vett komputációs szintű magyarázatára törekszik. Marr (1982) a kognitív folyamatok működésének leírását három lehetséges szinten képzelte el, amelyek a kognitív rendszerek működésének vizsgálata során a megértés minőségileg eltérő formáit teszik lehetővé. Az elemzés legabsztraktabb szintje a komputációs szint; itt a középpontban a kognitív rendszer által megoldani hivatott probléma és a lehetséges megoldási módok elemzése, valamint a kérdéses folyamatok funkciójának a megértése áll. A következő, algoritmus szinten az érdeklődés azokra a kognitív folyamatokra és mentális reprezentációkra irányul, amelyek a kérdéses probléma megoldásában szerepet játszanak. A leírások legalapvetőbb, implementációs szintjén annak a biológiai-, vagy fizikai rendszernek a felépítése áll az érdeklődés középpontjában, amely megvalósítja a kérdéses kognitív folyamatokat. Bár a Bayes-i megközelítés területén is felfedezhető némi törekvés a magyarázatok kevésbé absztrakt szinten való megfogalmazására (pl. Griffiths, Lieder és Goodman, 2015), jellemzően azonban a kognitív folyamatok magyarázata a megoldandó problémára fókuszál, és annak jellemzően egy optimális módját vázolja fel. Ez egy racionális, normatív alapot is biztosíthat az emberi teljesítmény megítéléséhez.

A komputációs magyarázatoknak, így a Bayes-i modelleknek is fontos alapja a kognitív rendszer által megoldandó probléma elemzése, amelynek Anderson (1990) által

javasolt módszere a racionális elemzés. A módszer abból a feltevésből indul ki, hogy a kérdéses kognitív ágens alkalmazkodott a környezetéhez. Ez az alkalmazkodás tekinthető az evolúciós fejlődés következményének is – bár a módszer nem igényli feltétlenül az evolúciós érvek elfogadását. A probléma egy optimális megoldásának felvázolásához ismernünk kell a kognitív ágens céljait, annak a környezetnek a releváns jellemzőit, amelyben az ágens ezen célok elérésére törekszik, valamint azokat a kognitív feldolgozásbeli korlátokat, amelyek a megismerési folyamatokat meghatározzák. Természetesen egy problémakör racionális elemzése iteratív folyamat, amely során a magyarázatot folyamatosan felülvizsgáljuk annak érdekében, hogy jobban illeszkedjen a tapasztalatokhoz.

A megismeréskutatás Bayes-i irányzatával szemben megfogalmazott kritikák jellemzően a magyarázatokban alkalmazott, a környezetre, valamint a kognitív folyamatokra vonatkozó feltevések empirikus megalapozottságának hiányára, az ugyanazon viselkedés magyarázatára vonatkozó magyarázatok sokféleségére, és ezek egzakt összevetésének hiányára, valamint a modellek bizonyos feltevéseinek némileg önkényes – a jobb illeszkedés érdekében történő – megválasztására vonatkoznak; ezeken túl fel szokott merülni az is, hogy a Bayes-i megközelítés tényleges pszichológiai mondanivalója csekély, és sokszor nem kellően világos (Bowers és Davis, 2012; Jones és Love, 2011; Marcus és Davis, 2013). A kritikák egy másik része a kognitív folyamatok mechanisztikus leírását, vagy a biológiai, és kognitív kapacitásbeli korlátok figyelembevételét kéri számon a megközelítéstől, holott az kifejezetten a magyarázatoknak egy absztrakt szintjét, a Marr-i komputációs szintet célozza meg, és itt elfogadott lehet a racionális, normatív elvek alkalmazása, valamint a tényleges kognitív folyamatok jellemzésének hiánya (Hahn, 2014). A feltevések önkényesnek tűnő megválasztása, a magyarázatok megfogalmazásának és alkalmazásának gyakorlata valóban finomításra szorulhat, de tekintettel arra, hogy a megközelítés még nagyon fiatal, és alapjait folyamatosan tisztázni igyekeznek, a jelenlegi ígéretes eredményektől még hosszú és nehéz út vezet addig, hogy a megismerés vizsgálatának igazán eredményes és hatékony eszközévé válhasson (Griffiths és mtsai., 2015).

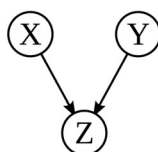
### **1.3. Az oksági Bayes-hálók formalizmus áttekintése**

Az oksági Bayes-hálók formalizmust a XX. század második felében dolgozták ki, és lényegében sztochasztikus változók közötti oksági kapcsolatokon alapuló komplex

valószínűségi következtetések megfogalmazásának, valamint az oksági rendszer jellemzőinek adatok alapján való megismerésének matematikai eszköze (Pearl, 1988, 2000; Spirtes, Glymour és Scheines, 2000). A formalizmus megjelenése teret nyitott a kognitív pszichológián belül az oksági gondolkodás egy újfajta szemléletének, és a területen megfogalmazott kérdések és hipotézisek körét megsokszorozta, ezáltal kifejezetten élnkítően hatott a téma kutatására (Glymour, 2003).

Ebben a fejezetben röviden bemutatjuk az oksági Bayes-hálókat, különös hangsúlyt fektetve a formalizmus azon aspektusaira, amelyek a ROT-modell működésének megértése szempontjából kritikusak lehetnek. A formalizmus részletekbe menő ismertetésétől itt eltekintünk, azt kimerítően tárgyalja ugyanakkor Jensen (2001), valamint Koller és Friedman (2009).

Az oksági Bayes-hálóknak megkülönböztethetjük a kvalitatív aspektusát (ami a kérdéses változók közötti oksági kapcsolatokat definiálja), valamint a kvantitatív aspektusát, amely ezen kapcsolatok funkcionális leírását tartalmazza.



3. ábra  
Oksági Bayes-háló, egy három változóból álló oksági rendszer ábrázolására

A kérdéses oksági rendszer szerkezetét (az oksági struktúrát) irányított körmentes gráf segítségével ábrázolják az oksági Bayes-hálókat (ld. 3. ábra). A gráf csomópontjai a változókat reprezentálják, amelyek lehetnek folytonosak, vagy diszkrét. A csomópontokat (változókat) összekötő élek lényegében nyilak, amelyek a változók közötti oksági kapcsolatokat ábrázolják; a nyíl kezdeténél található változó az ok, a végénél található változó pedig az okozat szerepét tölti be ebben a kapcsolatban. Fontos megjegyeznünk, hogy az oksági Bayes-hálókat esetében ezeknek az oksági kapcsolatoknak a metafizikai státusza nem rögzített (Woodward, 2003), tehát ábrázolhatnak valamilyen közvetlen, fizikai jellegű kapcsolatot (például a gázok nyomásának a megemelése a hőmérsékletük növekedését okozza), vagy akár valamilyen elvont szinten megfogalmazható, esetleg közvetett

kapcsolatot (például az elnökválasztás eredménye az értékpapírok értékének a csökkenését okozza). Az oksági kapcsolatok ilyen rugalmas megfogalmazása lehetővé teszi, hogy a formalizmust problémák széles körére alkalmazzassuk.

Az ábrázolt változók közötti viszonyokra a klasszikus gráfelméleti fogalmakkal utalhatunk. A változók tehát lehetnek szülő-gyerek viszonyban, ahol a szülő-változó az ok, a gyerek-változó pedig az okozat szerepét viseli. Egy változó felmenője alatt azokat a változókat értjük, amelyek vagy közvetlen, vagy közvetett okai annak a változónak, a leszármazottak pedig egy változó közvetlen, vagy közvetett következményei. A gráf körmentes jellege arra utal, hogy nincs olyan csomópont, amelyből kiindulva, és a nyilak irányába haladva, visszatérhetünk ugyanahhoz a csomóponthoz, tehát egyik változó sem lehet önmagának leszármazottja, vagy felmenője.

Az oksági Bayes-hálók kvantitatív aspektusa a változók közötti oksági kapcsolatok funkcionális meghatározása, azaz a háló paramétereinek a megadása. Ez az egyes változók feltételes valószínűségi eloszlásának a meghatározását jelenti, azaz minden csomópontra meg kell adnunk  $p(Z|\text{szülő}(Z))$ -t, ahol  $\text{szülő}(Z)$  a  $Z$  változó összes felmenőjét (közvetlen, vagy közvetett okát) jelenti. Azon változók esetében, amelyeknek nincs felmenője, a feltételes valószínűségi eloszlás egyszerű valószínűségi eloszlásként adható meg, mint  $p(X)$ , illetve  $p(Y)$ . Elterjedt az egyes változók közötti oksági kapcsolatok erejét jelző súlyok (oksági súlyok) alkalmazása is, amelyeket felhasználva, bizonyos – akár a különböző változók hatását integráló – függvényekkel is leírhatók a feltételes valószínűségi eloszlások. Ilyen, elterjedten használt függvények például a zajos-vagy, illetve a zajos-és-nem függvények, amelyek több, egymástól független hatással bíró változó hatásának integrálását végzik el, előbbi generatív oksági kapcsolatok, utóbbi pedig preventív oksági kapcsolatok esetén.

A formalizmus lényegében alkalmas arra, hogy az ábrázolt oksági rendszer változóinak együttes valószínűségi eloszlását írja le tömören (kevés paraméterrel), a változók között feltételezett oksági kapcsolatok jelentette megszorítások mellett. Az ábrázolt változók együttes valószínűségi eloszlása az úgynevezett láncszabály segítségével felírható:  $p(X, Y, Z) = p(Z|X, Y)p(X)p(Y)$ , vagy általánosabb formában,  $X_1, X_2, \dots, X_n$  változók

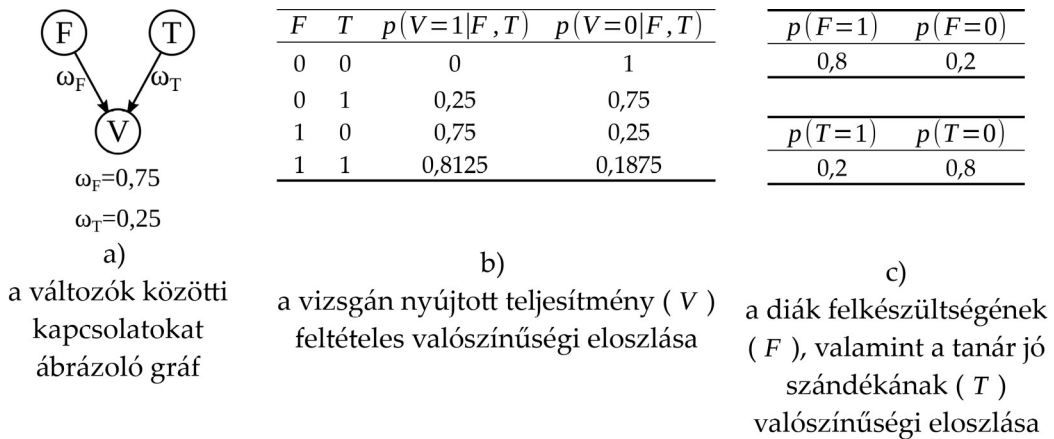
közötti kapcsolatokat ábrázoló oksági Bayes-háló esetében:

$$p(X_1, X_2, \dots, X_n) = \prod_{i=1}^n p(X_i | \text{szülő}(X_i)).$$

A formalizmus segítségével levezethető valószínűségi következtetéseket részletesen tárgyalja Pearl (1988, 2000), Spirtes és mtsai. (2000), Jensen (2001), valamint Koller és Friedman (2009), az oksági Bayes-hálók adatok alapján való tanulását pedig Jordan (1999), valamint Glymour és Cooper (1999). Itt annak az egyszerű következtetés-fajtának az ismertetésére szorítkozunk, amely során néhány változó együttes valószínűségi eloszlását úgy szeretnénk meghatározni, hogy egyéb változókat figyelmen kívül hagyunk (ami lényegében ezen változók kiintegrálását jelenti az együttes valószínűségi eloszlásból), mivel ez az a következtetés-fajta, amelyet a ROT-modell esetében is alkalmazunk. A kérdéses változók együttes valószínűségi eloszlása, bizonyos változók figyelmen kívül hagyása mellett meghatározható (3) módon.

$$\begin{aligned} p(X, Z) &= \sum_Y p(Z|X, Y) p(X) p(Y) \\ &= p(Z|X, Y=0) p(Y=0) + p(Z|X, Y=1) p(Y=1) \end{aligned} \quad (3)$$

Az oksági Bayes-hálók megadására, és egy egyszerű következtetésre tekintünk itt egy példát, amelyben leegyszerűsítve ábrázoltunk egy oksági rendszert, nevezetesen egy diák vizsgán nyújtott teljesítményét befolyásoló tényezőket (4. ábra).



4. ábra

Egy egyszerű oksági Bayes-háló, egy diák vizsgán nyújtott teljesítményének modellezésére

Az oksági rendszer kvalitatív aspektusát a három változót tartalmazó gráf ábrázolja (4.a ábra). Itt  $V$  jelöli a diák vizsgán nyújtott teljesítményét, amelyet dichotóm változóként kezelünk; ennek megfelelően a vizsga lehet sikeres ( $V=1$ ), vagy sikertelen ( $V=0$ ). Kettő, a vizsga eredményét meghatározó tényezőt emeltünk ki, a diák felkészülését a vizsgára ( $F$ ), valamint a tanár jó szándékát ( $T$ ). Mindkét tényezőt dichotómnak tekintjük, tehát a diák vagy felkészült a vizsgára ( $F=1$ ), vagy nem készült fel a vizsgára ( $F=0$ ), a tanár pedig vagy jó szándékú ( $T=1$ ), vagy nem ( $T=0$ ). A példában ezen tényezők oksági erejét – tehát azt, hogy azok milyen megbízhatósággal határozzák meg a vizsga eredményét – oksági súlyként adtuk meg, amely a felkészülésnek egy viszonylag markáns pozitív ( $\omega_F=0,75$ ), a tanár jó szándékának pedig egy viszonylag gyenge pozitív hatását ( $\omega_T=0,25$ ) jelenti a vizsga eredményére.

A vizsga eredményének feltételes valószínűségi eloszlása (4.b ábra), valamint a diák felkészülésének, és a tanár jó szándékának valószínűségi eloszlása (4.c ábra) alkotja a modell kvantitatív aspektusát. Ami a diák felkészültségét illeti, annak valószínűségét 0,8-nak tekintjük a példában, annak a valószínűsége pedig, hogy a tanár jószándékú lesz a vizsgán, 0,2. A vizsga eredményének feltételes valószínűségi eloszlását a zajos-vagy függvény segítségével számoltuk ki, amely szerint  $p(V|F,T)=1-(1-\omega_F)^F(1-\omega_T)^T$ . Érdemes megfigyelni, hogy ilyen módon a három dichotóm változót tartalmazó rendszer együttes valószínűségi eloszlását a 8 helyett 4 paraméterrel tudtuk leírni, az ugyanis a láncszabály segítségével megadható a következő módon:  $p(F,T,V)=p(V|F,T)p(F)p(T)$ .

A következtetés, amit a példa alapján levezetünk annak a valószínűségére vonatkozik, hogy a diák felkészült a vizsgára és a vizsga sikeres lesz (arról nincs információnk, hogy a tanár jó szándékú-e a vizsgán), azaz  $p(F=1,V=1)$ . Ennek meghatározásához a tanár jó szándékának kiintegrálására van szükség, tehát:  $p(F=1,V=1)=\sum_T p(V|F=1,T)p(F=1)p(T)$ , amely szerint 0,6 annak a valószínűsége, hogy a diák felkészült a vizsgára és a vizsga sikeres lesz, függetlenül attól, hogy a tanár jó szándékú, vagy sem. A számítás menetét részletezve (4)-ben ábrázoltuk.

$$\begin{aligned}
 p(F=1,V=1) &= \sum_T p(V=1|F=1,T)p(F=1)p(T) \\
 &= p(V=1|F=1,T=0)p(F=1)p(T=0) + \\
 &\quad p(V=1|F=1,T=1)p(F=1)p(T=1) \\
 &= 0,75 \cdot 0,8 \cdot 0,8 + 0,8125 \cdot 0,8 \cdot 0,2 \\
 &= 0,6
 \end{aligned} \tag{4}$$

## 2. A ROT-modell bemutatása

A ROT-modell Griffiths és Tenenbaum (2005) support-modelljének egy módosított változata, és ahhoz hasonlóan elemi oksági kapcsolatok (egy bizonyos előzmény és következmény közötti oksági kapcsolat) kovariancia-információn alapuló tanulásának Bayes-i modellje. Ebben az értelemben az oksági tanulást valószínűségi következtetésként modellezi, és eközben az oksági Bayes-hálókat formalizmust használja. A modell a Marr-i komputációs szintet célozza meg, és ilyen tekintetben egy normatív, racionális nézőpontot képvisel, amennyiben az oksági tanulásnak egy lehetséges optimális módját írja le. Jelen formájában a modell a generatív oksági kapcsolatok magyarázatára korlátozódik, és a becslése az oksági kapcsolat meglétében való bizonyosságra vonatkozik (az oksági kapcsolat súlya/ereje helyett). Ahogy az elnevezése is utal rá, a modell fontos szerepet tulajdonít az oksági tanulásban a tanulási helyzetben nem megfigyelhető (rejtett) oksági tényezőknek. Az, hogy ezekről az oksági tényezőkről nem áll rendelkezésre közvetlen információ, alapvető bizonytalanságot eredményez az oksági kapcsolat meglétére vonatkozó következtetés során.

A modell működését azon lépések, szakaszok mentén mutatjuk be, amelyeken keresztül az oksági kapcsolatra vonatkozó becslést meghatározza. Ezek a lépések a következők:

1. Magyarázatok generálása a megfigyelt esetekre.
2. A rejtett oksági tényezők jelenlétére vonatkozó bizonytalanság felülvizsgálata.
3. Az oksági kapcsolatra vonatkozó hipotézisek felülvizsgálata.

A becslés első szakaszában a modell oksági magyarázatokat generál a megfigyelt kovariancia-információ alapján. Ezekben a magyarázatokban központi szerepet töltenek be a tanulási helyzetben nem megfigyelt (rejtett) oksági tényezők, amelyek között a modell megkülönböztet generatív- (a következményt előidézni képes), illetve preventív (a következmény bekövetkezését megakadályozni képes) oksági tényezőket. Ez a ROT-modell egyik jellegzetes tulajdonságára mutat rá, nevezetesen, hogy explicitté teszi a kovariancia-információt alkotó eseménytípusokra adott oksági magyarázatok szerepét az oksági tanulás folyamatában.

A következő szakaszban a rejtett oksági tényezők egyes esetekben való jelenlétére vonatkozó bizonytalanság felülvizsgálata történik, a korábban generált oksági magyarázatok

alapján – ez ennek a bizonytalanságnak a csökkenését fogja jelenteni. A ROT-modell ugyancsak explicit módon reprezentálja azt a bizonytalanságot, amely a rejtett oksági tényezők jelenlétére vonatkozik az egyes megfigyelt esetekben. Ez a szakasz lényegében a rejtett oksági tényezők jelenlétére vonatkozó következtetésként is felfogható.

Az utolsó, harmadik szakaszban az oksági kapcsolatra (a kapcsolat meglétére, illetve hiányára) vonatkozó hipotézisek kiértékelése, és a rendelkezésre álló kovariancia-információnak az oksági kapcsolat meglétére vonatkozó hipotézis melletti, valamint az oksági kapcsolat hiányára vonatkozó hipotézis melletti valószínűségének összevetése történik. A modell alapján annál nagyobb fokú a bizonyosság az oksági kapcsolat meglétében, minél hihetőbbek a megfigyelések az oksági kapcsolat meglétére vonatkozó hipotézis mellett, mint az oksági kapcsolat hiányára vonatkozó hipotézis mellett.

A következőkben részletesen ismertetjük a ROT-modell formális leírását, amely lehetőséget ad annak számítógépes program formájában való megvalósítására is. Ahogy McClelland (2009) hangsúlyozza, amikor egy számítógépes modellt építünk, akkor kénytelenek vagyunk a számítások minden egyes lépést egzakt módon, és pontosan megadni – ez szükségessé teheti bizonyos alapfeltevések alkalmazását, amelyek vonatkozhatnak az emberek képességeire, ismereteire, hiedelmeire, vagy bizonyos kognitív folyamatok jellegzetességeire. A feltevések között lehetnek olyanok, amelyek több-kevesebb empirikus megalapozottsággal bírnak, valamint olyan alkalmi feltevések, amelyeknek ugyan nincs empirikus alapjuk, de alkalmazásuk szükséges ahhoz, hogy a modell működőképes legyen. A ROT-modell építése során igyekeztünk explicitté tenni, hogy mely feltevéseink bírnak empirikus megalapozottsággal (ezek a modell központi alapfeltevései), és melyek pusztán alkalmi, ad hoc feltevések.

Elsőként a ROT-modell központi alapfeltevéseit ismertetjük, amelyek a modell építése során számos döntést meghatároztak, majd bemutatjuk azt a két – a kérdéses oksági kapcsolat meglétére, valamint annak hiányára vonatkozó – hipotézist, amelyeket a modell kiértékel, ezután pedig ismertetjük a becslés részletes menetét, a fent említett három szakaszt. A modellt R szkript formájában implementáltuk – amit az A melléklet tartalmaz.

A ROT-modell bemutatása során egy példán keresztül mutatjuk be a számítások menetét. A kidolgozott példa több részletben, az egyes fejezetek végén, önálló dobozban fog szerepelni.

### **Segít a kávé felébredni?**

A példában a ROT-modellnek a kávé és az éberség közötti oksági kapcsolatra vonatkozó becslését fogjuk levezetni, egy néhány esetből álló kovariancia-információ alapján.

A rendelkezésre álló adatok a következőképpen néznek ki (több embert kérdeztünk meg, hogy ivott-e kávé az elmúlt 1 órában, illetve hogy ébernek érzi-e magát):

		<b>ébernek érzi magát?</b>	
		<i>igen</i>	<i>nem</i>
<b>ivott kávé?</b>	<i>igen</i>	12 fő	6 fő
	<i>nem</i>	6 fő	12 fő

Az adatok szerint tehát abból a 18 főből, akik ittak kávé, 12 fő ébernek érezte magát (a-típusú esemény), míg 6 fő nem érezte magát ébernek (b-típusú esemény), abból a 18 főből pedig, akik nem ittak kávé, 6 fő érezte magát ébernek (c-típusú esemény), 12 fő pedig nem érezte magát ébernek (d-típusú esemény).

A kovariancia-információt alkotó adatok az egyes eseménytípusoknak megfelelően:

$$N_a = 12$$

$$N_b = 6$$

$$N_c = 6$$

$$N_d = 12$$

## **2.1. A ROT-modell központi alapfeltevései**

A ROT-modell megépítése során három olyan alapfeltevést alkalmaztunk, amelyek kellő empirikus megalapozottsággal rendelkeznek. Ezek az oksági kapcsolatok determinisztikus felfogása, a magyarázatok – azokon belül is az oksági magyarázatok – markáns szerepe az oksági tanulásban, valamint az oksági magyarázatok jellegzetes szerkezete, amennyiben azok a tanulási helyzetben nem megfigyelt (rejtett) generatív- és preventív oksági tényezők jelenlétére, vagy hiányára utalhatnak. A következőkben ezeket az alapfeltevéseket részletesebben is áttekintjük.

A ROT-modell egyik központi feltevése szerint az emberek az oksági gondolkodás során az oksági kapcsolatokról mint determinisztikus kapcsolatokról gondolkodnak. Többen kritizálják azt az elképzelést, amely szerint az emberek az oksági kapcsolatokról mint eleve probabilisztikus kapcsolatokról gondolkodnának, a kevés empirikus bizonyíték inkább arra utal, hogy az emberek az oksági kapcsolatokat determinisztikus kapcsolatnak képzelik el

(Goldvarg és Johnson-Laird, 2001; Schulz és Sommerville, 2006; Frosch és Johnson-Laird, 2011). Közvetetten azok az eredmények is erre utalhatnak, amelyek szerint jellemzően úgy képzeljük el az oksági kapcsolatokat, amelyek nagyon erősek, tehát nagy valószínűséggel létrehozzák a következményeiket (Yeung és Griffiths, 2011; Yeung és Griffiths, 2015).

Mit is jelent az oksági determinizmus? Enyhébb formájában arra utal, hogy minden eseménynek van oka, semmi sem történik ok nélkül. Erősebb formájában azt jelenti, hogy az ok szükségszerűen létrehozza a következményt (okozatot). A probabilisztikus okság feltételezése ezzel szemben azt állítja, hogy az események közötti oksági kapcsolat révén az ok jelenlétében nagyobb valószínűséggel fog bekövetkezni a következmény, mint annak hiányában. Az erős determinizmus egyik képviselője a jeles matematikus Pierre-Simon de Laplace, aki szerint csak akkor beszélhetünk egy oksági kapcsolatról probabilisztikus kapcsolatként, ha a szóban forgó oksági rendszernek (amelynek a kérdéses kapcsolat is részét képezi) számos változóját ignoráljuk. Ugyanezen az alapon Griffiths és Tenenbaum (2009) javasolja, hogy az oksági tanulás során az oksági kapcsolatok probabilisztikus vagy determinisztikus felfogása összeegyeztethető – amikor az emberek oksági kapcsolatokról való gondolkodását probabilisztikusként kezeljük, akkor lényegében arra utalunk ezzel, hogy jellemzően ignorálják az adott oksági rendszer egyéb, számukra az adott helyzetben érdektelen változóit.

Milyen következményei lehetnek annak, ha az emberek esetében az oksági kapcsolatokról való determinisztikus gondolkodást feltételezünk? Várhatóan azokra a tapasztalatainkra, amelyek ellentmondanak várakozásaiknak (például egy ismert oksági összefüggés esetén az ok bekövetkezik, de elmarad annak következménye), igyekeznek valamilyen magyarázatot keresni, amely segít megérteniük az adott helyzetet az összefüggésre vonatkozó hiedelmeik teljes feladása helyett. A ROT-modell másik központi feltevése az ilyen magyarázatok oksági tanulásban betöltött alapvető szerepére vonatkozik.

Nem passzív feldolgozói vagyunk annak az adathalmaznak, amelyet tapasztalataink szolgáltatnak, hanem spontán módon interpretáljuk ezeket az adatokat, magyarázatokat alkotunk. Úgy tűnik, hogy ezek a magyarázatok nem csupán a tanulási folyamat melléktermékei, hanem sokkal inkább ezek a tanulás folyamatának lényegi részét képezik, amelyek a tanulás folyamatát jellegzetes módon meghatározhatják (Keil, 2006; Legare és Lombrozo, 2014; Lombrozo, 2006; Lombrozo és Vasilyeva, 2016).

Elmondható ugyanakkor a mindennapi gondolkodásunk során alkotott magyarázatokról, hogy sokszor akár rendkívüli mértékben leegyszerűsítőek, a valóságnak egy vázlatos és nem feltétlenül pontos képét adják vissza. Az is jellemző, hogy ennek ellenére hiszünk ezekben a magyarázatokban, „a valóság”-ként éljük meg ezeket, eltekintünk pontatlanságuktól - ezt a magyarázatok teljességének illúziójaként említi a szakirodalom (Keil, 2003; Rozenblit és Keil, 2002). A magyarázatok teljességének illúziója adaptív szerepet tölthet be, hiszen egy bizonytalansággal teli, komplex környezetben, amit lehetetlen a maga teljességében megismernünk, lehetővé teszi azt, hogy magabiztosan cselekedjünk (Keil, 2006). Az oksági magyarázatok során erre a korlátozott jellegre utalhat Mackie (1965) „okási mező” fogalmával, valamint Cheng (1997) a „fokális halmaz” fogalmával. Mindkét fogalom arra utal, hogy az oksági kapcsolatokról való gondolkodás során az összes oksági tényezőnek csak egy szűkebb körét mérlegeljük ahelyett, hogy minden lehetséges tényezőt tekintetbe vennénk. Hogy mely tényezőket mérlegeljük, azt meghatározhatják céljaink (Mackie, 1965), valamint kognitív erőforrásaink is behatárolhatják ezek körét (Cheng és Novick, 1992; Cheng, 1997).

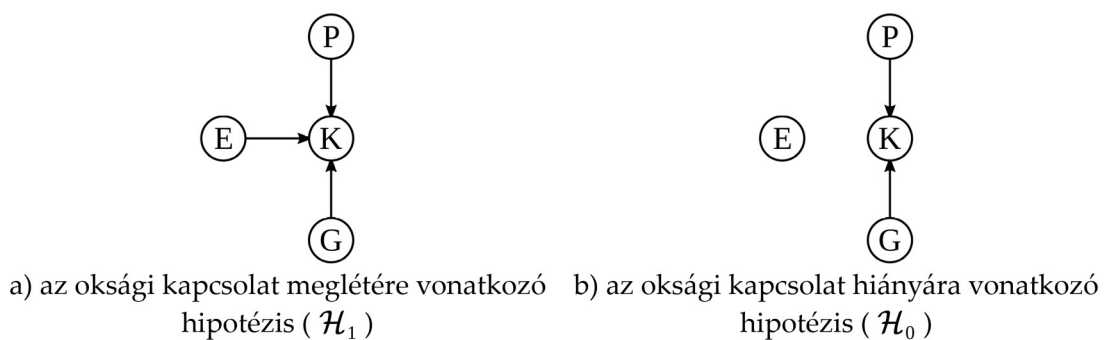
Az események magyarázata során hajlamosak vagyunk oksági magyarázatokra, tehát az események okaira, vagy éppen következményeire vonatkozó magyarázatokra támaszkodni (Keil, 2006; Lombrozo, 2006; Lombrozo és Vasilyeva, 2016), amelyek azokra az oksági tényezőkre vonatkoznak, amelyek ismereteink szerint szerepet játszhatnak a kérdéses helyzetben. A ROT-modell harmadik központi feltevése szerint ezek a tényezők kétfélek lehetnek: generatív oksági tényezők, amelyek képesek létrehozni a következményt, valamint preventív oksági tényezők, amelyek képesek megakadályozni a következmény bekövetkezését.

