

# BÜNTETŐI JOG

## MESTERSÉGES INTELLIGENCIA ÉS SZEREPMEGFORDÍTHATÓ ÍTÉLŐKÉPESSÉG

KIEL BRENNAN-MARQUEZ & STEPHEN E. HENDERSON†

*Az intelligens gépek egyre inkább felülmúlják az emberi szakértők teljesítményét, ami felveti a kérdést, hogy mikor (és miért) kell az embereknek a döntéshozatalban "benne maradniuk". Az egyik gyakori válasz az eredményekre összpontosít: az emberek az intuícóra és a tapasztalatra támaszkodva képesek felismerni az értelmezési hibákat - néha katasztrofális hibákat -, amelyek a gépektől elkerülik őket. Bár ez az érv ma még hihető, a technológia fejlődésével egyre inkább el fog kopni.*

*Ebben a cikkben szilárdabb alapot keresünk: az emberi ítélőképesség védelmét, amely a döntéshozatal normatív integritására összpontosít. Konkrétabban a demokratikus egyenlőség "szerep-visszafordíthatóságként" történő bemutatását javasoljuk. Egy demokráciában a döntések meghozatalával megbízottaknak kölcsönösen fogékonyaknak kell lenniük a döntések hatására; léteznie kell egy olyan értelmes értelemben, amelyben a résztvevők szerepe a döntési folyamatban mindig megfordulhat. A szerepek megfordíthatósága az ítélkezés aktusát a "csak Isten kegyelméből" dinamikájával ruhazza fel, és ezáltal az ítélkezést az önrendelkezés eredményének tekinti.*

*Miután megvédtük a szerepvisszafordíthatóság fogalmát, megmutatjuk, hogyan igazolódik ez a büntető esküdtszéki tárgyalások paradigmás eseteiben. Bár nem ez volt az esküdtszéki tárgyalás történeti mozgatórugója - legalábbis nem erős értelemben -, de*

---

† Kiel Brennan-Marquez a Connecticuti Egyetem jogászprofesszora és William T. Golden ösztöndíjasa; Stephen E. Henderson az Oklahomai Egyetem Haskell A. Holloman bíró, a jogtudományok professzora. Szeretnének köszönetet mondani azoknak az olvasóknak, akik javították ezt a cikket, és azoknak a sokaknak, akik részt vettek a témával kapcsolatos beszélgetésekben, köztük Mark Blitznek, Andy Coannak, Brian Choi-nak, Aloni Cohennek, Andrew Fergusonnak, David Graynek, Roger Fordnak, Paul Kahnnak, Karen Levynnek, Jonathan Manesnek, Paul Ohmnak, Frank Pasquale-nak, Richard Re, Andrew Selbst, Ric Simmons, Christopher Slobogin, Kelly Sorensen, Katherine Strandburg, Daniel Susser, Mark Verstraete, Tal Zarsky, valamint a Privacy Law Scholars Conference és a Fordham, a Penn State Law és az Arizona James E. Rogers College of Law.

*azzal érvelünk, hogy a szerepvisszafordíthatóság megmagyarázza az intézmény néhány alapvető jellemzőjét, és a legjobb okok között van a megőrzésére. Végül, a sci-fi rajongók számára a szerepvisszafordíthatóság receptet kínál arra, hogy mikor lesz a jogrendszer készen áll a robotbírók és robotbírák számára: amikor a jogrendszer robotvádoltakat is magában foglal.*

	TARTALOMJEGYZÉK BEVEZETÉS.....	138
I. AI ÉS AZ EMBERI ÍTÉLŐKÉPESSÉG .....		143
A. Mesterséges általános intelligencia.....		143
B. Emberek .....	a	146
hurokban .....		146
II. A DEMOKRATIKUS EGYENLŐSÉG MINT SZEREP-VISSZAFORDÍTHATÓSÁG.....		149
A. "Csak Istenkegyelméből.....		149
B. Ítéletalkotás és önérvényesítés.....		152
III. AZ ESKÜDTSZÉKI PEREK MINT PARADIGMÁS ESETEK .....		156
A. A "kortársaidesküdszéke .....		158
B. Az eltűnő bűnügyi esküdszék.....		160
KÖVETKEZTETÉS.....		163

*Az örült nemcsak egy koldus, aki azt hiszi magáról, hogy király, hanem egy király is, aki azt hiszi magáról, hogy király.*

-Jacques Lacan

#### BEVEZETÉS

Képzeljük el, hogy 2049-ben járunk, és a gépek átvették a jogrendszer nagy részét. A régi idők emberi döntéshozóihoz képest a gépek gyorsabbak és pontosabbak is, ami a rendszer elfogultságtól és kiszámíthatatlanságtól való megszabadítására tett több évtizedes erőfeszítés megkoronázása.<sup>1</sup> A "pontosságról" persze még mindig léteznek egymással versengő elképzelések - az emberek továbbra sem értenek egyet a jog különböző területeinek céljait és az ezzel járó prioritásokat illetően -, de a gépeket úgy képezték ki, hogy figyelembe vegyék ezeket a nézeteltéréseket. A

<sup>1</sup> Ahogy Max Tegmark fizikus leírta a célt:

Melyek az első asszociációk, amelyek eszébe jutnak, amikor az országa bírósági rendszerére gondol? Ha a hosszadalmas késedelmek, a magas költségek és az időnkénti igazságtalanság, akkor nincs egyedül. Nem lenne csodálatos, ha az első gondolatai helyett a "hatékonyság" és a "méltányosság" lennének? Mivel a jogi eljárás elvontan egy számításnak tekinthető, amely a bizonyítékokra és a jogszabályokra vonatkozó információkat beírja, majd döntést hoz, egyes tudósok arról álmodoznak, hogy *robotbírókkal* teljesen automatizálják azt: Olyan mesterséges intelligencia-rendszerek, amelyek fáradhatatlanul ugyanazokat a magas jogi normákat

alkalmazzák minden ítéletre anélkül, hogy az emberi hibáknak, például az elfogultságnak, a fáradtságnak vagy a legújabb ismeretek hiányának engednének.

MAX TEGMARK, LIFE (3.02017105).

valójában kiderül, hogy a gépek jobban fel vannak szerelve az emberi pluralizmus valóságának kezelésére, mint maguk az emberek: a gépek a harcon kívül állva könnyedén szintetizálják a különböző normatív nézőpontokat. Ráadásul a gépek kifogástalanul következetesek. A "hasonló eseteket egyformán kell kezelni" eszménye, amely a decentralizált emberi ítélkezés világában mindig is bizonytalan volt, végre igazolást nyert.<sup>2</sup> A hiperkomplex modellezési technikák alkalmazásával a gépi döntéshozatal hatékonyan garantálja, hogy az értelmesen azonos jellemzőkkel rendelkező esetek mindig ugyanúgy végződnek.

Mi a baj ezzel a képpel - ha van egyáltalán -? Gyakorlati szempontból rengeteg okunk van arra, hogy szkeptikusak legyünk a hipotetikus feltételezésekkel kapcsolatban.<sup>3</sup> De az érvelés kedvéért tegyük fel, hogy a feltételezések helytállóak; tegyük fel, hogy a gépek valóban képesek lesznek pontosabb és következetesebb döntésekre. Akkor is van valami baj azzal, ha bizonyos döntéseket gépekre bízunk? Az embereknek "benne kellene maradniuk" bizonyos döntési folyamatokban?

---

<sup>2</sup> Sir Patrick Devlin 1956-os híres zsűrielőadásában írta le a "következetesség" eszményét és annak elérésének nehézségeit:

Van tehát többféle igazságosság? Igen: az igazságosság a legjobb kompromisszum, amely a törvény követelményei és az érdemi ítélet között elérhető. E két pontot mindig valamilyen távolság választja el egymástól, legyen az hosszú vagy rövid, és az igazságosság bárhol megállhat közöttük.

az érdemhez nem feltétlenül tesznek igazságot. Így lenne, ha csak egy bíró lenne, és ő mindig emlékezne arra, hogy mindent ugyanúgy ítéljen meg, és örökké élne - röviden, ha a földi igazságszolgáltatás isteni lenne, nem pedig emberi. De az emberi igazságossághoz elengedhetetlen a törvény. Ez az igazságosság egyik alapvető tulajdonsága egy

közösség, hogy hasonló esetekben hasonló döntéseket kell hozni. Ez azonban csak a törvény betartásával lehetséges, tehát a törvény adja meg az általános szabályt. Nincs általános szabály

soha nem illik pontosan egy adott esetre.

SIR PATRICK DEVLIN, ESKÜDTSZÉKI TÁRGYALÁS 151-52, (1531956).

Meg kell jegyeznünk, hogy Ben Johnson és Richard Jordan provokatíván érvelt a "hasonló esetek" eszménye ellen, amely valódi (jó) munkát végez. *Lásd* Ben Johnson & Richard Jordan, *Should Like Cases Be Decided Alike?: A Formal Analysis of Four Theories of Justice* (megjelenés előtt) (a tervezet elérhető a [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3127737](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3127737) [https://perma.cc/94B3-N9ZG] oldalon). Mi azonban továbbra is megelégszünk azzal a meglehetősen minimális igényünkkel, hogy következetességre törekszünk a büntető igazságszolgáltatásra vonatkozó szabályaink alkalmazása során.

<sup>3</sup> Például a büntető igazságszolgáltatásban sokkal nehezebb lesz tudni, hogy egy gép "helyes eredményt" ér el, mint például a minta- vagy biometrikus felismerésben - két jelenlegi fejlesztési területen -. Még ha a gép pontosan meg is jósolja a büntető igazságszolgáltatás korábbi eredményeit, akkor is bőven van okunk gyanakodni a múltbeli eredmények jogszerűségére, mind tényszerűen, mind jogilag, ugyanakkor nincsenek jobb adataink, amelyek alapján a rendszer tanulhatna. A történeti igazságot önmagában is nehéz megállapítani - különösen a büntetőjog területén, ahol megpróbálunk belelátni a vádolt megértésébe és motivációiba -, de a büntető igazságszolgáltatás eredményei olyan rendszer-korrektív és igazságossági elveket adnak hozzá, amelyek az egyedi ügyekben a pontossággal szemben állnak, beleértve a releváns bizonyítékok elhallgatását. Ugyanez vonatkozik az ítélethozatalra is, ahol nehéz elképzelni, hogy történelmileg jól megjósoltuk

volna a sivatagot vagy a vízhiányt. De ismét, a mi projektünk az, hogy a tudományt a legjobb esetnek tekintjük.

még akkor is, ha ez nem növeli - és talán csökkenti - a pontosságot és a következetességet?

Úgy gondoljuk, és gyanítjuk, hogy ez az intuíció széles körben elterjedt. A kérdés az, hogy racionalizálható-e. Ebben a cikkben a mesterséges intelligencia (AI) tudományának és szókincsének rövid áttekintése után az egyenlőség elvein alapuló intuícioról adunk számot. Leegyszerűsítve az érvelésünk az, hogy egy liberális demokráciában bizonyos ítéletek esetében szükség van a "szerep-visszafordíthatóság" szempontjára. Bizonyos kontextusokban azoknak, akik ítéletet hoznak, fordítva is sebezhetőnek kell lenniük a folyamatokkal és a hatásokkal szemben. És azoknak, akik ki vannak téve a hatásoknak, kölcsönösen képesnek kell lenniük az ítékezés gyakorlására. Más szóval, van, és kell is, hogy legyen olyan értelemben, hogy a résztvevők szerepe a folyamatban mindig megfordulhat: egy másik világban, de egy esetleges eseménysorozat miatt a döntéshozó lehet a sebezhető helyzetben, nem pedig a hatalomban. Minél kevesebb az esetlegesség ebben a megfordításban, annál inkább megfordítható a döntés.

A szerepvisszafordíthatóság lehetővé teszi a döntéshozók számára, hogy tiszteletben tartsák a döntéshozatal súlyát az érintett felek szemszögéből. Ez viszont lehetővé teszi, hogy az ítélethozatal egy szélesebb erkölcsi közösség által osztott értékek igazolásaként értelmezhető legyen - egy olyan egyenlő közösség, amely magában foglalja mind a döntéshozót, mind az érintett felet, valamint sok más embert, akik nem vettek részt a döntésben, de ugyanúgy részt vehettek volna, és akik mindenesetre osztoznak a döntés következményeire viselt felelősségben.

A központi gondolat az, hogy egy szabály alkalmazásának aktusa - pontosan ugyanaz a szabály - más, ha azt egy olyan fél hajtja végre, aki maga is a szabály hatálya alá tartozik, szemben egy olyan féllel, aki mentes a szabály hatálya alól. A demokráciában a polgárok nem állnak kívül az ítélezési folyamaton, mintha félelemmel vagy félelemmel válaszolnának egy orákulum kijelentéseire. Inkább kollektívan vagyunk felelősek minden egyes ítéletért. Így az ítékezéssel megbízott félnek - aki végül is bármelyikünk lehetett volna - képesnek kell lennie arra, hogy kimondhassa: *Ez a döntés olyan korlátokat tükröz, amelyeket mi magunknak szabtunk meg, és ebben az esetben történetesen nem nekem, hanem egy másik személynek kell felelnie értük.* És az elítélt félnek - aki szintén bármelyikünk lehetett volna - azt kell tudnia mondani: *A döntési folyamatot mi magunk gyakoroljuk, és ebben az esetben történetesen nem én, hanem egy másik személy hajtja végre.*

A szerepek visszafordíthatóságának kérdéseit tehát nem kizárólag az erős mesterséges intelligencia ösztönzi. Éppen ellenkezőleg, a demokratikus legitimitás mindig is a népszuverenitás problémája volt: az ítékezés közös vállalkozásként való ítékezés a

amelyeknek a közösség minden tagja igényt tarthat a szerzősége.<sup>4</sup> Ehhez nem elég, ha a szabályalkotási folyamatok inkluzívak, vagy ha az emberek a választások napján a szavazófülkékbe özönlenek (bár ez csodálatos lenne). A közösség nevében végrehajtott ítélezési aktusoknak a kollektív alkalmazás pecsétjét kell viselniük; bár minden ítélezési aktus egyedi, és *ebben a* konkrét esetben *erre a* konkrét félre vonatkozik, minden ilyen aktusnak valamilyen szinten mindannyiunkra vonatkoznia kell. Így, még ha ugyanazokat a kritériumokat alkalmazzák is, és ugyanazokra az eredményekre jutnak is, alapvetően más, ha egy király vagy királynő a törvények felett állva ítélezik valamelyik alattvalója felett, vagy ha egy kasztrendszerben a magas származásúak döntenek az alacsony származásúak sorsáról.<sup>5</sup> És ugyanezen alapvető okból alapvetően más lenne, ha egy gép mondaná ki a végső szót a döntések felett. A szabályok és az eredmények funkcionálisan azonosak lehetnek, de a döntés aktusa más lenne. Nem lennének demokratikus aktusok.

