

Egyenlő bánásmód az önvezető járműközlekedés szabályozásánál*

I. BEVEZETÉS

Bár az önvezető autókról széles körben feltételezik, hogy idővel sokkal biztonságosabbak lesznek, mint az emberi irányítás alatt állók, lehetnek olyan váratlan események, például hirtelen akadály megjelenése vagy szabálytalan gyalogos vagy kerékpáros viselkedés, amelyek veszélyt jelentenek az utasok számára. Az autonóm járművektől elvárnánk, hogy a baleseti helyzeteket úgy kezeljék, hogy a lehető legkisebb kár keletkezzon, különösen az emberi élet és fizikai épség szempontjából. Kiterjedt szakirodalom foglalkozik a kérdéssel, hogy szélsőséges esetekben mi vagy ki lenne inkább, illetve kevésbé feláldozható (ld. pl. az ezen cikkben szereplő szövegeket).

Az ilyen döntések, amennyiben személyek életére vonatkoznak, látszólag eleve akadályba ütköznek: az önvezető autók gyártásában élenjáró országok, például Németország és az Egyesült Államok alkotmánya *minden egyes* állampolgár élethez és méltósághoz való jogát védi (ld. Lin 2016. 70). Egyes személyek életének mások életével szemben való előnyben részesítése – bármilyen elveket is alkalmazzanak – megengedhetetlennek tűnik.¹

Ennek ellenére mind elméleti szakemberek, mind az autógyártók felkészülni látszanak ilyen döntésekre, amint azt bizonyos gyártók nyilatkozatai, az MIT jól ismert Moral Machine kísérlete, valamint számos alább hivatkozott szöveg tanúsítja. Gyakran felmerülő kérdés, hogy balesetek esetén kit kell felelősségre vonni, ha pl. az erkölcsi döntéseket a gyártó előre „beprogramozza”, vagy az autó használója egyénileg hozza meg (Marchant–Lindor 2012; Contissa –Lagioia – Sartor 2017).² Jelen tanulmány más megközelítést alkalmaz: bennünket az érde-

* A szöveg a *Culturally Sustainable Social Robotics* (szerk. Marco Norskov, Joanna Seibt, Oliver Santiago, IOS Press, 2020) kötetben megjelent fejezet magyar nyelvű és kissé átdolgozott változata.

¹ A szakirodalom jellemzően az életekre vonatkozó explicit mérlegelésre koncentrál, és ehelyütt ezt a megközelítést alkalmazzuk. Nem zárható ki azonban, hogy ez a módszer irrelevánssá válik, pl. a balesetekre vonatkozó gépi tanulás vagy az áldozatok véletlenszerű kiválasztása melletti döntés révén.

² A szakirodalom az önvezető autók működésére vonatkozó erkölcsi döntéseket vagy választási lehetőségeket általában explicit, emberi döntésekként fogja fel. Azzal a lehetőséggel,

kel, hogyan lehet olyan döntéseket hozni, amelyek a legjobban megőrzik a mai nyugati társadalmainkban egyik legfőbbnek tartott értéket, a különböző személyekkel szembeni egyenlő bánásmódot. Ugyanakkor az itt kifejtett koncepció a döntéshozatal átláthatósága és kiszámíthatósága (procedurális) értékeinek fenntartásához is hozzájárulna.

Kétségtelen, hogy a „nyugati társadalmak” kifejezés leegyszerűsítő és homályos, és bizonyos fokú heterogenitást és pluralitást takar az értékek vonatkozásában. Az is világosnak tűnik, hogy a mesterséges intelligenciát igénybe vevő eszközökkel szemben támasztott némely érték-elvárás (melyet pl. az Európai Bizottság Magas Szintű Szakértői Csoportjának etikai iránymutatásai (HLEG) rögzítenek), így például azok használatának magyarázható és átlátható volta, a különböző mesterséges intelligencián alapuló rendszerek esetében *nem* fognak kellőképpen érvényesülni; a kompromisszumok azonban sokak számára elfogadhatóak lesznek. Az itt tárgyalt érték azonban viszonylag vitathatatlan, és nem könnyű belátni, hogyan fogadható el kompromisszum annak vonatkozásában. Ha erkölcsi felfogásunknak vannak kétségségbevonhatatlan elemei, az erkölcsi-leg egyenlő bánásmód ezek közé tartozik.

Sajnálatos módon az önvezető járművek működése során bizonyos szélsőséges esetekben az egyenlő bánásmód nem jelentheti azt, hogy senkinek sem esik bántódása. Egy olyan balesetben, amelyben az egyik életet egy másik élet ellenében választják, ezeket az életeket nem kezelik egyformán. Az egyenlő bánásmód az ilyen helyzetekben azt jelentheti, hogy a *különböző, relevánsan hasonló tulajdonságokkal rendelkező személyek hasonló helyzetekben egyenlő védelemben részesülnek*. Ennek az elvnek az alkalmazásához a gyakorlatban, mint látni fogjuk, egységes szabályozásra van szükség. Az egyének védelme nem múlhat esetleges tényezőkön, beleértve az egyes gyártók vagy az autók használóinak döntéseit. (Ha maga a döntés elve a véletlenszerűség, akkor egyfajta esetlegesség mégiscsak szerepet kap. Ahhoz azonban, hogy az egyenlő bánásmód elve megvalósuljon, a véletlenszerűséget is egységesen kell bevezetni, és minden egyénnek egyenlő esélyt kell kapnia).