Amikor ismert oksági kapcsolatok alapján oksági következtetéseket alkotunk, akkor erősen támaszkodunk azokra az alternatív okokra, valamint akadályozó tényezőkre, amelyek befolyásolhatják még a következményt (Cummins, Lubart, Alksnis és Rist, 1991; Cummins, 1995, 2014). Ezen tényezők adott helyzetben való felhasználhatósága (akár elérhetőségük, akár ezek előhívásához rendelkezésre álló kognitív kapacitás) befolyásolja, hogy mennyire vagyunk hajlamosak elfogadni bizonyos prediktív- (azaz az előzményből a következményre vonatkozó), illetve diagnosztikus (azaz a következményből az előzményre vonatkozó)

következtetéseket (De Neys, Schaeken és dYdewalle, 2002; De Neys, Schaeken és dYdewalle, 2005; Kurucz, 2010b).

## 2.2. A mérlegelt hipotézisek formális leírása

A kérdéses oksági kapcsolat meglétére, illetve hiányára vonatkozó hipotézisek ábrázolására a ROT-modell az oksági Bayes-háló formalizmust használja. A két hipotézis közötti alapvető különbség az előzmény és a következmény közötti oksági kapcsolat meglétében van, amit a hipotézisekben feltételezett oksági rendszerekben az azokat reprezentáló nyilak megléte, illetve hiánya ábrázol (5. ábra).



5. ábra

A ROT-modell által vizsgált, az előzmény és következmény közötti oksági kapcsolat meglétére, illetve hiányára vonatkozó hipotézisek

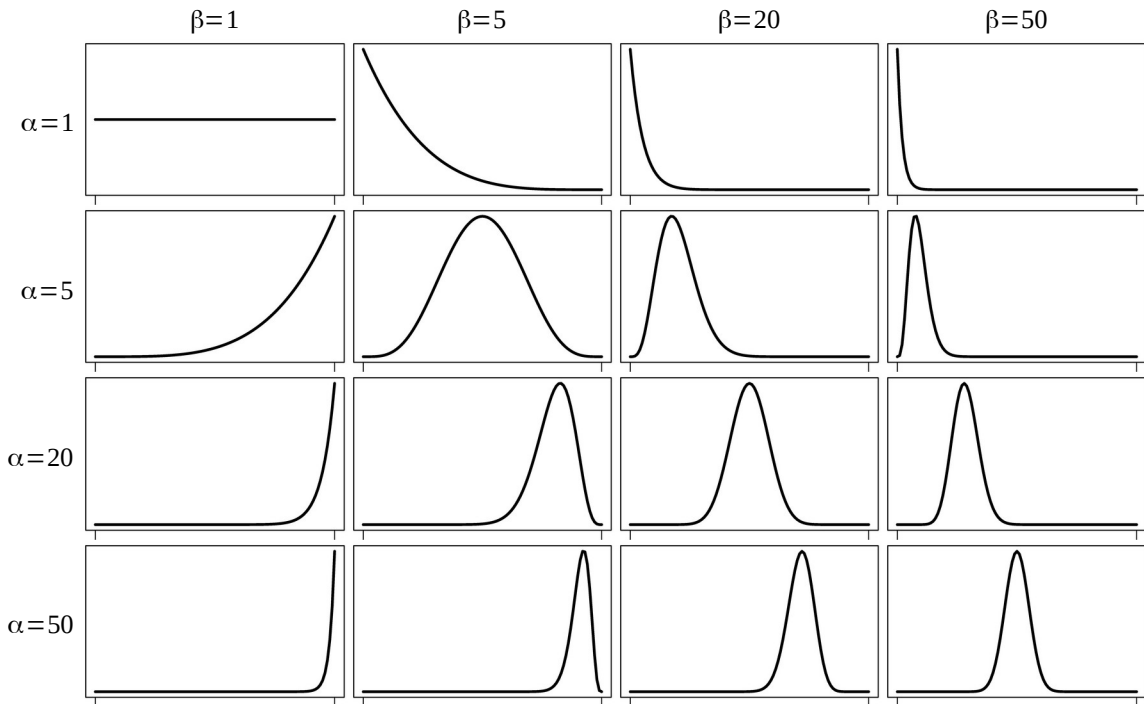
Mindkét hipotézis ugyanazokkal a változókkal dolgozik, amelyek között vannak a tanulási helyzetben megfigyelt változók (ezek az előzmény /  $E$  / és a következmény /  $K$  /), valamint olyan változók, amelyeket a tanulási helyzetben nem figyelhettünk meg – ezek olyan rejtett oksági tényezők, amelyek fontos szerepet játszanak az adott oksági rendszer működésében. A generatív oksági tényezők ( $G$ ) olyan események, amelyek képesek létrehozni a következményt, a preventív oksági tényezők ( $P$ ) pedig olyan események, vagy feltételek, amelyek megakadályozzák, hogy a következmény bekövetkezzen. Mindkét rejtett oksági tényező különböző eseményeknek és feltételeknek egyfajta konglomerátumát ábrázolja, tehát ahelyett, hogy egy-egy konkrét esemény jelenlétére, vagy hiányára utalnának, általában utalnak arra, hogy valamilyen, a következmény létrehozására, vagy annak megakadályozására képes tényező jelen van. Meg kell említenünk, hogy esetünkben mellékes, hogy milyen ezen tényezők metafizikai státusza, hogy azok valóban eseményeknek

tekinthetők, esetleg bizonyos akadályozó feltételekre, vagy megengedő feltételek hiányára utalnak (a preventív tényezők esetében), a lényeg, hogy a vizsgált oksági rendszerben a hatásuk a következményre milyen lehet. A hipotézisekben ábrázolt oksági rendszerek ilyen formában teljesekek, az adott következmény szempontjából minden lényeges oksági tényezőt tartalmaznak.

A hipotézisekben alkalmazott változók mindegyikét dichotóm változóként kezeljük, tehát a kérdéses előzmény, illetve következmény vagy jelen van ( $E=1$ , illetve  $K=1$ ), vagy nincs ( $E=0$ , illetve  $K=0$ ) a tanulási helyzetben, illetve vagy jelen van valamilyen egyéb generatív-, illetve preventív oksági tényező ( $G=1$ , illetve  $P=1$ ) vagy nincs ( $G=0$ , illetve  $P=0$ ).

A hipotéziseket ábrázoló oksági Bayes-hálók kvantitatív aspektusa jelenti egyrészt a különböző oksági tényezők (ide értve a kérdéses előzményt is) valószínűségi eloszlását, valamint a következmény feltételes valószínűségi eloszlását. Ami az előbbit illeti, a modell azzal az ad hoc feltevessel él, hogy teljes a bizonytalanság ezek bekövetkezését illetően, ami az előzmény esetében azt jelenti, hogy  $p(E=1)=p(E=0)=0,5$ .

A generatív-, és preventív oksági tényezők esetében azok jelenlétére, pontosabban a jelenlétük valószínűségére vonatkozó bizonytalanságot a modellben a béta-eloszlás segítségével ábrázoltuk. A béta-eloszlást ( $Beta(\alpha, \beta)$ ) gyakran használják erre a célra a Bayesi statisztikában (Gelman, Carlin, Stern, Dunson, Vehtari és Rubin, 2014), mivel értelmezési tartománya a valószínűségeknek megfelel (0 és 1 között). Az eloszlás alakját annak két paramétere ( $\alpha$  és  $\beta$ ) határozza meg. A 6. ábrán megfigyelhető, hogy amikor a két paraméter értéke ugyanakkora, akkor az eloszlás szimmetrikus, és a módusza 0,5-nél, az értelmezési tartomány közepénél található, viszont amikor a két paraméter eltérő, az eloszlás aszimmetrikussá válik (a módusza jobbra, vagy balra tolódik). Ugyancsak érdemes megfigyelni, hogy minél nagyobb a paraméterek értéke, az eloszlás annál „karcsúbbá” válik, míg a paraméterek minimális értéke (1) esetén az egyenletes eloszlásnak felel meg a  $[0,1]$  tartományban. Az eloszlás alakja jelzi számunkra a bizonytalanság mértékének változását: minél karcsúbb az eloszlás, annál kisebb a bizonytalanság.



6. ábra

Példa a béta-eloszlás alakjának (sűrűségfüggvényének) változására a két paraméterének ( $\alpha$  és  $\beta$ ) függvényében

A generatív oksági tényezők jelenlétére vonatkozó bizonytalanság a fentieknek megfelelően  $\pi_G = p(G=1) \sim \text{Beta}(\alpha_G, \beta_G)$ , a preventív oksági tényezők esetében pedig  $\pi_P = p(P=1) \sim \text{Beta}(\alpha_P, \beta_P)$ . A modell kezdetben maximális bizonytalanságot tételez fel a rejtett oksági tényezők jelenlétének valószínűségére vonatkozóan, tehát  $\alpha_G = \beta_G = 1$  a generatív oksági tényezők, és  $\alpha_P = \beta_P = 1$  a preventív oksági tényezők esetében – ahogy említettük, a béta-eloszlások ilyen paraméterezése egyenletes eloszlást eredményez a  $[0,1]$  tartományban.

A következmény feltételes valószínűségi eloszlását megadhatjuk, mint  $p(K|E, G, P, \mathcal{H}_1)$  az oksági kapcsolat meglétére vonatkozó hipotézis ( $\mathcal{H}_1$ ) mellett, illetve mint  $p(K|G, P, \mathcal{H}_0)$  az oksági kapcsolat hiányára vonatkozó hipotézis ( $\mathcal{H}_0$ ) mellett. Ezek meghatározásához determinisztikus oksági kapcsolatokat, és az oksági tényezők közötti interakciókat figyelmen kívül hagyó integrációs függvényt választottunk. A függvény azt az oksági hatásokra vonatkozó intuitív elképzelést képviseli, amely szerint a következmény biztosan bekövetkezik, ha jelen van valamilyen generatív oksági tényező, és ugyanakkor nincs jelen preventív oksági tényező, és biztosan nem következik be, ha nincs jelen generatív

oksági tényező, és/vagy jelen van valamilyen preventív oksági tényező. Mivel a kérdéses előzmény csak az egyik hipotézisben tölt be oksági szerepet, ezért a két hipotézis esetében a függvény némileg eltérő formát ölt (5).

$$\begin{aligned} p(K=1|E, G, P, \mathcal{H}_1) &= [1-(1-E)(1-G)](1-P) \\ p(K=1|G, P, \mathcal{H}_0) &= G(1-P) \end{aligned} \quad (5)$$

A determinisztikus oksági kapcsolatok feltevésének megfelelően a következmény feltételes valószínűsége csak 0, vagy 1 lehet, tehát vagy biztosan bekövetkezik, vagy biztosan nem következik be – ezeket a feltételes valószínűségeket a különböző oksági tényezők lehetséges értékei mellett az 1. táblázatban ábrázoltuk..

1. táblázat

A következmény feltételes valószínűségi eloszlásának alakulása a vizsgált hipotézisek mellett

$E$	$G$	$P$	$p(K=1 E, G, P, \mathcal{H}_1)$	$p(K=1 G, P, \mathcal{H}_0)$
0	0	0	0	0
0	0	1	0	0
0	1	0	1	1
0	1	1	0	0
1	0	0	1	0
1	0	1	0	0
1	1	0	1	1
1	1	1	0	0

#### ***A vizsgált – oksági kapcsolatra vonatkozó – hipotézisek leírása***

*Az egyik hipotézis szerint a kávé éberebbé teszi az embert ( $\mathcal{H}_1$ ), a másik szerint a kávének nincs hatása az éberségre ( $\mathcal{H}_0$ ).*

***előzmény ( E ):*** kávé ivása (ivott kávé:  $E=1$ , nem ivott kávé:  $E=0$ )

***következmény ( K ):*** ébernek érzi-e magát (igen:  $K=1$ , nem:  $K=0$ )

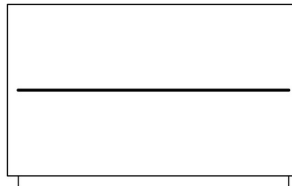
***generatív oksági tényezők ( G ):*** pl. fekete teát ivott, tornázott, stb. (bármelyik igaz volt:  $G=1$ , egyik sem volt igaz:  $G=0$ )

***preventív oksági tényezők ( P ):*** pl. koffeinmentes volt a kávé, gyorsan lebontja a szervezet a koffeint, stb. (bármelyik igaz:  $P=1$ , egyik sem igaz:  $P=0$ )

*folyt.köv.*

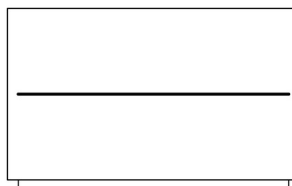
A generatív oksági tényezők jelenlétének valószínűségére vonatkozó bizonytalanságot ábrázoló béta-eloszlás paraméterei és sűrűségfüggvénye:

$$\begin{aligned}\pi_G = p(G=1) &\sim \text{Beta}(\alpha_G, \beta_G) \\ &\sim \text{Beta}(1, 1)\end{aligned}$$



A preventív oksági tényezők jelenlétének valószínűségére vonatkozó bizonytalanságot ábrázoló béta-eloszlás paraméterei és sűrűségfüggvénye:

$$\begin{aligned}\pi_P = p(P=1) &\sim \text{Beta}(\alpha_P, \beta_P) \\ &\sim \text{Beta}(1, 1)\end{aligned}$$



Annak valószínűsége, hogy valaki ivott, illetve nem ivott kávé:

$$p(E=1) = p(E=0) = 0,5$$

Az oksági tényezők hatásának integrálására vonatkozó szabály egy lehetséges megfogalmazása: csak akkor lesz valaki éber, ha iszik/eszik/tesz valamit (kávé, teát iszik, vagy tornázik), ami éberebbé teheti, és annak hatását nem akadályozza meg semmi (pl. nem túlságosan kialvatlan).

## 2.3. Az oksági kapcsolat meglétére vonatkozó becslés lépéseinek részletezése

### 2.3.1. Magyarázatok generálása a megfigyelt esetekre

A ROT-modell, a kovariancia-információt alkotó egyes esetek mindegyikére magyarázatokat generál, amelyekben a rejtett oksági tényezőknek meghatározó szerepe van. Ez a modell azon két központi alapfeltevéséből származik, amelyek szerint az oksági jellegű magyarázatok fontos szerepet töltenek be a tanulás során, illetve hogy ezek a magyarázatok jellemzően tartalmaznak utalást generatív-, valamint preventív oksági tényezőkre.

Azt fel kell ismernünk, hogy a magyarázatok minden egyes eseménytípus esetében többfélék lehetnek, és hogy ezek között vannak olyanok, amelyek valid magyarázatok, és vannak olyanok, amelyek nem valid magyarázatok. Akkor tekinthetünk egy magyarázatot validnak, ha az az adott oksági rendszert (és annak determinisztikus kapcsolatait) feltételezve elképzelhető szituációt ír le. Egy b-típusú esemény elfogadható magyarázata

lehet például, hogy bár jelen volt az előzmény, de valamilyen preventív tényező is jelen volt, ami megakadályozta a következmény bekövetkezését. Ugyanezen eseményre viszont nem valid az a magyarázat, hogy jelen volt még valamilyen egyéb generatív tényező, és nem volt jelen preventív tényező (mivel ez esetben a következménynek be kellett volna következnie).

2. táblázat

A kovariancia-információ egyes eseménytípusaira adott lehetséges magyarázatok

eseménytípus	<i>E</i>	<i>K</i>	<i>G</i>	<i>P</i>	$\mathcal{H}_0$ esetén valid?	$\mathcal{H}_1$ esetén valid?	magyarázat
<i>a</i>							
	1	1	1	1	nem	nem	
	1	1	1	0	igen	igen	$a_1$
	1	1	0	1	nem	nem	
	1	1	0	0	nem	igen	$a_2$
<i>b</i>							
	1	0	1	1	igen	igen	$b_1$
	1	0	1	0	nem	nem	
	1	0	0	1	igen	igen	$b_2$
	1	0	0	0	igen	nem	$b_3$
<i>c</i>							
	0	1	1	1	nem	nem	
	0	1	1	0	igen	igen	$c_1$
	0	1	0	1	nem	nem	
	0	1	0	0	nem	nem	
<i>d</i>							
	0	0	1	1	igen	igen	$d_1$
	0	0	1	0	nem	nem	
	0	0	0	1	igen	igen	$d_2$
	0	0	0	0	igen	igen	$d_3$

A lehetséges magyarázatokat a különböző eseménytípusok esetében a 2. táblázatban foglaltuk össze. A modell által generált magyarázatokat az is befolyásolja, hogy mit feltételezünk a rejtett oksági tényezők jelenlétéről, azaz mennyire tartjuk valószínűnek azok jelenlétét a kovariancia-információt alkotó egyes esetekben. Ezeket a valószínűségeket a

generatív- és preventív oksági tényezők jelenlétére vonatkozó bizonytalanságot ábrázoló béta-eloszlások alapján, azok várható értékével (6) becsüli a modell.

$$\begin{aligned}\Pi_G &= E(\pi_G) = \frac{\alpha_G}{\alpha_G + \beta_G} \\ \Pi_P &= E(\pi_P) = \frac{\alpha_P}{\alpha_P + \beta_P}\end{aligned}\tag{6}$$

A generált valid magyarázatok gyakorisági eloszlását egy adott kovariancia-információ ( $\{N_a, N_b, N_c, N_d\}$ ), és azon belül az egyes eseménytípusok esetében a multinomiális eloszlás segítségével adjuk meg, ahol a magyarázatok egyes típusaihoz rendelt valószínűségeket a két rejtett oksági tényező jelenlétének valószínűségét ábrázoló mutatók ( $\Pi_G$ , és  $\Pi_P$ ) alapján határozzuk meg. Itt minden magyarázattípust mérlegelünk, amely a vizsgált hipotézisek legalább egyike esetében validnak tekinthető. Az a-típusú esemény esetében ez kétféle, a b-típusú és d-típusú események esetében háromféle, a c-típusú esemény esetében pedig egyféle lehet.

Az a-típusú esemény (ahol az előzmény és a következmény is bekövetkezett) esetében az egyik lehetséges magyarázat ( $a_1$ ) szerint jelen van egyéb generatív oksági tényező, és nincs jelen preventív tényező, míg a másik magyarázat ( $a_2$ ) szerint nincs jelen sem egyéb generatív, sem preventív oksági tényező. A magyarázatok gyakorisági megoszlását itt felírhatjuk, mint  $\{N_{a_1}; N_{a_2}\} \sim Multinomial(N_a, \{\pi_{a_1}; \pi_{a_2}\})$ , ahol  $\{\pi_{a_1}; \pi_{a_2}\} \propto \{\Pi_G; 1 - \Pi_G\}$ .

A b-típusú esemény (ahol az előzmény bekövetkezett, de a következmény nem), három valid magyarázat képzelhető el. Az egyik magyarázat ( $b_1$ ) szerint jelen voltak egyéb generatív tényezők, ugyanakkor preventív oksági tényezők is. Egy másik magyarázat ( $b_2$ ) szerint nem volt ugyan jelen egyéb generatív oksági tényező, viszont jelen volt valamilyen preventív oksági tényező. A magyarázatok harmadik típusa ( $b_3$ ) szerint nem volt jelen semmilyen generatív-, vagy preventív oksági tényező (ez csak az előzmény és következmény közötti oksági kapcsolat hiánya esetén számít valid magyarázatnak). A magyarázatok gyakorisági eloszlását felírhatjuk, mint  $\{N_{b_1}; N_{b_2}; N_{b_3}\} \sim Multinomial(N_b, \{\pi_{b_1}; \pi_{b_2}; \pi_{b_3}\})$ , ahol  $\{\pi_{b_1}; \pi_{b_2}; \pi_{b_3}\} \propto \{\Pi_G \Pi_P; (1 - \Pi_G) \Pi_P; (1 - \Pi_G)(1 - \Pi_P)\}$ .

A c-típusú esemény (ahol az előzmény nem következett be, a következmény bekövetkezett), csak egyféleképpen képzelhető el ( $c_1$ ): amikor jelen volt valamilyen egyéb generatív oksági tényező és nem volt jelen preventív tényező. Ebben az esetben a generált

magyarázatok száma az eseménytípus megfigyelt gyakoriságával egyezik meg, tehát  $N_{c_i} = N_c$ .

A d-típusú eseményre (ahol az előzmény és a következmény sem következett be) a b-típusú eseményhez hasonlóan három valid magyarázat létezik ( $d_1, d_2, d_3$ ), amelyek a rejtett oksági tényezők jelenlétét illetően megegyeznek a b-típusú események esetében alkalmazott magyarázatokkal. A magyarázatok gyakorisági eloszlása a b-típusú eseményhez hasonlóan megadható mint  $\{N_{d_1}; N_{d_2}; N_{d_3}\} \sim \text{Multinomial}(N_d, \{\pi_{d_1}; \pi_{d_2}; \pi_{d_3}\})$ , ahol  $\{\pi_{d_1}; \pi_{d_2}; \pi_{d_3}\} \propto \{\Pi_G \Pi_P; (1 - \Pi_G) \Pi_P; (1 - \Pi_G)(1 - \Pi_P)\}$ .

Az egyes magyarázattípusok gyakoriságának a megfelelő multinomiális eloszlás alapján várható értéket tekinti a modell  $N_{x_i} = \pi_{x_i} N_x$ , ahol  $x$  a megfelelő eseménytípust (a, b vagy d),  $x_i$  pedig a megfelelő magyarázattípust jelöli (a fentiek alapján a c-típusú események esetén  $N_{c_i} = N_c$ ).

### **Magyarázatok generálása a megfigyelt esetekre**

*Lehetséges magyarázatok az egyes eseménytípusok esetén:*

**a-típusú** események (ivott kávé és ébernek érezte magát):

- $a_1$  : más is éberre tehetette (pl. fekete teát ivott) és nem volt semmi akadályozó tényező
- $a_2$  : nem volt semmi akadályozó tényező jelen, ezért kifejtette a kávé a hatását

**b-típusú** események (ivott kávé, nem érezte magát ébernek):

- $b_1$  : pl. túlságosan kialvatlan volt az illető, ezért semmi nem tehetette éberre
- $b_2$  : pl. túlságosan kialvatlan volt az illető ezért nem hatott már a kávé
- $b_3$  : ugyan ivott kávé, de mivel annak nincs hatása az éberségre, ezért nem is lett éber

**c-típusú** események (nem ivott kávé, ébernek érezte magát):

- $c_1$  : pl. az illető épp a kérdezősködésünk előtt tornázott, és ezért volt teljesen éber

**d-típusú** események (nem ivott kávé, nem volt éber):

- $d_1$  : pl. ivott ugyan fekete teát, de nagyon gyorsan lebontja a szervezete a koffeint
- $d_2$  : nem ivott/csinált semmit, amitől felébredt volna (ugyanakkor nagyon kialvatlan volt)
- $d_3$  : nem ivott/csinált semmit, amitől felébredt volna

*A rejtett oksági tényezők jelenlétének valószínűségére vonatkozó becslés, a megfelelő béta-eloszlások paraméterei alapján:*

$$\Pi_G = \frac{\alpha_G}{\alpha_G + \beta_G} = \frac{1}{1+1} = 0,5$$

$$\Pi_P = \frac{\alpha_P}{\alpha_P + \beta_P} = \frac{1}{1+1} = 0,5$$

*folyt.köv.*

A magyarázatok gyakorisági eloszlását a multinomiális eloszlás alapján becsüljük, mint  $N_{x_i} = \pi_{x_i} N_x$ , ahol  $N_x$  a kérdéses eseménytípus gyakorisága,  $N_{x_i}$  egy bizonyos magyarázattípus gyakorisága,  $\pi_{x_i}$  pedig ehhez a magyarázattípushoz tartozó valószínűség a multinomiális eloszláson belül.

Magyarázatok gyakorisági eloszlása az **a-típusú** események esetén:

$$\{N_{a_1}; N_{a_2}\} \sim \text{Multinomial}(N_a, \{\pi_{a_1}; \pi_{a_2}\}), \text{ ahol } \{\pi_{a_1}; \pi_{a_2}\} \propto \{\Pi_G; 1 - \Pi_G\}$$

$$\pi_{a_1} = \frac{\Pi_G}{\Pi_G + (1 - \Pi_G)} = \frac{0,5}{0,5 + 0,5} = 0,5$$

$$\pi_{a_2} = \frac{1 - \Pi_G}{\Pi_G + (1 - \Pi_G)} = \frac{0,5}{0,5 + 0,5} = 0,5$$

$$N_{a_1} = N_a \pi_{a_1} = 12 \cdot 0,5 = 6$$

$$N_{a_2} = N_a \pi_{a_2} = 12 \cdot 0,5 = 6$$

Magyarázatok gyakorisági eloszlása a **b-típusú** események esetén:

$$\{N_{b_1}; N_{b_2}; N_{b_3}\} \sim \text{Multinomial}(N_b, \{\pi_{b_1}; \pi_{b_2}; \pi_{b_3}\}), \text{ ahol } \{\pi_{b_1}; \pi_{b_2}; \pi_{b_3}\} \propto \{\Pi_G \Pi_p; (1 - \Pi_G) \Pi_p; (1 - \Pi_G)(1 - \Pi_p)\}$$

$$\pi_{b_1} = \frac{\Pi_G \Pi_p}{\Pi_G \Pi_p + (1 - \Pi_G) \Pi_p + (1 - \Pi_G)(1 - \Pi_p)} = \frac{0,25}{0,25 + 0,25 + 0,25} = \frac{1}{3}$$

$$\pi_{b_2} = \frac{(1 - \Pi_G) \Pi_p}{\Pi_G \Pi_p + (1 - \Pi_G) \Pi_p + (1 - \Pi_G)(1 - \Pi_p)} = \frac{0,25}{0,25 + 0,25 + 0,25} = \frac{1}{3}$$

$$\pi_{b_3} = \frac{(1 - \Pi_G)(1 - \Pi_p)}{\Pi_G \Pi_p + (1 - \Pi_G) \Pi_p + (1 - \Pi_G)(1 - \Pi_p)} = \frac{0,25}{0,25 + 0,25 + 0,25} = \frac{1}{3}$$

$$N_{b_1} = N_b \pi_{b_1} = 6 \cdot \frac{1}{3} = 2$$

$$N_{b_2} = N_b \pi_{b_2} = 6 \cdot \frac{1}{3} = 2$$

$$N_{b_3} = N_b \pi_{b_3} = 6 \cdot \frac{1}{3} = 2$$

Magyarázatok gyakorisági eloszlása a **c-típusú** események esetén:

$$N_{c_1} = N_c = 6$$

Magyarázatok gyakorisági eloszlása a **d-típusú** események esetén:

$$\{N_{d_1}; N_{d_2}; N_{d_3}\} \sim \text{Multinomial}(N_d, \{\pi_{d_1}; \pi_{d_2}; \pi_{d_3}\}), \text{ ahol } \{\pi_{d_1}; \pi_{d_2}; \pi_{d_3}\} \propto \{\Pi_G \Pi_p; (1 - \Pi_G) \Pi_p; (1 - \Pi_G)(1 - \Pi_p)\}$$

$$\pi_{d_1} = \frac{\Pi_G \Pi_p}{\Pi_G \Pi_p + (1 - \Pi_G) \Pi_p + (1 - \Pi_G)(1 - \Pi_p)} = \frac{0,25}{0,25 + 0,25 + 0,25} = \frac{1}{3}$$

$$\pi_{d_2} = \frac{(1 - \Pi_G) \Pi_p}{\Pi_G \Pi_p + (1 - \Pi_G) \Pi_p + (1 - \Pi_G)(1 - \Pi_p)} = \frac{0,25}{0,25 + 0,25 + 0,25} = \frac{1}{3}$$

$$\pi_{d_3} = \frac{(1 - \Pi_G)(1 - \Pi_p)}{\Pi_G \Pi_p + (1 - \Pi_G) \Pi_p + (1 - \Pi_G)(1 - \Pi_p)} = \frac{0,25}{0,25 + 0,25 + 0,25} = \frac{1}{3}$$

$$N_{d_1} = N_d \pi_{d_1} = 12 \cdot \frac{1}{3} = 4$$

$$N_{d_2} = N_d \pi_{d_2} = 12 \cdot \frac{1}{3} = 4$$

$$N_{d_3} = N_d \pi_{d_3} = 12 \cdot \frac{1}{3} = 4$$

### 2.3.2. A rejtett oksági tényezők jelenlétére vonatkozó bizonytalanság felülvizsgálata

A ROT-modell azokat a magyarázatokat, amelyeket a korábbi szakaszban a kovariancia-információt alkotó esetekre generált, a rejtett oksági tényezők jelenlétére, illetve hiányára vonatkozó bizonyítékként kezeli a továbbiak során. Például egy  $a_2$  típusú magyarázatot, amely szerint nincs jelen sem egyéb generatív, sem preventív tényező, ezen oksági tényezők hiányára vonatkozó bizonyítékként értékeli. Hasonlóképpen egy  $c_1$  típusú magyarázatot a modell mint preventív oksági tényezők hiányára, és egyéb generatív oksági tényezők jelenlétére vonatkozó bizonyítékot értékeli.

A magyarázatok ilyen kiértékelését követően a ROT-modell összegzi ezeket a bizonyítékokat. Az összegzés során meghatározza a modell, hogy mennyi magyarázat utal generatív oksági tényező jelenlétére ( $G^+ = N_{a_1} + N_{b_1} + N_c + N_{d_1}$ ), illetve hiányára ( $G^- = N_{a_2} + N_{b_2} + N_{b_3} + N_{d_2} + N_{d_3}$ ), továbbá hogy mennyi magyarázat szól valamilyen preventív oksági tényező jelenléte ( $P^+ = N_{b_1} + N_{b_2} + N_{d_1} + N_{d_2}$ ), illetve hiánya ( $P^- = N_{a_1} + N_{a_2} + N_{b_3} + N_c + N_{d_3}$ ) mellett.