Miután ezeket az elveket elvontan megvédtük, a III. részben egy paradigmaticus esetre térünk rá: a büntetőjogi ítéletre. Még nem vagyunk biztosak abban - és nem is foglalnunk véglegesen állást -, *hogy* pontosan *mely* döntések demokratikus legitimitációja függ a szerepvisszafordíthatóságtól. Nem vagyunk biztosak abban sem, hogy pontosan milyen kontextusokban, még ha a szerepvisszafordíthatóság nem is szükséges szigorúan a legitimitáshoz, az ettől való eltekintésnek mégis erős ellenérveket kell tartalmaznia. Erős a gyanúnk azonban, hogy a büntető igazságszolgáltatással kapcsolatos döntések valószínűleg e két tábor egyikébe tartoznak, ami azt jelenti - legalábbis -, hogy alaposan át kell gondolnunk, mielőtt a büntetőjogi kontextusban feladnánk a szerepmegfordíthatóságot. Az elv diktálta elvek szerint tehát azoknak, akik büntetőjogi ítéletet hoznak, értelmesen kiszolgáltatottnak kell lenniük ugyanazoknak a büntetőjogi eljárásoknak, mint a büntetőjogi vádlottaknak. És

ugyanígy a büntetőjogi vádlottaknak is képesnek kell lenniük az ítélezés gyakorlására. De a közelmúltban bekövetkezett fordulat esetén a felek szerepei akár fel is cserélődhetnek volna: a büntetőjogi vádlott viselhetne volna a talárt vagy ülhetett volna az esküdtszék páholyában, a bíró vagy az esküdtek pedig vádlottként vonulhattak volna a bíróságra. Ez - érvelni fogunk - az angol-amerikai "egyenrangú esküdtszék" eszményének normatív alapja - ha nem is feltétlenül történelmi alapja -, és ez szól arról, hogy miért ünneplik az esküdtszéket régóta a "népi bölcsesség" szerveként. Az elképzelés nem az, hogy az esküdtek jobb érzékkel rendelkeznek a jó és a rossz viszonyáról.

---

<sup>4</sup> Lásd Kiel Brennan-Marquez & Paul W. Kahn, *Statutes and Democratic Self-Authorship*, 56 WM. & MARY L. REV. 115, 173-77 (2014).

<sup>5</sup> Mondanom sem kell, hogy az Egyesült Államok politikai-gazdasági viszonyai és a hagyományos kasztrendszer közötti funkcionális hasonlóságok - és hogy következképpen bizonyos intézményi ítélezési struktúráink már most is sértőek lehetnek-e a szerepvisszafordíthatóság elve szempontjából, teljesen függetlenül a gépi kérdéstől - olyan kérdés,

amelyet elemzésünk nem ~~írt meg~~, hanem felvet.

~~ÍTELEK~~



mint az intézményi szereplők. (Bár ez is igaz lehet.) Alapvetőbb, hogy az esküdtek *állampolgárként*, nem pedig hivatalnokokként reagálnak az ítékezés aktusára, és ebben a tekintetben az esküdtszéki perek modellértékűek arra nézve, hogy mit tesz lehetővé a szerep megfordíthatósága: még ha az esküdtszéki eljárás nem is vezet más eredményre, mint egy intézményes bíró (vagy más tényfeltáró eljárás) előtti tárgyalás, megkönnyíti az ítékezés emberi áldozatának rendszerszintű elismerését. És ami még ennél is alapvetőbb, a tárgyalást demokratikus aktussá alakítja át.

A mi beszámolóink egyik vonzó tulajdonsága - még ha minden más előnyt félreteszünk is -, hogy ellenáll az "emberiség-fetisizmusnak" (vagy "faji fetisizmusnak"), amely gyakran jelen van az emberekről és a gépekről szóló beszélgetésekben. Véleményünk szerint nincs semmi különleges az emberi döntéshozókban. Inkább az számít, hogy a döntéshozó képes-e pozíciót cserélni az érintett féllel. Ebben az értelemben a mi beszámolóink kész választ ad arra, hogy mikor válhat normatívan elfogadhatóvá, hogy a robotok bírói talárt öltsenek, esküdtszéki tagok legyenek, és más demokratikus döntéshozatali szerepeket töltsenek be: amikor felcserélhetően robot-vádlottakká válnak.

-

Három figyelmeztetés, mielőtt belevágnánk. Először is, azzal, hogy a szerep-visszafordíthatóságot olyan elvként írjuk le, amelynek értéke nem a hibák azonosításán és csökkentésén múlik, nem azt akarjuk mondani, hogy a szerep-visszafordíthatóság *nem képes* erre. Természetesen lehetséges. Sőt, okkal gondolhatjuk, hogy a szerep-visszafordíthatóság arra ösztönözheti a döntéshozókat, hogy eltérjenek az előírt szabályoktól - például amikor az esküdtszék érvénytelenít -, és ez potenciálisan üdvözítő módon történhet. Ezt a kérdést a jövőbeni munkára hagyjuk. Egyelőre csak annyit állítunk, hogy még ha feltételezzük is, hogy a szerepmegfordíthatóság nem javítja a döntéshozatal pontosságát, akkor is van belső értéke.<sup>6</sup>

Másodszor, még ha igazunk is van a szerepvisszafordíthatóság belső értékével kapcsolatban, ebből nem következik, hogy minden más szempontnak, például a hatékonyságnak és a pontosságnak, háttérbe kell szorulnia. Mint minden összetett normatív kérdés esetében, valószínűleg vannak egymással versengő értékdimenziók, amikor arról döntünk, hogy az embereket a körforgásban tartsuk-e. Elképzelhető például egy olyan jövő, amelyben a gépi döntéshozatalhoz kapcsolódó pontosságnövekedés - legalábbis bizonyos körülmények között - elég nagy ahhoz, hogy indokoltta tegye a szerepvisszafordíthatóság mellőzését, és természetesen a kompromisszumelemzés eltérhet attól függően, hogy milyen típusú döntésekről van szó - a döntések gyakoriságát tekintve,

---

<sup>6</sup> Analógia lehetne az Alkotmány által a büntetőügyekben biztosított védőügyvéd, amely garancia szinte biztosan segít elkerülni a hibákat. De még ha ez nem is így lenne - még ha például olyan tanulmányok látnának napvilágot, amelyek szerint a jogi képviselő kevésbé befolyásolja a tárgyalás kimenetelét -, akkor is lehetne azzal érvelni, hogy az egyenlő jogi képviselővel való egyenlő hozzáférés belső jótétemény, az eljárási integritás és az emberi méltóság iránti elkötelezettség, függetlenül a kimenetelektől.

komplexitás és gravitáció - is szerepet játszanak. A végkövetkeztetés egyszerűen az, hogy a szerepek visszafordíthatósága önmagában véve jó, és ugyanígy, hogy annak csökkenése (vagy megszüntetése) önmagában véve veszteséget jelentene. Ebben az értelemben a szerepvisszafordíthatóság fogalmilag szilárd alapot biztosít azokon a területeken, ahol a döntéshozatal demokratikus származása számít, az emberi ítélet értékéhez való ragaszkodáshoz - nem annyiban, hogy az emberi ítélet jobb eredményekhez vezet (valójában nem biztos, hogy ez így van), hanem annyiban, hogy az ember az, akit megítélnék.

Harmadszor, a szerep-visszafordíthatóság oroszlánszerepének hangoztatásával nem azt akarjuk sugallni, hogy a meglévő jogrendszerekben a döntéshozatal egésze, vagy akár csak a legtöbb döntési folyamat a gyakorlatban megfelel a normatív kritériumnak. Távolról sem. Bár ezt a jövőbeni munkára tartogatjuk, az a kérdés, hogy a mai Egyesült Államok - az osztályok közötti rétegződés és a büntető igazságszolgáltatási rendszer faji egyenlőtlenségei miatt - "általánosságban" nem felel meg a szerep-visszafordíthatóságnak, fontos kérdés, amelyet elemzésünk közvetlenül felvet (bár nyilvánvalóan nem old meg). A motiváló példa az erőteljes mesterséges intelligencia. Ami azonban a jogrendszerünk szervezésére gyakorolt gyakorlati következményeket illeti, a mesterséges intelligencia talán csak a jéghegy csúcsa - vagy ha úgy tetszik, a kanári az aknában.

## I. AI ÉS AZ EMBERI ÍTÉLŐKÉPESSÉG

Tézisünkben kiindulva - amely feltételezi, hogy a technológia messze túl fog fejlődni jelenlegi formájában - nincs szükség arra, hogy a mesterséges intelligencia jelenlegi állapotát részletezzük. Mindazonáltal egy rövid ismertetés arról, hogy hol tart a tudomány, hová tart, és meddig tarthatnak a céljai, eloszlatja azokat az aggodalmakat, hogy fantáziadús kérdést teszünk fel. Most van itt az ideje, mielőtt a tudomány a maga természetes útján megadná a kegyelemdőfést, hogy elkezdjük mérlegelni, milyen normatív és jogi szabályoknak kellene irányítani szép új világunkat.<sup>7</sup>

### A. MESTERSÉGES ÁLTALÁNOS INTELLIGENCIA

A mesterséges intelligencia egy hatalmas tudományterület, amely nem követ egyetlen dogmát sem.<sup>8</sup> És mint minden tudományág esetében, a definíciós kérdések már a kezdetektől fogva fennállnak. Alan Turing úttörő munkája például a "Turing-teszt", amely azt a kérdést teszi fel, hogy egy emberi kérdezőbiztos, miután

---

<sup>7</sup> Néhány ilyen szabály 1.0 verzióját lásd Stephen E. Henderson, *A Few Criminal Justice Big Data Rules*, OHIO15 ST. J. CRIM. L. (2018)527.

<sup>8</sup> A vezető tankönyv szavai szerint a mesterséges intelligencia "magában foglalja a logikát, a valószínűséget és a folyamatos matematikát; az észlelést, az érvelést, a tanulást és a cselekvést; és mindent a mikroelektronikai eszközöktől a robotizált bolygófelderítőig".

2019]

*SZEREPVISSZAFORDÍTHATÓ*

143

STUART J. RUSSELL & PÉTER NORVIG, MESTERSÉGES INTELLIGENCIA: A MODERN APPROACH  
vii (3d ed. 2010).

egy mesterséges intelligencia rendszerrel kapcsolatba lépve összetévesztené azt egy embertársával.<sup>9</sup> Ez a teszt megragadta az informatikusok egy generációjának fantáziáját, sőt, még néhány filozófus költőét is (amint azt Brian Christian is megírta csodálatos könyvében, *a The Most Human Human (A legemberibb ember)* címűben).<sup>10</sup>

A mesterséges intelligencia alapvető rejtélyei azonban jóval Turing előttiék; valójában legalább olyan messzire nyúlnak vissza, mint a nyugati filozófia kezdetei az ókori Görögországban. Amennyiben Platónnak vagy Arisztotelésznek igaza volt abban, hogy mi a racionális gondolkodás, akkor vitathatatlanul ők határozták meg a mesterséges intelligencia célját.<sup>11</sup> Másrészt, vajon az a cél, hogy olyan gépeket építsünk, amelyek (a) gondolkodnak vagy (b) racionálisan cselekszenek? Vagy az a cél, hogy olyan gépeket építsünk, amelyek (a) úgy gondolkodnak vagy (b) úgy viselkednek, mint az emberek? A szakértők nem értenek egyet.<sup>12</sup> Egy dolog azonban egyértelmű: az emberi szintű és emberi szintű intelligencia - gyakran "mesterséges általános intelligenciának" vagy AGI-nek nevezett - "szent gráljához" való eljutáshoz számos tudományágban további fejlesztésekre lesz szükség.<sup>13</sup> A mai "intelligens" gépek rendkívül jók bizonyos feladatokban, például a sakkozásban vagy a repülőgépek vezetésében. Az emberek ezzel szemben sok mindenben jók, a számítógépek programozásától kezdve a zeneszerzésen át a repülőgépek (remélhetőleg) leszállásáig, ha a repülési tervek vagy a rendszerek meghibásodnak. Mikorra várható tehát, hogy valódi AGI-t láthatunk?

A szakértők ismét nem értenek egyet. A "techno-szkeptikusoktól", akik szerint az AGI-t csak egy évszázad múlva érjük el, ha egyáltalán valaha is, a "techno-szkeptikusokig" terjednek a színes nézetek.

<sup>9</sup> *Lásd id.* at 2. Gondoljunk csak a Google Duplex (lenyűgözően, mégis hátborzongatóan) időpontot ad egy fodrászszalonban. *Lásd* Gary Marcus & Ernest Davis, *A.I. Is Harder Than You Think*, N.Y. TIMES (2018. május 18.), <https://www.nytimes.com/2018/05/18/opinion/artificial-intelligence-challenges.html> [<https://perma.cc/PA8Y-CLYR>].