Az érvelés három lépésből áll. Először azonosítom a szóban forgó szélsőséges erkölcsi forgatókönyvek általános jellemzőit, és szembeállítom az önvezető autók működése során előkerülő erkölcsi választások tulajdonságait döntések egy másik, szintén emberi életek egymással szembeni mérlegelését igénylő típusával: vészhelyzetben hozott orvosi döntésekkel. Ezután amellet érvelek, hogy ezeket a döntéseket nem egyéni vagy gyártói, hanem központi szinten kell meghozni,³ és inkább egységes, mint pluralista módon szabályozni (miköz-

hogy a releváns döntésekhez gépi tanulási folyamatok révén jutnak, amelyekkel eltérő és előre nem látható döntések születhetnek, itt nem foglalkozunk. Az érvelés egy része azonban az ezekre vonatkozó egységes szabályozásra is alkalmazható lenne.

³ Hogy az igazgatás *mely* szintje legyen ez a központi szint, további kérdés, melyet itt nem tárgyalok.

ben nem tagadom a releváns kérdésekre vonatkozó ésszerű erkölcsi álláspontok pluralitását). Harmadszor, megindoklom, miért érdemes fenntartani az egyenlő bánásmódot mint értéket ebben az összefüggésben.

II. ERKÖLCSI DÖNTÉSEK EXTRÉM, RITKA, POTENCIÁLISAN ÉLETVESZTÉSSEL JÁRÓ ESETEKBEN

Bár valószínű, hogy az emberi járművezetők hibáinak kiküszöbölésével jelentősen csökken a balesetek száma, a kockázatok különböző személyek közötti elosztására vonatkozó előzetes döntések, amelyek gyakorlatilag preferált potenciális áldozatokká tennének bizonyos csoportokat⁴, új típusú erkölcsi helyzetet teremtenének, amely súlyos buktatókkal járhat. A mód, ahogyan ezeket a kérdéseket kezeljük, nemcsak a közlekedésre és a közúti biztonságra nézve járhat komoly következményekkel, hanem más technológiák bevezetésére, használatára és annak társadalmi konzekvenciáira is hatással lehet.

Az önvezető autók működése során potenciálisan felmerülő extrém erkölcsi helyzetek újdonsága a kockázatok (illetve súlyos károk) kihelyezésének sajátos módjában áll, amely – az itt tárgyalt megközelítésekben – a megszokottnál kevésbé lenne általános. Gondoljunk például az ipari légszennyezéssel vagy a nagysebességű vonatok üzemeltetésével járó kockázatokra. A légszennyezés emberek óriási tömegeseit veszélyezteti: évente mintegy 8 millió ember korai halála következik be annak hatásai miatt. A nagysebességű vonatok, amelyek nem rendelkeznek olyan berendezéssel, amely egy bizonyos sebességhatár túllépésekor automatikusan fékez, szintén jelentős kockázatnak teszik ki az utasokat (Yang–Xu–Cheng 2010). Személyek bizonyos csoportjai érzékenyebbek lehetnek ezekre az ártalmakra, amelyek tehát számukra nagyobb kockázatot jelentenek, mint mások számára. Idősek és nagyon fiatalok például inkább ki vannak téve a környezetszennyezéssel járó ártalmaknak. Nem arról van azonban szó, hogy légszennyező gyártási folyamatokat vagy nagysebességű vonatokat úgy terveznének meg, hogy azok bizonyos csoportok számára kockázatot vagy potenciális kárt jelentsenek. Nem választanak ki bizonyos tulajdonságokkal rendelkező csoportokat, melyeket azután nagyobb mértékben tesznek ki ártalomnak. Nem próbálnak egyes csoportokat mások kárára védeni, nem tekintik megengedhetőbbnek, hogy ők szenvedjenek kárt, mint mások; pusztán arról van szó, hogy bizonyos tulajdonságaik történetesen, nem-szándékolt módon sebezhetőbbé teszik őket ebben a tekintetben.

⁴ Patrick Lin szerint az önvezető járművek morális beállításai tulajdonképpen „targeting” (bizonyos tulajdonságokkal rendelkező embereket áldozatként „mege célzó”) algoritmusoknak tekinthetők (Lin 2016).