A magyarázatok/bizonyítékok ilyen összegzését követően, azt felhasználva, a modell felülvizsgálja a különböző oksági tényezők jelenlétének vaslószerűségére vonatkozó bizonytalanságot. Ez az egyes rejtett oksági tényezőkhöz tartozó béta-eloszlások felülvizsgálatát jelenti, ami a Gelman és mtai. (2014) által javasolt módszer szerint történik. A módszer lényege, hogy az eloszlás  $\alpha$  paraméteréhez hozzáadjuk az adott oksági tényező jelenlétére utaló bizonyítékok mennyiségét ( $G^+$ , illetve  $P^+$ ), a második,  $\beta$  paraméteréhez pedig az adott oksági tényező hiányára utaló bizonyítékok mennyiségét ( $G^-$ , illetve  $P^-$ ). Ennek megfelelően a generatív oksági tényezők esetében a felülvizsgált béta eloszlás:  $\pi'_G \sim \text{Beta}(\alpha_G + G^+, \beta_G + G^-)$ , míg a preventív oksági tényezők esetében:  $\pi'_P \sim \text{Beta}(\alpha_P + P^+, \beta_P + P^-)$ .

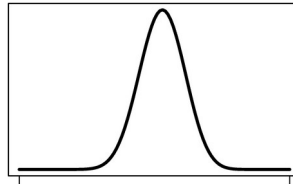
### *sA rejtett oksági tényezők jelenlétére vonatkozó bizonytalanság felülvizsgálata*

*Az előző szakasz eredményeként rendelkezésre áll az egyes esetekre generált magyarázatok gyakorisági eloszlása a különböző esemény-típusok esetében:*

$$\begin{aligned} \{N_{a_1}; N_{a_2}\} &= \{6; 6\} \\ \{N_{b_1}; N_{b_2}; N_{b_3}\} &= \{2; 2; 2\} \\ N_{c_1} &= N_{c_2} = 6 \\ \{N_{d_1}; N_{d_2}; N_{d_3}\} &= \{4; 4; 4\} \end{aligned}$$

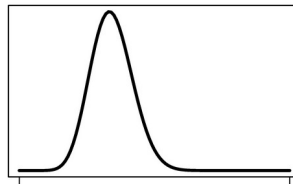
*A rejtett generatív oksági tényezők jelenlétére/hiányára utaló bizonyítékok összegzése, és azok jelenlétére vonatkozó bizonytalanság felülvizsgálata (a felülvizsgált béta-eloszlás sűrűségfüggvényével):*

$$\begin{aligned} G^+ &= N_{a_1} + N_{b_1} + N_{c_1} + N_{d_1} = 6 + 2 + 6 + 4 = 18 \\ G^- &= N_{a_2} + N_{b_2} + N_{b_3} + N_{d_2} + N_{d_3} = 6 + 2 + 2 + 4 + 4 = 18 \\ \pi'_G &\sim \text{Beta}(\alpha_G + G^+, \beta_G + G^-) \\ &\sim \text{Beta}(1 + 18, 1 + 18) \\ &\sim \text{Beta}(19, 19) \end{aligned}$$



*A rejtett preventív oksági tényezők jelenlétére/hiányára vonatkozó bizonyítékok összegzése, és azok jelenlétére vonatkozó bizonytalanság felülvizsgálata (a felülvizsgált béta-eloszlás sűrűségfüggvényével):*

$$\begin{aligned} P^+ &= N_{b_1} + N_{b_2} + N_{d_1} + N_{d_2} = 2 + 2 + 4 + 4 = 12 \\ P^- &= N_{a_1} + N_{a_2} + N_{b_3} + N_{c_1} + N_{d_3} = 6 + 6 + 2 + 6 + 4 = 24 \\ \pi'_P &\sim \text{Beta}(\alpha_P + P^+, \beta_P + P^-) \\ &\sim \text{Beta}(1 + 12, 1 + 24) \\ &\sim \text{Beta}(13, 24) \end{aligned}$$



### **2.3.3. Az oksági kapcsolatra vonatkozó hipotézisek felülvizsgálata (az oksági support kiszámítása)**

Utolsó lépésként a ROT-modell, Griffiths és Tenenbaum (2005) support-modelljéhez hasonlóan, kiszámítja az oksági supportot, amely lényegében a Bayesi statisztikában alkalmazott Bayes-faktornak felel meg (Kass és Raftery, 1995). A Bayes-faktort Jeffreys (1935, 1961) dolgozta ki annak a jellegzetes problémának a megoldására, amikor vannak versengő

elméleteink, hipotéziseink a megfigyeléseink magyarázatára, és szeretnénk számszerűen kifejezni azt, hogy az adatok mennyire erős bizonyítékként szolgálnak az egyik hipotézis javára, a másikkal szemben. Ez a kérdéses hipotézisek posterior valószínűségének hányadosát jelenti (7), amelynek – a könnyebb értelmezés érdekében – a logaritmusát szokták tekinteni.

$$\log \frac{p(\mathcal{H}_1|\mathcal{D})}{p(\mathcal{H}_0|\mathcal{D})} = \log \frac{p(\mathcal{D}|\mathcal{H}_1)}{p(\mathcal{D}|\mathcal{H}_0)} + \log \frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)} \quad (7)$$

Ennek megfelelően a pozitív Bayes-faktor arra utal, hogy az adatok jobban támogatják  $\mathcal{H}_1$  hipotézist, mint  $\mathcal{H}_0$  hipotézist, és vice versa a negatív Bayes-faktor esetén. A Bayes-faktor 0 értéke arra utal, hogy az adatok mindkét hipotézist hasonló mértékben támogatják. A 3. táblázat néhány példán keresztül szemlélteti a két hipotézis posterior valószínűségének hányadosa, és a Bayes-faktor közötti viszonyt.

3. táblázat

Néhány példa a Bayes-faktor értelmezésének szemléltetésére

$\frac{p(\mathcal{H}_1 \mathcal{D})}{p(\mathcal{H}_0 \mathcal{D})}$	Bayes-faktor	a hányados szöveges értelmezése
1	0	$\mathcal{H}_1$ ugyanolyan valószínű, mint $\mathcal{H}_0$
2	0,69	$\mathcal{H}_1$ kétszer annyira valószínű, mint $\mathcal{H}_0$
10	2,3	$\mathcal{H}_1$ tízszer annyira valószínű, mint $\mathcal{H}_0$
100	4,6	$\mathcal{H}_1$ százszor annyira valószínű, mint $\mathcal{H}_0$
0,5	-0,69	$\mathcal{H}_0$ kétszer annyira valószínű, mint $\mathcal{H}_1$
0,1	-2,3	$\mathcal{H}_0$ tízszer annyira valószínű, mint $\mathcal{H}_1$
0,01	-4,61	$\mathcal{H}_0$ százszor annyira valószínű, mint $\mathcal{H}_1$

A ROT-modell a fentieknek megfelelően az oksági kapcsolat meglétére vonatkozó hipotézis annak hiányára vonatkozó hipotézissel szembeni támogatottságát fejezi ki, azzal az ad hoc feltevéssel, hogy ezen hipotézisek előzetes támogatottsága (előzetes valószínűsége) azonos, tehát  $p(\mathcal{H}_1)=p(\mathcal{H}_0)=0,5$ . Ebben a speciális esetben az előzetes valószínűségek hányadosának logaritmus 0, és így a modell becslése felírható egyszerűbb formában (8).

$$\log \frac{p(\mathcal{D}|\mathcal{H}_1)}{p(\mathcal{D}|\mathcal{H}_0)} \quad (8)$$

A ROT-modell becslése tehát lényegében azt ábrázolja, hogy a rendelkezésre álló bizonyítékok milyen erős bizonyítékot jelentenek a kérdéses oksági kapcsolat megléte mellett, az oksági kapcsolat hiányával szemben.

A megfigyelt kovariancia-információ (  $\mathcal{D}$  ) valószínűségét az egyes hipotézisek mellett (9) módon számíthatjuk ki (itt azt feltételezzük, hogy a kovariancia-információt alkotó egyes esetek egymástól függetlenek).

$$p(\mathcal{D}|\mathcal{H}) = \binom{N_a + N_b}{N_a} \binom{N_c + N_d}{N_c} \prod_K \prod_E p(K, E|\mathcal{H})^{n(K, E)}$$

,ahol

$$n(K, E) = \begin{cases} N_a, & \text{ha } K=1 \text{ és } E=1 \\ N_b, & \text{ha } K=0 \text{ és } E=1 \\ N_c, & \text{ha } K=1 \text{ és } E=0 \\ N_d, & \text{ha } K=0 \text{ és } E=0 \end{cases} \quad (9)$$

A teljes kovariancia-információ a négyféle eseménytípusból tevődik össze – az egyes eseménytípusok valószínűségét, tehát az előzmény és a következmény együttes valószínűségi eloszlását (  $p(K, E|\mathcal{H})$  ), az oksági Bayes-hálóok láncszabályának felhasználásával határozzuk meg, amely során kiintegráljuk a rejtett oksági tényezőket (tekintve, hogy azok jelenlétéről nincs közvetlen információnk a tanulási helyzetben), valamint a rejtett oksági tényezők jelenlétének valószínűségére vonatkozó bizonytalanságot, amelyeket a korábbi lépésben felülvizsgált béta eloszlások reprezentálnak. Ez (10) módon történik az oksági kapcsolat meglétére vonatkozó hipotézis (  $\mathcal{H}_1$  ), és (11) módon az oksági kapcsolat hiányára vonatkozó hipotézis (  $\mathcal{H}_0$  ) esetében. Mindkét esetben  $\pi'_G$  a generatív oksági tényezők,  $\pi'_P$  pedig a preventív oksági tényezők jelenlétének valószínűségére vonatkozó felülvizsgált béta-eloszlásokat jelöli.

$$p(K, E|\mathcal{H}_1) = \int_0^1 \int_0^1 \sum_G \sum_P p(K|E, G, P) p(E) \pi'_G \pi'_P d\pi'_G d\pi'_P \quad (10)$$

$$p(K, E|\mathcal{H}_0) = \int_0^1 \int_0^1 \sum_G \sum_P p(K|G, P) p(E) \pi'_G \pi'_{Pd} \pi'_G d\pi'_P \quad (11)$$

**Az oksági kapcsolatra vonatkozó hipotézisek felülvizsgálata és az oksági support kiszámítása**

A kávéivás és az éberség közötti oksági kapcsolat meglétére vonatkozó hipotézis ( $\mathcal{H}_1$ ), valamint az oksági kapcsolat hiányára vonatkozó hipotézis ( $\mathcal{H}_0$ ) előzetes valószínűsége:

$$p(\mathcal{H}_1) = p(\mathcal{H}_0) = 0,5$$

**a-típusú** esemény valószínűsége  $\mathcal{H}_1$  esetén:

$$\begin{aligned} p(K=1, E=1|\mathcal{H}_1) &= \int_0^1 \sum_G \sum_P p(K=1|E=1, G, P) p(E=1) \pi'_G \pi'_P d\pi'_G d\pi'_P \\ &= \int_0^1 p(K=1|E=1, G=0, P=0) p(E=1) (1-\pi'_G) (1-\pi'_P) + \\ &\quad p(K=1|E=1, G=0, P=1) p(E=1) (1-\pi'_G) \pi'_P + \\ &\quad p(K=1|E=1, G=1, P=0) p(E=1) \pi'_G (1-\pi'_P) + \\ &\quad p(K=1|E=1, G=1, P=1) p(E=1) \pi'_G \pi'_P d\pi'_G d\pi'_P \\ &= \int_0^1 1 \cdot 0,5 (1-\pi'_G) (1-\pi'_P) + \\ &\quad 0 \cdot 0,5 (1-\pi'_G) \pi'_P + \\ &\quad 1 \cdot 0,5 \pi'_G (1-\pi'_P) + \\ &\quad 0 \cdot 0,5 \pi'_G \pi'_P d\pi'_G d\pi'_P \\ &= \int_0^1 0,5 (1-\pi'_G) (1-\pi'_P) + 0,5 \pi'_G (1-\pi'_P) d\pi'_G d\pi'_P \\ &= \int_0^1 0,5 (1-\pi'_P) d\pi'_P \approx 0,33 \end{aligned}$$

**b-típusú** esemény valószínűsége  $\mathcal{H}_1$  esetén:

$$\begin{aligned} p(K=0, E=1|\mathcal{H}_1) &= \int_0^1 \sum_G \sum_P [1-p(K=1|E=1, G, P)] p(E=1) \pi'_G \pi'_P d\pi'_G d\pi'_P \\ &= \int_0^1 [1-p(K=1|E=1, G=0, P=0)] p(E=1) (1-\pi'_G) (1-\pi'_P) + \\ &\quad [1-p(K=1|E=1, G=0, P=1)] p(E=1) (1-\pi'_G) \pi'_P + \\ &\quad [1-p(K=1|E=1, G=1, P=0)] p(E=1) \pi'_G (1-\pi'_P) + \\ &\quad [1-p(K=1|E=1, G=1, P=1)] p(E=1) \pi'_G \pi'_P d\pi'_G d\pi'_P \\ &= \int_0^1 0 \cdot 0,5 (1-\pi'_G) (1-\pi'_P) + \\ &\quad 1 \cdot 0,5 (1-\pi'_G) \pi'_P + \\ &\quad 0 \cdot 0,5 \pi'_G (1-\pi'_P) + \\ &\quad 1 \cdot 0,5 \pi'_G \pi'_P d\pi'_G d\pi'_P \\ &= \int_0^1 0,5 (1-\pi'_G) \pi'_P + 0,5 \pi'_G \pi'_P d\pi'_G d\pi'_P \\ &= \int_0^1 0,5 \pi'_P d\pi'_P \approx 0,17 \end{aligned}$$

folyt.köv.

**c-típusú** esemény valószínűsége  $\mathcal{H}_1$  esetén:

$$\begin{aligned}
 p(K=1, E=0|\mathcal{H}_1) &= \int_0^1 \sum_G \sum_P p(K=1|E=0, G, P) p(E=0) \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 p(K=1|E=0, G=0, P=0) p(E=0) (1-\pi'_G) (1-\pi'_P) + \\
 &\quad p(K=1|E=0, G=0, P=1) p(E=0) (1-\pi'_G) \pi'_P + \\
 &\quad p(K=1|E=0, G=1, P=0) p(E=0) \pi'_G (1-\pi'_P) + \\
 &\quad p(K=1|E=0, G=1, P=1) p(E=0) \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 0 \cdot 0,5 (1-\pi'_G) (1-\pi'_P) + \\
 &\quad 0 \cdot 0,5 (1-\pi'_G) \pi'_P + \\
 &\quad 1 \cdot 0,5 \pi'_G (1-\pi'_P) + \\
 &\quad 0 \cdot 0,5 \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 0,5 \pi'_G (1-\pi'_P) d\pi'_G d\pi'_P \approx 0,16
 \end{aligned}$$

**d-típusú** esemény valószínűsége  $\mathcal{H}_1$  esetén:

$$\begin{aligned}
 p(K=0, E=0|\mathcal{H}_1) &= \int_0^1 \sum_G \sum_P [1-p(K=1|E=0, G, P)] p(E=0) \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 [1-p(K=1|E=0, G=0, P=0)] p(E=0) (1-\pi'_G) (1-\pi'_P) + \\
 &\quad [1-p(K=1|E=0, G=0, P=1)] p(E=0) (1-\pi'_G) \pi'_P + \\
 &\quad [1-p(K=1|E=0, G=1, P=0)] p(E=0) \pi'_G (1-\pi'_P) + \\
 &\quad [1-p(K=1|E=0, G=1, P=1)] p(E=0) \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 1 \cdot 0,5 (1-\pi'_G) (1-\pi'_P) + \\
 &\quad 1 \cdot 0,5 (1-\pi'_G) \pi'_P + \\
 &\quad 0 \cdot 0,5 \pi'_G (1-\pi'_P) + \\
 &\quad 1 \cdot 0,5 \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 0,5 (1-\pi'_G) (1-\pi'_P) + 0,5 (1-\pi'_G) \pi'_P + 0,5 \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 0,5 [(1-\pi'_G) + \pi'_G \pi'_P] d\pi'_G d\pi'_P \approx 0,34
 \end{aligned}$$

**Teljes kovariancia-információ** valószínűsége  $\mathcal{H}_1$  esetén:

$$\begin{aligned}
 p(\mathcal{D}|\mathcal{H}_1) &= \binom{N_a+N_b}{N_a} \binom{N_c+N_d}{N_c} p(K=1, E=1|\mathcal{H}_1)^{N_a} p(K=0, E=1|\mathcal{H}_1)^{N_b} \cdot \\
 &\quad p(K=1, E=0|\mathcal{H}_1)^{N_c} p(K=0, E=0|\mathcal{H}_1)^{N_d} \\
 &= \binom{18}{6} \binom{18}{6} 0,33^{12} \cdot 0,17^6 \cdot 0,16^6 \cdot 0,34^{12} \\
 &\approx 5,58 \cdot 10^{-13}
 \end{aligned}$$

*folyt.köv.*

**a-típusú** esemény valószínűsége  $\mathcal{H}_0$  esetén:

$$\begin{aligned}
 p(K=1, E=1|\mathcal{H}_0) &= \int_0^1 \int_0^1 \sum_G \sum_P p(K=1|G, P) p(E=1) \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 \int_0^1 p(K=1|G=0, P=0) p(E=1) (1-\pi'_G) (1-\pi'_P) + \\
 &\quad p(K=1|G=0, P=1) p(E=1) (1-\pi'_G) \pi'_P + \\
 &\quad p(K=1|G=1, P=0) p(E=1) \pi'_G (1-\pi'_P) + \\
 &\quad p(K=1|G=1, P=1) p(E=1) \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 \int_0^1 0,5 \cdot (1-\pi'_G) (1-\pi'_P) + \\
 &\quad 0,5 \cdot (1-\pi'_G) \pi'_P + \\
 &\quad 1,0,5 \cdot \pi'_G (1-\pi'_P) + \\
 &\quad 0,0,5 \cdot \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 \int_0^1 0,5 \pi'_G (1-\pi'_P) d\pi'_G d\pi'_P \approx 0,16
 \end{aligned}$$

**b-típusú** esemény valószínűsége  $\mathcal{H}_0$  esetén:

$$\begin{aligned}
 p(K=0, E=1|\mathcal{H}_0) &= \int_0^1 \int_0^1 \sum_G \sum_P [1-p(K=1|G, P)] p(E=1) \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 \int_0^1 [1-p(K=1|G=0, P=0)] p(E=1) (1-\pi'_G) (1-\pi'_P) + \\
 &\quad [1-p(K=1|G=0, P=1)] p(E=1) (1-\pi'_G) \pi'_P + \\
 &\quad [1-p(K=1|G=1, P=0)] p(E=1) \pi'_G (1-\pi'_P) + \\
 &\quad [1-p(K=1|G=1, P=1)] p(E=1) \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 \int_0^1 1,0,5 \cdot (1-\pi'_G) (1-\pi'_P) + \\
 &\quad 1,0,5 \cdot (1-\pi'_G) \pi'_P + \\
 &\quad 0,0,5 \cdot \pi'_G (1-\pi'_P) + \\
 &\quad 1,0,5 \cdot \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 \int_0^1 0,5 [(1-\pi'_G) + \pi'_G \pi'_P] d\pi'_G d\pi'_P \\
 &\approx 0,34
 \end{aligned}$$

**c-típusú** esemény valószínűsége  $\mathcal{H}_0$  esetén:

$$\begin{aligned}
 p(K=1, E=0|\mathcal{H}_0) &= \int_0^1 \int_0^1 \sum_G \sum_P p(K=1|G, P) p(E=0) \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 \int_0^1 p(K=1|G=0, P=0) p(E=0) (1-\pi'_G) (1-\pi'_P) + \\
 &\quad p(K=1|G=0, P=1) p(E=0) (1-\pi'_G) \pi'_P + \\
 &\quad p(K=1|G=1, P=0) p(E=0) \pi'_G (1-\pi'_P) + \\
 &\quad p(K=1|G=1, P=1) p(E=0) \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 \int_0^1 0,0,5 \cdot (1-\pi'_G) (1-\pi'_P) + \\
 &\quad 0,0,5 \cdot (1-\pi'_G) \pi'_P + \\
 &\quad 1,0,5 \cdot \pi'_G (1-\pi'_P) + \\
 &\quad 0,0,5 \cdot \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 \int_0^1 0,5 \pi'_G (1-\pi'_P) d\pi'_G d\pi'_P \approx 0,16
 \end{aligned}$$

folyt.köv.

**d-típusú** esemény valószínűsége  $\mathcal{H}_0$  esetén:

$$\begin{aligned}
 p(K=0, E=0|\mathcal{H}_0) &= \int_0^1 \int_0^1 \sum_G \sum_P [1-p(K=1|G, P)] p(E=0) \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 [1-p(K=1|G=0, P=0)] p(E=0) (1-\pi'_G) (1-\pi'_P) + \\
 &\quad [1-p(K=1|G=0, P=1)] p(E=0) (1-\pi'_G) \pi'_P + \\
 &\quad [1-p(K=1|G=1, P=0)] p(E=0) \pi'_G (1-\pi'_P) + \\
 &\quad [1-p(K=1|G=1, P=1)] p(E=0) \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 1 \cdot 0,5 \cdot (1-\pi'_G) (1-\pi'_P) + \\
 &\quad 1 \cdot 0,5 \cdot (1-\pi'_G) \pi'_P + \\
 &\quad 0 \cdot 0,5 \cdot \pi'_G (1-\pi'_P) + \\
 &\quad 1 \cdot 0,5 \cdot \pi'_G \pi'_P d\pi'_G d\pi'_P \\
 &= \int_0^1 \int_0^1 0,5 [(1-\pi'_G) + \pi'_G \pi'_P] d\pi'_G d\pi'_P \\
 &\approx 0,34
 \end{aligned}$$

Teljes kovariancia-információ valószínűsége  $\mathcal{H}_0$  esetén:

$$\begin{aligned}
 p(\mathcal{D}|\mathcal{H}_0) &= \binom{N_a+N_b}{N_a} \binom{N_c+N_d}{N_c} p(K=1, E=1|\mathcal{H}_0)^{N_b} p(K=0, E=1|\mathcal{H}_0)^{N_d} \cdot \\
 &\quad p(K=1, E=0|\mathcal{H}_0)^{N_c} p(K=0, E=0|\mathcal{H}_0)^{N_a} \\
 &= \binom{18}{6} \binom{18}{6} 0,16^{12} \cdot 0,34^6 \cdot 0,16^6 \cdot 0,34^{12} \\
 &\approx 7,77 \cdot 10^{-15}
 \end{aligned}$$

A Bayes-faktor kiszámítása (a két hipotézis előzetes valószínűsége  $p(\mathcal{H}_1)=p(\mathcal{H}_0)=0,5$ ):

$$\begin{aligned}
 \log \frac{p(\mathcal{H}_1|\mathcal{D})}{p(\mathcal{H}_0|\mathcal{D})} &= \log \frac{p(\mathcal{D}|\mathcal{H}_1)}{p(\mathcal{D}|\mathcal{H}_0)} + \log \frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)} \\
 &= \log \frac{p(\mathcal{D}|\mathcal{H}_1)}{p(\mathcal{D}|\mathcal{H}_0)} + \log \frac{0,5}{0,5} \\
 &= \log \frac{p(\mathcal{D}|\mathcal{H}_1)}{p(\mathcal{D}|\mathcal{H}_0)} \\
 &\approx \log \frac{5,58 \cdot 10^{-13}}{7,77 \cdot 10^{-15}} = \log 71,91 = 4,28
 \end{aligned}$$

A ROT-modell fentebb levezetett becslése szerint a rendelkezésre álló kovariancia-információ inkább a kávéivás és éberség közötti oksági kapcsolat **meglétére** utal, szemben az oksági kapcsolat hiányával – az oksági kapcsolat meglétére vonatkozó hipotézis posterior valószínűsége kb. 71,91-szer akkora, mint az oksági kapcsolat hiányára vonatkozó hipotézisé.

### 3. A ROT-modell illeszkedésének vizsgálata

A dolgozat további részében bemutatjuk azokat a vizsgálatokat és elemzéseket, amelyeket a ROT-modell humán adatokra való illeszkedésének ellenőrzése céljából végeztünk. A célunk az volt, hogy a modell becsléseit emberektől származó, egyszerű oksági kapcsolat meglétére, illetve hiányára vonatkozó ítéletekkel hasonlítsuk össze. Ehhez két forrásból gyűjtöttünk adatokat: egyrészt a szakirodalomban kerestünk olyan vizsgálatokat, amelyek alkalmasak lehetnek a ROT-modell becsléseinek ellenőrzésére, másrészt több kísérletet is lefolytattunk, amelyek a célnak megfelelő adatokat szolgáltatottak.

A kísérletek egy részében a résztvevőknek összegzett kovariancia-információ állt rendelkezésre az oksági ítélet megalkotásához (1-3. kísérlet). Szerettük volna kipróbálni azt is, hogy mennyire terjeszthető ki a modell magyarázóereje olyan tanulási helyzetre, amelynek kezelésére a modellépítés során közvetlenül nem készítettük fel, ezért egy további kísérletet is végeztünk, amelyben a résztvevőknek online tanulási helyzetben (ahol tehát a kovariancia-információ egyesével érkezik) kellett oksági ítéleteket alkotniuk (4. kísérlet).

A kísérleteink mindegyikét interneten keresztül végeztük, a résztvevőket crowdsourcing szolgáltatás igénybevételével toboroztuk – a minta gyűjtésének hatékonyságát szem előtt tartva olyan országokban, ahol az elsődleges hivatalos nyelv az angol. A mintavétel ezen módjával kapcsolatban egyre több biztató tapasztalatunk van annak alkalmazhatóságát illetően, és számos, a kutatási gyakorlatban is jól használható javaslatot is találhatunk (pl. Reips, 2000; Skitka és Sargis, 2006; Behrend, Sharek, Meade és Wiebe, 2011; Stewart, Chandler és Paolacci, 2017). A kísérletek során kifejezetten törekedtünk arra, hogy az adatok megfelelő minősége biztosított legyen, ennek érdekében többféle módszerrel igyekeztünk monitorozni a résztvevők figyelmét, illetve részvételi motivációjának megfelelőségét (őszinteségét), valamint olyan kérdéseket is igyekeztünk alkalmazni, amelyek bizonyos kritikus készségek (pl. nyelvi készségek) mentén való szűrést is lehetővé tettek. A részvételért járó anyagi jutalmakat kifejezetten alacsony szinten határoztuk meg, amivel azt szerettük volna elérni, hogy a részvételi szándék mögött jelentéktelen szerepet töltsön be az anyagi motiváció, és nagyobbat a puszta érdeklődés, kíváncsiság. A kísérlet során az instrukciókat és a feladatokat a lehető legrövidebbre terveztük, hogy megelőzzük a

motivációvesztést, amely az online vizsgálatok esetében reális veszély lehet. Emiatt a kísérleteinkben a személyek közötti elrendezést részesítettük előnyben.

A ROT-modell illeszkedésének megítélésében fontosnak tartottuk, hogy annak megfelelőségét ne csak önmagában próbáljuk megítélni, ezért összevetettük, mintegy megversenyeztettük, a szakirodalomból ismert másik két modellel, amelyek ugyanolyan információ alapján (összegzett kovariancia-információ), ugyanolyan becsléseket (oksági kapcsolat meglétében való bizonyosság) tesznek.

A következőkben ismertetjük azokat az alternatív modelleket, amelyek illeszkedését a ROT-modell illeszkedésével párhuzamosan teszteltük, majd bemutatjuk a szakirodalomból feltárt vizsgálatok újraelemzése során kapott eredményeinket, végül pedig ismertetjük a kísérletek során gyűjtött adatok elemzését – először az összegzett kovariancia-információ alapján való tanulásra, majd az online tanulásra vonatkozóan.

### **3.1. Alternatív modellek**

Célunk volt, hogy a ROT-modell illeszkedését összevethessük más modellekkel, amelyek ugyancsak elemi oksági kapcsolatok meglétére vonatkozó ítéletek becslésére vállalkoztak. Ezekből nem volt túl nagy választék; két olyan modellt választottunk, amelyek számos hasonlóságot mutatnak a ROT-modellel, amennyiben mindegyik az oksági tanulás Bayes-i modellje, és a becslése az oksági struktúrára, tehát elemi oksági kapcsolatok meglétére/hiányára vonatkozik. Az egyik választott modellünk Griffiths és Tenenbaum (2005) support-modellje, a másik pedig Lu és mtsai. (2008) SS-modellje (az SS a „strong and sparse”, tehát erős és ritka rövidítésre utal). A három modell számos hasonlóságot mutat, tekintve, hogy az SS-modell és a ROT-modell is a support-modell egy-egy módosított változatának tekinthető. Ami közös bennük, az kettő, az oksági kapcsolat meglétére, illetve annak hiányára vonatkozó hipotézis összevetése a Bayes-faktor segítségével. Ugyanakkor vannak fontos különbségek a modellek között. A következőkben röviden bemutatjuk a support-modell, valamint az SS-modell, amely során kiemeljük, hogy mely jellemzőikben mutatnak jelentős eltérést a ROT-modelltől.

Griffiths és Tenenbaum (2005) a support-modell az elemi oksági tanulás magyarázatára alkotta – céljuk egy olyan keret-elmélet felvázolása volt, amelybe

beilleszthetők specifikus alapfeltevések, a support-modell ennek megfelelően rendkívül robusztus és felszínesen definiált ad hoc feltevéseket fogalmaz meg az oksági gondolkodást illetően.