<sup>10</sup> BRIAN CHRISTIAN, *A LEGEMBERIBB: MIT TANÍT NEKÜNK A MESTERSÉGES INTELLIGENCIA AZ ÉLETRŐL* (2011). A Turing-teszt kész kritikája, hogy *a mimikri* nem *lény*: "A repülőmérnöki szövegek nem úgy határozzák meg szakterületük célját, hogy "olyan gépeket készítsenek, amelyek olyan pontosan úgy repülnek, mint a galambok, hogy még a többi galambot is képesek becsapni"." RUSSELL & NORVIG, *Supra* note 8, 3. o., 3. o. Teljesen igaza van.

<sup>11</sup> RUSSELL & NORVIG, *Supra* note 8, at 4,5.

<sup>12</sup> *Lásd id.* 2-5. o.; *lásd még* TEGMARK, 1. l. ábjegyzet, 49-55. o.

<sup>13</sup> RUSSELL & NORVIG, *supra* note 8, 27. o.; TEGMARK, *supra* note 1, 30., 39., 52. o. ("Az AI-kutatás szent grálja az "általános AI" (ismertebb nevén *mesterséges általános intelligencia*, AGI) megalkotása, amely maximálisan széleskörű: gyakorlatilag bármilyen célt képes elérni, beleértve a tanulást is."); *lásd még: id.* 1.1. táblázat. Nick Bostrom így írja le az AGI-t: "Az emberi intelligenciának megfelelő gépek - vagyis olyanok, amelyek rendelkeznek józan ésszel és hatékony tanulási, érvelési és tervezési képességgel, hogy komplex információfeldolgozási kihívásoknak feleljenek meg a természetes és absztrakt területek széles skáláján". NICK BOSTROM, *SZUPERINTELLIGENCIA* (42014). Russel és Norvig hangsúlyozzák, hogy ehhez a matematika (a kiszámíthatóságtól a teljességen át a valószínűségig), a közgazdaságtan (a hasznosságtól a játékelméleten át a kielégítésig), az idegtudomány (hogyan végzik el a százmilliárd neuronunk a feladatukat?), a pszichológia (a

behaviorizmustól a kognitív tudományig), valamint a számítástechnika, az irányításelmélet, a kibernetika és a nyelvészet területén történő fejlesztésekre lesz szükség. RUSSELL & NORVIG, *fenti* jegyzet, 8,5-16. o.

optimista kevesek, akik szerint ez hamarosan bekövetkezhet.<sup>14</sup> A legtöbben úgy vélik, hogy valahol néhány évtizedtől egy évszázadig terjedő időszakra számíthatunk.<sup>15</sup> Akárhogy is, a végeredmény az, hogy ha robotzsidókat (vagy ennek megfelelőt) akarunk, akkor még a mi életünkben vagy hamarosan utána megkaphatjuk őket.

Meg kell jegyeznünk egy fontos fenntartást: lehetséges, hogy az "intelligencia robbanás" eléri az AGI-t, de aztán szinte azonnal túl is lép rajta. Sokan azt várják, hogy az AGI-t a "szingularitás" fogja bevezetni, vagyis az a pont, amikor az emberek félreállhatnak, mert a gépek a gépek javítására fognak törekedni.<sup>16</sup> (A Hitchhiker-rajongóknak: gondoljunk csak a Földet tervező Deep Thoughtra, csak sokkal gyorsabban.<sup>17</sup>) Ha mondjuk a gépek egy nap elérik az AGI-t, hogy aztán a következő nap már sokkal magasabb intelligenciát érjenek el, akkor a gépi urainknak sokkal több mondanivalójuk lesz a büntető igazságszolgáltatási rendszereinkről, mint ahogyan azt terveztük. Mégis, minden ilyen irányú ellenvetés - akár a mi érvelésünkkel, akár bármely más, az emberek és az intelligens gépek közötti kapcsolatra vonatkozó ellenvetés - túl soknak bizonyulna. Lehetséges, hogy a felsőbbrendű gépek, ha egyszer megérkeznek, gyorsan átveszik az irányítást szinte *minden* felett? Bizonyára, és ha ez bekövetkezik, akkor talán az, amit korábban normatív szinten elméletileg elmagyaráztunk, nem bizonyul majd felvilágosítónak, vagy legalábbis irrelevánsnak. De az is lehet, hogy nem következik be; és ha nem, akkor az emberi és a gépi képességek közötti megfelelő viszony megoldása kulcsfontosságú lesz. Szóval, haladjunk tovább.

---

<sup>14</sup> Lásd TEGMARK, *fenti* megjegyzés 1, 30-33., 40-42. Ahogy Bostrom megjegyzi, a mesterséges intelligencia története tele van túlbuzgó állításokkal, és elég gyakran,

[t]két évtized a radikális változások előrejelzői számára a legjobb pont: elég közel ahhoz, hogy figyelemfelkeltő és releváns legyen, de elég messze ahhoz, hogy feltételezhető legyen, hogy egy sor, jelenleg csak homályosan elképzelhető áttörés addigra bekövetkezhet Húsz év is közel lehet ahhoz a tipikus időtartamhoz, ami egy radikális változástól még hátravan

az előrejelző karrierjét, korlátozva a merész jóslatok hírnevének kockázatát.

BOSTROM, 13. lábjegyzet, 4-5. o. A mesterséges intelligencia kutatás történelmi "hullámvölgyeiről" lásd: RUSSELL & NORVIG, 8. lábjegyzet, 16-28. o. (részletes idővonal); TEGMARK, 8. lábjegyzet, 1, 40-42. o. (az "idővonal-mítoszok" leírása); BOSTROM, *supra* note. 13, 6-14. o. (a "remény és a kétségbeesés évszakai" ismertetve).

<sup>15</sup> Lásd TEGMARK, 1. lábjegyzet, 30-33, 40-42; BOSTROM, 13,22-25. lábjegyzet.

<sup>16</sup> Lásd TEGMARK, *fenti* megjegyzés 1, 54., 134-35. o.; BOSTROM, *supra* note. 13, 3-6. o. Két korai, az ügyvédek számára különösen fontos úti cél a gépi tanulási rendszerek voltak, amelyek bizonyos feladatokban (különösen az iratellenőrzésben) felülmúlták az ügyvédeket, valamint a bírósági eredmények helyes előrejelzése. Lásd Erin Winick, *Lawyer-Bots Are Shaking Up Jobs*, MIT TECH. REV., 2017. dec. 12. (dokumentumellenőrzés); Nikolas Aletras et al., *Predicting Judicating Decisions of the European Court of Human Rights: A Natural Language Processing Perspective*, PEER2 J. COMPUTER SCI., Oct. 24, 2016.

<sup>17</sup> Lásd DOUGLAS ADAMS, *THE HITCHHIKER'S GUIDE TO THE GALAXY* 181-83 (1. kiadás, 1980) (28. fejezet).

## B. EMBEREK A HUOKBAN

Bármilyen pontos időtávlatban is valósuljon meg az AGI, minden emberi szintű intelligenciával rendelkező gép hibás marad, ugyanazért, amiért az emberek maguk is hibásak; és minden, az emberi szintű intelligenciát *megközelítő* gép ugyanúgy hibás lesz, *a fortiori*. Ezért az általános nézet az, hogy az embereket "a körforgásban" kell tartanunk, és a végső szót kell gyakorolnunk a döntéshozatali folyamat felett.<sup>18</sup>

Az embereknek a körforgásban való megtartása mellett szóló érvek különböző formákat öltenek, de jellemzően arra az elképzelésre vezethetők vissza, hogy az emberek megértik a döntéshozatali *célok* és széles körű intuícióval rendelkeznek, ami lehetővé teszi számukra, hogy felismerjék a problémákat vagy hibákat, amelyek a gépektől elkerülik őket.<sup>19</sup> Egy drámai példával élve, gondoljunk Sztanyiszlav Petrov 1983. szeptember 26-i döntésére, amikor felülbírált egy szovjet rendszert, amely tévesen észlelt egy nukleáris támadást.<sup>20</sup> Amikor Petrov a műholddal összekapcsolt számítógépektől értesítést kapott arról, hogy egy amerikai rakéta közeledik, az ösztöne azt súgta neki, hogy valami nincs rendben, ezért jelentette a rendszer hibáját.<sup>21</sup> És amikor hasonló értesítések érkeztek - egy második, harmadik, negyedik és ötödik nukleáris rakétáról, amelyek mindegyike tizenkét percen belül érkezett volna - Petrov megérzései ugyanezt súgták neki, ezért ismét hibát jelentett.<sup>22</sup> Petrovnak igaza volt: a műholdakat megzavarták a napfény visszaverődései, bár akkor még egyáltalán nem volt tudatában ennek a részletnek.<sup>23</sup> Egyszerűen csak *érezte*, hogy valami "nem stimmel", és ez az érzés talán elkerülte a katasztrofális atomháborút.<sup>24</sup>

---

<sup>18</sup> Az is elképzelhető, hogy a gépi intelligencia *gonosz*, és nem osztja az emberiség normatív értékeit. *Lásd pl.* BOSTROM, *Supra* note. 13, 130. o. (az "ortogonalitási tézis" felvetése, amely szerint "[az] intelligencia és a végső célok ortogonálisak: az intelligencia többé-kevésbé bármilyen szintje elvileg kombinálható többé-kevésbé bármilyen végső céllal"). Mi nem vagyunk ebben olyan biztosak. Mindenesetre ez az eshetőség olyan nyilvánvalóan ellene szólna annak, hogy az emberi ítélőképességet bármilyen ilyen géppel helyettesítsük, hogy nem foglalkozunk vele.

<sup>19</sup> E nézet kiváló filozófiai kifejtését lásd: HUBERT DREYFUSS, WHAT COMPUTERS STILL CAN'T DO (1992).

<sup>20</sup> *Lásd: Midnight and Counting*, ECONOMIST, 2017. szeptember 30., 2017. 86.

<sup>21</sup> *Id.*

<sup>22</sup> *Id.*

<sup>23</sup> *Id.*; *lásd még:* TEGMARK, 1. lábjegyzet, 112-13. o.

<sup>24</sup> Egy másik ilyen szovjet hős Vaszilij Arkhipov, aki a kubai rakétaválság idején megtagadta egy nukleáris torpedó kilövését. *Lásd* Yasmeen Serhan, *When the World Lucked Out of a Nuclear War*, ATLANTIC (2017. október 31.), <https://www.theatlantic.com/international/archive/2017/10/when-the-world-lucked-out-of-nuclear-war/544198/> [https://perma.cc/JRY4-4RYM]; *lásd még* TEGMARK, *supra* note. 1, a Tegmark-féle 112. *Future of Life Institute* 2017 végén posztumusz kitüntette Arkhipovot. *Lásd* Tucker Davey, *55 Years After Preventing Nuclear Attack, Arkhipov Honored With Inaugural Future of Life Award*, FUTURE OF LIFE INSTITUTE (2017. október 27.), <https://futureoflife.org/2017/10/27/55-years-preventing-nuclear-attack-arkhipov-honored-inaugural-future-life-award/> [https://perma.cc/Q5G8-THKQ].

Az ehhez hasonló esetek - valamint azok tragikus párhuzamai, amelyekben a gépek emberi beavatkozás vagy felülbírálat nélkül katasztrofálisan rossz döntéseket<sup>25</sup> hoztak - az AI-politikai körökben számos kommentátort arra készítettek, hogy amellet érveljenek, hogy az embereket *mindig* "a hurokban kell tartani", különösen, ha olyan súlyos és visszafordíthatatlan döntésekről van szó, mint a fegyverrendszerek használata.<sup>26</sup> Hasonló érvek merültek fel a bűnüldözés, az<sup>27</sup> orvostudomány, a<sup>28</sup> pénzügyek<sup>29</sup> és a szociális programok igazgatása kapcsán.<sup>30</sup> A oldalon

---

<sup>25</sup> Például öt évvel a Petrov-epizód után egy hasonlóan téves értesítés érkezett a USS Vincennes irányított rakétacirkálóhoz, de senki sem találta ki a gépet - ami egy "megtorló" csapáshoz vezetett, amely megsemmisítette az Iran Air 655-ös járatát, megölve mind a 290 embert a fedélzeten, és felerősítve a több évtizedes nemzetközi bizalmatlanságot. *Lásd* Max Fisher, *The Forgotten Story of Iran Air Flight 655*, WASH. POST, 2013. október 16., A13. o.; *lásd még* TEGMARK, *fenti* 111. o., 1,111. o.

<sup>26</sup> *Lásd* TEGMARK, *fenti* megjegyzés 1, 111-13. o.; Rebecca Crootof, *The Killer Robots Are Here*: CARDOZO 36 L. REV. 1837, 1872-83 (2015) (összefoglalva az ezzel kapcsolatos különféle érveket). Ugyanezen irányvonalak mentén általánosabb érveket *lásd* például Bryan Casey et al.: *The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise* (megjelenés előtt) (az Európai Unió 2016-os "kizárólag automatizált feldolgozáson alapuló döntés meghozatalához való jog" magyarázata, beleértve az "emberi beavatkozáshoz való jogot"); Tamara Capeta, *Of Judges and Robots* in CHALLENGES OF LAW IN LIFE REALITY 129-42 (2017) (azzal érvelve, hogy a robotokból hiányoznak a szükséges "érzelmek"); A. Michael Froomkin et al., *When AIs Outperform Doctors: The Dangers of a Tort-Induced Over-Reliance on Machine Learning and What (Not) To Do About It* (hamarosan megjelenik) (azzal érvelve, hogy a gépi tanulás egészségügyi algoritmusai emberi közreműködő nélkül az ellátás minőségének csökkenéséhez vezetnének).

<sup>27</sup> A rendőrséggel kapcsolatban *lásd* például Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, 164 U. PA. L. REV. 871, 897-98 (2015) (a megalapozott gyanú automatizálása ellen érvelve azon az alapon, hogy "az emberi lények mindig képesek legalább potenciálisan egy új releváns információ bevonására az elemzésbe", míg "az adatbázis[ok] nem tartalmazhatják az összes olyan tényt, amely minden esetben releváns", ami azt jelenti, hogy a gépi tanuló algoritmusok "nem tudják figyelembe venni a 'teljes képet' egy személy lehetséges bűnelkövetésére vonatkozóan, ahogy azt a negyedik módosítás megköveteli"). Az ítélethozatal tekintetében *lásd* például Sonja Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803 (2014). *Lásd általában* Kiel Brennan-Marquez, *"Plausible Cause": Magyarázati normák a hatalmi gépek korában*, 70 VAND. L. REV. 1249 (2017) (azzal érvelve, hogy a büntető igazságszolgáltatási rendszerben a magyarázó normák dacolnak az automatizálással, mivel az embereknek "a hurokban" kell maradniuk, hogy megmagyarázzák, *miért* hoznak egy döntést úgy, ahogyan az van).

<sup>28</sup> *Lásd pl.* Andreas Holzinger, *Interactive Machine Learning for Health Informatics: When Do We Need A Human in the Loop*, 3 BRAIN INFORMATICS 119 (2016). Az emberi ítélőképesség értékének szemléletes példája az orvostudományban a tüdőgyulladásos betegek optimális osztályozása. Még az 1990-es évek végén a tüdőgyulladásos betegek magasabb és alacsonyabb kockázati kategóriákba történő válogatására tervezett algoritmus az asztmás betegeket hazaküldte ahelyett, hogy kórházi kezelést javasolt volna. Ennek az ellenkező értelmű (és helytelen) javaslatnak az volt az oka, hogy történelmileg az asztmás tüdőgyulladásos betegek, mivel nagyobb kockázatnak voltak kitéve, általában több ellátásban részesültek. Így jobb volt az általános egészségi állapotuk. Az algoritmus úgymond "csak ezeket az eredményeket látta", és így az asztmás tüdőgyulladásos betegeket a következő csoportokba sorolta



E "katasztrófa-központú" érvek mellett egyes megfigyelők hangsúlyozták az emberi felügyelet egyéb előnyeit is - például azt, hogy az érintett felek könnyebben tiszteletben tartják a nehezen lenyelhető eredményeket, ha tudják, hogy egy ember is részt vett a döntéshozatali folyamatban;<sup>31</sup> és hogy az emberek képesek lehetnek olyan mintákat azonosítani, amelyek elemzése után "visszailleszthetők" a döntéshozatali rendszerbe, és idővel finomíthatják az eredményeket.<sup>32</sup>

Nem kicsinyítjük le az emberi ítélőképesség mellett szóló műszeres érveket. Ezek ma is érvényesek, és talán túlélnek a mesterséges intelligencia következő generációját is. A cikk további részét azonban annak vizsgálatának szenteljük, hogy mi történjen, ha az ilyen érvek *nem* maradnak fenn. Ha a technológia olyan módon fejlődik, amely ténylegesen elvitatja az emberi felsőbbrendűség igényét a döntéshozatal céljainak megítélésében - és a hibák bizonyos osztályainak felismerésében -, akkor az emberi ítélőképesség továbbra is védhető-e?

---

alacsonyabb kockázatúnak. Ezt a hibát csak az emberi felügyeletnek köszönhetően sikerült felismerni és megoldani. *Lásd* Paul Voosen, *How AI Detectives are Cracking Open the Black Box of Deep Learning*, SCIENCE (2017. július 6.).

<sup>29</sup> *Lásd pl.* Yesha Yadav, *How Algorithmic Trading Undermines Efficiency in Capital Markets*, 68 VAND. L. REV. 1607 (2015). Figyelmeztető példa a 2010-es "Flash Crash", amelyben az automatizált tőzsdei kereskedési algoritmusok egy trillió dolláros tőzsdei összeomlást okoztak azzal, hogy emberi kereskedő számára érthetetlen módon cselekedtek. *Lásd*: BOSTROM, *Supra* note, 13,20. o., BOSTROM, *Supra* note, 20. o..

21. Egy csavaros módon azonban egy másik automatizált rendszer volt az, amelyik valóban észlelte a problémát.

*Lásd id.*

<sup>30</sup> *Lásd pl.*, Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1256-57 (2008) (felsorolva a hibákra hajlamos algoritmikus kormányzási eszközöket, beleértve (1) "[b]enefit [m]anagement [s]ystems", amelyek "több százezer hibás Medicaid, élelmiszerjegy és jóléti juttatás jogosultságát határozták meg," (2) a "gyermektartással tartozók" felkutatására szolgáló algoritmusok, amelyek sok nem bűnözőt söpörnek be, és automatikus bérlefoglalást indítanak el, és (3) a terrorizmus elleni eszközök, amelyek a "bonyolult algoritmusok és a hibás adatok" miatt "nagy arányú hamis pozitív eredményeket" hoznak létre, amelyek súlyos bűnüldözési következményekkel járnak).

<sup>31</sup> *Lásd pl.*, TEGMARK, *supra* note 1, at 106-07 (arra utalva, hogy megmagyarázható mesterséges intelligencia nélkül az érintett felek nem biztos, hogy "úgy érzik, hogy eléggé megértik a logikai érvelését ahhoz, hogy tiszteletben tartsák az ítéletét"); Ric Simmons, *Big Data and Procedural Justice: Legitimizing Algorithms in the Criminal Justice System*, 5 OHIO ST. J. CRIM. L. 573 (2018) (azzal érvelve, hogy a túl sok gépi döntéshozatal - legalábbis a büntetőjogi kontextusban - hátrányosan befolyásolhatja az emberek "eljárásai igazságosságról" alkotott képét).

<sup>32</sup> *Lásd pl.* Brennan-Marquez, *Supra* note. 277, 1280-97. o. Ennek a dinamikának egyik gyakori oka a "túlillesztés", amelynek felismerése (és - *a fortiori* - korrekciója) gyakran emberi értelmezést igényel. Brian Christian és Tom Griffiths *Algorithms to Live By* című könyvében szerepel egy emlékeztető eset, amelyben egy kilencfaktoros modell szinte tökéletesen illeszkedik az étellel való elégedettségi adatokhoz, de abszurd módon a házasságkötés előtt szinte nulla boldogságról drámai emelkedést jósol, majd tizenegy évvel később hasonló zuhanást a nulla boldogságig. *Lásd* BRIAN CHRISTIAN & TOM GRIFFITHS, *ALGORITHMS TO LIVE BY: AZ EMBERI DÖNTÉSEK SZÁMÍTÓGÉPES TUDOMÁNYA* 152-53 (2016).

II. A DEMOKRATIKUS EGYENLŐSÉG MINT SZEREP-  
VISSZAFORDÍTHATÓSÁG

Képzelnék el egy olyan világot, amelyben a csak gépi döntések pontosabbak és következetesebbek az esetek között, mint az ember plusz gépi döntések.<sup>33</sup> Egy ilyen világban lenne-e értelme az embereket a körforgásban tartani? Vajon lenne-e okunk ragaszkodni az emberi ítélőképességhez, mint bizonyos típusú döntések alapvető szempontjához, teljesen függetlenül - sőt, annak *ellenére is* -, *hogymilyen* eredményeket szül?

Úgy gondoljuk. Ebben a fejezetben azt vizsgáljuk, hogy miért, a döntéshozatalban a "szerepvisszahozhatóság" fogalmára összpontosítva - arra az elképzelésre, hogy az ítéletalkotás gyakorlóinak kölcsönösen sebezhetőnek kell lenniük a folyamatokkal és a hatásokkal szemben. Más szóval, van, és kellene is, hogy legyen olyan értelemben, hogy a résztvevők szerepe a folyamatban mindig megfordulhat: egy kissé más világban, de egy esetleges eseménysorozat esetében a döntéshozó lehet az, aki ítélkezik, nem pedig az, aki ítélkezik.

Ami tehát számít, az nem önmagában az emberség ténye. Ami számít, az az, hogy a döntéshozók képesek-e beleképzelni magukat egy érintett fél szerepébe, és fordítva - hogy mindkét résztvevő, és bizonyos értelemben az egész erkölcsi közösség demokratikus aktusként értelmezhesse az ítélkezést.

## A. "CSAK ISTEN KEGYELMÉBŐL"

Az a gondolat, hogy az embernek alázatosnak kell lennie annak a lehetősége előtt, hogy az élet egészen másképp is alakulhatott volna - és különösen, hogy könnyen elfoglalhatta volna az "elítelt" pozícióját -, ősi gondolatként hat ránk. Talán egyidős az emberi fajjal. A gondolat modern változata a tizenhatodik századig nyúlik vissza, amikor John Bradford angol reformátor állítólag kijelentette, miután látta az akasztófán felakasztott elítélteket: "de Isten kegyelméből John Bradford is elmenne".<sup>34</sup> Nemrégiben, 1964-ben Phil Ochs folkzenész ugyanezt az érzést fogalmazta meg vallási felhang nélkül: "Csak a szerencsének köszönhetjük, hogy te vagy én."<sup>35</sup>

---

<sup>33</sup> Ahogy az elején mondtuk, tisztában vagyunk vele, hogy sok okunk van arra, hogy szkeptikusak legyünk azzal kapcsolatban, hogy ezek a feltételek hamarosan - vagy talán soha - nem válnak valóra. De mi tudatosan arra a feltételezésre alapozzuk ezt az *elvi* érvelést, hogy az emberi ítélőképesség instrumentális értékei eltűntek. Bármilyen mértékben is működnek még ezek az értékek a gyakorlatban, ezek "extra cukormázatot tesznek a már megmázásodott tortára". *Yates v. United States*, 135 S. Ct. 1074, 1093 (2015) (Kagan, J., különvéleményt nyilvánítva).

<sup>34</sup> THE WRITINGS OF JOHN BRADFORD, M.A. xliii (Aubrey Townsend szerk., 1853), <https://archive.org/stream/writingsofjohnbr02braduoft#page/n5/mode/2up> [<https://perma.cc/HA8J-TKM6>].

<sup>35</sup> PHIL OCHS, *There But For Fortune*, a NEW FOLKS VOL. 2 (Vanguard 1964).

Az érzés visszhangzik. Azt állítják, hogy hasonlóságaink eltörpülnek a különbözőségeink mellett; hogy az élet nagyrészt a szerencsétől, vagy vallási nyelven szólva a kegyelemtől függ. Ha a szerencse szele másképp fújna, sokunk egyéni élete - és bizonyos szinten az egész társadalmi élet - is másképp alakulna.

Ez a gondolat azért erős, és azért kötődik erőteljesen a demokráciához, mert összeköti a sorsunkat. Ez a teológiai előfutára John Rawls igazságosságelméletének, T. M. Scanlon méltányosságelméletének és más filozófiai rendszereknek, amelyek ragaszkodnak a normatív kérdések *kontextuson kívüli* megoldásához - tekintet nélkül arra, hogy mely konkrét emberek veszítenek vagy nyernek.<sup>36</sup> Rawls híres módon a "tudatlanság fátyolára" hivatkozott, hogy megragadja ezt a gondolatot, Scanlon<sup>37</sup> pedig az "általános érvelés" analóg leírását dolgozta ki.<sup>38</sup> A végeredmény mindkét esetben ugyanaz: az elosztási döntések csak akkor legitimek, ha általánosságban igazolhatóak, függetlenül attól, hogy történetesen ki van a "nyertes" és a "vesztes" pozícióban.

A szabályok megfogalmazására alkalmazva - a rendszerszintű kérdések olyan módon történő megoldása, amely az általános érdekeket, nem pedig egyes csoportok érdekeit szolgálja - ez az elv meglehetősen ismerős. Sőt, már-már magától értetődőnek tűnik. Egy demokratikus társadalomban természetesen az "általános" indoklások felé húzunk, amikor a különböző politikák előnyeit vitatjuk meg.<sup>39</sup> Vagy fordítva, az olyan indoklások, amelyek *nem* általános formát öltenek, intuitív módon nem illenek a demokratikus érzékenységhez. Ha valaki megpróbál

<sup>36</sup> T.M. SCANLON, WHAT WE OWE TO EACH ANOTHER 189-90 (1998) (különböző filozófiai álláspontok leírása - köztük az ő, valamint Rawls, Kant és Habermas álláspontja -, amelyek az erkölcsi kérdésekkel kapcsolatos "kontextuson kívüli" érvelés valamilyen változatát vallják).

<sup>37</sup> Lásd JOHN RAWLS, A THEORY OF JUSTICE (1971); lásd még John E. Roemer & Alan Trannoy, "Equality of Opportunity," in HANDBOOK OF INCOME DISTRIBUTION (Atkinson & Bourgoignon szerk.), (2014) (kifejtve, hogy Rawls központi törekvése az volt, hogy amellet érveljen, hogy "a döntéshozókat megvédjé a helyzetükre vonatkozó olyan információk ismeretétől, amelyek [erkölcsileg] "önkényesek", hogy az igazságos elosztásra vonatkozó döntéseik [a döntéseik] pártatlanok legyenek . . . . [így] az igazságosság elvei a racionalitásból és a pártatlanságból származtathatók").

<sup>38</sup> Lásd: SCANLON, *Supra* 36. lábjegyzet; Rahul Kumar, *Reasonable Reasons in Contractualist Moral Argument*, 114 ETHICS 6, 6-7 (2003) (Scanlon álláspontját a következőképpen foglalja össze: a szabályozás indokoltsága azon múlik, hogy "a [] szabályozásra javasolt konkrét elv ... olyan-e, amelyet megfelelően motiváltan senki sem tudna ésszerűen elutasítani a tájékozott, kényszerítés nélküli, általános megállapodás alapjaként").