A support-modell és ROT-modell között egyik különbség a vizsgált hipotézisek szerkezete (7. ábra). A support-modell esetében – a ROT-moddal ellentétben – egyetlen rejtett oksági tényezőt ábrázolnak a hipotézisek, amelyek alternatív generatív oksági tényezők (A), és a következmény létrehozására képes különböző események konglomerátumának tekinthetők. Ugyancsak egy eltérés, hogy a support-modell ezen tényezőket konstansnak tekinti, tehát ezek az alternatív oksági tényezők minden esetben jelen vannak. Az oksági kapcsolatokat a support-modell mint probabilisztikus kapcsolatokat kezeli. A modell mind az alternatív okokhoz, mind a vizsgált előzményhez az oksági kapcsolat erejét ábrázoló súlyokat rendel ( $\omega_0, \omega_1$ ), ezek mértékében ugyanakkor maximális mértékű bizonytalanságot feltételez, amit az egyenletes eloszlás segítségével ábrázolja a  $[0,1]$  tartományban, tehát  $\omega_0 \sim Uniform(0,1)$ , és  $\omega_1 \sim Uniform(0,1)$ . A két oksági tényező hatásának integrálására a zajos-vagy függvényt használja, amely szerint  $p(K|E, A, \omega_0, \omega_1) = 1 - (1 - \omega_0)^A (1 - \omega_1)^E$ .



a) az oksági kapcsolat meglétére vonatkozó hipotézis ( $\mathcal{H}_1$ )      b) az oksági kapcsolat hiányára vonatkozó hipotézis ( $\mathcal{H}_0$ )

7. ábra

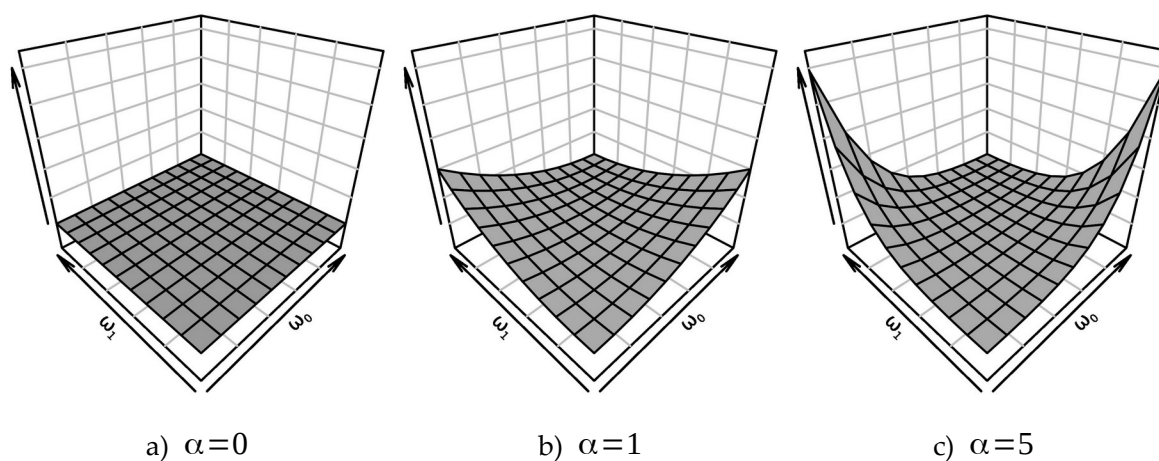
A support-modell által vizsgált, az előzmény és következmény közötti oksági kapcsolat meglétére, illetve hiányára vonatkozó hipotézisek

A Lu és mtsai. (2008) által készített SS-modell a support-modellhez képest az oksági tényezők oksági erejére vonatkozó bizonytalanság szerkezetében mutat jelentős eltérést. Azt a feltevést fogalmazza meg, hogy az oksági kapcsolatokat jellemzően erős kapcsolatoknak gondoljuk, és hogy nem feltételezünk túl sok alternatív okot a háttérben. A feltevés első része némi összhangot mutat a ROT-modell determinisztikus oksági kapcsolatokra vonatkozó alapfeltevésével, a második része pedig az események megértése érdekében megfogalmazott

magyarázatok egyszerűségére utal. Az SS-modell a oksági súlyok értékét illető bizonytalanságot (12)-nek megfelelő eloszlással ábrázolja.

$$p(\omega_0, \omega_1) \propto (e^{-\alpha\omega_0 - \alpha(1-\omega_1)} + e^{-\alpha(1-\omega_0) - \alpha\omega_1}) \quad (12)$$

Az  $\alpha$  paraméter meghatározza az eloszlás szélsőségségét – alacsony értéke mellett az oksági súlyok eloszlása az egyenletes eloszlásnak felel meg (és teljes mértékben megegyezik a support-moddal), magas értékei mellett viszont nagyobb valószínűséget rendelünk az oksági súlyok azon párosaihoz, ahol az egyik érték magas, a másik pedig alacsony, mint azokhoz, amelyeknél vagy mind a két érték magas, vagy mind a két érték alacsony (ld. 8. ábra). Lu és mtsai. (2008) a modell ellenőrzése során  $\alpha=5$  értéket választott – mi is így tettünk a jelen dolgozatban.



8. ábra

Az oksági súlyok  $(\omega_0, \omega_1)$  bizonytalanságát ábrázoló együttes valószínűségi eloszlás alakulása az  $\alpha$  paraméter különböző értékei mellett

Az SS-modell valamint a support-modell R szkriptként való implementációját a B valamint a C melléklet tartalmazza.

## 3.2. A modellek illeszkedésének vizsgálata összegzett kovariancia-információ alapján való tanulás esetén

### 3.2.1. A vizsgált modellek predikciói

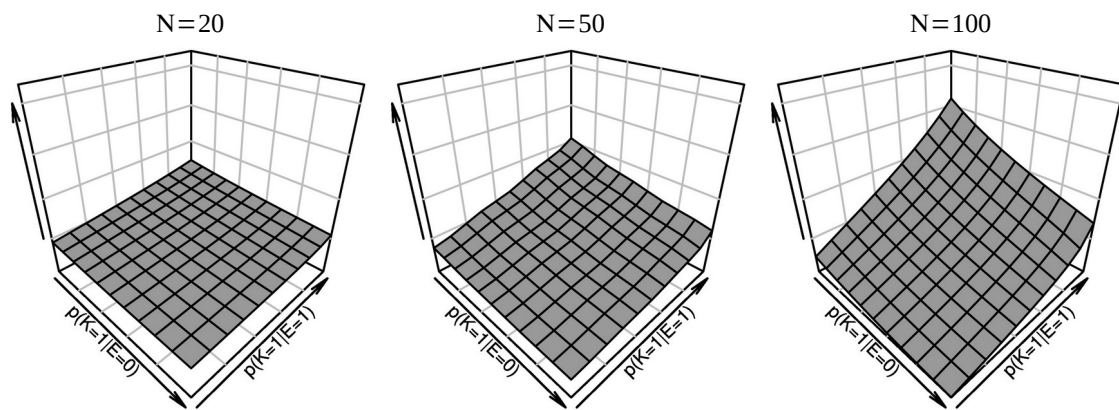
A ROT-modell, valamint a két alternatív modell – a support-modell és az SS-modell – predikcióit a kovariancia-információ három jellemzőjének, a következmény előzmény hiányában tapasztalt valószínűségének  $(p(K=1|E=0))$ , a következmény előzmény

jelenlétében tapasztalt valószínűségének ( $p(K=1|E=1)$ ), valamint a virtuális mintanagyságnak ( $N$ ) a szisztematikus variálása mellett vizsgáltuk meg. A virtuális mintanagyság esetében három szintet különböztettünk meg ( $N=20$ ,  $N=50$ , valamint  $N=100$ ), a két valószínűséget pedig 0 és 1 között változtattuk 0,1-es lépésekben. A predikciókat a 9. ábrán ábrázoltuk.

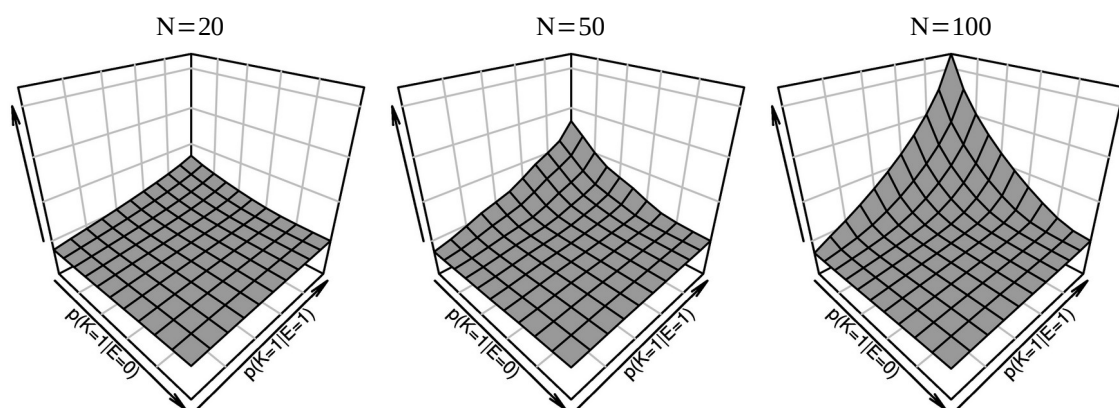
A három modell predikciói több hasonlóságot mutatnak. Mindhárom modellnél megjelenik a következmény előzmény jelenlétében tapasztalt valószínűségének pozitív hatása, akárcsak a következmény előzmény hiányában tapasztalt valószínűségének negatív hatása mindhárom virtuális mintanagyság esetén. Ugyancsak megfigyelhető az interakció a két valószínűség között – a következmény előzmény hiányában tapasztalt valószínűségének növekedésével csökken a következmény előzmény jelenlétében tapasztalt valószínűségének hatása.

A virtuális mintanagyság polarizáló hatása is megjelenik mindhárom modell esetében: ahogy nő a virtuális mintanagyság, az oksági ítéletek egyre szélsőségesebbé válnak. Egy érdekes eltérés a modellek becsléseiben arra a helyzetre vonatkozik, amikor a következmény előzmény hiányában tapasztalt valószínűsége meghaladja a következmény előzmény jelenlétében tapasztalt valószínűségét (ami inkább egy prediktív oksági kapcsolatra utal). A ROT-modell becslése ebben a tartományban csökkenő tendenciát mutat, ahogy a két valószínűség különbsége nő, míg a support-modell és SS-modell becslései alig mutatnak változást.

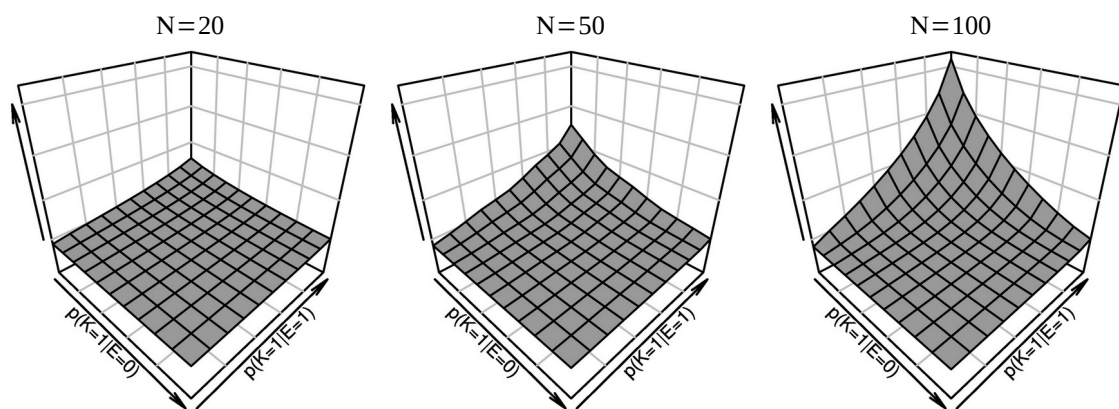
A modellek becslései – kvalitatív szinten – összhangban vannak az empirikus kutatások eredményeivel, amennyiben az oksági kapcsolat meglétében való magasabb fokú bizonyosságot jósolnak a következmény előzmény jelenlétében tapasztalt magas, és a következmény előzmény hiányában tapasztalt alacsony valószínűsége esetén. A virtuális mintanagyság ilyen polarizáló hatása (az oksági kapcsolat meglétében való nagyobb fokú bizonyosság több rendelkezésre álló adat esetén) – legalább normatív szempontból – ugyancsak plauzibilisnek tűnik.



a) a ROT-modell predikciói



b) az SS-modell predikciói



c) a support-modell predikciói

9. ábra

A vizsgált modellek predikcióinak alakulása (függőleges tengely) a következmény előzmény jelenlétében ( $p(K=1|E=1)$ ), illetve hiányában ( $p(K=1|E=0)$ ) tapasztalt valószínűségének, illetve a virtuális mintanagyságnak ( $N$ ) a különböző szintjei mellett

### 3.2.2. Mások vizsgálatai

Ahogy Perales és Shanks (2007) rámutat, elterjedt gyakorlat, hogy egy-egy kognitív modellt olyan vizsgálatokból származó adatokon tesztelnek, amelyek eleve kedveznek a vizsgált modell illeszkedésének. Mi a ROT-modell illeszkedésének reális megítélésére törekedtünk, ezért igyekeztünk a szakirodalomban olyan vizsgálatokat keresni, amelyek alkalmasak lehetnek az illeszkedés ellenőrzésére. A keresés során két szempontot határoztunk meg, amelyek a ROT-modell becsléseinek a jellegéből fakadnak. Az egyik szempont az volt, hogy a vizsgálat kovariancia-alapú oksági tanulásról szóljon, tehát a vizsgálati személyek a vizsgálatban egy vagy több kovariancia-információ alapján alkossanak ítéletet. A másik szempont, hogy az ítélet valamilyen módon két esemény közötti oksági kapcsolat meglétében való bizonyosságra vonatkozzon. Az egyszerű események közötti, oksági struktúra tanulására vonatkozó vizsgálatok száma nem bizonyult túl jelentősnek. A továbbiakban három olyan tanulmányban található vizsgálatokat ismertetünk, amelyek megfeleltek a szempontjainknak.

Thomas Griffiths és Josh Tenenbaum tanulmánya (Griffiths és Tenenbaum, 2005) a ROT-modell alapját képező support-modell első ismertetése, valamint a tanulmányban több kísérletet is közöltek a szerzők, amelyeket a javasolt modelljük becsléseinek ellenőrzésére terveztek. Itt két kísérletet találtunk megfelelőnek, ahol a vizsgálati személyeknek különböző kovariancia-információkat tartalmazó kártyákat kellett sorba rendezniük, aszerint, hogy azok alapján mennyire biztosak az oksági kapcsolatok meglétében. Ezek az eredmények a ROT-modell becslésének egy viszonylag robusztus megítélését tették lehetővé.

Lu és mtsai. (2008) tanulmányában is találtunk egy kísérletet, amelynek eredményei megfelelőek lehettek a ROT-modell illeszkedésének árnyaltabb megítéléséhez. Itt a vizsgálati személyeknek az oksági kapcsolat meglétében való bizonyosságukat egyszerű értékelőskálán kellett megítélniük, és így a modell becslésének egy finomabb skálán való megítélését is lehetővé tették az adatok.

A harmadik kiválasztott vizsgálatot Saito és Shimazaki (2013) közölte. A vizsgálatban a vizsgálati személyek az oksági kapcsolat egyszerű meglétét, vagy éppen hiányát állapíthatták meg, tehát az adatok nem a kapcsolat meglétében való bizonyosság mértékét tükrözték, hanem azokra vonatkozó dichotóm döntéseknek tekinthetők. Azért döntöttünk mégis az eredmények felhasználása mellett, mert azok lehetőséget biztosítottak a ROT-

modell illeszkedésének vizsgálatára kvázi-valós viselkedéses adatok (az említett dichotóm döntés) alapján, szemben a megszokott szubjektív ítéletekkel.

A továbbiakban részletesen ismertetjük a kiválasztott vizsgálatokat, valamint azok eredményét, továbbá elemezzük a vizsgált modellek illeszkedését az eredményekre.

### ***Griffiths és Tenenbaum (2005) 2. és 3. kísérlete***

Griffiths és Tenenbaum (2005) az elemi oksági kapcsolatok tanulására vonatkozó support-modelljük bemutatása mellett, a modell illeszkedésének ellenőrzése céljából végzett kísérleteik eredményeit is ismerteti. Ezek közül a 2. és 3. kísérletük felelt meg a kritériumainknak, a többi kísérletükben vagy nem az oksági kapcsolat meglétében való bizonyosságukat kellett megítélni a résztvevőknek (1. kísérlet), vagy nem kovariancia-információ állt a rendelkezésükre a feladat során (4. és 5. kísérlet).

Mindkét kísérletben az volt a résztvevők feladata, hogy kovariancia-információkat tartalmazó kártyákat állítsanak sorrendbe. A kovariancia-információk különböző vegyi anyagok, illetve bizonyos gének kifejeződéseinek együttes előfordulására vonatkoztak. Az instrukció szerint a sorrendben első kártya az kellett legyen, amelyik esetében a vizsgálati személyek a legkevésbé tartják valószínűnek, hogy van oksági kapcsolat a kérdéses vegyi anyag és az adott gén kifejeződése között, az utolsó pedig, amelyik esetében a legvalószínűbbnek tartják, hogy van oksági kapcsolat.

A 2. kísérletben 9 kártya állt a résztvevők rendelkezésére. A kísérletben manipulálták a kovariancia-információ által ábrázolt asszociáció mértékét ( $\Delta P=0,5$ , vagy  $\Delta P=0$ ), valamint a virtuális mintanagyságot ( $N=8$ ,  $N=40$ , vagy  $N=100$ ). A kártyák közül három olyan kovariancia-információt tartalmazott, amelyek esetében mérsékelt pozitív asszociáció volt ( $\Delta P=0,5$ ), három olyat, amelyek esetében nem volt asszociáció a kérdéses vegyi anyag és az adott gén kifejeződése között ( $\Delta P=0$ ), további három kártya pedig csupán a kísérleti manipuláció elfedését szolgálta. Mindkét feltétel esetében a virtuális mintanagyság eltérő volt az egyes kártyákon.

A kísérlet résztvevőinek jelentős többsége a support-modell becslésének megfelelően a virtuális mintanagyság szerint növekvő sorrendbe rendezte azokat a kártyákat, amelyek generatív kapcsolatot jeleztek ( $\Delta P=0,5$ ), míg a virtuális mintanagyság szerint csökkenő sorrendbe rendezték azokat, amelyek a kapcsolat hiányára utaltak ( $\Delta P=0$ ). Az eredmények a virtuális mintanagyság polarizáló hatására utalnak, azaz arra, hogy amikor az adatok az

oksági kapcsolat mellett szólnak, akkor annál bizonyosabbak vagyunk az oksági kapcsolat meglétében, minél több adat áll rendelkezésünkre, és ha az adatok az oksági kapcsolat hiányára utalnak, akkor annál bizonyosabbak vagyunk az oksági kapcsolat hiányában, minél több adat áll rendelkezésünkre.

A 3. kísérletben a hiányos kovariancia-információ hatását vizsgálták a kutatók, ahol a résztvevők számára prezentált kártyák csak az oksági kapcsolat meglétére utaló adatokat tartalmaztak. Összesen 5 kártyát kellett sorba rendezni, amelyeken a kutatók a virtuális mintanagyságot manipulálták. A résztvevők jelentős többsége a virtuális mintanagysággal megegyező sorrendbe rendezte a kártyákat. Ez részben ugyancsak demonstrálja a virtuális mintanagyság polarizáló hatását maximális asszociáció, és részleges kovariancia-információ mellett.

A két kísérlet eredményét és a modellek becsléseit a kísérletben alkalmazott konkrét feltételekre a 5. táblázat tartalmazza. Az eredmények alapján a vizsgált modellek illeszkedését ordinális szinten volt lehetőségünk ellenőrizni, tehát azt tudtuk vizsgálni, hogy a modellek által adott becslések az egyes kovariancia-információkra a résztvevők többsége által adott sorrendekhez hasonlóan rangsorolnak-e. A vizsgált modellek mindegyike sikeresen előrejelezte a vizsgálati személyek többsége által alkotott sorrendet, tehát sikeresen jósolták be a virtuális mintanagyság polarizáló hatását.

#### 4. táblázat

A kovariancia-információk résztvevők általi rangsorolása Griffiths és Tenenbaum (2005) 2. és 3. kísérletében, valamint a vizsgált modellek predikciói a kísérletekben alkalmazott kovariancia-információkra

						a vizsgált modellek becslései, és a hozzájuk tartozó rangok (zárójelben)		
					kísérlet résztvevői által alkotott sorrend	ROT- modell	SS-modell	support- modell
N	N <sub>a</sub>	N <sub>b</sub>	N <sub>c</sub>	N <sub>d</sub>				
2. kísérlet, $\Delta P=0$								
8	0	4	0	4	3.	-0,97 (3)	-3,47 (3)	-1,62 (3)
40	0	20	0	20	2.	-4,13 (2)	-5,59 (2)	-3,07 (2)
100	0	50	0	50	1.	-10,04 (1)	-6,58 (1)	-3,81 (1)
2. kísérlet, $\Delta P=0,5$								
8	2	2	0	4	1.	1,02 (1)	0,58 (1)	0,62 (1)
40	10	10	0	20	2.	5,23 (2)	6,23 (2)	6,17 (2)
100	25	25	0	50	3.	13,15 (3)	18,36 (3)	18,01 (3)
3. kísérlet								
0	0	0	0	0	1.	0 (1)	0 (1)	0 (1)
1	1	0	0	0	2.	0,69 (2)	0,52 (2)	0,4 (2)
3	3	0	0	0	3.	1,67 (3)	0,73 (3)	,74 (3)
20	20	0	0	0	4.	9,1 (4)	0,9 (4)	1,28 (4)
40	20	0	0	20	5.	17,4 (5)	26,24 (5)	23,28 (5)

#### **Lu és mtsai (2008) 3. kísérlete**

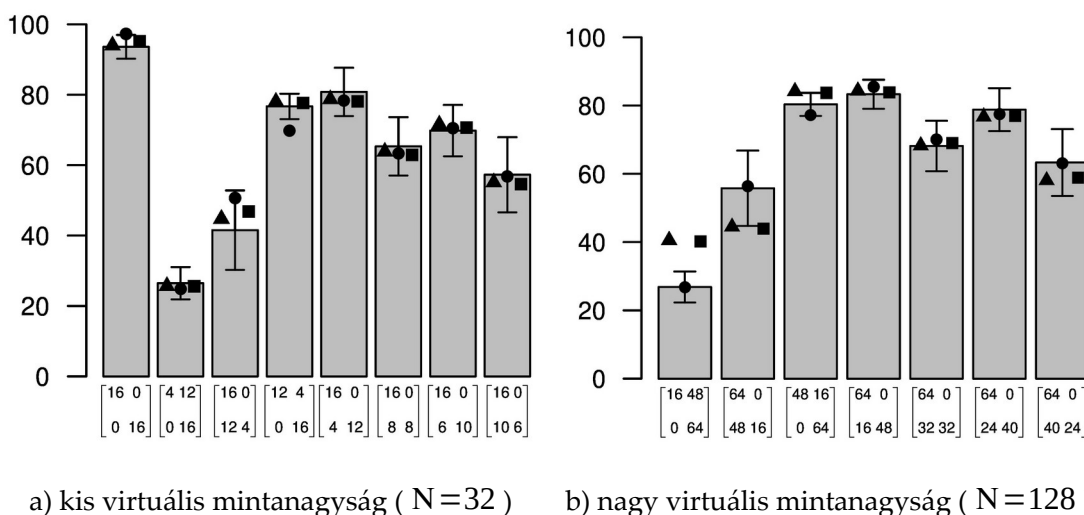
Lu és mtsai. (2008) 3. kísérletüket az SS-modell oksági struktúrára vonatkozó becsléseinek az ellenőrzésére tervezték. A kísérleti személyek számára az instrukcióban egy gyógyszergyártó cég kutatásainak a kontextusát vázolták fel, ahol is a gyógyszergyártó cégek azt vizsgálják, hogy különböző ásványi anyagok kiváltanak-e fejfájást. A vizsgálatok során több ásványi anyagot is teszteltek, és különböző számú kísérleti alanyt alkalmaztak.

A kísérlet során több változót is manipuláltak a kutatók. Egyrészt volt egy generatív és egy preventív feltétel. Míg a generatív feltételben az ásványi anyagok fejfájást kiváltó hatása volt a vizsgálat tárgya, addig a preventív feltételben az ásványi anyagok fejfájást megszüntető hatásáról kellett ítéletet hozni. A további manipulációk a kísérleti személyek

rendelkezésre álló kovariancia-információkat érintették; összesen 16 különböző kovariancia-információ állt rendelkezésre, amelyek változatosak voltak a következmény előzmény hiányában tapasztalt valószínűségét, az előzmény oksági erejét, valamint a virtuális mintanagyságot illetően. Az oksági kapcsolat irányát személyek között manipulálták, a kontingencia-információkat személyen belül. Technikai problémákból kifolyólag a kutatóknak csak 15 kontingencia-információra vonatkozó becsléseket sikerült rögzíteniük.

A kísérleti személyek a válaszaikat egy 0-100 tartó skálán adhatták meg, ahol a 0 azt jelentette, hogy nagyon nagy valószínűséggel nincs oksági kapcsolat, míg a 100 azt jelentette, hogy nagyon nagy valószínűséggel van oksági kapcsolat (az oksági kapcsolat irányára való tekintet nélkül).

Mivel a ROT-modell csak a generatív oksági kapcsolatok tanulására vonatkozóan fogalmaz meg becsléseket, ezért itt a vizsgálatnak csak azokat az eredményeit elemezzük újra, amelyek a generatív oksági kapcsolatok tanulására vonatkoznak. A kutatók - kérésünkre - rendelkezésünkre bocsátották a kísérlet eredményeinek leíró statisztikáit: a kísérleti személyek által adott értékelések átlagait és sztenderd hibáit, ami jelentősen megkönnyítette a modellek illeszkedésének ellenőrzését.



10. ábra

Lu és mtsai. (2008) 3. kísérletének eredményei (átlagos értékelések) a generatív feltételben, és a vizsgált modellek lineáris regresszióanalízis segítségével előkészített becslései (kör:

ROT-modell, négyzet: SS-modell, háromszög: support-modell).

A hibásávok a sztenderd hiba alapján számolt 95%-os konfidencia-intervallumok

Az eredmények (10. ábra) arról árulkodnak, hogy mind a következmény előzmény hiányában tapasztalt valószínűsége, mind a következmény előzmény jelenlétében tapasztalt valószínűsége, mind a virtuális mintanagyság hatással volt az értékelésekre. Az értékelések annál magasabbak voltak, minél közelebb volt a következmény előzmény hiányában tapasztalt valószínűsége az optimálishoz (generatív oksági kapcsolatok esetén ez 0). Minél magasabb volt az előzmény oksági ereje, annál magasabbak voltak az értékelések. A virtuális mintanagyság értékeléseket polarizáló hatása is megjelent - azaz a mintanagyság növekedésével az értékelések is növekedtek (megjegyezzük, hogy csak pozitív asszociációkat jelző kovariancia-információk szerepeltek a generatív feltételben).

A vizsgált modellek meglehetősen gyenge illeszkedést mutattak a vizsgálati személyek értékeléseire, ha minden feltételt tekintetbe vettünk,  $R_{ROT}^2=0,62$ ,  $R_{SS}^2=0,59$ ,  $R_{support}^2=0,53$ . Amikor a feltételeket ketté bontottuk a virtuális mintanagyság szerint, és a két virtuális mintanagyság esetén külön-külön illesztettük a modelleket, akkor azok illeszkedése jelentősen javult mind a kis,  $R_{ROT}^2=0,95$ ,  $R_{SS}^2=0,98$ ,  $R_{support}^2=0,99$ , mind a nagy virtuális mintanagyság esetén,  $R_{ROT}^2=0,99$ ,  $R_{SS}^2=0,85$ ,  $R_{support}^2=0,84$ ; ezen belül a ROT-modell némileg jobban teljesített a másik két modellnél a nagy virtuális mintanagyság esetén, és kicsivel rosszabbul a kis virtuális mintanagyság esetén. Ezek az eredmények arra utalhatnak, hogy a vizsgált modellek az emberekhez képest túlhangsúlyozzák a virtuális mintanagyság hatását, nagyobb mértékű polarizációt jósolnak, mint amik ténylegesen megjelentek az adatokban.

A virtuális mintanagyság kontrollálásával nagyon jó illeszkedési mutatókat kaptunk, ugyanakkor még egy tényező árnyalhatja a modellek illeszkedésének megítélését. A vizsgálati személyek különböző kovariancia-információkra adott értékelései nem minden esetben különböztethetőek meg megbízhatóan a becslési hiba miatt. Egy jól illeszkedő modelltől legalább annyit biztosan elvárhatunk, hogy amennyiben a vizsgálati személyek két kovariancia-információ alapján eltérő ítéletet alkotnak, akkor a modell is eltérő becsléseket ad ezekre, az eltérés iránya pedig megfelel a résztvevők által adott ítéletek közötti különbségnek. Ennek ellenőrzéséhez az egyes kovariancia-információkra adott értékeléseket párosával vizsgáltuk (minden lehetséges párt tekintetbe vettünk), és kiválasztottuk azokat a párokat, amelyeknél a résztvevők értékelései empirikus alapon megbízhatóan megkülönböztethetőek (azaz a 95%-os konfidencia-intervallumuk között nincs átfedés). Kis

virtuális mintanagyság esetén összesen 9 ilyen párt azonosítottunk a lehetséges 28-ból, míg a nagy virtuális mintanagyság esetén 8-at a lehetséges 21-ből. A vizsgált modellek a kiválasztott értékelés párok mindegyike esetében a résztvevők értékeléseikhez hasonlóan rangsoroltak, tehát tökéletes illeszkedést mutattak az adatokra.

### ***Saito és Shimazaki (2013)***

Saito és Shimazaki (2013), kísérletükben arra kérték vizsgálati személyeiket, hogy újonnan felfedezett baktériumok közötti oksági kapcsolatokat értékeljenek, ahol az oksági kapcsolat arra vonatkozott, hogy az egyik baktérium (X) „bekebelezi” a másik baktériumot (Y). A kísérleti instrukció szándékosan nem tisztázta az oksági szerepeket, mindegyik baktérium bekebelezhetette a másikat. Összesen 15 baktérium-párosról láttak kovariancia-információt a kísérleti személyek, amelyekben variálták az Y baktérium előfordulásának a valószínűségét X baktérium jelenlétében ( $p(Y=1|X=1)$ ), valamint az Y baktérium előfordulásának valószínűségét az X baktérium hiányában ( $p(Y=1|X=0)$ ), úgy hogy ezek különbsége pozitív maradjon.

Az egyes kovariancia-információk megtekintése után a résztvevőknek azt kellett megítélnie, hogy vajon X baktérium bekebelezi-e Y baktériumot, illetve, hogy Y baktérium bekebelezi-e X baktériumot. A két kérdésre egymástól függetlenül válaszolhattak a résztvevők igennel, vagy nemmel. Ez összesen négy lehetséges válaszkombinációt tett lehetővé minden egyes kovariancia-információra (zárójelben az adott válaszkombináció rövid jelölése látható):

- X bekebelezi Y-t és Y bekebelezi X-et ( $x \rightarrow y \quad y \rightarrow x$ )
- X bekebelezi Y-t, de Y nem kebelezi be X-et ( $x \rightarrow y$ )
- Y bekebelezi X-et, de X nem kebelezi be Y-t ( $y \rightarrow x$ )
- sem X, sem Y nem kebelezi be a másikat (-)

5. táblázat

Saito és Shimazaki (2013) kísérletében alkalmazott kovariancia-információk, a résztvevők által adott leggyakoribb válaszok, valamint a vizsgált modellek által adott becslések

$N_a$	$N_b$	$N_c$	$N_d$	résztevők leggyakoribb válasza	vizsgált modellek becslései <sup>a</sup>					
					ROT-modell		SS-modell		support- modell	
8	0	0	8	$x \rightarrow y$ $y \rightarrow x$	$x \rightarrow y$	$y \rightarrow x$	$x \rightarrow y$	$y \rightarrow x$	$x \rightarrow y$	$y \rightarrow x$
8	0	2	6	$y \rightarrow x$	$y \rightarrow x$		$x \rightarrow y$		$x \rightarrow y$	
6	2	0	8	$x \rightarrow y$	$x \rightarrow y$		$x \rightarrow y$		$x \rightarrow y$	
8	0	4	4	$y \rightarrow x$	$y \rightarrow x$		$y \rightarrow x$		$y \rightarrow x$	
6	2	2	6	-	$y \rightarrow x$		$y \rightarrow x$		$y \rightarrow x$	
4	4	0	8	$x \rightarrow y$	$x \rightarrow y$		$x \rightarrow y$		$x \rightarrow y$	
8	0	6	2	$y \rightarrow x$	$y \rightarrow x$		$y \rightarrow x$		-	
6	2	4	4	-	-		-		-	
4	4	2	6	-	-		-		-	
2	6	0	8	$x \rightarrow y$	-		-		-	
8	0	8	0	$y \rightarrow x$	$y \rightarrow x$		-		-	
6	2	6	2	-	-		-		-	
4	4	4	4	-	-		-		-	
2	6	2	6	-	-		-		-	
0	8	0	8	-	-		-		-	

<sup>a</sup> a modellek becsléseit ebben a speciális esetben lineáris diszkriminancia-analízis segítségével vezettük le

A kísérletben alkalmazott kovariancia-információkat, valamint a résztvevők által adott leggyakoribb válaszokat az egyes feltételekben a 5. táblázat tartalmazza.

A vizsgált modellek becsléseit lineáris diszkriminancia-analízis segítségével határoztuk meg. Ennek menete a következő volt. Minden kovariancia-információ esetében a modellek két-két becslését vettük alapul (az egyik esetben X-et tekintettük előzménynek, Y-t következménynek, a másik esetben felcseréltük a szerepeket). Ezt követően ezt a két becslést a lineáris diszkriminancia-analízis modelljében prediktor változóként alkalmaztuk, míg a kísérlet résztvevőinek leggyakoribb válasza volt a célváltozó. Az így kapott diszkriminancia-függvény segítségével becsültük a résztvevők leggyakoribb válaszát.

A többváltozós varianciaanalízis eredménye alapján szignifikáns eltérést találtunk a modellek becsléseiben a résztvevők leggyakoribb válasza mentén, a ROT-modellnél,  $Wilks-\lambda=0,24$   $p<0,05$ , az SS-modellnél,  $Wilks-\lambda=0,23$   $p<0,05$ , valamint a support-modellnél is,  $Wilks-\lambda=0,18$   $p<0,01$ . A legpontosabb becsléseket a ROT-modell esetében kaptuk, ahol a modell becslései erős egyezést mutattak a résztvevők válaszaival,  $Cohen\kappa=0,8$ , 13 találat a 15-ből; gyenge egyezést találtunk az SS-modell esetében,  $Cohen\kappa=0,48$ , 10 találat mellett, továbbá szinte mérsékelt egyezést a support-modell esetében,  $Cohen\kappa=0,59$ , a 15-ből 11 találat mellett. Az egyezés mértékének értelmezéséhez egy konzervatívnak tekinthető kritériumot használtunk, amelyet McHugh (2012) javasolt.

### ***Következtetések***

A ROT-modell – a másik két modellhez hasonlóan – jó illeszkedést mutatott a fentebbi vizsgálatok eredményeire, a résztvevők válaszaira, három különböző helyzetben: ahol a résztvevőknek rangsorolni kellett a kovariancia-információkat, ahol egy skálán kellett jelezniük a bizonyosságukat az oksági kapcsolat meglétében, valamint ahol az oksági kapcsolat pusztá meglétét, vagy hiányát kellett jelezni (megengedve a kétirányú kapcsolatokat is).

Amennyiben Perales és Shanks (2007) az irányú megállapítását tekintetbe vesszük, hogy az egyes modellek készítésekor, specifikusan azok ellenőrzésére tervezett vizsgálatokban ezek jól is teljesítenek, akkor nem lepődhetünk meg azon, hogy a support-modell jól teljesített Griffiths és Tenenbaum (2005) kísérleteiben, illetve azon sem, hogy az SS-modell jól teljesített Lu és mtsai. (2008) kísérletében – ami számunkra érdekes, hogy a ROT-modell ezekhez hasonlóan jó, vagy akár jobb becsléseket adott. Saito és Shimazaki (2013) kísérlete, különösen annak speciális jellege miatt, valódi teszterep mindhárom modell számára. Bár mindhárom modell viszonylag jól teljesített ezen a teszten, a legjobb illeszkedést a ROT-modell mutatta, alkalmasabbnak bizonyult a másik két modellnél egy olyan területen, amelyhez hasonlóat még csak nem is mérlegeltünk a modell készítése során.

A modellek illesztése során ugyanakkor arra a következtetésre kellett jutnunk, hogy az emberek a vizsgált normatív modellekhez képest, nem tulajdonítanak túl nagy szerepet a rendelkezésre álló adatok mennyiségének (a virtuális mintanagyságnak) – erre utalhat, hogy Lu és mtsai. (2008) kísérletében a modellek illeszkedései nagy mértékben romlottak, amikor a virtuális mintanagyság hatását nem kontrolláltuk.

### 3.2.3. Első kísérlet: $p(K=1|E=1)$ és $p(K=1|E=0)$ hatásának vizsgálata, kis virtuális mintanagyság esetén

Ebben a kísérletben azt szeretnénk volna megvizsgálni, hogy a következmény előzmény jelenlétében ( $p(K=1|E=1)$ ), illetve előzmény hiányában ( $p(K=1|E=0)$ ) tapasztalt valószínűsége hogyan befolyásolja a generatív oksági kapcsolat meglétére vonatkozó ítéleteket. Az itt vizsgált hatásokat már számos vizsgálat demonstrálta az oksági kapcsolatok erősségére vonatkozó ítéletekre (ld. Buehner és Cheng, 1997; Buehner, Cheng, és Clifford, 2003; Lober és Shanks, 2000), valamint néhány vizsgálat az oksági kapcsolat meglétére vonatkozó ítéletekre (ld. Griffiths és Tenenbaum, 2005; Lu és mtsai., 2008). Jelen kísérletben ezeket a hatásokat - talán elsőként, személyek közötti elrendezés keretén belül vizsgáltuk, illetve olyan kontingencia-információt is alkalmaztunk, amelyet a generatív oksági kapcsolatok tanulásának vizsgálata során jellemzően nem szokás alkalmazni: erős preventív oksági kapcsolatra utaló kontingencia-információt.

#### ***Minta***

Összesen 81-en vettek részt a kísérletben, amelyek közül 70 résztvevő adatait tartottuk elemzésre alkalmasnak. A résztvevők 71,4%-a (50 fő) nő, 28,6%-a (20 fő) férfi. Az átlagos életkor 42,2 év (19-70 év,  $SD = 12,37$ ).

A résztvevők \$0,01 fizetségben részesültek a kísérletben való részvételért, és további \$0,10 jutalmat kaptak a válaszaik ellenőrzését követően. Mindenki megkapta ezt a jutalmat, aki minden feladatot megoldott, és minden kérdésre válaszolt, függetlenül attól, hogy felhasználtuk-e a válaszait az adatelemzés során.

#### ***Módszerek és eljárás***

A kísérletet a Qualtrics platformján készítettük el. A résztvevők először egy tájékoztatót olvashattak, amely információkat tartalmazott a kutatás tárgyára, a feladat jellegére, a várható fizetésre, valamint a válaszok validitásának ellenőrzésére vonatkozóan. Az oldalról az egyértelműen deklarált beleegyezésük után léphettek tovább a vizsgálati személyek a kísérlet instrukciós oldalára.

Az instrukciós oldalon egy fedőszöveget olvashattak, amely felkészítette őket arra, hogy egy oksági kapcsolat megismerése lesz a feladat. A fedőszöveg a következő volt:

Képzeld el, hogy misszionárius vagy egy bennszülöttek által lakott faluban. A faluban számos bennszülöttnek furcsa kiütése van. Te azt

próbálsz kideríteni, hogy mi okozta a kiütéseket. Miután felfedezed, hogy a falu egyik kútja furcsa szagot áraszt, arra következtetsz, hogy annak szennyezett a vize, ezért ez lesz az első gyanúsítottad. Hogy meggyőződj róla, hogy valóban a szennyezett víz okozza-e a kiütéseket, meglátogatsz néhány bennszülöttet, és megkérdezed tőlük, hogy ittak-e a kút vizéből, és megvizsgálod őket, hogy van-e kiütésük.

Az instrukciós oldalt követően a vizsgálati személyek a vizsgálódásaik eredményét mutató kovariancia-információt láthatták, összegzett formában, listaként. A listában feltüntettük, hogy hány olyan eset volt, amelynél:

- a bennszülött ivott a vízből és volt kiütése (a-típusú esemény),
- a bennszülött ivott a vízből és nem volt kiütése (b-típusú esemény),
- a bennszülött nem ivott a vízből és volt kiütése (c-típusú esemény), illetve amelynél
- a bennszülött nem ivott a vízből és nem volt kiütése (d-típusú esemény).

Az egyes eseménytípusok gyakoriságára vonatkozó információk a fenti sorrendben szerepeltek a listán, és az előzményre (ivott-e a vízből) és a következményre (volt-e kiütése) vonatkozó megállapítás ki volt emelve félkövér betűtípussal, hogy megkönnyítsük annak felismerését, hogy eltérő eset-típusokról van szó. Arra is törekedtünk, hogy a megfogalmazás ne sugalljon az oksági kapcsolatra vonatkozó elvárásokat, ennek megfelelően minden esetben az „és” kötőszót használtuk az előzmény és következmény együttes előfordulásának jelölésére az „ennek ellenére”, vagy „ezért”, illetve „de” kötőszavak helyett.

Mind az instrukciós-, mind a kovariancia-információkat tartalmazó oldalon eltöltött időt mértük, annak érdekében, hogy a feladatokon való esetleges átklikkelést feltárhassuk.

A kovariancia-információ megtekintése után a résztvevők (új oldalon) megadhatták az oksági kapcsolat meglétére vonatkozó ítéletüket. A kérdés így hangzott: „A vizsgálódásaid eredménye alapján mit gondolsz, okoz-e a szennyezett víz bőrkiütéseket?”. Válaszukat egy ötfokú skálán adhatták meg a vizsgálati személyek, ahol az egyes szinteket szövegesen határoztuk meg, a következők szerint:

- 1 = teljesen biztos vagyok benne, hogy nem okoz bőrkiütéseket
- 2 = némileg biztos vagyok benne, hogy nem okoz bőrkiütéseket
- 3 = nem tudom eldönteni, hogy okoz-e vagy sem bőrkiütéseket

- 4 = némileg biztos vagyok benne, hogy okoz bőrkiütéseket
- 5 = teljesen biztos vagyok benne, hogy okoz bőrkiütéseket

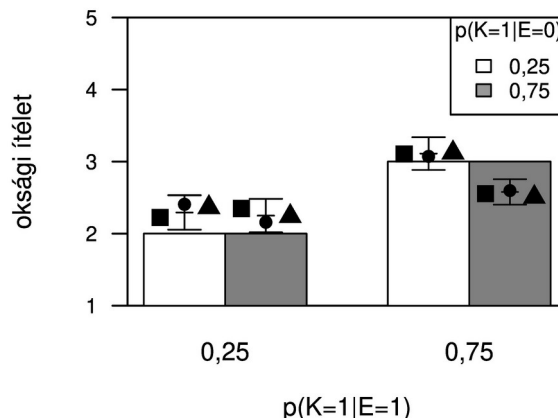
Az ítélet megadása után egyéb kérdések következtek, amelyek a résztvevők nemére, életkorára, legmagasabb iskolai végzettségére, valamint a nyelvi készségekre vonatkoztak. Ami a nyelvi készségeket illeti, a résztvevőknek az angol nyelvű olvasási szövegértési képességük szintjét kellett megbecsülniük egy ötfokú skálán, a „nagyon rossz”-tól a „nagyon jó”-ig.

A kísérlet során két változót manipuláltunk: a következmény előzmény jelenlétében tapasztalt valószínűségét ( $p(K=1|E=1)$ ), valamint a következmény előzmény hiányában tapasztalt valószínűségét ( $p(K=1|E=0)$ ). Mindkét változónak két-két szintjét különböztettük meg (0,25 és 0,75), a virtuális mintanagyság 32 volt (az egyes feltételekben szereplő kovariancia-információkat ld. 6. táblázatban). Az ebből származó négy kísérleti feltételt személyek között manipuláltuk, így minden résztvevő egyetlen kovariancia-információt látott, és egyetlen oksági ítéletet kellett alkotnia a feladat során. A résztvevőket véletlenszerűen rendeltük az egyes feltételekhez.

### ***Eredmények***

Az adatok megfelelő validitásának biztosítása érdekében az instrukciós oldalon, illetve a kovariancia-információkat tartalmazó oldalon töltött idő alapján, valamint az angol olvasási szövegértés szintje alapján szűrtük az adatokat. A kísérlet két kritikus oldalán töltött idő esetében Box-Cox transzformációt végeztünk az eloszlás normalizálása céljából; a kilógó értékek azonosításához pedig az első- és harmadik kvartilistől vett, az interkvartilis terjedelem 1,5-szeresénél nagyobb távolságot tekintettük kritériumnak. Az olvasási szövegértésnél a közepesnél (a kérdésben szereplő megfogalmazás szerint se nem rossz, se nem jó) magasabb szintet határoztuk meg kritériumként. Ezek alapján összesen 11 résztvevő adatait hagytuk figyelmen kívül az adatelemzés során.

A kísérleti csoportok közel azonos méretűek (17, 16, 18, illetve 19 fő), nem találtunk jelentős különbséget közöttük sem a nem,  $X^2(3)=3,11$   $p=0,375$ , sem az életkor, *Kruskal–Wallis*  $X^2(3)=1,88$   $p=0,598$ , sem a nyelvi készségek szintje,  $X^2(3)=6,56$   $p=0,088$  tekintetében. Az oksági ítéletek ordinális jellege miatt az adatok elemzése során nemparaméteres eljárásokat alkalmaztunk.



11. ábra

Az első kísérlet egyes feltételeiben adott ítéletek alakulása  
(az oszlopok az ítéletek mediánját, a hibasávok az átlagot, és az átlag körüli  $\pm 1$  sztenderd hibát jelölik; a négyzet az SS-modell, a kör a ROT-modell, a háromszög a support-modell lineáris regresszióanalízis segítségével előkészített becslése)

Az eredményeken (11. ábra) megfigyelhető, hogy a résztvevők meglehetősen konzervatív véleményeket alkottak még abban a feltételben is, ahol a kovariancia-információ határozottan az oksági kapcsolat megléte mellett szól (3. feltétel), a vélemények inkább arra utalnak, hogy nem tartották egyértelműen pozitív bizonyítéknak az adatokat az oksági kapcsolat mellett. A többi feltételben jellemzően inkább az oksági kapcsolat meglétével szemben fogalmaztak meg ítéleteket a résztvevők – bár itt sem utasították el egyértelműen annak lehetőségét. Az egymintás Wilcoxon-próba eredményei alapján az értékelések nem térnek el jelentősen a teljes bizonytalanságot jelző skála közepétől a 3. feltételben,  $V=52$   $p=0.644$ , viszont lényegesen alacsonyabbak annál az 1. feltétel,  $V=9$   $p<0,05$ , a 2. feltétel,  $V=11$   $p<0,05$ , és a 4. feltétel,  $V=3,5$   $p<0,05$  esetében.

A Scheirer-Ray-Hare próba alapján nem volt jelentős interakció a két változónk között,  $X^2(1)=0,63$   $p=0,427$ . Nem volt szignifikáns főhatása a következmény előzmény hiányában tapasztalt valószínűségének,  $X^2(1)=1,35$   $p=0,245$ , viszont a következmény előzmény jelenlétében tapasztalt valószínűségének szignifikáns hatását találtuk,  $X^2(1)=6,13$   $p<0,05$ , amely szerint kevésbé voltak hajlamosak a résztvevők elutasítani az oksági kapcsolat lehetőségét következmény és előzmény között, ha az előzmény jelenlétében a következmény nagy valószínűséggel megjelent.

A vizsgált modellek predikciói kvalitatív szinten elemezve megfelelő hasonlóságokat mutatnak a résztvevők ítéleteihez (6. táblázat). Azokban a feltételekben, ahol a következmény előzmény jelenlétében tapasztalt valószínűsége magas (3. és 4. feltétel), a modellek becslései is magasabbak, mint a másik két feltételben. Ezen a szinten találhatunk ugyanakkor eltéréseket a modellek becslései között, nevezetesen az SS-modell eltérően rangsorolja az 1. és 2. feltételben szereplő kovariancia-információkat, mint a ROT-modell, vagy a support-modell. Tekintve, hogy a résztvevők értékeléseiben igen kicsi a különbség, és empirikusan nem tudjuk azokat szétválasztani, itt nem tudunk egyértelmű ítéletet alkotni arról, hogy melyik modell becslése tekinthető megfelelőbbnek.

6. táblázat

A résztvevők által adott válaszok átlaga, valamint a vizsgált modellek becslései az első kísérlet egyes feltételeiben

	$p(K=1 E=1)$	$p(K=1 E=0)$	$N_a$	$N_b$	$N_c$	$N_d$	$M^a$	modellek predikciói		
								ROT-modell	SS-modell	support-modell
1.	0,25	0,25	4	12	4	12	2,29	-1,62	-2,63	-1,45
2.	0,25	0,75	4	12	12	4	2,25	-4,15	-1,72	-2,23
3.	0,75	0,25	12	4	4	12	3,11	5,14	3,81	3,38
4.	0,75	0,75	12	4	12	4	2,58	0,31	-0,21	-0,54

<sup>a</sup> a résztvevők által adott értékelések átlaga

Azt találtuk, hogy vizsgált modellek becslései meglehetősen jól magyarázzák a kapott eredményeket; a legjobb illeszkedést a support-modell mutatta,  $R^2=0,98$ , aztán az SS-modell,  $R^2=0,97$ , végül pedig a ROT-modell,  $R^2=0,95$ .

### 3.2.4. Második kísérlet: $p(K=1|E=0)$ és virtuális mintanagyság ( $N$ ) hatásának vizsgálata

A második kísérlet számos hasonlóságot mutatott az első kísérlethez, ugyanakkor itt szerettük volna a virtuális mintanagyság oksági ítéletet polarizáló hatását direkt módon tesztelni, és ellenőrizni arra a vizsgált modellek becsléseinek illeszkedését.

### ***Minta***

A kísérletben összesen 92-en vettek részt, akik közül 71 résztvevő adatait találtuk elemzésre alkalmasnak. A mintában szereplők 63,4%-a (45 fő) nő, 36,6%-a (26 fő) férfi. Az átlagos életkor 39,7 év (18-73 év,  $SD = 10,71$ ).

A résztvevők \$0,03-t kaptak a kísérletben való részvételért, amelyet kiegészítettünk \$0,15 bónusszal, az eredményeik ellenőrzését követően (minden résztvevő megkapta ezt a bónuszt, aki befejezte a feladatot).

### ***Módszerek és eljárás***

A kísérletet a Qualtrics platformján tettük elérhetővé, és ugyanazt a képzeletbeli helyzetet (misszionárius-feladatot) alkalmaztuk, ugyanakkor néhány ponton hasznosnak tartottunk bevezetni bizonyos változtatásokat, amelyek esetleg segíthetnek hatékonyabban elkülöníteni a különböző kovariancia-információkra adott értékeléseket.

Az egyik változtatás a kovariancia-információ prezentációját érintette. Ahelyett, hogy listaként mutattuk volna be az összegzett kovariancia-információt, azt egy mátrixba rendeztük, amely könnyebben összehasonlíthatóvá tette az egyes eseménytípusok gyakoriságára vonatkozó információkat. Az egyes eseménytípusokra vonatkozó adatok megfogalmazásánál ugyanazokat az elveket követtük, mint az első kísérletben, ennek megfelelően kerültük az oksági kapcsolat meglétére, vagy hiányára akár csak közvetett módon utaló kötőszavakat.

Változtattunk az ítéletek megadásának módján is. Egyrészt az ítéletek megadására egy kvázi-analóg skálát használtunk, amely lényegében egy 101 fokú skála (0-100) volt. A résztvevők a válaszukat egy csúszka segítségével tudták megadni, amelyen az egyes értékeket nem, csak a két végpontot jelöltük címkével. A 0 pozícióban a címke „teljesen biztos vagyok benne hogy nem okoz bőrkiütéseket”, a 100 pozícióban pedig „teljesen biztos vagyok benne, hogy bőrkiütéseket okoz”. A csúszka használatához rövid instrukciót is adtunk, amelyben arra is felhívtuk a figyelmet, hogy amennyiben nem tudja eldönteni, hogy van-e oksági kapcsolat a szennyezett víz fogyasztása és a bőrkiütések között, akkor a csúszkát állítsa a skála közepére. A másik változtatás az ítélet megadásában az volt, hogy a kovariancia-információ nem tűnt el, azt a résztvevők egészen addig látták, amíg jóvá nem hagyták a válaszukat.

Az ítéletek megadását követően ugyancsak néhány általános demográfiai jellegű kérdés (nem, életkor), illetve az olvasási szövegértésre vonatkozó kérdés következett. Egy további újítás itt az első kísérlethez képest, hogy egy validáló kérdést is alkalmaztunk, amelyben azt kérdeztük meg, hogy mi volt a történetben szereplő falunak a neve (ezt az információt az instrukcióban feltüntettük, azzal a kéréssel, hogy jegyezzék azt meg)? Itt hét lehetőségből választhattak a résztvevők.

A kísérletben manipuláltuk a következmény előzmény hiányában tapasztalt valószínűségét ( $p(K=1|E=0)$ ), amelynek két szintje volt, 0,25 és 0,75, valamint a virtuális mintanagyságot ( $N$ ), aminek szintjei 40 és 280 voltak. A következmény előzmény jelenlétében tapasztalt valószínűségét 0,8-as szinten rögzítettük (az egyes feltételekben szereplő kovariancia-információkat ld. 7. táblázatban). Személyek közötti elrendezést használtunk, a résztvevőket véletlenszerűen soroltuk az egyes kísérleti csoportokba.

### ***Eredmények***

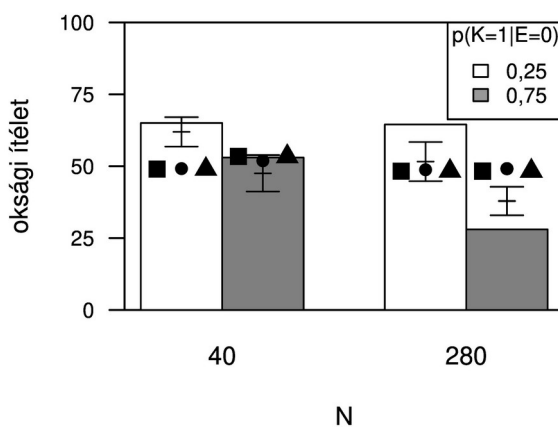
Az adatok validitásának biztosítása érdekében az instrukciós oldalon, valamint az oksági ítélet megadására (és egyben a kovariancia-információ megtekintésére) szolgáló oldalon töltött idő, az olvasási szövegértés szintje, valamint a validáló kérdésre adott válaszok alapján szűrtünk. Az egyes oldalakon töltött időket Box-Cox transzformáció segítségével alakítottuk át, majd a kilógó értékek azonosításához az első- és harmadik kvartilistól számolt, az interkvartilis terjedelem 1,5-szeresénél nagyobb távolságot határoztuk meg kritériumként. Az olvasási szövegértésnél a közepesnél (se nem rossz, se nem jó) magasabb szintet határoztunk meg a bekerülés kritériumaként, a validáló kérdés esetében pedig figyelmen kívül hagytuk azoknak az eredményeit, akik azt helytelenül válaszolták meg. További egy résztvevő adatait azért nem vehettük tekintetbe, mert nem töltötte be 18. életévét. Összesen 21 résztvevő adatait hagytuk ki az elemzésből a fenti szűrés kritériumok miatt.

Az egyes kísérleti feltételekben hasonló számú résztvevő szerepelt (18, 17, 18 és 18 fő). A feltételek között nem találtunk szignifikáns különbséget a nemek megoszlása,  $X^2(3)=5,54$   $p=0,136$ , az életkorok, *Kruskal–Wallis*  $X^2(3)=5,46$   $p=0,141$ , vagy a nyelvi készségek,  $X^2(3)=3,55$   $p=0,314$  tekintetében.

A Shapiro-Wilks próba szerint az oksági ítéletek megfelelnek a normál eloszlásnak az 1. feltételben,  $W=0,91$   $p=0,115$ , a 2. feltételben,  $W=0,96$   $p=0,656$ , valamint a 4.

feltételben,  $W=0,94$   $p=0,261$ , viszont eltérnek attól a 3. feltételben,  $W=0,86$   $p<0,05$ . Az adatok elemzésére emiatt nemparaméteres eljárásokat használtunk.

A kovariancia-információk normatív értelmezése alapján az 1. és 3. feltételben várunk az oksági kapcsolat mellett szóló ítéleteket, a 2. és 4. feltételben pedig inkább azok ellen szóló ítéleteket. A virtuális mintanagyság polarizáló hatása ráadásul ezen feltétel-párokban eltérően kellene, hogy megjelenjen: a több esetet tartalmazó kovariancia információk esetében azoknak szélsőségesebbnek kellene lenniük. Az egymintás Wilcoxon próba szerint a 4. feltételben az ítéletek valóban jelentősen eltérnek a skála közepétől (50),  $V=35,5$   $p<0,05$ , ugyanakkor a 2. feltételben nem találtunk jelentős eltérést,  $V=74$   $p=0,925$ . A következmény előzmény hiányában tapasztalt valószínűségének alacsony értéke mellett a 3. feltételben nem,  $V=91,5$   $p=0,811$ , csupán az 1. feltételben találtunk az elvárt irányú, szignifikáns eltérést a skála közepétől,  $V=120,5$   $p<0,05$ . Ez az első kísérlethez hasonlóan arra utalhat, hogy a résztvevők meglehetősen konzervatívák voltak az ítéleteikben, különösen azokban a feltételekben, ahol az oksági kapcsolat mellett szóló ítéleteket vártunk (12. ábra).



12. ábra

A második kísérlet egyes feltételeiben adott ítéletek alakulása  
(az oszlopok az ítéletek mediánját, a hibasávok az átlagot, és az átlag körüli  $\pm 1$  sztenderd hibát jelölik; a négyzet az SS-modell, a kör a ROT-modell, a háromszög a support-modell lineáris regresszióanalízis segítségével előkészített becslése)

Scheirer-Ray-Hare próba segítségével teszteltük a változóink főhatását, és a köztük lévő interakciót. Nem találtunk szignifikáns interakciót a virtuális mintanagyság és a következmény előzmény hiányában tapasztalt valószínűsége között,  $X^2(1)=0$   $p=0,993$ . A

virtuális mintanagyság főhatása sem bizonyult szignifikánsnak,  $X^2(1)=2,71$   $p=0,1$ , viszont a következmény előzmény hiányában tapasztalt valószínűségének szignifikáns főhatását találtuk,  $X^2(1)=6,73$   $p<0,01$ , amely szerint az értékelések magasabbak voltak akkor, amikor a következmény előzmény hiányában tapasztalt valószínűsége alacsony volt. Ezek az eredmények ez utóbbi főhatást tekintve megfelelnek a korábban megfogalmazott várakozásainknak, ugyanakkor az interakció hiánya a virtuális mintanagyság elvártaktól eltérő hatását jelezte.

A vizsgált modellek becslései az egyes kovariancia-információkra hasonlóan rangsorolnak egy kivételtől eltekintve: a support-modellnél a másik kettőtől eltérően rangsorol a 4. és 2. feltételben alkalmazott kovariancia-információra vonatkozó becslés. Míg a ROT-modell és az SS-modell a 4. feltételnél a gyenge összefüggés megléte miatt magasabb becslést ad, mint a 2. feltételnél, a support-modellnél ez megfordul. Ugyanakkor azt is érdemes megfigyelnünk, hogy a virtuális mintanagyság polarizáló hatása kifejezettebb formában jelenik meg a ROT-modellnél, mint az SS-modellnél, illetve a support-modellnél. Ugyan a vizsgálati személyeinknél is megjelent a polarizáló hatás, az inkább az oksági kapcsolat hiánya irányában hatott függetlenül attól, hogy a következmény előzmény hiányában tapasztalt valószínűsége alacsony, vagy magas volt.