<sup>39</sup> A gyakorlatban persze a tervezési döntéseink előnyei inkább egyes csoportokat fognak érinteni, mint másokat; a *valóságban* mindannyian meghatározott társadalmi pozíciókat foglalunk el, még akkor is, ha normatív érvelésünk során megpróbálunk e pozíción túlmutatóan általánosítani. Valójában a meggyőzés feladata nagyon gyakran az, hogy meggyőzzünk valakit, hogy elvonatkoztasson saját társadalmi pozíciójától - hogy elfogadjon egy olyan normatív álláspontot, amely ellentétes a saját közvetlen érdekeivel (vagy, hasonlóképpen, hogy igazolja a saját közvetlen érdekeinek kedvező normatív álláspontot annak bizonyításával, hogy az álláspont elfogadásának *oka* általános, nem pedig személyes).

egy politikát azzal az indokkal véd, hogy az a saját zsebeit tömné meg, vagy hogy az hátrányos helyzetbe hozna egy bizonyos csoportot, az érv nem csak hogy nem meggyőző, de még csak megfontolásra sem érdemes.

Az "általános" érvelés folyamata, az "ott van, de a szerencsétlenség miatt" megközelítés alkalmazása az ítékezésben azonban a szabályok konkrét esetekre való alkalmazására is kihat - ha lehet, még élesebben. Amikor a szabályok megfogalmazására kerül sor, a közösség minden tagja "megfordítható", oly módon, hogy mindenkinek képesnek kellene lennie azt mondani: *az e szabály mögött álló érvelés olyan, amelyet társadalmi helyzetemtől függetlenül elvárható, hogy elfogadjak*. Természetesen ez egy nagyon tág fogalom, amely végtelen elosztási világokat felölelő képzeletbeli bravúrt igényel. Mégis az érvelésnek ez a stílusa is teljesen természetes. Éppen a tágassága és az általános jóléti hatások megjósolásának nehézsége miatt - vagyis hogy végül mely konkrét emberek viselik majd egy adott politika terheit - gyakran alapértelmezésből az általános érvelés felé húzunk.

Ezzel szemben a szabályok konkrét ügyekre való alkalmazása során teljesen világos, hogy mely konkrét személyek viselik a vonatkozó terheket - a döntésre felkért fél, akinek el kell viselnie az "ítélet gyötrelmeit",<sup>40</sup> és az a fél, akinek a sorsát a döntés végül alakítja. Az absztraktból a konkrét felé haladva a "szerep-visszafordíthatóság" többé válhat, mint az érvelés egyszerű metaforikus eszköze. Pszichológiai valósággá válhat, lehetővé téve a döntéshozatali folyamat résztvevői számára, hogy *szó szerint* elképzeljék, hogy helyet cserélnek. És amikor ez megtörténik, a résztvevők valamilyen szinten arra kényszerülnek, hogy elfoglalják az "általános" pozíciót.

Kicsit költőibben fogalmazva, elképzelhető, hogy az érintett fél - egy kedvezőtlen döntés esetén - a következő gondolatokra gondol:

Nyilvánvalóan szeretném ezt elkerülni. Nem akarom ezt a megbélyegzést, ezt a fájdalmat. Mégis... Nem mondhatom, hogy ha esküdtnem kellett volna lennem - és tényleg, néhány körülménytől eltekintve lehetett volna -, másképp vagy "jobban" cselekedtem volna. Ebben az értelemben, azt hiszem, van *némi* elfogadásom. Legalábbis ez a döntéshozó - az én konkrét "bíró" - nem hibáztatható.

Hasonlóképpen, fordítva: elképzelhető, hogy a döntéshozó - az a személy, akit arra kérnek, hogy ítélkezzen saját erkölcsi közösségének<sup>41</sup> egy tagja felett - ilyen gondolatmenetet folytat:

---

<sup>40</sup> Owen M. Fiss, a *megegyezés ellen*, YALE93 L. J. 1073, 1086 (1984).

<sup>41</sup> *Vö.* Máté 7:1-2 ("Ne ítéljétek, hogy meg ne ítéljenek benneteket. Mert amilyen ítélettel ítéltek, azzal ítéltetek meg, és amilyen mértékkel mérték, azzal mérnek vissza nektek."); Ávot 2:4 ("És ne ítélkeztek embertársatok felett, amíg nem álltatok a helyére.").

Nem ezt kértem; tényleg, inkább csinálnék bármi mást. Nem állítok semmiféle erkölcsi vagy egyéb felsőbbrendűséget. De néhány véletlen körülmény miatt inkább az én sorsom, mint az övé lehetett volna a tét. És ha így lenne, ha fordult volna a kocka, akkor vélhetően ő is ugyanezt az ítéletet hozná rólam, és nekem is ugyanezt a végeredményt kellene elfogadnom. Úgyhogy, azt hiszem, elsétálhatok a dolog elől. Nem azért, mert "helyesen" cselekedtem, bármit is jelentsen ez. Hanem azért, mert hiszem, hogy ő - az elítélt fél - ugyanígy cselekedett volna.

Ebben az értelemben a szabályok alkalmazásának aktusa közvetlenebbül emlékeztet a társadalmi esetlegességre, mint a szabályok megfogalmazásának (ritkább) aktusa. Amikor a "szerepvisszafordítható" döntéshozók bizonyos esetekben ítéletet hoznak, szembesülnek azzal a valósággal, hogy az események nagyrészt véletlenszerű sorozata nélkül ők is az érintett fél helyében lehettek volna. Értelemszerűen saját maguk felett ítélekeznek.

## B. ÍTÉLŐKÉPESSÉG ÉS ÖNSZERZŐSÉG

Az ilyen általános érvelés - mind a szabályok megfogalmazása, mind azok konkrét esetekre való alkalmazása során - a demokrácia szerves része. Ahhoz, hogy demokratikusan legitimek legyenek, a szabályoknak és az ítéleteknek a népszuverenitás eredményének kell lenniük, valaminek, ami felett a nép gyakorolja a szerzőséget. A szavazás ennek egyik eleme, de önmagában nem elegendő.<sup>42</sup> Jóval azután, hogy hazaérünk a szavazóurnából, a kérdés az, hogy el tudjuk-e fogadni a szabályok - jogszabályok, irányelvek és egyéb hivatalos aktusok - megfogalmazását és alkalmazását is sajátunknak.<sup>43</sup>

Rendszerszinten a kérdés az, hogy a szabályok értelmezhetők-e úgy, mint olyan *korlátok*, amelyeket mi magunk szabunk meg magunknak, és nem mint olyanok, amelyeket kívülről kényszerítenek ránk. Ez az eszme az, ami megkülönbözteti a demokráciákat a monarchiáktól, a kasztrendszerektől és a gyarmati berendezkedésektől. Ahhoz, hogy az ideálnak megfeleljen, nem kell mindenkinek mindig és mindenkor egyetértenie az összes szabály tartalmával - ez teljesen megvalósíthatatlan törekvés. De a demokratikus származás még mindig számít. A szabályoknak végső soron általános okokra kell visszavezetniük; többről kell szólniuk, mint a haszonszerzésről, a nepotizmusról, a plutokráciáról és más csupasz hatalomgyakorlásról.

És ugyanez igaz - ugyanezen alapvetően demokratikus okból - az egyéni ítéletekre is. Ahogyan mindannyiunknak képesnek kell lennünk arra, hogy a szabályokat a magunkra vonatkozó korlátozásoknak tekintsük, a döntéshozóknak képesnek kell lenniük arra, hogy lássák

---

<sup>42</sup> Innen ered a "bizalmi politikai elmélet" vonzereje - az az elképzelés, hogy a törvényhozók, bírák és más köztisztviselők a legjobban úgy értelmezhetők, mint akik az emberek ügynökeiként működnek. *Lásd pl.* „Ethan J. Lieb et al., *A Fiduciary Theory of Judging*, 101 CAL. L. REV. 699 (2013); D. Theodore Rave, *Politicians as Fiduciaries*, HARV126. L. REV. 671 (2013).

<sup>43</sup> *Lásd* Brennan-Marquez & Kahn, 4. lábjegyzet, 115., 163-77. lábjegyzet. A politikai elméletben ez a gondolat végső soron Hobbesra vezethető vissza. *Lásd id.* 163-73. o.; THOMAS HOBBS, THE LEVIATHAN (1668).

az ítélezés aktusát, mint olyasvalamit, amit valójában önmagukkal tesznek. Nem szó szerint, mert természetesen az érintett fél, nem pedig a döntéshozó az , akire az ítélet terhe nehezedik

.De képletesen éstörekvésszerűen - mivel a döntéshozó ugyanolyan könnyen lehetett volna (de szerencsés esetben) az érintett fél helyzetében, és a konkrét ítélet alapjának, csakúgy, mint a szabály alapjának, amelyből az következik, olyan okokon kell alapulnia, amelyeknek semmi közük a társadalmi véletlenhez. Ebben az értelemben az érvelés fentebb elképzelt stilizált formái - amelyekben az érintett fél fontolóra veszi egy döntés meghozatalát, a döntéshozó pedig fontolóra veszi, hogy kölcsönösen viselje annak terhét - még tovább fejleszthetők. Nem pusztán arról van szó, hogy a döntéshozónak el kell tudnia képzelni, hogy ugyanazt az ítéletet fordítvaelfogadja. Azt kell tudnia mondani: *ez a döntés a demokrácia eredménye; egy olyan kényszert tükröz, amelyet úgy döntöttünk, hogy magunkra kényszerítünk, és ebben az esetben történetesen nem nekem, hanem egy másik személynek kell felelnie érte.* Hasonlóképpen, az érintett félnek nem pusztán el kell tudnia képzelni, hogy fordítva ítélezik, hanem képesnek kell lennie azt mondani: *ez a döntés a demokrácia eredménye; egy olyan kényszert tükröz, amelyről úgy döntöttünk, hogy magunkra kényszerítjük, és ebben az esetben történetesen nem én, hanem egy másik személy az, aki....*

*végrehajtva.*

Más szóval, az egyéni ítélezési aktusokat, nem kevésbé, mint az általános szabályokat, önmeghatározásként kell értelmezni. Bár történetesen *ez a* konkrét döntéshozó az, akinek ítéletet kell hoznia, és *ez az* érintett fél az, akinek el kell viselnie annak eredményét, ez bármelyikünk lehetett volna - és egy kontrafaktuális világban lehet, hogy bármelyikünk lehetett volna, bármelyik szerepben. Ismétlem, ez különbözteti meg a demokratikus kormányzást a szigorúbb hierarchikus alternatíváktól. A gyakorlatban tudjuk, hogy a formálisan demokratikus intézmények gyakran nem felelnek meg ennek az elvnek, és ilyenkor a döntéshozók sokban hasonlítanak a királyokra vagy az imperialistákra. A gyakorlati hiányosságok azonban nem veszítik el az elv jelentőségét. Az elv továbbra is törekvés marad - egy normatív alapelv.

Természetesen abban is reménykedünk, hogy a demokratikus struktúrák üdvözlendő eredményeket hoznak. A demokrácia melletti érv - történelmileg és gyakran koncepcionálisan is - részben az, hogy a hatalmat a nép elszámoltathatóságának veti alá, és ezáltal a hatalom kevésbé lesz fogékony a visszaélésekre. Hogy ez empirikusan igaz-e, az attól függ, kit kérdezzünk. A mi céljaink szempontjából azonban az a fontos, hogy a demokrácia értéke valójában nem az általa elért eredményektől függ. A szörnyű eredmények, mint minden más dolog esetében is, bizalomvesztésre adhatnak okot. De az alapvető állítás nem az, hogy a demokráciák mindig jobban teljesítenek - mindig jobb döntéseket hoznak a szabályokról -, mint az alternatív kormányzati rendszerek. Ez inkább egy állítás.

az önrendelkezésről. Még ha a demokratikus eredmények politikai szinten el is maradnak, akkor is *a miénk*.<sup>44</sup>

Ugyanez vonatkozik az egyes ítéletek esetében is. Mint a demokratikus jogalkotás esetében általában, remélhetjük, hogy a szerepvisszafordíthatóság elősegíti a jobb döntéseket - ha nem is a konkrét eredmények tekintetében (mivel ezt a változót az elemzés kedvéért állandónak tartjuk),<sup>45</sup> de legalábbis az ítélethozatal módját illetően: a hangnemet, a textúrát, az ítélet néha tragikus jellegeinek figyelembevételét. Valószínűnek tűnik például, hogy egy szerepmegfordítható döntéshozó valószínűleg úgy *gyötrődne* a nehéz döntéseken, ahogyan egy, a szabály hatókörétől mentes döntéshozó nem tenné. Az is hihetőnek tűnik, hogy a szerepét megfordítható döntéshozó hajlamos lenne a hivatalos ítélet mellett bocsánatkérést vagy más, alázatot és megértést kifejező szavakat is kiadni, vagy - az ügy típusától függően - esetleg a hivatalos ítélet mellett elítélő szavakat is.<sup>46</sup>

Röviden, akár számítanak (vagy egyáltalán előfordulnak) a szerepvisszafordítható ítéletnek ezek a perifériás aspektusai, akár nem, nem ezek igazolják a szerepvisszafordíthatóságot. A szerepvisszafordítható ítéletek előnyben részesítésének alapvető indoklása a demokratikus hűségük.

-

Mielőtt továbblépnénk, szeretnénk megelőlegezni egy kifogást - és remélhetőleg a válaszával tovább erősíthetjük állításunkat. Az ellenvetés a következő. *Miért lenne aggályos - kérdezhetnénk - demokratikus okokból a szabályok végrehajtásának gépekre való átruházása? Egy dolog azzal érvelni, hogy emberi közösségként képesnek kell lennünk arra, hogy saját sorsunkról döntsünk. De ha már egyszer ez a sors megválasztásra került - ha már a szabályokat előírták -, mi olyan gyenge abban, hogy*

---

<sup>44</sup> Hasonlót állíthatunk az autonómia elvéről is. Lehet, hogy jobb eredményekhez vezet, de lehet, hogy nem. Akárhogy is, a döntések *az önké*.