A modellek illeszkedése mind a négy kovariancia-információt tekintetbe véve kifejezetten rossznak mutatkozott; a ROT-modell esetében,  $R^2=0,02$ , az SS-modell esetében,  $R^2=0,06$ , és a support-modell esetében is,  $R^2=0,06$ .

#### 7. táblázat

A résztvevők által adott válaszok átlaga, valamint a vizsgált modellek becslései a második kísérlet egyes feltételeiben

	N	$p(K=1 E=0)$	$N_a$	$N_b$	$N_c$	$N_d$	$M^a$	modellek predikciói		
								ROT-modell	SS-modell	support-modell
1.	40	0,25	16	4	5	15	61,94	7,48	6,12	5,43
2.	40	0,75	16	4	15	5	47,53	1,18	-0,07	-0,28
3.	280	0,25	112	28	35	105	51,61	51,78	43,79	42,97
4.	280	0,75	112	28	105	35	37,89	7,08	0,12	-0,4

<sup>a</sup> a résztvevők által adott értékelések átlaga

Érdeemes megfigyelnünk azonban, hogy a becslések rangsorolása a virtuális mintanagyság kontrollálása mellett mindegyik modell esetében megfelel annak, ahogy a résztvevők ítéletei rangsorolnak (2. majd 1. feltétel kis virtuális mintanagyság mellett, valamint 4. majd 3. feltétel nagy virtuális mintanagyság mellett). Ez erős bizonyítéknak tekinthető amellet, hogy a modellek becslései a virtuális mintanagyság eltérő kezelése miatt mutatnak jelentős eltéréseket a résztvevők értékeléseitől, nevezetesen a résztvevőkhöz képest túlbecsülik annak polarizáló hatását.

### **3.2.5. Harmadik kísérlet: $p(K=1|E=1)$ és $p(K=1|E=0)$ hatásának vizsgálata nagy virtuális mintanagyság esetén**

A harmadik kísérletben az első kísérletet ismételtük meg, ugyancsak rögzítve a virtuális mintanagyságot – annak ezúttal egy magasabb szintjét választottuk. A második kísérletben ismertetett módosításokat itt is alkalmaztuk, továbbá kiegészítettük a feladatot a rejtett oksági tényezők jelenlétére vonatkozó valószínűség becslésére vonatkozó kérdésekkel annak érdekében, hogy a modell azok szerepére vonatkozó központi feltevésének megfelelőségéről is képet kaphassunk.

#### ***Minta***

Összesen 103-an vettek részt a kísérletben, amelyből 72 fő adatait ítéltük elemzésre alkalmasnak. A minta 63,9%-a (46 fő) nő, 36,1%-a (26 fő) férfi. Az átlagos életkor 38,6 év (21-70 év,  $SD = 11,03$ ).

A részvételért \$0,03-t fizettünk a vizsgálati személyeknek, és mindenki, aki befejezte a feladatot további \$0,15 bónuszt kapott.

#### ***Módszerek és eljárás***

Mint a korábbi kísérletekben, a kísérletet a Qualtrics platformján publikáltuk, és a misszionárius-feladatot alkalmaztuk. A kovariancia-információt a második kísérletben ismertetett módon, mátrix-ként mutattuk be, és ugyanolyan 101 fokú, kvázi-analóg skálát alkalmaztunk, amelyen a résztvevők egy csúszka segítségével adhatták meg válaszaikat.

A korábbi kísérletekhez képest újítás volt, hogy az oksági ítélet megadását követően megkérdeztük a résztvevőktől, hogy mennyire tartják valószínűnek, hogy jelen volt valamilyen egyéb tényező, ami okozhatott bőrkiütéseket (generatív oksági tényező), illetve ami megakadályozhatta, hogy a bennszülötteknek bőrkiütése legyen (preventív). Mindkét

kérdésre egy 7 fokú skálán adhatták meg a válaszaikat a vizsgálati személyek – a skála egyes pontjait címkékkel jelöltük, a következők szerint:

- 1 = nagyon valószínűtlen
- 2 = valószínűtlen
- 3 = némileg valószínűtlen
- 4 = nem tudom eldönteni
- 5 = némileg valószínű
- 6 = valószínű
- 7 = nagyon valószínű

Az adatok megfelelő minőségének biztosítása céljából néhány módosítást vezettünk be az eljárásban, amely azonban annak korábban is alkalmazott felépítését lényegében nem érintette. Ilyen volt a kísérlet záróoldalán szereplő kérdések között két további kérdés feltüntetése. Az egyik kérdés arra vonatkozott, hogy a válaszadó vett-e már részt ehhez hasonló vizsgálatban az elmúlt néhány hónap folyamán (a dolgozatban ismertetett egyéb kísérletek közül többel, valamint egyéb, hasonló témában folytatott kísérleteinkkel is találkozhattak a résztvevők). A másik pedig arra, hogy tapasztalt-e valamilyen furcsaságot, vagy technikai problémát a csúszka használata során, amely a válasza megadására szolgált (erre vonatkozóan ugyanis kaptunk néhány megjegyzést ugyanezen platformon végzett egyéb vizsgálataink során).

Akárcsak az első kísérletben, itt is két változót manipuláltunk, a következmény előzmény jelenlétében ( $p(K=1|E=1)$ ), és annak hiányában tapasztalt valószínűségét ( $p(K=1|E=0)$ ). Mindkét változónak két-két szintjét határoztuk meg, amelyek 0,25 és 0,75 voltak. A virtuális mintanagyság minden esetben 340 volt (az egyes kísérleti feltételekben alkalmazott kovariancia-információt ld. 8. táblázatban). Személyek közötti elrendezést használtunk, a résztvevőket véletlenszerűen rendeltük a kísérleti feltételekhez.

### ***Eredmények***

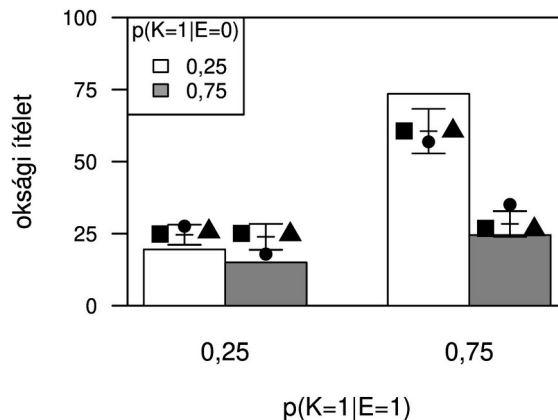
Csak azoknak a résztvevőknek az eredményeit használtuk fel az elemzés során, akiknek az instrukciós oldalon, valamint a kovariancia-információt prezentáló, és oksági ítéletet bekérő oldalon töltött idejük nem különbözött lényegesen a többségtől (itt a korábbi kísérletekben alkalmazott módon transzformált változókat, és a kvartilisek, valamint interkvartilis terjedelem alapján meghatározott kritériumot használtuk), a validáló kérdésre helyes választ

adatok, az olvasási szövegértésük meghaladta a közepes (se nem rossz, se nem jó) szintet, nem jeleztek problémát az értékelés megadására szolgáló csúszka működésében, valamint azt jelezték, hogy biztosan nem vettek részt ehhez hasonló vizsgálatban az elmúlt néhány hónapban. Összesen 72-en feleltek meg ezeknek a kritériumoknak, 31 résztvevő adatait nem vettük figyelembe az elemzés során.

Az egyes kísérleti csoportok mérete mutatott némi eltérést (18, 20, 14 és 20 fő), ugyanakkor nem találtunk közöttük szignifikáns eltérést sem a nemek megoszlása,  $X^2(3)=3,56$   $p=0,313$ , sem a nyelvi készségek,  $X^2(3)=3,7$   $p=0,295$ , sem életkor tekintetében, *Kruskal–Wallis*  $X^2(3)=3,67$   $p=0,299$ .

A Shapiro-Wilk próba alapján jelentősen eltértek a válaszok eloszlásai a normál eloszlástól az 1. feltételben,  $W=0,81$   $p<0,01$ , a 2. feltételben,  $W=0,85$   $p<0,01$ , valamint a 3. feltételben,  $W=0,82$   $p<0,01$ , és megfelelték annak a 4. feltételben,  $W=0,94$   $p=0,22$ . Ennek megfelelően az adatok elemzése során nemparaméteres eljárásokat használtunk.

A nagy virtuális mintanagyság ellenére – ahogy a korábbi kísérletekben – itt is konzervatívabb értékeléseket találtunk (13. ábra). A 3. feltételben, ahol a legmagasabb értékeléseket vártuk (és a többi feltételhez képest ez így is alakult), az értékelések nem tértek el jelentősen a skála közepétől (50) az egymintás Wilcoxon-próba alapján,  $V=73$   $p=0,217$ . Az értékelések lényegesen alacsonyabbak voltak a skála közepétől az 1. feltételben,  $V=0$   $p<0,001$ , a 2. feltételben,  $V=10$   $p<0,001$ , valamint a 4. feltételben,  $V=15,5$   $p<0,001$ .



13. ábra

A harmadik kísérlet egyes feltételeiben adott ítéletek alakulása (az oszlopok az ítéletek mediánját, a hibásávok az átlagot, és az átlag körüli  $\pm 1$  sztenderd hibát jelölik; a négyzet az SS-modell, a kör a ROT-modell, a háromszög a support-modell lineáris regresszióanalízis segítségével előkészített becslése)

A Scheirer-Ray-Hare próba alapján nem bizonyult szignifikánsnak a független változók közötti interakció,  $X^2(1)=2,48$   $p=0,116$ , ugyanakkor mind a következmény előzmény jelenlétében,  $X^2(1)=5,67$   $p<0,05$ , mind annak hiányában tapasztalt valószínűsége,  $X^2(1)=5,99$   $p<0,05$ , szignifikáns hatással volt az okszági ítéletekre. Ezek szerint tehát itt a két változó egyszerű additív hatását támasztják alá az eredmények.

A vizsgált modellek által adott predikciók az egyes feltételekre mind egymáshoz képest, mind a résztvevők átlagos értékeléseikhez képest hasonlóan rangsoroltak (8. táblázat), tehát kvalitatív szinten nagyon jól illeszkednek az adatokra. Finomabb skálán vizsgálva megfelelő illeszkedést találtunk mindhárom modellnél, ugyanakkor a ROT-modell illeszkedése,  $R^2=0,89$ , némileg elmarad az SS-modell,  $R^2=0,99$ , és support-modell,  $R^2=0,99$ , illeszkedésétől. Ez magyarázható lehet a szélsőségesebb becslésekkel az 1-3. feltételben, amely utalhat a következmény előzmény jelenlétében, illetve hiányában tapasztalt valószínűségének markánsabb hatására a ROT-modell esetében.

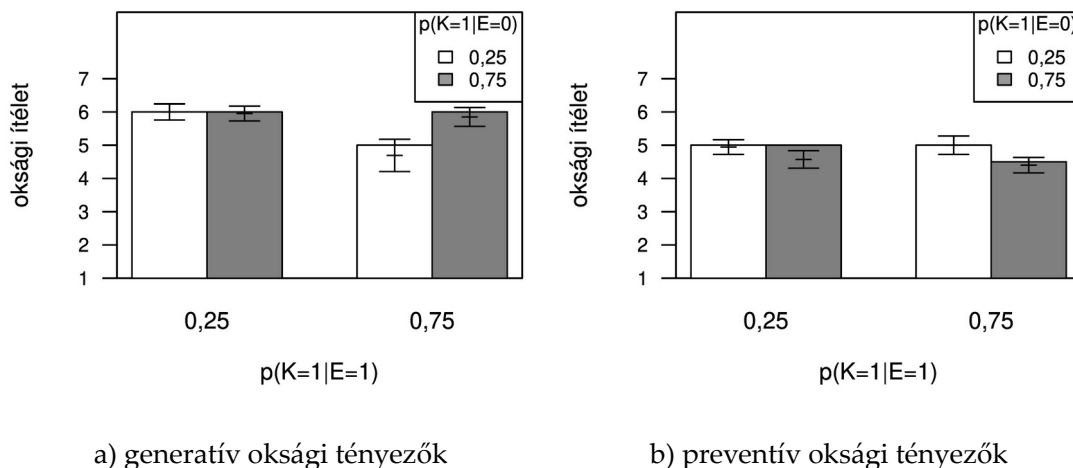
8. táblázat

A résztvevők által adott válaszok átlaga, valamint a vizsgált modellek becslései a harmadik kísérlet egyes feltételeiben

	$p(K=1 E=1)$	$p(K=1 E=0)$	$N_a$	$N_b$	$N_c$	$N_d$	$M^e$	modellek predikciói		
								ROT-modell	SS-modell	support-modell
1.	0,25	0,25	34	136	34	136	24,61	-19,08	-4,17	-2,71
2.	0,25	0,75	34	136	136	34	23,81	-48,34	-3,76	-4,64
3.	0,75	0,25	136	34	34	136	59,15	69,54	64,47	63,32
4.	0,75	0,75	136	34	136	34	28,35	3,69	-0,49	-1,39

<sup>a</sup> a résztvevők által adott értékelések átlaga

A rejtett generatív- és preventív oksági tényezők jelenlétére vonatkozó ítéletek alakulását a 14. ábrán ábrázoltuk. A legtöbb feltétel esetében a résztvevők inkább valószínűnek gondolták az alternatív generatív okok jelenlétének a valószínűségét, az értékelések szignifikánsan magasabbak voltak, mint a skála közepe (ami a teljes bizonytalanságot jelzi) az 1. feltétel,  $V=136$   $p<0,001$ , a 2. feltétel,  $V=210$   $p<0,001$ , a 4. feltétel,  $V=185$   $p<0,001$  esetében, és nem tértek el szignifikánsan a 3. feltétel,  $V=62$   $p=0,255$  esetében. A preventív oksági tényezők jelenlétére vonatkozó ítéletek a teljes bizonytalanságot tükrözték a 4. feltételben, ahol a következmény valószínűsége mind az előzmény hiánya, mind jelenléte esetén magas volt,  $V=57$   $p=0,154$ , ugyanakkor a résztvevők inkább biztosak voltak azok jelenlétében az 1.,  $V=66$   $p<0,01$ , a 2.,  $V=98$   $p<0,05$ , és a 3. feltételben,  $V=73,5$   $p<0,01$ .



14. ábra

A rejtett okszági tényezők jelenlétének valószínűségére vonatkozó ítéletek alakulása a harmadik kísérletben  
(az oszlopok az ítéletek mediánját, a hibasávok az átlagot, és az átlag körüli  $\pm 1$  sztenderd hibát jelölik)

A Scheirer-Ray-Hare próba alapján nem találtunk jelentős eltéréseket a következmény előzmény jelenlétében tapasztalt valószínűsége mentén sem a generatív,  $X^2(1)=2,84$   $p=0,092$ , sem a preventív,  $X^2(1)=0,04$   $p=0,841$ , okszági tényezők jelenlétére vonatkozó ítéletekben. Hasonló eredményeket kaptunk a következmény előzmény hiányában tapasztalt valószínűségére vonatkozóan, generatív okszági tényezők esetén,  $X^2(1)=2,54$   $p=0,111$ , és preventív okszági tényezők esetén,  $X^2(1)=2,95$   $p=0,086$ . Nem találtunk interakciót sem a generatív,  $X^2(1)=2,63$   $p=0,11$ , sem a preventív okszági tényezők jelenlétére vonatkozó ítélet estében,  $X^2(1)=0,58$   $p=0,448$ .

Mivel azokat a ROT-modell a számítások során explicit módon ábrázolja, tudunk becsléseket tenni a generatív- és preventív okszági tényezők felülvizsgált valószínűségére. Ezek alapján jellegzetes eltéréseket várhatunk azokban (9. táblázat). Fontos látnunk, hogy ezen adatok alapján lehetőségünk van a ROT-modell egyik köztes lépésének – a rejtett okszági tényezők jelenlétére vonatkozó bizonytalanság felülvizsgálatának – adekvát jellegét közvetlenül ellenőrizni. A modell becslései mutatnak jellegzetes eltéréseket a résztvevők ítéleteitől, amely különösen az 1. feltételben szembeötlő, ahol a modell mind közül a legalacsonyabb valószínűséget becsüli a generatív okszági tényezőkre, és a legmagasabbat a

preventív oksági tényezőkre. A résztvevők ítéletei ugyanakkor itt meglehetősen magasak a generatív oksági tényezőkre, és nem mutat markáns eltérést a többi feltételtől a preventív oksági tényezőkre. A modell kifejezetten rosszul magyarázza az értékelések alakulását mind a generatív,  $R^2=0,14$ , mind a preventív oksági tényezőkre esetében,  $R^2=0,54$ .

#### 9. táblázat

A generatív-, és preventív oksági tényezők jelenlétének valószínűségére vonatkozó ítéletek, és a ROT-modell ez irányú becslései a harmadik kísérlet egyes feltételeiben

	$p(K=1 E=1)$	$p(K=1 E=0)$	$N_a$	$N_b$	$N_c$	$N_d$	$M_{gen.}^a$	$M_{prev.}^b$	<i>ROT-modell becslése<sup>c</sup></i>	
									generatív	preventív
1.	0,25	0,25	34	136	34	136	6	4,94	0,42	0,53
2.	0,25	0,75	34	136	136	34	6,1	4,65	0,62	0,33
3.	0,75	0,25	136	34	34	136	4,5	5,07	0,47	0,33
4.	0,75	0,75	136	34	136	34	5,85	4,4	0,67	0,14

<sup>a</sup> A résztvevők átlagos ítéletei a generatív oksági tényezők jelenlétének valószínűségére

<sup>b</sup> A résztvevők átlagos ítéletei a preventív oksági tényezők jelenlétének valószínűségére

<sup>c</sup> A becsléseket az egyes oksági tényezők jelenlétének valószínűségében való bizonytalanságot reprezentáló, felülvizsgált béta eloszlások móduszaként határoztuk meg

Annak érdekében, hogy feltárjuk ezen eltérések okait, mindegyik feltételben külön-külön megvizsgáltuk, hogy a rejtett oksági tényezőkre vonatkozó becslések segítségével milyen hatékonyan magyarázhatók az oksági kapcsolatra vonatkozó ítéletek – ez az elemzés rámutathat a rejtett oksági tényezők szerepére az oksági ítéletek alkotásában. Az elemzés során az egyes rejtett oksági tényezők jelenlétére vonatkozó ítéletek és az oksági ítéletek közötti összefüggéseket vizsgáltuk meg, továbbá többszörös lineáris regresszióanalízis segítségével ellenőriztük, hogy van-e interakció a két rejtett oksági tényezőre vonatkozó ítélet között.

Az 1. feltételben, ahol mind a következmény előzmény jelenlétében, mind annak hiányában tapasztalt valószínűsége alacsony, nem találtunk szignifikáns összefüggést az oksági ítéletek és generatív oksági tényezőkre vonatkozó ítéletek,  $r=-0,11$   $p=0,657$ , sem az oksági ítéletek és a preventív oksági tényezőkre vonatkozó ítéletek,  $r=0,14$   $p=0,583$ , között. A lineáris regresszióanalízis eredménye alapján viszont úgy tűnik a két tényező között van interakció – csak az interakciót tartalmazó modell illeszkedett az adatokra,  $F(3,14)=3,77$   $p<0,05$ , annak magyarázóereje nem volt elhanyagolható,  $R^2=0,33$ . Az

interakció szerint a generatív oksági tényezők jelenlétének valószínűsége pozitív hatással bír az oksági ítéletekre, ha a preventív oksági tényezők valószínűsége magas, míg inkább negatív hatással, ha a preventív oksági tényezők valószínűsége alacsonyabb.

A 2. feltételben szignifikáns negatív összefüggést találtunk a generatív oksági tényezők megítélt valószínűsége és az oksági ítéletek között,  $r = -0,56$   $p < 0,05$ , és nem találtunk szignifikáns összefüggést a preventív oksági tényezőkre vonatkozó ítéletek és az oksági ítéletek között,  $r = -0,36$   $p = 0,117$ . Itt csak az additív modell mutatott megfelelő illeszkedést az adatokra,  $F(2,17) = 3,89$   $p < 0,05$   $R^2 = 0,23$ , amelyen belül csak a generatív oksági tényezőknél volt jelentős (negatív) hatása az oksági ítéletekre.

3. feltételben szereplő kovariancia-információ tekinthető normatív szempontból a legerősebb bizonyítéknak az oksági kapcsolat megléte mellett. Itt úgy találtuk, hogy mind a generatív oksági tényezőkre vonatkozó ítéletek,  $r = -0,65$   $p < 0,05$ , mind a preventív oksági tényezőkre vonatkozó ítéletek,  $r = -0,55$   $p < 0,05$  szignifikáns negatív összefüggést mutattak az oksági ítéletekkel. Bár az adatokra jó illeszkedést mutatott mind az egyszerű additív modell,  $F(2,11) = 11,52$   $p < 0,01$   $R^2 = 0,62$ , mind az interakciót tartalmazó modell,  $F(3,9) = 12,02$   $p < 0,01$   $R^2 = 0,72$ , az interakciós modell magyarázóereje – annak ellenére, hogy az interakció mértéke nem volt teljesen elhanyagolható – nem volt szignifikánsan nagyobb,  $F(1,9) = 4,89$   $p = 0,052$ . Ezek szerint ha a résztvevők bármely rejtett oksági tényező valószínűségét nagyobbobbnak ítélték, akkor kevésbé valószínűnek találták az oksági kapcsolat meglétét.

A 4. feltételben szereplő kovariancia-információban a következmény valószínűsége magas volt függetlenül attól, hogy az előzmény jelen volt, vagy sem. Az oksági ítéletekkel szignifikáns negatív összefüggést találtunk a generatív oksági tényezőkre vonatkozó ítéleteknél,  $r = -0,78$   $p < 0,001$ , és nem találtunk szignifikáns összefüggést a preventív oksági tényezőkre vonatkozó ítéleteknél,  $r = 0,34$   $p = 0,148$ . Mind az additív modell,  $F(2,17) = 15,25$   $p < 0,001$   $R^2 = 0,6$ , mind az interakciót tartalmazó modell,  $F(3,16) = 10,28$   $p < 0,001$   $R^2 = 0,59$  jól illeszkedett az adatokra, viszont az interakciós modell magyarázóereje nem volt szignifikánsan jobb, mint az additív modellé,  $F(1,16) = 0,76$   $p = 0,396$ .

A kísérletben szereplő egyes kovariancia-információkra vonatkozó oksági ítéletek, valamint az azokban szerepet játszó rejtett oksági tényezők jelenlétére vonatkozó ítéletek

fenti vizsgálata egy érdekes dologra mutathat rá, nevezetesen, hogy a rejtett oksági tényezők hatása az oksági ítéletekre bizonyos esetekben interakciós hatás is lehet – ez alapvetően eltérhet a ROT-modellben feltételezett additív hatástól. Egyetlen feltételben sikerült szignifikánsan kimutatnunk az interakciós hatást (az 1. feltételben). Amennyiben ezt a feltételt figyelmen kívül hagyjuk, akkor a ROT-modell oksági tényezők jelenlétének valószínűségére vonatkozó becslései jobban magyarázzák a résztvevők ezekre vonatkozó tényleges ítéleteit a generatív oksági tényezők esetében,  $R^2=0,86$ , és a korábbihoz némileg hasonlóan a preventív oksági tényezők esetében,  $R^2=0,62$ . Meg kell ugyanakkor jegyeznünk, hogy ezek az eredmények, bár a modell bizonyos feltevéseire vonatkozó érdekes következtetésre adnak lehetőséget, további megerősítésre szorulnak, hiszen itt meglehetősen kevés adattal dolgozhattunk (az egyik feltételben mindössze 14-en vettek részt), és a rejtett oksági tényezőkre vonatkozó becslések is inkább a magasabb tartományban mozogtak.

### **3.3. Illeszkedés vizsgálata online tanulás esetében**

A ROT-modellt az egyszerű oksági kapcsolatok összegzett kovariancia-információ alapján való tanulására készítettük fel. Itt megvizsgáljuk, hogy a modell által képviselt magyarázat alkalmazható-e az online tanulásra, ahol a kovariancia-információ nem összegzett formában áll a tanuló rendelkezésére, hanem szekvenciálisan, egyesével érkezik. Ennek során a Danks, Griffiths és Tenenbaum (2003) által javasolt eljárást használtuk a ROT-modell predikcióinak a levezetésére. Az eljárás lényege, hogy a kovariancia-információt, amelyet a modell a becsléséhez felhasznál, a tanulás minden egyes lépésében bővítjük az újonnan érkező adattal – a tanulás utolsó lépésében, amikor már a kovariancia-információ teljes, a modell becslése megegyezik az eredeti, összegzett adatokra vonatkozó becsléssel.

#### **3.3.1. Negyedik kísérlet: valószínűségi kontraszt ( $\Delta P$ ) hatásának a vizsgálata**

A kísérletben a célunk az volt, hogy képet kapjunk arról, hogy az oksági ítéletek hogyan változnak abban az esetben, amikor az oksági kapcsolat, amelyet éppen vizsgálunk egy mérsékelt erős generatív kapcsolat, vagy egy mérsékelt erős preventív kapcsolat, vagy esetleg nincs is oksági kapcsolat. Mivel a ROT-modell az oksági kapcsolat meglétében való bizonyosságra vonatkozóan ad becsléseket, itt is az ezekre vonatkozó becslések alakulását

vizsgáltuk, míg a szakirodalomban jellemzően az oksági kapcsolatok erősségére, súlyára vonatkozó becslések állnak a vizsgálódások középpontjában.

### ***Minta***

A kísérletet összesen 150-en fejezték be, akik közül 96 résztvevő adatait tartottuk elemzésre alkalmasnak. A mintában szereplők 75%-a (72 fő) nő, 25%-a (24 fő) férfi. Az átlagos életkor 35,1 év (18-64 év,  $SD=10,01$ ).

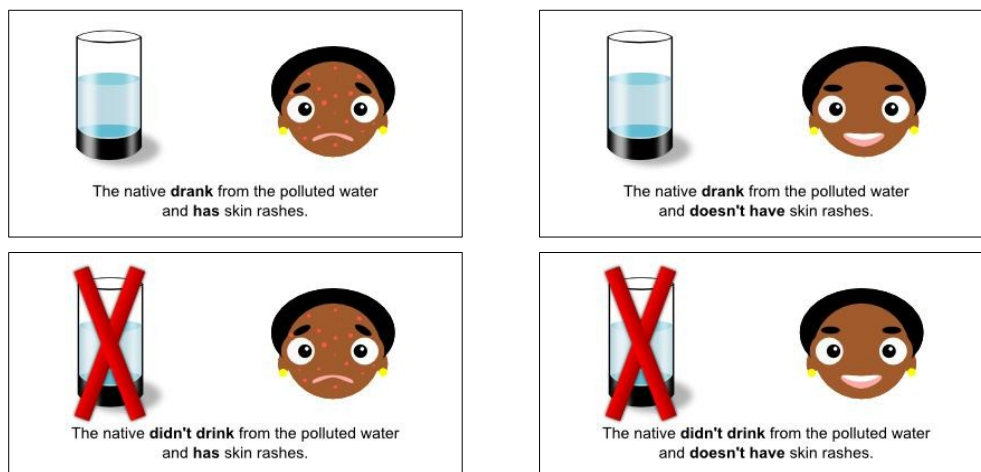
A résztvevőknek \$0,10-et fizettünk a vizsgálatban való részvételért, és azoknak, akik befejezték az összes feladatot, és válaszoltak minden kérdésre, további \$0,25-t fizettünk jutalomként.

### ***Módszerek és eljárás***

A kísérletet a Qualtrics platformján publikáltuk, a korábbi kísérletekben már megismert misszionárius feladatot használtuk, ahol a vizsgálati személynek az volt a feladata, hogy egy bennszülöttek által lakott faluban megállapítsa, hogy a több lakosnál megjelenő bőrkiütéseket okozhatta-e a faluban található egyik kút vizének a fogyasztása, amely feltehetően szennyezett volt. A résztvevő itt a vizsgálódásai eredményét nem összegzett formában, hanem egyesével ismerhette meg, és mindegyik eset után meg kellett ítélnie, hogy mennyire van meggyőződve arról, hogy a szennyezett víz fogyasztása bőrkiütéseket okoz.

A résztvevők az első instrukciós oldalon megismerkedhettek a háttértörténettel, és megismerkedtek azzal, hogy mi a feladatuk a kísérlet során. A második instrukciós oldalon részletesen bemutattuk nekik, hogy milyen lehetséges adatokkal fognak majd találkozni a feladat során (a négy eseménnytípusnak megfelelően).

Ezt követően kértünk egy ítéletet a szennyezett víz fogyasztása és a bőrkiütések közötti oksági kapcsolatra vonatkozóan mielőtt még bármilyen bennszülöttre vonatkozó adatot láttak volna a résztvevők, majd egyesével mutattuk egy-egy bennszülött vizsgálatának eredményét, és a skálát, amelyen megadhatták a résztvevők az oksági kapcsolatra vonatkozó értékelésüket. Az egyes eseteket szövegesen, valamint egy-egy kártya formájában is prezentáltuk (15. ábra). Az értékeléshez kapcsolódó instrukcióban felhívtuk a résztvevők figyelmét arra, hogy az értékelésüket az addig látott összes eset figyelembevételével adják meg. Minden résztvevő összesen 16 esetet látott, így a kezdeti ítéletet is beleszámolva, 17 alkalommal kellett oksági ítéletet alkotniuk; mindegyik esetben a 2. és 3. kísérletből ismerős 101 fokú, kvázi-analóg skálát használtuk.



15. ábra

A negyedik kísérletben alkalmazott, az egyes eseménytípusokat ábrázoló „kártyák”

Az online tanulási feladatot követően kértünk becsléseket az egyes eseménytípusok gyakoriságára vonatkozóan, itt a résztvevők szabadon írhattak bármilyen értéket, majd további kérdéseket kellett megválaszolniuk a nemükkel, életkorukkal, iskolai végzettségükkel, olvasási szövegértési képességükkel kapcsolatban. Az adatok minőségének ellenőrizhetősége céljából továbbá megkérdeztük, hogy rendben működött-e az értékelés megadására szolgáló csúszka a feladat során, valamint egy validáló kérdést is használtunk, amelyben az instrukcióban elhelyezett információra (a falu nevére) kérdeztünk rá.

A kísérlet során manipuláltuk a valószínűségi kontrasztot, amely lényegében a következmény előzmény jelenlétében tapasztalt valószínűségének ( $p(K=1|E=1)$ ) és a következmény előzmény hiányában tapasztalt valószínűségének ( $p(K=1|E=0)$ ) a különbsége, azaz  $\Delta P = p(K=1|E=1) - p(K=1|E=0)$ . Három szintet különböztettünk meg a független változókban, az egyik szint egy mérsékelt negatív kapcsolatot jelzett a szennyezett víz fogyasztása és a bőrkiütések között (1. feltétel,  $\Delta P = -0,5$ ,  $\{2,6,6,2\}$ ), a másik az oksági kapcsolat hiányára utalt (2. feltétel,  $\Delta P = 0$ ,  $\{4,4,4,4\}$ ), a harmadik pedig mérsékelt pozitív kapcsolatot jelzett (3. feltétel,  $\Delta P = 0,5$ ,  $\{6,2,2,6\}$ ). A résztvevők számára a kovariancia-információt alkotó eseteket véletlenszerű sorrendben prezentáltuk.

Személyek közötti elrendezést használtunk, a résztvevőket véletlenszerűen soroltuk a kísérleti csoportokba.

## ***Eredmények***

Annak érdekében, hogy az adatok minőségét magas szinten tartsuk, figyelmen kívül hagytuk azon résztvevők adatait, akikre az alábbiak legalább egyike igaz volt:

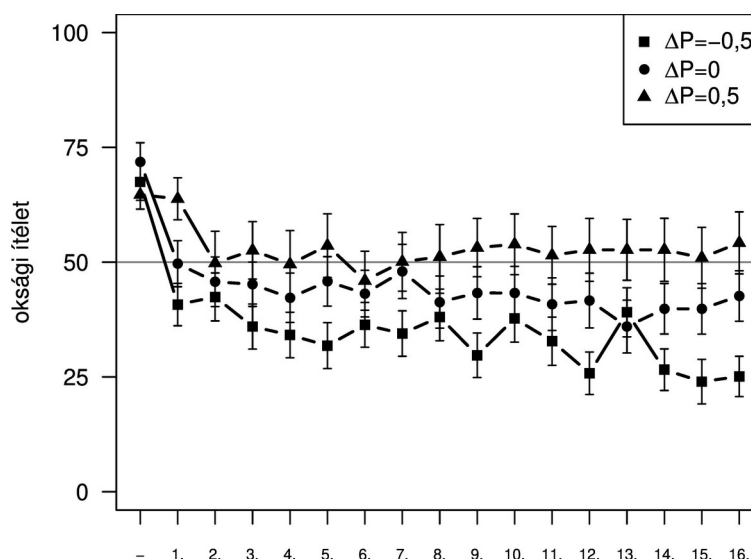
- az instrukciós oldalak bármelyikén eltöltött ideje jelentősen eltért a többségtől (ehhez az egyes oldalakon töltött időket Box-Cox transzformációval alakítottuk át, majd kilógó értéknek tekintettük azokat az értékeket, amelyek az első-, vagy harmadik kvartilistól számolt interkvartilis terjedelem 1,5-szeresénél is távolabb voltak az értékek többségétől)
- saját bevallása szerinti olvasási szövegértési képessége nem haladta meg a közepes (se nem rossz, se nem jó) szintet
- jelezte, hogy a válasza megadásakor problémát okozott a csúszka használata
- a validáló kérdésre rossz választ adott
- minden esetre ugyanazt a választ adta (például végig 0 értékelést adott)
- az egyes eseménytípusok gyakoriságának a becslésekor nem számot adott meg válaszként, vagy bármely eseménytípusnál jelentősen túlbecsülte a gyakoriságot (nagyobb gyakoriságot adott meg, mint a ténylegesen látott esetek száma)

Ezen okok miatt összesen 54 résztvevő adatait kellett figyelmen kívül hagynunk.

A kísérleti csoportok méretében volt némileg jelentősnek tűnő eltérés (38, 33 és 25 fő). A kísérleti csoportok között nem találtunk lényeges eltérést a nemek szerinti megoszlás,  $X^2(2)=2,7$   $p=0,259$ , a nyelvi készség,  $X^2(2)=1,86$   $p=0,395$ , illetve az életkor, *Kruskal–Wallis*  $X^2(2)=2,53$   $p=0,282$ , tekintetében.

Az oksági kapcsolatra vonatkozó első, adatok hiányában alkotott ítéletek szignifikáns eltérést mutattak a skála közepétől (50) az 1. feltétel,  $V=619,5$   $p<0,001$ , a 2. feltétel,  $V=494$   $p<0,001$ , és a 3. feltétel esetében is,  $V=220$   $p<0,001$ , az egymintás Wilcoxon-próba eredménye alapján. Ugyanakkor a kezdeti értékelésekben nem volt különbség az egyes feltételek között, *Kruskal–Wallis*  $X^2=4,01$   $p=0,135$ . A résztvevők mindegyik feltételben inkább az oksági kapcsolat megléte mellett foglaltak állást, ami arra utal, hogy annak ellenére, hogy nem rendelkeztek konkrét adatokkal, amelyek az oksági kapcsolat meglétére, vagy hiányára utaltak volna, jellegzetes előzetes elvárással (az oksági kapcsolat megléte) rendelkeztek azt illetően.

Az utolsó esetre (tehát a teljes kovariancia-információ alapján) adott értékelések szignifikáns eltérést mutattak a skála közepétől az 1. feltételben,  $V=86,5$   $p<0,001$ , ahol az értékelések inkább az oksági kapcsolat elutasítása mellett szóltak. Nem tértek el jelentősen a skála közepétől az értékelések a 2. feltételben,  $V=212,5$   $p=0,228$ , valamint a 3. feltételben,  $V=166,5$   $p=0,647$ . A három feltétel között találtunk szignifikáns eltéréseket, *Kruskal–Wallis*  $X^2(2)=13,1$   $p<0,01$ ; a Mann-Whitney próba alapján az 1. feltételben adott értékelések szignifikánsan alacsonyabbak mint a 2. feltételben,  $W=410$   $p<0,05$ , valamint a 3. feltételben,  $W=240$   $p<0,001$  adott értékelések. A 2. és 3. feltételben adott értékelések között nem volt szignifikáns a különbség,  $W=321,5$   $p=0,155$ . Az utolsó értékelések a kovariancia-információ normatív értelmezésének megfelelően alakultak, amennyiben a  $\Delta P=0,5$  esetében voltak a legmagasabbak és a  $\Delta P=-0,5$  esetében a legalacsonyabbak.

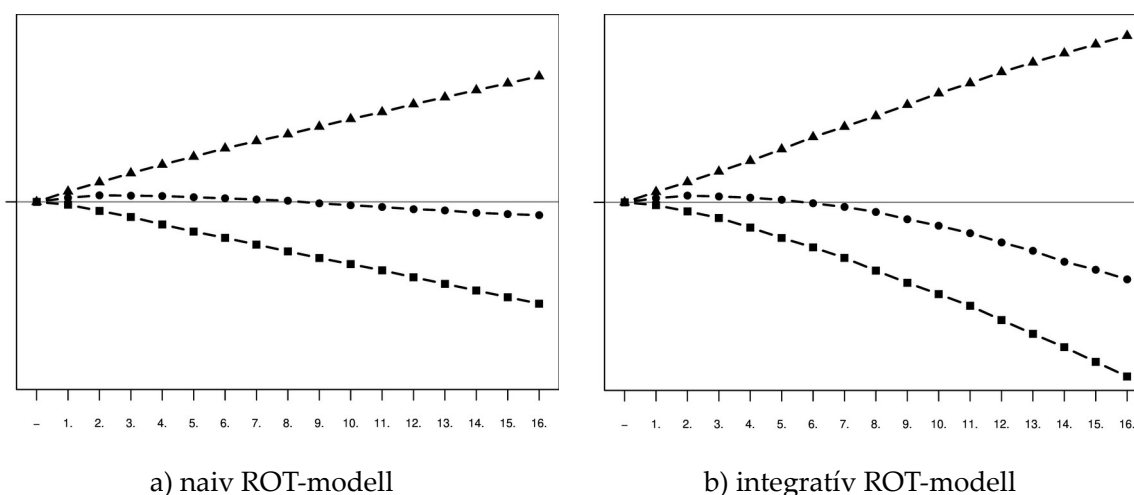


16. ábra

A negyedik kísérlet egyes feltételeiben adott értékelések átlagai az egyes kísérleti feltételekben (a hibásávok az átlag körüli  $\pm 1$  sztenderd hiba intervallumot jelölik)

A ROT-modell becsléseit az egyes feltételekben a Danks és mtsai. (2003) által javasolt módon vezettük le. Ehhez a kovariancia-információt minden egyes lépésnél egy további esettel bővítettük, és erre az új, több esetet tartalmazó kovariancia-információra határoztuk meg a modell becslését. Ahhoz, hogy az esetek konkrét sorrendjétől független becsléseket kapjunk, minden feltételnél 1000-szer ismételtük meg az eljárást; mindig az esetek egy új, random sorrendjét használtuk, majd az így kapott becsléseket minden lépésben átlagoltuk. A ROT-modell két változatát próbáltuk ki: az egyik változatot naiv ROT-modellnek, a másikat

integratív ROT-modellnek neveztük; az alapvető különbség a két változat között az volt, hogy hogyan kezelték a rejtett oksági tényezők jelenlétére vonatkozó elvárásokat. A naiv modell minden egyes lépésnél külön-külön maximálisnak tételezte fel ezt a bizonytalanságot, míg az integratív modell a korábbi lépésben látott kovariancia-információ alapján már felülvizsgált bizonytalanságot alkalmazta; ez utóbbi esetben tehát a rejtett oksági tényezőkre vonatkozó közvetett információ akumulálódhatott a tanulás során.



17. ábra

A ROT-modell két változata által becsült tanulási görbék a negyedik kísérlet egyes feltételeire (háromszög:  $\Delta P=0,5$ , kör:  $\Delta P=0$ , négyzet:  $\Delta P=-0,5$ )

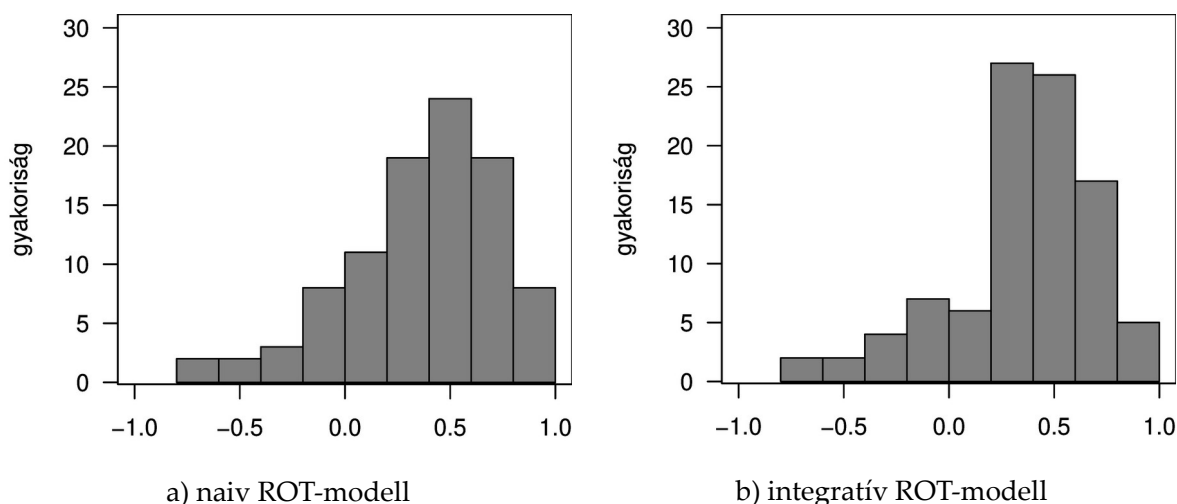
A ROT-modell két változata által becsült tanulási görbék (17. ábra) lefutása alapvető jellegét tekintve hasonló, ugyanakkor az integratív modell esetében nagyobb mértékű az értékelések polarizálódása. Érdekes, hogy az oksági kapcsolat hiányára utaló kovariancia-információ esetében ( $\Delta P=0$ ) megjelenik az értékelések enyhén pozitív (okszági kapcsolat meglétére utaló) jellege a kezdeti eseteknél, amely aztán később negatív irányba tart (inkább már a kapcsolat hiányára utal). Ezt a hatást már sikerült kimutatni hasonló feltételek mellett, igaz nem az oksági kapcsolat meglétére vonatkozó értékelések, hanem az oksági kapcsolat erősségére vonatkozó ítéletek esetében (pl. Shanks, 1995).

A naiv- és integratív ROT-modell becsléseinek illeszkedését a kísérlet során gyűjtött adatokra megítélhetjük a tanulási görbék lefutásának jellegét vizsgálva. Ebben a tekintetben azt mondhatjuk, hogy a kísérlet résztvevőinél tapasztalt tanulási görbék mutatnak hasonlóságokat a modellek által becsülthöz: a 3. feltétel tanulási görbéje végig felül fut, az 1. feltétel pedig végig alul, míg a 2. feltétel tanulási görbéje a kettő között helyezkedik el. A

jellegzetes eltérések ugyanakkor a résztvevők ítéleteinek konzervatív jellegére utalnak, azok még a 3. feltételben is végig inkább az oksági kapcsolat meglétében való bizonytalanságra utalnak a tanulás teljes tartama alatt.

Kereszt-korreláció segítségével ítéltük meg az egyes feltételekben a modell illeszkedését a résztvevők válaszainak egyes lépésekben számolt átlagára (az átlagos értékelésekre). Ez az 1. feltétel esetében mérsékelt erősségű mind a naiv,  $CCF=0,69$ , mind az integratív ROT-modellnél,  $CCF=0,66$ , és némileg gyengébb a 2. feltétel esetében a naiv,  $CCF=0,34$ , és az integratív modellnél is,  $CCF=0,45$ . A 3. feltételnél a modellek becslései kifejezetten rossz illeszkedést mutatnak a naiv,  $CCF=-0,38$ , és az integratív változat,  $CCF=-0,35$ , esetében is, mivel a résztvevők értékelései inkább az oksági kapcsolat meglétében való bizonytalanság felé kezdenek el tolódni a tanulás során, míg a modell az oksági kapcsolat meglétében való bizonyosság növekedését jósolja.

A ROT-modell illeszkedéséről árnyaltabb képet kaphatunk, ha az átlagos értékelések helyett egyéni szinten próbáljuk azt ellenőrizni. Ehhez minden résztvevő számára egyedi becslést készítettünk, amely a kovariancia-információt alkotó egyes esetek általa tapasztalt konkrét sorrendjén alapult. A becslések elkészítése után kereszt-korrelációt számoltunk az egyes résztvevők válaszai, és a számukra készített becslések között; ezek eloszlásának alakulása a 18. ábrán látható.



18. ábra

A negyedik kísérlet résztvevőinek válaszai, és a számukra egyedileg készített, becsült tanulási görbék közötti kereszt-korrelációk értékeinek eloszlása a ROT-modell két változatánál

A naiv ROT-modell esetében a résztvevők egy jelentős részénél (70 fő) 0,2 feletti, és egy kisebb részükénél (27 fő) 0,6 feletti kereszt-korrelációkat találtunk. Nyolc résztvevőnél a kereszt-korreláció értéke kifejezetten magasnak ( $>0,8$ ) bizonyult. Az integratív ROT-modell esetében 75 főnél találtunk 0,2-nél magasabb, 22 főnél 0,6-nál magasabb, és 5 főnél 0,8-nál magasabb kereszt-korrelációkat. Ezek az eredmények összességében a ROT-modell gyenge, de nem teljesen elhanyagolható illeszkedésére utalnak az online tanulási helyzetre.

## 4. Diszkusszió

A dolgozatban bemutatottuk, és empirikusan teszteltük a ROT-modellt, ami az egyszerű oksági kapcsolatok kovariancia-alapú tanulásának Bayes-i modellje. Elfogadva Sun (2009) érvelését, a modellt, mint az oksági tanulásra vonatkozó tudományos elméletet kezelhetjük, amely egyértelmű elképzeléseket fogalmaz meg a rejtett oksági tényezők oksági tanulásban betöltött szerepét illetően, amelyeket a modell különböző alapfeltevésein keresztül tudunk megragadni. Lévén egy számítógépes-matematikai modell, a ROT-modell esetében számos alapfeltevést kellett használnunk ahhoz, hogy képes legyen a modell megfelelő predikciók levezetésére – ezek között vannak olyanok, amelyeket empirikus alapon határoztunk meg, és vannak alkalmi, ad hoc feltevések is, amelyek azt az egyszerű célt szolgálják, hogy működőképes legyen a modell. A ROT-modell központi (empirikusan megalapozott) feltevései szerint az oksági kapcsolatokról az események közötti determinisztikus kapcsolatokként gondolkodunk, és a tanulás folyamatában fontos szerepe van a megfigyelt jelenségek magyarázatának, amelyben központi szerepet töltenek be azok az oksági (generatív, és preventív) tényezők, amelyekről nem áll rendelkezésre közvetlen információ. A modell Bayes-i modell, a Marr-i komputációs szintet célozza meg, tehát lényegében az oksági tanulásnak a megfogalmazott központi feltevések szerinti normatív modellje, és így egy lehetséges optimális megoldását nyújtja a problémának.

Több kísérletet is végeztünk, valamint a szakirodalomban talált adatokat is újraelemztünk, amiktől azt vártuk, hogy képet kapunk arról, a modell mennyire illeszkedik a humán adatokra. Az empirikus adatokkal való összevetés jó lehetőség volt arra is, hogy a modell becsléseinek humán adatoktól való eltéréseinek elemzésén keresztül felülvizsgálhassuk azokat az alapfeltevéseket, amelyeket a modell alkalmaz. Ahogy arra McClelland (2009) figyelmeztet, nem reális várakozás, hogy az empirikus adatokkal való összevetés során a ROT-modellt az oksági tanulás legjobb magyarázatának kiáltssuk ki, ugyanakkor megerősíthet minket abban, hogy a modellt az oksági tanulás egy lehetséges és adekvát magyarázatának tekintsük. Az eredményeink szerint a ROT-modell megfelelő magyarázatot szolgáltat az oksági tanulás bizonyos jelenségeire.

Mit tudunk az eredményeink alapján összefoglalóan megállapítani az egyszerű oksági kapcsolatok tanulásáról? A modellt normatív kritériumként alkalmazva azt mondhatjuk,

hogy az emberek az oksági ítéleteik alkotása során kifejezetten konzervatívak, ami megjelenik egyrészt abban, hogy az egyértelműnek tűnő jeleket sem hajlamosak egyértelműen az oksági kapcsolat mellett szóló bizonyítékként kezelni, továbbá a rendelkezésre álló adatok mennyiségét sem tartják túlzottan meggyőzőnek, alulértékelik annak jelentőségét. Az ítéletek ilyen polarizációjának enyhe mértéke különösen azokban az esetekben szembetűnő, amikor az adatok az oksági kapcsolat megléte mellett szólnak.

A ROT-modell összevetése a vizsgálatok eredményeivel arra is alkalmas lehet, hogy rámutasson a modell gyengeségeire, és utalhat arra, hogy a modell bizonyos alapfeltevései esetleg nem állják meg a helyüket. Ez a modellépítés szempontjából rendkívül hasznos információ. A Bayes-i modellek alapvetően moduláris jellegűek, ami azt jelenti, hogy viszonylag könnyen lecserélhetőek a modell által alkalmazott alapfeltevések. A megközelítéssel szemben megfogalmazott kritikák egy része pontosan erre a jellemzőjükre vonatkozik, amely szerint ez a moduláris jelleg lehetővé teszi, hogy az alapfeltevések önkényes megválasztásával bármilyen jelenséget képesek legyenek magyarázni (Jones és Love, 2011). Bár ennek a veszélye valóban nem elhanyagolható, azonban elkerülhető, ha figyelünk arra, hogy a modell a választott alapfeltevéseknek egy konzisztens készletét, és ne egyszerű konglomerátumát képviselje – úgy gondoljuk, hogy ez a cél praktikusán elérhető akkor, ha a modellt olyan központi alapfeltevések köré építjük, amelyek kellő empirikus megalapozottsággal bírnak, illetve empirikus alapon ellenőrizhetőek – a ROT-modell megalkotásánál erre különös hangsúlyt fektettünk. A vizsgálódásaink során arra is törekedtünk, hogy képet kapjunk arról, hogy a ROT-modell központi feltevései, és némely ad hoc feltevése valóban megállhatja-e a helyét. Noha az erre vonatkozó következtetéseink meggyőző erejét nem tekinthetjük teljesnek, de mindenképpen érdekesek lehetnek, amennyiben motiválhatják ezen feltevések alaposabb ellenőrzését.

Az egyik ilyen érdekes megfigyelésünk a 3. kísérlet eredményei alapján a különböző oksági tényezők hatásának integrálására vonatkozik. A ROT-modell itt egy rendkívül egyszerű, intuitív feltevést alkalmaz, nevezetesen azt, hogy egy következmény bekövetkezik, ha van jelen valamilyen generatív oksági tényező, és ugyanakkor nincs jelen preventív oksági tényező, és nem következik be, ha nincs jelen generatív oksági tényező és/vagy van jelen preventív oksági tényező. A vizsgálatok bizonyos eredményei ugyanakkor arra utalnak, hogy az oksági tényezők jelenlétére vonatkozó hiedelmeink, elképzeléseink esetleg ennél

komplikáltabb, és esetleg a rendelkezésünkre álló adatok, megfigyelések jellegétől függően eltérő módokon határozhatják meg az oksági ítéleteinket.

Úgy tűnik, hogy a tanulási helyzetben nem megfigyelhető – rejtett – oksági tényezők jelenlétére vonatkozó bizonytalanságra vonatkozó feltevések is felülvizsgálatra szorulhatnak, azok vélhetően túlbecsülik ezt a bizonytalanságot, ugyanakkor alulbecsülik ezen tényezők jelenlétének valószínűségét – erre is utalhatnak a 3. kísérlet eredményei. Itt mérlegelhető annak lehetősége is, hogy az erre vonatkozó hiedelmeink, elvárásaink változatosságot mutathatnak a szóban forgó téma szerint – ahogy erre utal Yeung és Griffiths (2015) eredménye. Ezek szerint sokkal erősebbnek tételezzük fel azokat az oksági kapcsolatokat, amelyek egyszerű fizikai jelenségekben játszanak szerepet, mint a biológiai-, vagy társadalmi jelenségek esetében. Ez magyarázhatja a kísérleteinkben tapasztalt jellemzően konzervatív értékeléseket, ugyanakkor, mivel következetesen egy témakört alkalmaztunk a vizsgálatainkban, az eredményeink nem adnak lehetőséget arra, hogy ezt a hatást tetten érhessük.

Az oksági kapcsolat meglétére vonatkozó előzetes hiedelmek, elvárások ugyancsak fontos szerepet játszhatnak a tanulás során – a ROT-modell ezekkel kapcsolatban egyszerű ad hoc feltevéseket alkalmazott, ami egyrészt az oksági kapcsolat meglétében való teljes bizonytalanságra vonatkozott, másrészt az előzetes hiedelmek, elvárások additív szerepére az oksági ítéletek alkotása során. A 4. kísérlet egyik tanulsága volt, hogy a teljes bizonytalanság feltételezése valószínűleg nem állja meg a helyét; egy teljesen új helyzetben is hajlamosak az emberek határozott állást foglalni az oksági kapcsolat mellett, vagy ellen.

Mik lehetnek a következő lépések a ROT-modell továbbfejlesztése érdekében? A modell alapfeltevéseinek (különösen a központi alapfeltevéseinek) az ellenőrzése, esetleg bizonyos ad hoc feltevések empirikus tesztelése, és felülvizsgálata, meg tudja erősíteni a modellt, mint tudományos elméletet azáltal, hogy stabilabb empirikus alapokra helyezi. Ez további, specifikus vizsgálatokat igényel arra vonatkozóan, hogyan alakul a rejtett oksági tényezők jelenlétére vonatkozó bizonytalanság – akár különböző, eltérő témakörök esetében, hogy milyen módon történik a különböző oksági tényezők hatásának integrálása – esetleg milyen interakciókat tételeznek fel az emberek ezen tényezők között.

Érdekes kihívást jelenthet az oksági tanulás és oksági gondolkodás területén az egyéni különbségek vizsgálata, amely a pszichometria hagyományostól eltérő megközelítését

igényelheti. A különböző képességek és készségek mérése területén bevett szokás a megfigyelt viselkedés/teljesítmény valamilyen kritérium szerinti megítélése, mint például a logikai deduktív gondolkodás során a klasszikus logika, mint kritérium alkalmazása (pl. Kurucz és Münnich, 2008). Bizonyos esetekben dinamikus helyzetben is van lehetőségünk a tesztalany teljesítményét/képességének szintjét megítélni, amennyiben egyértelműen el tudjuk dönteni, hogy az adott helyzetben adott válasza helyesnek, vagy helytelennek tekinthető (pl. Münnich és Kurucz, 2015). Ugyanakkor fontos felismernünk, hogy különösen az oksági kapcsolatok tanulásának mérése esetén egy ilyen specifikus kritérium sok esetben félrevezető lehet, mivel többfélék lehetnek azok a feltevések, amelyeket ezáltal elfogadunk a bizonytalanság kezelésének racionális módjaként, és a viselkedés alapvetően eltérő megítéléséhez vezethetnek (Gigerenzer és Selten, 2001). A kognitív modellezés területén a modellek aggregált adatokra való illesztése helyett azok egyéni válaszokra, vagy kisebb – lényeges jellemzőiket tekintve homogén – csoportok aggregált adataira való illesztése megfelelő szabad paraméterek megválasztásával, nem csak a valós kognitív folyamatokat hívebben reprezentáló modellek alkotásához vezethet (ahogy azt hangsúlyozza pl. Lee és Webb, 2005, valamint Navarro, Griffiths, Steyvers és Lee, 2006), hanem alkalmassá teheti a modell alkalmazását az adott tevékenységben szerepet játszó egyéni különbségek becslésére is, mint azt láthatjuk például Steyvers és mtsai. (2003) modelljében, ahol a modell szabad paraméterei bizonyos kapcsolódó kognitív folyamatokban (emlékezeti folyamatok, az adatok integrálására való hajlam) voltak képesek megragadni releváns különbségeket a vizsgálati személyek csoportjai között.

Érdekes iránynak tartjuk a hierarchikus kognitív modelleket (pl. Griffiths és Tenenbaum, 2009), amelyek az oksági kapcsolatokra vonatkozó absztrakt ismeretek, hiedelmek és elvárások hatását (például a különböző oksági tényezők hatásának integrálására vonatkozó szabályokat, azok erősségére vonatkozó bizonytalanságot) közvetlenül képesek modellezni. A ROT-modell esetében a hierarchikus irányba való elmozdulás segíthet a különböző – a kérdéses oksági kapcsolatra vonatkozó előzetes elvárásokra, a rejtett oksági tényezők jelenlétének valószínűségére vonatkozó bizonytalanságra, illetve a különböző oksági tényezők hatásának integrálására vonatkozó – alapfeltevéseinek egzakt módon való kezelésére, és lehetőséget biztosíthatnak a modell olyan szabad paraméterekkel való kiegészítésére, amelyek lehetőséget adnak az egyéni

különbségek vizsgálatára, konkrét szinten (egyres kognitív folyamatok, mint például emlékezeti folyamatok sajátosságai), vagy elvontabb szinten (általában az oksági kapcsolatok jellegéről vallott elképzelés, az oksági tényezők integrálásának lehetséges módjai).

## Felhasznált irodalom

- Ahn, W., & Dennis, M. J. (2000). Induction of casual chains. In L. R. Gleitman, A. K. Joshi (szerk.), *Proceedings of the 22th Annual Meeting of the Cognitive Science Society*, Erlbaum.
- Anderson, J. R. (1990). *The adaptive character of thought*. Lawrence Erlbaum Associates.
- Baetu, I., & Baker, A. G. (2009). Human judgments of positive and negative causal chains. *Journal of Experimental Psychology: Animal Behavioral Processes*, 35(2), 153–168.
- Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Method*, 43(3), 800-813.
- Blanco, F., Baeyens, F., & Beckers, T. (2014). Blocking in human causal learning is affected by outcome assumptions manipulated through causal structure. *Learning and Behavior*, 42(2), 185–199.
- Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, 138(3), 389–414.
- Bramley, N. R., Lagnado, D. A., & Speekenbirk, M. (2015). Conservative forgetful scholars: How people learn causal structure through sequences of interventions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(3), 708–731.
- Buehner, M. J. (2005). Contiguity and covariation in human causal inference. *Learning and Behavior*, 33(2), 230–238.
- Buehner, M. J., Cheng, P. W. (1997). Causal induction: the power PC theory versus the Rescorla-Wagner theory. In M. Shafto, P. Langley (szerk.), *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, (pp. 55-61), Lawrence Erlbaum Associates.
- Buehner, M. J., Cheng, P. W., Clifford, D. (2003). From covariation to causation: a test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1119-1140.