<sup>45</sup> Amint azt korábban megjegyeztük, ez nem jelenti azt, hogy a szerepvisszafordíthatóság *nem* feltétlenül befolyásolja a döntési eredményeket. Valójában ez nagyon is lehetséges. Egy kísérő tanulmányban azt tervezzük, hogy megvizsgáljuk azt a tételt, hogy a szerepvisszafordítható döntéshozók nagyobb valószínűséggel térnek el az előírt szabályoktól és az ortodoxiától, mint a nem szerepvisszafordítható társaik, az esküdtszék érvénytelenítésével mint paradigma esettel. Más szóval, leíróan hihetőnek tűnik számunkra - és valószínűsíthetően a normatív érték forrása -, hogy egy olyan döntéshozó, aki képes értékelní az ítélezés "ott van, de Isten kegyelméből/szerencséjére" aspektusát, hajlamosabb lesz arra, hogy a szabályok alkalmazását nehéz esetekre erkölcsileg kirívónak tekintse. Egyelőre azonban a lényeg az, hogy érvelésünkben semmi sem függ attól, hogy a szerepvisszafordíthatóság vonzerőt gyakorol-e a döntési eredményekre. Akár így van, akár nem, a demokratikus legitimitációra vonatkozó állítás megmarad.

<sup>46</sup> Például Chin bíró, a New York-i déli körzet bírása figyelemre méltó beszédet tartott, amikor Bernard Madoffot elítélte, és Madoff bűncselekményeit "rendkívüli gonoszságnak" minősítette, amelyek maximális büntetésért kiáltanak. *Lásd Benjamin Weiser, Madoff Judge Recalls Rationale For Imposing 150-Year Sentence*, N.Y. TIMES, June 29, 2011, at A1. Hogy ez a fajta közösségi büntető érzés megváltoztatja-e az ítélethozatalt - jó vagy rossz irányba -, azt a jövőbeni munkára hagyjuk.

*a végrehajtást erős mesterséges intelligenciára bízva? Végül is az ilyen delegálásnak legalább egy egyértelmű előnye van: megszabadít a döntéshozatal gyötrelmeitől.*

Ez az ellenvetés ismerős lehet. Először is, a jogi és ténybeli kérdések közötti különbséget követi nyomon. Az általános bölcsesség szerint az előbbiekhez normatív érvelésre van szükség, míg az utóbbiakhoz csak empirikus érvelésre - ami az emberek és a gépek közötti hihető munkamegosztást vetíti előre.<sup>47</sup> Ráadásul ez egy olyan érv, amelyet a technológia-rajongók is elő szoktak hozni. Az algoritmikus kormányzási körökben a géppárti érvek középpontjában gyakran az értékek és célok (amelyekbe az embereknek kell végső szót mondaniuk) és a végrehajtási eredmények (amelyek - a logika szerint - delegálhatók a gépekre) közötti különbségtétel áll.<sup>48</sup>

A probléma ezzel az ellenvetéssel az, hogy a logikája túl sokat bizonyít. Ha hajlandóak vagyunk nem demokratikus megközelítést alkalmazni a szabályok végrehajtásában, miért ne alkalmazhatnánk nem demokratikus megközelítést a szabályok megfogalmazásában is? Ha - ahogyan az ellenvetés feltételezi - nyugodtan ruházzuk fel gépeket az értékeink megvalósításának hatáskörével, miért ne engednénk meg nekik azt is, hogy ezeket az értékeket *megalkossák*?

A válasz nem lehet technológiai. Még ha a jelenlegi mesterséges intelligencia nem is képes "értékeket alkotni", nem számíthatunk arra, hogy ez örökké így marad.<sup>49</sup> A válasznak inkább a demokratikus származás fontosságáról kell szólnia. Egyszerűen illegitimnek tűnik - egy kasztrendszerhez hasonlóan -, ha az értékeket kívülről alakítják ki, és a szabályokat kívülről szabják meg. A fentiekben kifejtett okok miatt azonban a szabályok konkrét esetekre való alkalmazása *szintén* a demokratikus legitimitást érinti, és ennek folytán az önrendelkezés elveinek betartását is megköveteli. Végző soron tehát a végrehajtási hatalom intelligens gépekre való átruházásának nincs több (vagy kevesebb) értelme, mint az emberek egy kizárólagos részhalmazának - mint például a papoknak - átruházásának.

---

<sup>47</sup> *Lásd pl.*, Emad H. Atiq, *Legal vs. Factual Normative Questions & The True Scope of Ring*, 32 NOTRE DAME J.L. ETHICS & PUB. POLY 101, 103 (2018) ("A jogelméleti szakemberek körében uralkodó nézet, hogy a jog/tény megkülönböztetés követi vagy leképezi a normatív és empirikus kérdések közötti megkülönböztetést.").

<sup>48</sup> *Lásd pl.* Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147 (2017) (egy ilyen irányú érvelés kidolgozása a közigazgatási jogból vett szabályalkotás/bírászkodás dichotómiájának felhasználásával).

<sup>49</sup> *Lásd az alábbi I. részt.* Az érvelés kedvéért feltételezzük, hogy a gépek a mesterséges általános intelligencia (AGI) szintjén vagy azon túl vannak, vagyis képesek az emberi agy sokféle feladatának elvégzésére. Nincs konszenzus arról, hogy ez a technológia milyen formát ölthet, mivel nincs konszenzus arról, hogy mi az intelligencia. De minden olyan állítás, miszerint egy intelligens gép nem lesz képes "x" feladatra, mert x "emberi érzés", sok kívánnivalót hagy maga után. Képzljük el például, hogy az idegtudósok képesek megfejteni ideghálózatos agyunk rejtélyeit, és az informatikusok képesek pontosan lemásolni azokat szilíciumban. Akkor az ilyen gépek képesek lesznek "értékeket csinálni", legalább olyan jól,





és a múlt lovagjai. Meg tudnánk ezt csinálni? Igen, persze, hogy meg tudnánk. De a demokratikus értékek mást tanácsolnak.

### III. AZ ESKÜDTSZÉKI PEREK MINT PARADIGMÁS ESETEK

A szerepvisszafordíthatóság absztrakt vizsgálatát és védelmét követően vizsgáljuk meg a paradigmátikus szerepvisszafordítható ítékezés egy sajátos formáját: az esküdszéket. Sir Patrick Devlin az *esküdszéki tárgyalásról* szóló híres Hamlyn Lecture on *Trial By Jury* című előadásában így vezet:

Az esküdszéki tárgyalás nem olyan téma, amelyről valami nagyon újszerűt vagy nagyon mélyreható lehetne mondani. Az angol esküdszék nem azért olyan, amilyen, mert egy törvényhozó így rendelkezett, hanem mert így nőtt fel. Valójában elképzelhetetlen, hogy egy törvényhozó találta ki. Megszoktuk, és tudjuk, hogy működik; ha nem így lenne, azt kellene mondanunk, hogy nevetséges és megvalósíthatatlan ötletet testesít meg.<sup>50</sup>

Nem értünk egyet - mondhatni.

Devlin munkája az egyik legfontosabb mű a zsűri munkái közül; csodáljuk a hozzászólásait, és sokukat elfogadhatónak találjuk. Történelmi elemzése például kifogástalannak tűnik számunkra: Devlin ügyesen tekereg az első közigazgatási esküdszékektől (vizsgálat) a prezentációs esküdszékeken (a mi nagyesküdszékünk elődjén) át az ítélező esküdszékekig.<sup>51</sup> A történelemnek ezeket a háborgó hullámzásait nem egyetlen - pláne nem racionális - elme irányította, és mindig felüdülés olvasni Devlin kritikáját azokkal szemben, akik misztikus pontossággal akarnák a történelmi valóság gyötrelmes útját járni.<sup>52</sup> A mi érvelésünk nem történelmi érv.

<sup>50</sup> DEVLIN, *fenti* 2,3-4. lábjegyzet.

<sup>51</sup> *Id.* 5-14.

<sup>52</sup> Vegyük például az esküdtek hagyományos számát, tizenkettőt:

A tizenkettes számra sok romantikus magyarázatot adtak már - Izrael tizenkét törzse, a tizenkét pátriárka, Salamon tizenkét tisztje, akiről a Királyok könyvében olvashatunk, és a tizenkét apostol. Nem mindegyik felvetés egyformán szerencsés; az első azt sugallja, hogy lehet egy tizenharmadik esküdt, aki valahol a folyosón eltévedt, az utolsó pedig azt, hogy minden esküdszékben van egy Júdás. Nyilvánvaló, hogy olyan számot akartak, amely elég nagy ahhoz, hogy a győztes oldal javára félelmetes véleményt alkosson; és kétségtelen, hogy a tíz, tizenegy vagy tizenhárom helyett tizenkettőnek nagyjából ugyanaz volt az oka, mint amiért tizenkét pennyt adnak a shillinghez, és ami a korai angolok tizedes rendszerrel szembeni ellenszenvét mutatja.

*Id.* 8. o. Vagy gondoljunk csak azokra a régóta fennálló szabályokra, amelyek lehetővé tették a bírák számára, hogy az esküdteket étkezés nélkül bezárva tartsák, amíg egyhangú ítéletet nem hoznak:

Mindez nagyon emlékeztet az ítélet eredetére. Amit kerestek, az nem egy racionális következtetés volt, hanem egy jel, valami, ami a megpróbáltatás eredményéhez vagy a csatában aratott győzelemhez hasonlított; a folyamatot nem lehetett meghatározni, amíg meg nem szerezték, és ha már megvan, a megszerzésének módját kevésbé tartották fontosnak, mint azt a tény, hogy megvan.

*Id.* 51. o. Az esküdszéki tárgyalásról az esküdszéki eljárásra való áttérésről lásd: Elizabeth



És egyetértünk Devlin állításával, miszerint esküdszékeink egyes aspektusai továbbra is meglehetősen "nevetségesek és kivitelezhetetlenek", vagy legalábbis tökéletlenek.<sup>53</sup> Érvelésünk nem az esküdszéki perek dicshimnusza, különösen nem a gyakorlatban való megvalósításuk miatt.<sup>54</sup> Csak Devlin azon állításával vitatkozunk, hogy az esküdszék "nevetséges és kivitelezhetetlen ötletet" testesít meg.

De még itt sem mondunk igazából ellent Devlinnek. Az esküdszéki rendszerben sok minden megvalósíthatatlannak, sőt kissé nevetségesnek tűnik.<sup>55</sup> Tézisünk talán legpontosabb megfogalmazása tehát az, hogy meg kellene őriznünk az esküdszéket - és különösen a büntető esküdszéket - annak hiányosságai ellenére. Hogy miért? Mert az esküdszék erőteljesen megtestesíti a demokratikus szellemet.

---

*Bűnözés a JOG ÉS TÁRSADALOM A KÉSŐKORI KÖZÉPKORI ANGOLORSZÁGON ÉS ÍRORSZÁGON:*  
ESSAYS IN

HONOUR OF PAUL BRAND (Travis Baker szerk., 2018).

<sup>53</sup> Például miért van az, hogy esküdszékeink nem adnak magyarázatot vagy indoklást az ítéleteikhez, amikor általában tudjuk, hogy az indoklás megkövetelése jobban megalapozott döntéshez vezet? Devlin szerint pusztán a történelmi anakronizmus miatt: "Ez a kétértelműség jogától megfosztott orákulum. Az esküdszék eredetét tekintve éppúgy orákulum volt, mint a próbatétel - egyiket sem az értelemben fogalmazták meg -, az ítélet, éppúgy, mint a próbatétel eredménye, racionális kritikára nyitott volt." DEVLIN, *supra* note 2, 14. o. (enyhe pontozási változtatás). Különösen szimpatikus esetek nyomulnak végre rá erre az évszázados gyakorlatra. *Lásd* Pena- Rodriguez v. Colorado, 137 S. Ct. 855 (2017) (alkotmányosan megköveteli, hogy egy állam engedélyezze az esküdszéki ítélet megtámadását, ha a tanácskozás során faji előítéletességgel vádolják); *lásd még pl.* , Andrew E. Taslitz & Stephen E. Henderson, *Reforming the Grand Jury to Protect Privacy in Third Party Records*, 64 AM. U. L. REV. 195 (2014) (kritizálva a nagyesküdszék védelmének hiányát).

<sup>54</sup> Büntető esküdszékünk átgondolt, modern védelmét és a polgárok "cselekvésre való felhívását" *lásd* ANDREW GUTHRIE FERGUSON, WHY JURY DUTY MATTERS: A CITIZEN'S GUIDE TO CONSTITUTIONAL ACTION (2012). Végso soron Ferguson érvei, akárcsak a miénk, az esküdszéki bíráskodás rendszerét védik, nem pedig a végrehajtásával kapcsolatos esetleges hibákat.

<sup>55</sup> Gondoljunk csak Devlin csodálatos leírására:

Tizenkét (miért tizenkettő?) férfit és nőt véletlenszerűen kell kiválasztani; korábban soha nem volt tapasztalatuk a bizonyítékok mérlegelésében, és talán abban sem, hogy bírói gondolkodással bármilyen problémára alkalmazzák elméjüket; gyakran - ahogyan az 1853-as Common Law Commissioners tapintatosan fogalmazott - "nem szoktak hozzá a komoly szellemi gyakorlathoz vagy az elhúzódo gondolkodáshoz". Az ügy lehet bonyolult, hetekig is eltarthat, és az ügyvédek előttük tarthatják az iratok halmát, amelyek közül az esküdtek csak néhányat tekinthetnek meg. Lehet, hogy napokig hallgatják a szóbeli bizonyítékokat anélkül, hogy jegyzetelnének - legalábbis senki sem várja el tőlük, hogy jegyzetljenek, és az esküdtek páholyában nincs erre lehetőség, még könyöklőhely sem. Mégis azt mondják, hogy ők az egyedüli bírái az összes ténynek. Az ügy végén elvárják tőlük, hogy egy-két órán belül ugyanarra a következtetésre jussanak. Egyhangú ítéletük nélkül senkit sem lehet megbüntetni a nagyobb bűncselekmények egyikéért sem. Elméletileg nem lenne lehetséges a büntetőjogot ilyen eszközökkel sikeresen érvényesíteni.