- Buehner, M. J., & May, J. (2002). Knowledge mediates the timeframe of covariation assessment in human causal induction. *Thinking&Reasoning*, *8(4)*, 269–295.
- Buehner, M. J., & May, J. (2004). Abolishing the effect of reinforcement delay on human causal learning. *Quarterly Journal of Experimental Psychology*, *57B(2)*, 179–191.
- Buehner, M. J., & McGregor, S. (2006). Temporal delays can facilitate causal attribution: Towards a general timeframe bias in causal induction. *Thinking&Reasoning*, *12(4)*, 353–378.
- Burns, P., & McCormack, T. (2009). Temporal information and children's and adults' causal inferences. *Thinking&Reasoning*, *15(2)*, 167–196.
- Campbell, J. (2007). An interventionist approach to causation in psychology. In A. Gopnik, L. Schulz (szerk.), *Causal learning: Psychology, Philosophy, and Computation* (pp. 58–66), Oxford University Press.
- Catena, A., Maldonado, A., Megías, J., & Frese, B. (2002). Judgement frequency, belief revision, and serial processing of causal information. *The Quarterly Journal of Experimental Psychology: Section B*, *55(3)*, 267–281.
- Catena, A., Perales, J. C., & Maldonado, A. (2004). Judgment Frequency Effects in Generative and Preventative Causal Learning. *Psicológica*, *25(1)*, 67–85.
- Chater, N., & Oaksford, M. (2008). *The probabilistic mind: prospects for Bayesian cognitive science*. Oxford University Press.
- Cheng, P. W. (1997). From covariation to causation: a causal power theory. *Psychological Review*, *104(2)*, 367–405.
- Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, *99(2)*, 365–382.
- Cobos, P. L., López, F. J., Cano, A., Almaraz, J., Shanks, D. R. (2002). Mechanisms of predictive and diagnostic causal induction. *Journal of Experimental Psychology: Animal Behavior Processes*, *28(4)*, 331–346.
- Coenen, A., Rehder, B., & Gureckis, T. M. (2015). Strategies to intervene on causal systems are adaptively selected. *Cognitive Psychology*, *79*, 102–133.

- Collins, D. J., & Shanks, D. R. (2002). Momentary and integrative response strategies in causal judgment. *Memory & Cognition*, *30*(7), 1138–1147.
- Collins, D. J., & Shanks, D. R. (2006). Conformity to the power PC theory of causal induction depends on the type of probe question. *The Quarterly Journal of Experimental Psychology*, *59*(2), 225–232.
- Cummins, D. D. (1995). Naïve theories and causal deduction. *Memory & Cognition*, *23*(5), 646–658.
- Cummins, D. D. (2014). The impact of disablers on predictive inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1638–1655.
- Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, *19*(3), 274–282.
- Danks, D. (2009). The psychology of causal perception and reasoning. In H. Beebe, C. Hitchcock, P. Menzies (szerk.), *Oxford handbook of causation* (pp. 447–470), Oxford University Press.
- Danks, D., Griffiths, T. L., Tenenbaum, J. B. (2003). Dynamical causal learning. In S. Becker, S. Thrun, K. Obermayer (szerk.), *Advances in Neural Information Processing Systems 15* (pp. 67-74), The MIT Press.
- De Neys, W., Schaeken, W., & dYdewalle, G. (2002). Causal conditional reasoning and semantic memory retrieval: a test of the semantic memory framework. *Memory & Cognition*, *30*(6), 908–920.
- De Neys, W., Schaeken, W., & dYdewalle, G. (2005). Working memory and counterexample retrieval for causal conditionals. *Thinking & Reasoning*, *11*(2), 123–150.
- De Pierris, G., & Friedman, M. (2013). Kant and Hume on causality. In E. N. Zalta (szerk.), *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition).
- Dennis, M. J., & Ahn, W. K. (2001). Primacy in causal strength judgments: the effect of initial evidence for generative versus inhibitory relationships. *Memory & Cognition*, *29*(1), 152–164.
- Dickinson, A. (2001). The 28th Bartlett Memorial Lecture. Causal learning: an associative analysis. *Quarterly Journal of Experimental Psychology*, *54*(1), 3–25.

- Eberhardt, F. (2009). Introduction to the epistemology of causation. *Philosophy Compass*, 4(6), 913–925.
- Einhorn, H. J., & Hogarth, R. M. (1985). Prediction, Diagnosis, and Causal Thinking in Forecasting. In G. Wright (szerk.), *Behavioral Decision Making* (pp. 311–328), Springer
- Escobar, M., & Miller, R. R. (2012). Associative accounts of causality judgments. In E. A. Wasserman, T. R. Zentall (szerk.), *The Oxford Handbook of Comparative Cognition* (pp. 157–174), Oxford University Press.
- Frosch, C. A., & Johnson-Laird, P. N. (2011). Is everyday causation deterministic or probabilistic. *Acta Psychologica*, 137(3), 280–291.
- Galea, S., Riddle, M., & Kaplan, G. A. (2010). Causal thinking and complex system approaches in epidemiology. *International Journal of Epidemiology*, 39(1), 97–106.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis* (Third edition). CRC Press.
- Gigerenzer, G., & Selten, R. (2001). Rethinking rationality. In G. Gigerenzer, R. Selten (szerk.), *Bounded rationality: The adaptive toolbox* (pp. 1–12), The MIT Press.
- Glautier, S. (2004). Asymmetry of generalization decrement in causal learning. *Quarterly Journal of Experimental Psychology*, 57(4), 315–329.
- Glymour, C. (2003). Learning, prediction and causal Bayes nets. *Trends in Cognitive Sciences*, 7(1), 43–48.
- Glymour, C., & Cooper, G. F. (1999). *Computation, causation and discovery*. AAAI Press.
- Goedert, K. M., Ellefson, M. R., & Rehder, B. (2014). Differences in the weighting and choice of evidence for plausible versus implausible causes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(3), 683–702.
- Goldvarg, E., & Johnson-Laird, P. N. (2001). Naive causality: a mental model theory of causal meaning and reasoning. *Cognitive Science*, 25(4), 565–610.
- Gopnik, A., & Bonawitz, E. (2014). Bayesian models of child development. *Wiley Interdisciplinary Reviews*, 6(2), 75–86.

- Greville, W. J., & Buehner, M. J. (2007). The influence of temporal distributions on causal induction from tabular data. *Memory & Cognition*, *35*(3), 444–453.
- Greville, W. J., & Buehner, M. J. (2016). Temporal predictability enhances judgements of causality in elemental causal induction from both observation and intervention. *Quarterly Journal of Experimental Psychology*, *69*(4), 678–697.
- Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229.
- Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*(4), 334–384.
- Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, *116*(4), 661–716.
- Gupta, P. (2008). The role of computational models in investigating typical and pathological behaviors. *Seminars in Speech and Language*, *29*(3), 211–225.
- Hagmayer, Y., Meder, B., von Sydow, M., & Waldmann, M. R. (2011). Category transfer in sequential causal learning: the unbroken mechanism hypothesis. *Cognitive Science*, *35*(5), 842–873.
- Hagmayer, Y., Sloman, S., Lagnado, D., & Waldmann, M. R. (2007). Causal reasoning through intervention. In A. Gopnik, L. Schulz (szerk.), *Causal learning: Psychology, Philosophy, and Computation* (pp. 86–100), Oxford University Press.
- Hahn, U. (2014). The Bayesian boom: good thing or bad? *Frontiers in Psychology*, *5*:765.
- Hewstone, M. (1989). *Causal attribution: from cognitive processes to collective beliefs*. Basil Blackwell.
- Hilton, D. J., & Slugoski, R. (1986). Knowledge-based causal attribution: the abnormal conditions focus model. *Psychological Review*, *93*(1), 75–88.
- Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: the new synthesis. *Annual Review of Psychology*, *62*, 135–163.
- Jeffreys, H. (1935). Some tests of signifiacnce, treated by the theory of probability. *Mathematical Proceedings of the Cambridge Philosophical Society*, *31*(2), 203–222.

- Jeffreys, H. (1961). *Theory of probability*. Oxford University Press.
- Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied*, 79(1), 1–17.
- Jensen, F. V. (2001). *Bayesian networks and decision graphs*. Springer.
- Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or elightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34(4), 169–188.
- Jordan, M. I. (1999). *Learning in graphical models*. MIT Press.
- Joyce, J. (1999). *The foundations of causal decision theory*. Cambridge University Press.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795.
- Keil, F. C. (2003). Folkscience: coarse interpretations of a complex reality. *Trends in Cognitive Sciences*, 7(8), 368–373.
- Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, 57, 227–254.
- Kelley, H. H. (1973). The process of causal attribution. *American Psychologist*, 28(2), 107–128.
- Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2010). Learning to learn causal models. *Cognitive Science*, 34(7), 1185–1243.
- Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT Press.
- Kurucz Gy. (2010). Történelmi jelenségekkel kapcsolatos oksági gondolkodás – áttekintés. In Münnich Á., Hunyady Gy. (szerk.), *A nemzeti emlékezet vizsgálatának pszichológiai szempontjai* (pp. 171-190), ELTE Eötvös Kiadó.
- Kurucz Gy. (2010b). Diagnosztikus következtetések vizsgálata ismert oksági kapcsolatok esetén. In Vass Z., Tekler V. (szerk.), *I. Országos Alkalmazott Pszichológiai Konferencia* (pp. 129-134), Papyrusz Book Kiadó.
- Kurucz Gy. (2014). Rendszerszemlélet, lényeglátás. In Münnich Á., Balázs K., Hőgye-Nagy Á. (szerk.) *Vezető-kiválasztást szolgáló kompetencia alapú tesztek leírása: KVKteszt Komplette vezető-kiválasztási tesztrendszer* (pp. 105–110), Belügyminisztérium.

- Kurucz Gy., & Münnich Á. (2008). A deduktív gondolkodási képesség mérése: a Logikai Betűsorok teszt. *Alkalmazott Pszichológia, 10(1-2)*, 115–129.
- Kushnir, T., Gopnik, A., Lucas, C., & Schulz, L. (2010). Inferring hidden causal structure. *Cognitive Science, 34(1)*, 148–160.
- Lagnado, D. A., & Sloman, S. A. (2002). Learning causal structure. In W. D. Gray, C. Schunn (szerk.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 560–565). Lawrence Erlbaum.
- Lagnado, D. A., & Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(4)*, 856–876.
- Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32(3)*, 451–460.
- Lagnado, D. A., & Speekenbirk, M. (2010). The influence of delays in real-time causal learning. *The Open Psychology Journal, 3(2)*, 184–195.
- Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (2007). Beyond covariation: cues to causal structure. In A. Gopnik, L. Schulz (szerk.), *Causal learning: Psychology, Philosophy, and Computation* (pp. 154–172), Oxford University Press.
- Lee, M. D., & Webb, M. R. (2005). Modeling individual differences in cognition. *Psychonomic Bulletin & Review, 12(4)*, 605–621.
- Legare, C. H. (2012). Exploring explanation: explaining inconsistent evidence informs exploratory, hypothesis-testing behavior in young children. *Child Development, 83(1)*, 173–185.
- Legare, C. H., & Lombrozo, T. (2014). Selective effects of explanation on learning during early childhood. *Journal of Experimental Child Psychology, 126*, 198–212.
- Lober, K., Shanks, D. (2000). Is causal induction based on causal power? Critique of Cheng(1997). *Psychological Review, 107(1)*, 195-212.
- Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences, 10(10)*, 464–470.
- Lombrozo, T., & Vasilyeva, N. Y. (2016). Causal explanation. In M. R. Waldmann (szerk.), *The Oxford Handbook in Causal Reasoning* (pp. 415-432), Oxford University Press.

- Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review*, *115*(4), 955–984.
- Luhmann, C. C., & Ahn, W. K. (2011). Expectations and interpretations during causal learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 568–587.
- Mackie, J. L. (1965). Causes and conditions. *American Philosophical Quarterly*, *2*(4), 245–264.
- Mandel, D. R., & Vartanian, O. (2009). Weighting of contingency information in causal judgement: Evidence of hypothesis dependence and use of a positive-test strategy. *Quarterly Journal of Experimental Psychology*, *62*(12), 2388–2408.
- Marcus, G. F., & Davis, E. (2013). How robust are probabilistic models of higher-level cognition? *Psychological Science*, *24*(12), 2351–2360.
- Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. W. H. Freeman and Company.
- Marsh, J. K., & Ahn, W. (2003). Interpretation of ambiguous information in causal induction. In R. Alterman, D. Kirsch (szerk.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (pp. 775–780), Psychology Press.
- Marsh, J. K., & Ahn, W. (2009). Spontaneous assimilation of continuous values and temporal information in causal induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 334–352.
- McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, *1*(1), 11–38.
- McCormack, T., Frosch, C., Patrick, F., & Lagnado, D. (2015). Temporal and statistical information in causal structure learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(2), 395–416.
- McCormack, T., Bramley, N., Frosch, C., Patrick, F., & Lagnado, D. (2016). Children’s use of interventions to learn causal structure. *Journal of Experimental Child Psychology*, *141*, 1–22.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

- Mendelson, R., & Shultz, T. R. (1976). Covariation and temporal contiguity as principles of causal inference in young children. *Journal of Experimental Child Psychology, 22*(3), 408-412.
- Michotte, A. (1963). *The perception of causality*. Methuen.
- Münnich Á., & Kurucz Gy. (2015). Munkaköri alkalmasságvizsgálat a Diagramok teszttel: rendszerkezelők tesztelése. *Alkalmazott Pszichológia, 15*(2), 133-146.
- Navarro, D. J., Griffiths, T. L., Steyvers, M., & Lee, M. D. (2006). Modeling individual differences using Dirichlet processes. *Journal of Mathematical Psychology, 50*(2), 101-122.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. Morgan Kaufman.
- Pearl, J. (2000). *Causality: models, reasoning, and inference*. Cambridge University Press.
- Perales, J. C., & Shanks, D. R. (2007). Models of covariation-based causal judgment: a review and synthesis. *Psychonomic Bulletin & Review, 14*(4), 577-596.
- Perales, J. C., & Shanks, D. R. (2008). Driven by power? Probe question and presentation format effects on causal judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1482-1494.
- Piaget, J. (1930). *The child's conception of physical causality*. K. Paul, Trench, Trubner & Company Limited.
- Rahman, S., & Ogilvie, J. (szerk.). (2006). Probabilistic models of cognition [Special issue]. *Trends in Cognitive Sciences, 10*(7).
- Reips, U. D. (2000). The web experiment method. In M. H. Birnbaum (szerk.), *Psychological experiments on the Internet* (pp. 89-117), Academic Press.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, W. F. Prokasy (szerk.), *Classical conditioning II: current research and theory* (pp. 64-99), Appleton-Century-Crofts.
- Roser, M. E., Fugelsang, J. A., Dunbar, K. N., Corballis, P. M., & Gazzaniga, M. S. (2005). Dissociating processes supporting causal perception and causal inference in the brain. *Neuropsychology, 19*(5), 591-602.

- Rottman, B. M., & Keil, F. C. (2012). Causal structure learning over time: observations and interventions. *Cognitive Psychology*, *64*(1–2), 93–125.
- Rottman, B. M., Kominsky, J. F., & Keil, F. C. (2014). Children use temporal cues to learn causal directionality. *Cognitive Science*, *38*(3), 489–513.
- Rozenblit, L., & Keil, F. C. (2002). The misunderstood limits of folk science: an illusion of explanatory depth. *Cognitive Science*, *26*(5), 521–562.
- Saito, M., & Shimazaki, T. (2012). Strategy changes in causal structure learning: the role of task complexity. In N. Miyake, D. Peebles, R. P. Cooper (szerk.), *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 2264–2269), Cognitive Science Society.
- Saito, M., & Shimazaki, T. (2013). Interpreting covariation in causal structure learning. In M. Knauff, M. Pauen, N. Sebanz, I. Wachsmuth (szerk.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 3348–3353), Cognitive Science Society.
- Saito, M., & Shimazaki, T. (2014). Differences between observation and intervention in causal learning. In P. Bello, M. Guarini, M. McShane, B. Scassellati (szerk.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2865–2870), Cognitive Science Society.
- Schaffer, J. (2016). The metaphysics of causation. In E. N. Zalta (szerk.), *The Stanford Encyclopedia of Philosophy* (Fall 2016 Edition).
- Schlottmann, A., & Shanks, D. R. (1992). Evidence for a distinction between judged and perceived causality. *Quarterly Journal of Experimental Psychology*, *44*(2), 321–342.
- Schultz, T. R. (2007). The Bayesian revolution approaches psychological development. *Developmental Science*, *10*(3), 357–364.
- Schulz, L. E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science*, *10*(3), 322–332.
- Schulz, L. E., & Sommerville, J. (2006). God does not play dice: causal determinism and preschoolers' causal inferences. *Child Development*, *77*(2), 427–442.

- Schulz, L., Kushnir, T., & Gopnik, A. (2007). Learning from doing: intervention and causal inference. In A. Gopnik, L. Schulz (szerk.), *Causal learning: Psychology, Philosophy, and Computation* (pp. 67–85), Oxford University Press.
- Shanks, D. R., & Dickinson, A. (1988). Associative accounts of causality judgment. *Psychology of Learning and Motivation*, 21, 229–261.
- Shanks, D. R., Pearson, S. M., & Dickinson, A. (1989). Temporal contiguity and the judgement of causality by human subjects. *Quarterly Journal of Experimental Psychology*, 41(2), 139–159.
- Shou, Y., & Smithson, M. (2015a). Adapting to an Uncertain World: Cognitive Capacity and Causal Reasoning with Ambiguous Observations. *PLoS One*, 10(10).
- Shou, Y., & Smithson, M. (2015b). Effects of question formats on causal judgments and model evaluation. *Frontiers in Psychology*, 6:467.
- Skitka, L. J., & Sargis, E. G. (2006). The internet as psychological laboratory. *Annual Review of Psychology*, 57, 529–555.
- Sloman, S. A. (2005). *Causal models: how people think about the world and its alternatives*. Oxford University Press.
- Sloman, S. A., & Hagmayer, Y. (2006). The causal psycho-logic of choice. *Trends in Cognitive Sciences*, 10(9), 407–412.
- Sobel, D. M., & Kushnir, T. (2006). The importance of decision making in causal learning from interventions. *Memory & Cognition*, 34(2), 411–419.
- Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search*. Springer-Verlag.
- Stewart, N., Chandler, J., & Paolacci, G. (2017). Crowdsourcing Samples in Cognitive Science. *Trends in Cognitive Sciences*, 21(10), 736–748.
- Steyvers, M., Tenenbaum, J. B., Wagenmakers, E., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, 27(3), 453–489.
- Sun, R. (2008). Introduction to computational cognitive modeling. In R. Sun (szerk.), *The Cambridge Handbook of Computational Psychology* (pp. 3–20), Cambridge University Press.

- Sun, R. (2009). Theoretical status of computational cognitive modeling. *Cognitive Systems Research, 10*(2), 124–140.
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science, 331*(6022), 1279–1285.
- Vallée-Tourangeau, F., Payton, T., & Murphy, R. A. (2008). The impact of the format of covariation information on causal inferences. *European Journal of Cognitive Psychology, 20*(1), 177–194.
- Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: the role of nonrepresentation of compound stimulus elements. *Learning and Motivation, 25*(2), 127–151.
- Waismeyer, A., Meltzoff, A. N., & Gopnik, A. (2015). Causal learning from probabilistic events in 24-month-olds: an action measure. *Developmental Science, 18*(1), 175–182.
- Waldmann, M. R. (1996). Knowledge-based causal induction. In D. R. Shanks, K. Holyoak, D. L. Medin (szerk.), *Causal learning* (pp. 47–88), Academic Press.
- Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: evidence from an overshadowing paradigm. *Psychonomic Bulletin & Review, 8*(3), 600–608.
- Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition, 82*(1), 27–58.
- Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 216–227.
- Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: asymmetries in cue competition. *Journal of Experimental Psychology: General, 121*(2), 222–236.
- Ward, W. C., & M, J. H. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology, 19*(3), 231–241.

- Wasserman, E. A., & Berglan, L. R. (1998). Backward blocking and recovery from overshadowing in human causal judgment: the role of within compound associations. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, *51*(2), 121–138.
- White, P. A. (1995). *The understanding of causation and the production of action*. Erlbaum.
- White, P. A. (1998). Causal Judgement: Use of Different Types of Contingency Information as Confirmatory and Disconfirmatory. *European Journal of Cognitive Psychology*, *10*(2), 131–170.
- White, P. A. (2000). Causal judgment from contingency information: relation between subjective reports and individual tendencies in judgment. *Memory & Cognition*, *28*(3), 415–426.
- White, P. A. (2003). Effects of wording and stimulus format on the use of contingency information in causal judgment. *Memory & Cognition*, *31*(2), 231–242.
- Williamson, J. (2004). *Bayesian nets and causality: philosophical and computational foundations*. Oxford University Press.
- Woodward, J. (2003). *Making things happen: a theory of causal explanation*. Oxford University Press.
- Yarritu, I., & Matute, H. (2015). Previous knowledge can induce an illusion of causality through actively biasing behavior. *Frontiers in Psychology*, *6*:389.
- Yarritu, I., Matute, H., & Luque, D. (2015). The dark side of cognitive illusions: when an illusory belief interferes with the acquisition of evidence-based knowledge. *British Journal of Psychology*, *106*(4), 597–608.
- Yeung, S., & Griffiths, T. (2011). Estimating human priors on causal strength. In L. Carlson (szerk.) *Proceedings of the 33rd Annual conference of the Cognitive Science Society* (pp. 1709-1714), Cognitive Science Society.
- Yeung, S., & Griffiths, T. L. (2015). Identifying expectations about the strength of causal relationships. *Cognitive Psychology*, *76*, 1–29.
- Young, M. E. (2014). Sex differences in the inference and perception of causal relations within a video game. *Frontiers in Psychology*, *5*:926.

## **Mellékletek**

## A. A ROT-modell forrása

```
model.rot <- function(
  D,
  alfa.G.prior = 1, beta.G.prior = 1,
  alfa.P.prior = 1, beta.P.prior = 1)
{
  require(cubature)

  L.pigpip <- function(pig, pip, alfa.G, beta.G, alfa.P, beta.P)
    dbeta(pig, alfa.G, beta.G) * dbeta(pip, alfa.P, beta.P)

  p.ok_pigpipH1 <- function(pig, pip, o, k, po = 0.5)
    if (o == 1) {
      if (k == 1) {
        return( (1-pig)*(1-pip)*po+pig*(1-pip)*po )
      } else if (k == 0) {
        return( (1-pig)*pip*po+pig*pip*po )
      }
    } else if (o == 0) {
      if (k == 1) {
        return( pig*(1-pip)*(1-po) )
      } else if (k == 0) {
        return( (1-pig)*(1-pip)*(1-po)+(1-pig)*pip*(1-po)+pig*pip*(1-po) )
      }
    }
  }

  p.D_pigpipH1 <- function(pig, pip, data)
    p.ok_pigpipH1(pig, pip, o = 1, k = 1)^data[1] *
    p.ok_pigpipH1(pig, pip, o = 1, k = 0)^data[2] *
    p.ok_pigpipH1(pig, pip, o = 0, k = 1)^data[3] *
    p.ok_pigpipH1(pig, pip, o = 0, k = 0)^data[4]

  wrap.H1 <- function(x, data, alfa.G, beta.G, alfa.P, beta.P)
    p.D_pigpipH1(x[1], x[2], data) *
    L.pigpip(x[1], x[2], alfa.G, beta.G, alfa.P, beta.P)

  p.D_H1 <- function(data, alfa.G, beta.G, alfa.P, beta.P)
    choose(
      sum(data[3:4]),
      data[3]) * choose(sum(data[1:2]), data[1]) *
    adaptIntegrate(
      wrap.H1,
      lowerLimit = c(0,0),
      upperLimit = c(1,1),
      data = data,
      alfa.G = alfa.G,
      beta.G = beta.G,
      alfa.P = alfa.P,
      beta.P = beta.P)$integral

  p.ok_pigpipH0 <- function(pig, pip, o, k, po = 0.5)
    if (o == 1) {
      if (k == 1) {
        return( pig*(1-pip)*po )
      } else if (k == 0) {
        return( (1-pig)*(1-pip)*po+(1-pig)*pip*po+pig*pip*po )
      }
    }
  }
```

```

    }
  } else if (o == 0) {
    if (k == 1) {
      return( pig*(1-pip)*(1-po) )
    } else if (k == 0) {
      return( (1-pig)*(1-pip)*(1-po)+(1-pig)*pip*(1-po)+pig*pip*(1-po) )
    }
  }
}

p.D_pigpipH0 <- function(pig, pip, data) {
  p.ok_pigpipH0(pig, pip, o = 1, k = 1)^data[1] *
  p.ok_pigpipH0(pig, pip, o = 1, k = 0)^data[2] *
  p.ok_pigpipH0(pig, pip, o = 0, k = 1)^data[3] *
  p.ok_pigpipH0(pig, pip, o = 0, k = 0)^data[4]
}

wrap.H0 <- function(x, data, alfa.G, beta.G, alfa.P, beta.P) {
  p.D_pigpipH0(x[1], x[2], data) *
  L.pigpip(x[1], x[2], alfa.G, beta.G, alfa.P, beta.P)
}

p.D_H0 <- function(data, alfa.G, beta.G, alfa.P, beta.P)
  choose(
    sum(data[3:4]),
    data[3]) * choose(sum(data[1:2]), data[1]) *
  adaptIntegrate(
    wrap.H0,
    lowerLimit = c(0,0),
    upperLimit = c(1,1),
    data = data,
    alfa.G = alfa.G,
    beta.G = beta.G,
    alfa.P = alfa.P,
    beta.P = beta.P)$integral

apply(D, 1, function(x) {
  a = x[1]
  b = x[2]
  c = x[3]
  d = x[4]
  n = sum(x)

  p.G <- alfa.G.prior/(alfa.G.prior+beta.G.prior)
  p.P <- alfa.P.prior/(alfa.P.prior+beta.P.prior)

  p.a <- c( (1-p.G)*(1-p.P), p.G*(1-p.P))
  p.a <- p.a/sum(p.a)
  fr.a <- a*p.a

  p.b <- c((1-p.G)*(1-p.P), (1-p.G)*p.P, p.G*p.P)
  p.b <- p.b/sum(p.b)
  fr.b <- b*p.b

  p.d <- c((1-p.G)*(1-p.P), (1-p.G)*p.P, p.G*p.P)
  p.d <- p.d/sum(p.d)
  fr.d <- d*p.d

  alfa.G <- alfa.G.prior+(fr.a[2]+fr.b[3]+c+fr.d[3])
  beta.G <- beta.G.prior+(fr.a[1]+fr.b[1]+fr.b[2]+fr.d[1]+fr.d[2])

```

```
alfa.P <- alfa.P.prior+(fr.b[2]+fr.b[3]+fr.d[2]+fr.d[3])
beta.P <- beta.P.prior+(a + fr.b[1]+c+fr.d[1])

c(
  pred=log(p.D_H1(x, alfa.G, beta.G, alfa.P, beta.P) /
           p.D_H0(x, alfa.G, beta.G, alfa.P, beta.P)),
  alfa.G = alfa.G, beta.G= beta.G,
  alfa.P = alfa.P, beta.P = beta.P)
})
}
```

## B. SS-modell forrása

```
model.ss <- function(
  D,
  alpha = 5)
{
  require(cubature)

  p.D_w0w1H1 <- function(w0, w1, data)
    return(w0^data[3]*(1-w0)^data[4]*(1-(1-w0)*(1-w1))^data[1]*((1-w0)*(1-
w1))^data[2])

  L.w0w1_H1 <- function(w0, w1, alpha)
    exp(-alpha*w0-alpha*(1-w1))+exp(-alpha*(1-w0)-alpha*w1)

  p.D_H1 <- function(data, alpha)
    choose(sum(data[3:4]), data[3]) *
    choose(sum(data[1:2]), data[1]) *
    adaptIntegrate(
      function(x) { p.D_w0w1H1(x[1], x[2], data)*L.w0w1_H1(x[1], x[2],
alpha) },
      lowerLimit = c(0,0), upperLimit = c(1,1))$integral /
    adaptIntegrate(
      function(x) { L.w0w1_H1(x[1], x[2], alpha) },
      lowerLimit = c(0,0), upperLimit = c(1,1))$integral

  p.D_w0H0 <- function(w0, data)
    w0^(data[1]+data[3])*(1-w0)^(data[2]+data[4])

  L.w0_H0 <- function(w0, alpha)
    exp(-alpha*w0)+exp(-alpha*(1-w0))

  p.D_H0 <- function(data, alpha)
    choose(sum(data[3:4]), data[3]) *
    choose(sum(data[1:2]), data[1]) *
    integrate(
      function(x) { p.D_w0H0(x, data)*L.w0_H0(x, alpha) },
      lower = 0,
      upper = 1)$value/
    integrate(
      function(x) { L.w0_H0(x, alpha) },
      lower = 0,
      upper = 1)$value

  apply(D, 1, function(x) {
    log(p.D_H1(x, alpha) / p.D_H0(x, alpha))
  })
}
```

## C. Support-modell forrás

```
model.support <- function(
  D)
{
  apply(D, 1, function(data) {
    p.D.H1 <- function(x, data)
      ( (1-(1-x[1])*(1-x[2]))*0.5 )^data[1] *
      ( (1-x[1])*(1-x[2])*0.5 )^data[2] *
      ( x[1]*0.5 )^data[3] *
      ( (1-x[1])*0.5 )^data[4]

    p.D.H0 <- function(x, data)
      ( x[1]*0.5 )^data[1] *
      ( (1-x[1])*0.5 )^data[2] *
      ( x[1]*0.5 )^data[3] *
      ( (1-x[1])*0.5 )^data[4]

    require(cubature)

    p.D.H1.int <- adaptIntegrate(
      p.D.H1,
      lowerLimit = c(0,0),
      upperLimit = c(1,1),
      data = data)$integral

    p.D.H0.int <- adaptIntegrate(
      p.D.H0,
      lowerLimit = c(0,0),
      upperLimit = c(1,1),
      data = data)$integral

    log(p.D.H1.int / p.D.H0.int)
  })
}
```