DEVLIN, *fenti* megjegyzés, 2,4-5. pont.

a szerepvisszafordíthatóság.<sup>56</sup> Az egyik résztvevő szerint a zsűri "a legtisztább példája a demokrácia működésének, amit valaha is tapasztaltam."<sup>57</sup>

#### A. A "KORTÁRSAID ESKÜDTSZÉKE

Az esküdtszék védelmében el kell választanunk az ideálist a történelmitől. Az esküdtszék, enyhén szólva, mindig is tökéletlen volt. A legnagyobb igazságtalanságok némelyikét ők hozták,<sup>58</sup> és minden voltak, csak nem szerepük megfordítható, kizárva azokat, akiknek nincs bőséges vagyonuk, vagy akik nem egy kedvezőtlen fajhoz, nemhez vagy valláshoz tartoztak.<sup>59</sup> Ennek ellenére az idő múlásával az intézmény egyre demokratikusabbá vált, és egyre közelebb került a szerepvisszafordítható kialakításhoz.

Az esküdtszéki tárgyaláshoz való jognak lenyűgöző pedigréje van: ez az egyetlen olyan rendelkezés az amerikai Bill of Rights-ban, amely kifejezetten felesleges volt, mivel a III. cikk már előírta, hogy "[a]z összes bűncselekmény tárgyalása ... esküdtszék által történik".<sup>60</sup> A mai értelmezés szerint ezek a jogok megkövetelik, hogy az esküdtek egy tisztességes

---

<sup>56</sup> Lehetségesnek tartjuk, hogy maga Devlin is szimpatizált volna ezzel az állítással. Minden bizonnyal úgy vélte, hogy bár "az [esküdtszék] eredete véletlen ... a megtartása [szándékos]". *Id.* 154. o. És noha időnként egy eléggé nem elkötelezett, következetességen alapuló érvelést sürget az esküdtszék megőrzése mellett (*lásd pl.*, *id.* 57. o.), ha az egész érvelését tekintjük, talán a szerep-visszafordíthatóság korai támogatója lehetett, bár elsősorban annak a semmissé nyilvánításon keresztül történő eltérés összetevője miatt. *Lásd id.* 155-65. o. Az esküdtszéki döntéshozatal e lenyűgöző aspektusát a jövőbeli munkára hagyjuk.

<sup>57</sup> Raymond J. Brassard, *Juries Help Keep Our Democracy Working*, BOSTON GLOBE, 2003. május 1., A19. o. (idézi egy esküdtnak egy tárgyalást követően írt levelét). Az esküdtt így folytatta: "Mindenki részt vett. Nem tudom, hogy eljutottunk-e az igazságra, de úgy érzem, jó munkát végeztünk". *Id.* Ugyanennek a pontnak egy elméletibb kifejtését lásd: Jenny Carroll, *The Jury as Democracy*, ALA66. L. REV. 825 (2015).

<sup>58</sup> *Lásd pl.*, Hector (A Slave) v. State, Mo.2 166, 166 (1829) (az esküdtszék által hozott ítéletet figyelembe véve, ahol a vádlott azután tett vallomást, hogy "egész éjjel korbácsolták, ... sikoltozva az ostor alatt"); Brown v. Mississippi, 297 U.S. 278, 281 (1936) (az esküdtszék által hozott ítéletet figyelembe véve, ahol a vádlott azután tett vallomást, hogy "egy kötéllel [készer] egy fa ágához akasztották, majd [utána] ... egy fához kötözték és korbácsolták"). Figyelemre méltó, hogy a Brown-ügyben ezt a képviselőt az állam a tanúk padjára ültette a cáfolat során, és elismerte a korbácsolást. Érdekes megjegyezni, hogy a vallomásában az [egyik] vádlott megkorbácsolására való hivatkozással ... , és arra a kérdésre, hogy mennyire keményen korbácsolták meg, a helyettes kijelentette: "Nem túl sokat egy négerhez képest; nem annyit, amennyit én tettem volna, ha rám bízna". Két másik személyt is bemutatnak, akik részt vettek ezekben a korbácsolásokban, és elismerték ezt - egyetlen tanút sem mutattak be, aki tagadta volna ezt. *Id.* 284-85.

<sup>59</sup> *Lásd* Albert W. Alschuler & Andrew G. Deiss, *A Brief History of the Criminal Jury in the United States*, 61 U. CHI. L. REV. 867, 876-901 (1994) (az amerikai esküdtszék összetételének történelmi változásainak krónikája).

<sup>60</sup> U.S. CONST. art. III § 2; *lásd még* U.S. CONST. amend. VI. Amint Alschuler és Deiss rámutat, ez a redundancia nem meglepő - ez az egyetlen közös jog mind a tizenkét állam alkotmányában, amelyet a szövetségi alkotmányegyezmény előtt írtak. *Lásd* Alschuler &

Deiss, *supra* note at59 Az egyik kerületi bíró szavaival élve870., az esküdtszék "olyan amerikai, mint a rock 'n' roll".

a közösség keresztmetszete, egyetlen "megkülönböztető csoport" sincs szisztematikusan kizárva.<sup>61</sup> Még a "korlátlan" peremptorikus kihívás sem engedi meg többé a faji elfogultságot.<sup>62</sup> Az esküdtek pedig kölcsönösen kiszolgáltatottak az ítélkezési folyamatoknak és hatásoknak: nem csupán ugyanabban a joghatóságban kell élniük, ugyanazon törvények hatálya alá tartoznak,<sup>63</sup> hanem büntetőjogi szankciókkal is sújthatók az előttük lévő ügy eldöntése során tanúsított helytelen magatartásukért.<sup>64</sup> Az ítélet alávetettek - a büntetőjogi vádlottak - kölcsönösen részt vehettek egy korábbi esküdtszékben, és továbbra is alkalmasak arra, hogy egy jövőbeli esküdtszékben is részt vegyenek.<sup>65</sup>

Legalábbis *elméletben* tehát a résztvevők szerepe az esküdtszéki eljárásban könnyen felcserélhető. Aligha távoli az a kontrafaktuális világ, amelyben bármelyik esküdt ehelyett a vádlott székében ül - de egy kisebb véletlen eseménysorozat miatt könnyen lehet, hogy ez lenne a valóság. Úgy véljük, hogy ez a megfordítható szerepdinamika a "kortársak esküdtszéke" eszményének fogalmi alapja - bár nem a történelmi származása. Ez magyarázza, hogy miért

---

roll." William G. Young, *Vanishing Trials, Vanishing Juries, Vanishing Constitution*, 40 SUFFOLK U.L. REV. 67 (2006).

<sup>61</sup> *Lásd* Taylor kontra Louisiana, 419 U.S. 522, 530 (1975) (megállapítva, hogy a bírósági üléseknek a közösség tisztességes keresztmetszetét kell alkotniuk, ami azt jelenti, hogy egyetlen megkülönböztető csoportot sem zárnak ki szisztematikusan).

<sup>62</sup> *Lásd* Batson v. Kentucky, 476 U.S. 79 (1986) (a faji alapú peremptory challenge tiltása).

<sup>63</sup> *Lásd*: U.S. CONST. amend. VI ("Minden büntetőeljárásban a vádlottnak joga van ahhoz, hogy ... a bűncselekmény elkövetésének helye szerinti állam és körzet pártatlan esküdtszéke előtt tárgyalják, amely körzetet a törvény által előzetesen meghatározták.")

<sup>64</sup> *Lásd pl.* Molly McDonough, *Rogue Jurors*, ABA JOURNAL, 2006. október (ismertetve többek között egy olyan esetet, amelyben egy esküdt hat és fél év börtönbüntetésre ítélték összeesküvésért, bírói tiszteletlenségért és az igazságszolgáltatás akadályozásáért).

<sup>65</sup> A szövetségi esküdtszéki szolgálatra vonatkozó képesítések reprezentatívak és minimálisak:

Az ilyen meghatározás során a ... bíró ... bármely személyt alkalmasnak tekint a kerületi bíróság nagy- és kisesküdtzékében való részvételre, kivéve, ha-

(1) nem az Egyesült Államok tizennyolcadik életévét betöltött állampolgára, aki egy éve a bírósági körzetben lakik;

(2) nem képes az angol nyelvet olyan szinten olvasni, írni és érteni, hogy az esküdtképesítő űrlapot kielégítően ki tudja tölteni;

(3) nem beszél angolul;

(4) szellemi vagy fizikai fogyatékosága miatt nem képes kielégítő esküdtszéki szolgálatot teljesíteni; vagy

(5) egy évnél hosszabb szabadságvesztéssel büntetendő bűncselekmény elkövetése miatt vádat emeltek ellene, vagy egy állami vagy szövetségi bíróság előtt elítélték, és állampolgári jogait nem állították helyre.

28 U.S.C. § 1865(b)(2015). Néhány további személy mentesül (28 U.S.C. § 1863(b)(6)(2015)) vagy felmentés (28 U.S.C. § 1863(b)(5)).

még a vádlott sem mondhat le egyoldalúan az esküdtszéki felülvizsgálatról.<sup>66</sup> És bár egyetlen tizenkét fős esküdtszék aligha fogja jól (vagy legalábbis teljes mértékben) képviselni a "társait", a büntető esküdtszék demokratikus, szerepvisszafordítható normát testesít meg.<sup>67</sup> Alexis de Tocqueville szavaival élve, az amerikai esküdtszék "politikai intézmény", "a nép szuverenitásának egyik formája", és annak is szánták.<sup>68</sup>

## B. AZ ELTŰNŐ BŰNÜGYI ESKÜDTSZÉK

Nem meglepő tehát, hogy az esküdtszéki büntetőperек eltűnése miatti aggodalom nem újdonság, legalábbis akkor, ha a helyébe valószínűleg - ahogy a gyakorlat is mutatja - a vádalkuk kevésbé visszafordítható, kevésbé demokratikus rendszere lép, amelyet látszólag inkább a hatékonyság (és akár a karrierépítés), mint az egyes ügyek megfelelő eldöntése iránti aggodalom vezérel.<sup>69</sup> Több mint hatvan évvel ezelőtt az esküdtszék hanyatlása volt a mozgatórugó.

---

<sup>66</sup> *Lásd* Singer kontra Egyesült Államok, 380 U.S. 24 (1965) (megállapítva, hogy "az Alkotmány nem biztosít és nem ismeri el a büntetőjogi vádlottak azon jogát, hogy ügyüket kizárólag bíró előtt tárgyalják"). John Langbein szerint,

Az angol forrásokból származó történeti feljegyzésekkel rendelkezünk néhány olyan XVIII. századi esetről, amelyben néhány szánalmas vádlott, akit tetten értek, vagy más módon érezte ügye reménytelenségét, megpróbálta bűnösnek vallani magát. . . . A bíró újra és újra sürgette a vádlottat, hogy vallja magát "nem bűnösnek", és vigye ügyét az esküdtszék elé.

John H. Langbein, *Az írott alkotmányok mítoszáról: The Disappearance of Criminal Jury Trial*, 15 HARV. J. L. & PUB. POL'119, Y 120(1992).

<sup>67</sup> Miközben úgy véljük, hogy ez a szerepvisszafordíthatóság az esküdtszék legitimitásának záloga, megjegyezzük, hogy a laikus esküdtek "rendszeren kívüli" jellegükkel is rendelkeznek, ami további előnyökkel járhat. Devlin szavaival élve: "[a]z esküdtek úgy hallják a tanút, mint aki éppúgy nem ismeri az ügyvédek gondolkodásmódját, mint ők; [ez] a nagy előnye annak, hogy . . . a [saját] társai ítélkeznek". DEVLIN, *supra* note at2, Vagy140., Sir William Holdsworth szavaival élve, "az esküdtszéki rendszer . . . arra törekszik, hogy a törvényt érthetővé tegye azáltal, hogy az élet általános tényeivel tartja kapcsolatban."  
*Id.* 148.

<sup>68</sup>1 ALEXIS DE TOCQUEVILLE, DEMOKRÁCIA AMERIKÁBAN (3041835).

Az esküdtszék a nemzetnek az a része, amelyre a törvények végrehajtását bízzák, ahogyan a parlament házai a nemzetnek azt a részét alkotják, amely a törvényeket hozza; és annak érdekében, hogy a társadalmat következetesen és egységesen lehessen kormányozni, az esküdtszéki tagságra jogosult polgárok listájának a választók listájával együtt kell növekednie és csökkennie.

*Id;* lásd még Carroll, *supra* note 57; Jocelyn Simonson, *The Criminal Court Audience in a Post-Trial World*, 127 HARV. L. REV. 2173, 2184-90 (2014); Herbert Mitgang, *Inside the Jury Room*, N.Y. TIMES, 1986. április 8., 00017. o.; vö. Jason Kreag, *The Jury's Brady Right*, 98 B.U. L. REV. 345 (2018) (azzal érvelve, hogy az esküdtszéknek a büntető igazságszolgáltatási rendszerben még a vádlott egyéni érdekeitől elkülönülve is megkülönböztető-legitimáló szerepe van).

<sup>69</sup> Az esküdtszéki tárgyalás eltűnését sirató szakirodalom terjedelmes. *Lásd pl.* Robert J. Conrad, Jr. & Katy L. Clements, *The Vanishing Criminal Jury Trial: From Trial Judges to Sentencing Judges*, GEO86. WASH. L. REV. 99 (2018) (bírói nézőpont bemutatása a



Devlin gondolkodását.<sup>70</sup> Még az ő idejében is a büntetőügyek vádlottjainak mintegy 85%-a a bírói döntést választotta, és még az esküdszék elé kerülő maradék 15% közül is kétharmaduk bűnösnek vallotta magát.<sup>71</sup> Tehát száz alkalmas büntetőeljárásból csak öt esetben döntöttek esküdszék előtt. Messze van tehát a két évszázaddal korábbi állapot, amikor "[a] bűncselekményi ügyekben nem volt vádalku".<sup>72</sup>

Az eltűnés idővel fokozódott. Ahogy John Langbein két évtizeddel ezelőtt érvelt (és a kérdés azóta csak még sürgetőbbé vált),

a szövetségi büntetőeljárások drámai csökkenése); NACDL, THE TRIAL PENALTY: THE SIXTH AMENDMENT RIGHT TO TRIAL ON THE VERGE OF EXTINCTION AND HOW TO SAVE IT (2018),

elérhető: [www.nacdl.org/trialpenaltyreport](http://www.nacdl.org/trialpenaltyreport) [<https://perma.cc/U9XE-3J7H>] (a védelem álláspontját ismerteti); Michael Vitiello, *Bargained-for-Justice: Lessons from the Italian?*, 48 U. PAC. L. REV. 247 (2017) (elemezve Olaszország tétova, de egyre nagyobb mértékben a vádalkukra való támaszkodást); STEPHANOS BIBAS, THE MACHINERY OF CRIMINAL JUSTICE xvi-xvii (2012).

(érvelve amellet, hogy "[a] büntető igazságszolgáltatás korábban individualizált, erkölcsös, átlátható és részvételi jellegű volt, mára azonban személytelenné, erkölctelenné, rejtetté és az emberektől elszigetelté vált"); Jonathan A. Rapping, *Who's Guarding the Henhouse? How the American Prosecutor Came to Devour Those He Is Sworn to Protect*, 51 WASHBURN L. J. 513, 514 (2012) (a problémát egy kirendelt védő szemszögéből vizsgálja); Young, *supra* note 60 (egy másik bírói nézőpont bemutatása); Ronald F. Wright, *Trial Distortion and the End of Innocence in Federal Criminal Justice*, 154 U. PA. L. REV. 79, 83 (2005) (a bűnösségre való hivatkozás "tárgyalás torzítás elméletének" kidolgozása); GEORGE FISHER, PLEA BARGAINING'S TRIUMPH: A HISTORY OF PLEA BARGAINING IN AMERICA (2003) (aprólékos krónika a vádalku Massachusettsben való elterjedéséről); Albert W. Alschuler, *Plea Bargaining and Its History*, 79 COLUM. L. REV. 1 (1979) (a vádalku történetének kidolgozása); *lásd még* United States

v. Stevenson, 2018 U.S. Dist. LEXIS 61988 at \*1 (S.D.W.Va. 2018. ápr. 12.) (a többszörös vádalkuk elutasítását azzal magyarázza, hogy "nem közérdekű[,] ... .. a büntetőjogi ítélkezés a nyilvánosságtól az ügyészséghez kerül a célszerűség érdekében, a büntető igazságszolgáltatási rendszerbe vetett bizalom és annak hatékonysága árán"); Emily Yoffe, *Innocence Is Irrelevant*, ATLANTIC, Sept. 2017 ("Azzal, hogy az ország elfogadta minden dolog kriminalizálását, a büntetőjogi rendszer felduzzasztását és a vádalkuk térnyerését, garantálta, hogy állampolgárok millióinak nem lesz tisztességes esélye arra, hogy hétköznapi életet éljenek."); vö. Andrew Manuel Crespo, *The Hidden Law of Plea Bargaining*, 118 COLUM. L. REV. 1303, 1305-06 (2018) (azzal érvelve, hogy a tudósok nem értékelték a vádalkuk szabályozó "alkotmány alatti eljárási jogot"). Az "eltűnő tárgyalás" irodalma hatalmas - a Westlaw-on a kifejezés egyszerű változatai több száz másodlagos forrásban jelennek meg. Így csak néhány darabot emelünk ki, amelyek az érdeklődő olvasót sok máshoz vezetik.

<sup>70</sup> *Lásd* DEVLIN, 2. lábjegyzet, 129-33. o. A korábbi amerikai tanulmányok az 1920-as és 30-as évekből származnak. FISHER, 69. lábjegyzet, 6-8. o. Így történt, hogy 1928-ban Raymond Moley kiadta "The Vanishing Jury" (Az eltűnő esküdszék) című könyvét. Raymond Moley, *The Vanishing Jury*, S2. CAL. L. REV. 97 (1928).

<sup>71</sup> DEVLIN, *Supra* note 2, 129-30, 176-77 n.1. Az eredmény az lett, hogy "napjainkban évente körülbelül öt-hatezer esküdszéki tárgyalást tartanak". *Id.* 130.

<sup>72</sup> J. M. BEATTIE, CRIME AND THE COURTS IN ENGLAND, 1660-1800, 336-37. o. (1986). Még ha kissé túlzó is (*lásd pl.* Alschuler & Deiss, *Supra* 59. lábjegyzet, 922-23. o.), és még ha természetesen kritikus különbség is van a bűnösségre való hivatkozás elfogadása és a vádalku között, a számok sokatmondóak.

míg a hatodik módosítás azt parancsolja, hogy "*minden* büntetőeljárásban a vádlottnak joga van a gyors és nyilvános, pártatlan esküdtszék által lefolytatott tárgyaláshoz".<sup>73</sup>

az esküdtszéki tárgyalásnak a büntetőügyek elbírálásában való alkalmazására pontosabb kifejezés lenne a "gyakorlatilag nincs". Az emberi jogok csodálatos garanciáihoz hasonlóan, amelyek a totalitárius államok állítólagos alkotmányait díszítik, az esküdtszéki büntetőeljárások rutinszerűségére vonatkozó garanciánk is csalás Megtalálható-e vízilo Bronxban?

Igen, van egy a bronxi állatkertben, de ennek semmi köze a bronxi élethez. Ez már a múlté. És ugyanígy, megdöbbentő módon, a büntető esküdtszéki tárgyalás is.<sup>74</sup>

Hogy ne tűnjünk túl költőinek, teljes mértékben elismerjük, hogy az esküdtszéki büntetőeljárás alapításunk idején "összefoglaló eljárás" volt, amely hihetetlenül hiányzik a modern érzéseknek.<sup>75</sup> Aligha vágyunk visszatérni azokhoz az orákulumos időkhöz, amikor a büntetőperek alig voltak megbízhatóbbak, mint egy kockadobás, és ésszerű, hogy a sokkal több információval - beleértve a sokkal jobb törvényszéki vizsgálatokat is - rendelkező világ több tárgyalásos megegyezéshez vezetne.<sup>76</sup>

<sup>73</sup> U.S. CONST. amend. VI (kiemelés hozzáadva). *Lásd* AKHIL REED AMAR, AMERICA'S CONSTITUTION: A BIOGRAPHY 236 (2005) ("A büntető esküdtszék nélkül ülő büntetőbíró egyszerűen nem volt szabályszerűen felállított szövetségi bíróság, amely képes volt ügyeket tárgyalni, ahogy a Ház nélkül ülő Szenátus sem volt szabályszerűen felállított szövetségi törvényhozás, amely képes volt törvényeket alkotni.")

<sup>74</sup> Langbein, *Supra* 66. lábjegyzet, 120-21. o. A Legfelsőbb Bíróság a *Missouri kontra Frye* ügyben Robert Scott és Bill Stuntz tudósokat idézve úgy fogalmazott, hogy "a vádalku ... nem a büntető igazságszolgáltatási rendszer kiegészítője; ez a büntető igazságszolgáltatási rendszer". 566 U.S. 134, 144 (2012). George Fisher szavaival élve, a vádalku "győzedelmeskedett", "a legyőzött esküdtszékünket az ellenállás kis zsebeibe hajtotta". FISHER, *supra* note 69, at 1. Laura Appleman szavaival élve pedig "mi, a nép" "száműzetésbe[kerültünk] az igazságszolgáltatásból". LAURA I. APPLEMAN, DEFENDING THE JURY: CRIME, COMMUNITY, AND THE CONSTITUTION 3 (2015). Appleman így folytatja: "Az esküdtszéki büntetőeljárás virtuális megszüntetése olyan rendszert eredményezett, amelyben az igazságszolgáltatást bíróságok, bürokraták és ügyészek osztják ki, és a közösség hangjának kevés hely jut". *Id.* Hasonlóképpen, Stephanos Bibas megkérdőjelezte, hogy "a büntető igazságszolgáltatási rendszer hogyan távolodott el ennyire a néptől, amely névlegesen a felelős". BIBAS, *Supra* note 69, at xvi, xvii.

<sup>75</sup> *Lásd* Langbein, *supra* note at 66,122-23; *lásd még* John H. Langbein, *Understanding the Short History of Plea Bargaining*, 13 LAW & SOC'Y REV. 261 (1979) (e tézis továbbfejlesztése); John H. Langbein, *Torture and Plea Bargaining*, 46 U. CHI. L. REV. 3 (1978) (ugyanaz).

<sup>76</sup> *Lásd*: LAWRENCE M. FRIEDMAN & ROBERT V. PERCIVAL, AZ Igazságszolgáltatás gyökerei: A BŰNÖZÖSSÉG ÉS BŰNTETÉS A KALIFORNIAI ALAMEDA MEGYÉBEN, 1870-1910 (1981). A következő álláspont szerint Friedman és Percival,

Talán soha nem is volt a perek aranykora. Az igazán nagy ügyek kivételével az angol perek a tizenharmadik és tizenkilencedik században . . . [gyors.] Mindenesetre a rendőrség és a főállású ügyészek megjelenése megváltoztatta a büntető igazságszolgáltatás feltételeit; ahogy a "rendőrségi tudomány" is - az ujjlenyomatvétel, a érvizsgálat és így tovább. Egy olyan rendszerben, amelyet amatőrök, vagy az idejük és energiájuk kis darabkái ráfordító ügyvédek vezettek, a felderítés vagy a bizonyítás technológiája nélkül, a tárgyalás talán a legmegfelelőbb módja volt annak, hogy a bűnösöket kiszűrjék az ártatlanok közül. A tizenkilencedik

században azonban a büntető igazságszolgáltatás eltolódott

Más szóval, nem azt akarjuk sugallni, hogy az esküdtszéki perek visszaszorulása csak és kizárólag negatív, sem azt, hogy elméletileg a bírák nem tudnák biztosítani a kívánatos, visszafordítható etikát, csakúgy, mint a választott törvényhozók, akik a törvényeinket alkotják.<sup>77</sup> Az ilyen kérdésekben való mélyebb elmélyülést a jövőbeli munkára hagyjuk. Itt az állításunk szerényebb, de mégis alapvető: a büntető esküdtszék megőrzésének egyik oka az, hogy az erőteljesen szerepmegfordítható.

#### KÖVETKEZTETÉS

A több évtizedes fejlődés ellenére az "ideális értelemben vett" mesterséges intelligencia - amely széles körben intelligens és képes "gondolkodó gépekben" testesül meg - továbbra is elérhetetlen, és talán még évtizedekre van innen. Mindazonáltal a legtöbb szakértő meg van győződve arról, hogy ez a jövő közeleg, és ha ez bekövetkezik, a döntéshozatal hagyományos dinamikája meg fog nehezedni.

De a válságban lehetőség: a nagy teljesítményű mesterséges intelligencia hajnala új ablakot nyit demokratikus hagyományainkra, lehetővé téve számunkra, hogy jobban megkülönböztessük a megőrzésre érdemes hagyományokat, és feltehessük a kérdést, hogy melyek azok, amelyek ismertségük ellenére a gyakorlatban nem váltak be. Ebben a szellemben a demokratikus legitimitás egyik aspektusaként azonosítottuk a szerepvisszafordítható ítélkezést - a tanácskozás, a szavazás és más hagyományos mutatók mellett. A szerep-visszafordíthatóság helyzetbe hozza a döntéshozót, hogy értékelje a kegyelem vagy a sors által érkező szerencsét, lehetővé téve számára, hogy internalizálja az ítélkezés hatásait. Ez a testtartás javíthatja az eredményeket. De még ha nem is, akkor is központi szerepet tölt be a demokratikus önrendelkezés eszméiben, és ezért belső értékkel bír.

A "de a kegyelemért" orientáció így az emberiségről, mint olyanról, a megfordíthatóságra helyezi át a hangsúlyt. A lényeges kérdés nem az, hogy ki vagy milyen lényt bízunk meg az ítélethozatallal, hanem az, hogy ő (vagy az) kölcsönösen alá van-e vetve a fordított ítéletnek. A végkövetkeztetés nem az, hogy az intelligens robotok soha nem vehetnének részt a mi ítélkezési funkcióinkban. Hanem az, hogy ahhoz, hogy ezt megtehessék, maguknak is ítélőkészesnek kellene lenniük. Ami azt jelenti, hogy,

---

az amatőröktől és a rész munkaidősöktől a főállású bűnüldözők felé. Amint ez a változás bekövetkezett, többé nem lehetett feltételezni, hogy az esküdtszéki eljárás (az idealizált értelemben vett) büntetőügyek kezelésének normális módja. Végül is a rendőrség és az ügyészek már "bíróság elé állították" a vádlottat. Miért ne bízunk rájuk a dolgot?

*Id.* 194.

<sup>77</sup> A büntetőjogi vádlottak sokkal könnyebben megkérdőjelezhetik egy bíró "megfordíthatóságát", mint egy esküdttét: "Ja, persze, nőj fel te is úgy, ahogy én felnőttem, és

mevlátjuk, hogy bíróa tesznek-e!". Ismét a jövö munkájára hagyjuk, hogy az esküdszék szükségyszerűen nélkülözhetetlen-e egy demokratikusan legitim büntetőeljáráshoz, de kétségtelen, hogy a szerepe megfordíthatóbb.

valódi egyenrangúaknak kellene lenniük, nem pedig főuraknak vagy alárendelteknek - ahogyan bármelyik polgártársunknak is kellene.