Úgy tűnhet azonban, hogy más esetekben is pontosan ez történik. Egy vílágjárvány során a korlátozott számú rendelkezésre álló lélegeztetőgépet például a rászorulóknak életkora és egészségi állapota alapján osztják el. Ezekben az esetekben is prioritást élveznek bizonyos csoportok, míg mások nem. Szem előtt kell azonban tartanunk a két helyzet típus közötti különbségeket. Először is, az önvezető járművek bevezetése – szemben egy járvánnyal vagy természeti katasztrófával – társadalmi döntés. Ha előre látható, hogy az új fejlesztés egyes egyénekre elviselhetetlen terhet ró, bevezetésétől el lehet tekinteni. Másodszor, egy váratlan, súlyos járványhelyzet nem teszi lehetővé, hogy előzetesen optimális döntéseket hozzunk. Az önvezető autók esetében ezzel szemben sokkal nagyobbak az elvárások a kritikus helyzetek előrejelzésével és erkölcsi megoldásával kapcsolatban. Míg az előbbi esetben az orvosoknak esetleg olyan döntéseket kell hozniuk, melyek normális körülmények között megengedhetetlenek, az önvezető járművek tervezésének elhúzódó folyamata elvileg elegendő időt és alkalmat biztosít arra, hogy olyan döntések szülessenek, melyek normális körülmények között is racionálisnak tekinthetők. A vészhelyzeti intézkedések nem indokolják hasonló intézkedések bevezetését és fenntartását vészhelyzet hiányában. Az önvezető autók esetében a döntéseknek elfogadhatóknak, vagy akár optimálisnak kell lenniük, hiszen bőven van idő a mérlegelésre, és bőséges erőforrások állnak rendelkezésre.

Az ilyen döntések egyik várható jellemzője tehát a prioritások meghatározásának specifikus volta – azaz olyan döntések születhetnek, melyek értelmében meghatározott tulajdonságokkal rendelkező, pl. bizonyos korcsoportokhoz tartozó vagy egészségi állapotú személyeknek több kockázat jut (Leben 2018; Smilansky 2019). Másodszor, a prioritások előzetesen meghatározhatóak. Emiatt a balesetek során hátrányosan kezelt személyek (képviselői) azt állíthatják majd, hogy a méltánytalan bánásmódot megfelelőbb tervezéssel ki lehetett volna küszöbölni. Harmadszor, az erkölcsi tervezés kérdése – jelen diszkusszió keretein belül – nem megkerülhető: kénytelenek vagyunk választani, mégpedig nagyon korlátozott számú lehetőség közül. Az az opció, hogy „folytassuk a mérlegelést”, vagy nem áll rendelkezésre, vagy annak választása önmagában egy bizonyos cselekvési irány mellett való döntést jelentene (Bhargava – Wan Kim 2017. 8). E választások utolsó jellemzője a megismételhetőség, melynek erkölcsi jelentősége a továbbiakban fog kibontakozni. A döntéshozatal alapelveinek ugyanis a potenciálisan releváns módon hasonló helyzetekre is érvényesnek kell lennie. Ezeket a tulajdonságokat figyelembe kell venni abból a szempontból, hogy mely megoldásokat tartjuk elfogadhatónak.

III. EGYÉNI, ILLETVE GYÁRTÓI DÖNTÉSEK ÉS AZ EGYENLŐ BÁNÁSMÓD ELVÉNEK MEGVALÓSÍTÁSA

Amikor a fent jellemzett döntések elfogadhatóságáról beszélünk, előzetes kérdés, hogy *ki* hozza meg ezeket a döntéseket. Egyik lehetőség, hogy az ilyen döntéseket az érintett egyénekre, azaz az önvezető autó tulajdonosára vagy utasára bízuk (Sandberg–Bradshaw–Martin 2013; Millar 2014 a és b). Ezt a lehetőséget a jól ismert „morális gomb” ötletével próbálták érzékletessé tenni, amely szerint a vezető a járművet egy gomb segítségével „egoista”, „altruista” vagy „semleges” üzemmódba állíthatná (Contissa–Lagioia–Sartor 2017). Ezen elképzelés arra vonatkozó javaslatokkal egészült ki, hogy hogyan lehetne ösztönözni a nem egoista beállításokat.

Mi történne, ha az autó utasai vagy tulajdonosa megválaszthatnák a beállításokat? Bonnefon és munkatársai (Bonnefon–Shariff–Rahwan 2016) azt találták, hogy az önvezető autó utasának szerepében az emberek hajlamosak lennének egoista döntéseket hozni; *más* úthasználóként azonban a „pártatlan”, azaz az autó saját utasait nem előtérbe helyező beállításokat preferálnák. Hogyan próbálnák az egyéni választás hívei megakadályozni, hogy mindenki egoista beállításokat alkalmazzon? Az önző döntésekre való hajlam problémájának megoldására pénzügyi eszközöket javasoltak (uo.), például speciális biztosítást lehetne kötni egoista beállítások használatára. Ezt azonban aligha tekinthetjük igazságosnak: az élet és a testi épség védelme nem kellene, hogy többet fizető személyek számára inkább elérhető legyen. Az anyagi kompenzáció nem lehet erkölcsileg eléggé releváns ahhoz, hogy igazolja valamelyik élet választását egy másikkal szemben. A szélsőséges esetekben potenciálisan figyelembe vehető tulajdonságok inkább a személyek életkilátásaihoz kapcsolódnak, melyek alapján például az életkorról és az egészségi állapotról úgy gondolhatjuk a lélegeztetőgépek elosztásánál, hogy ezek erkölcsi szempontból szerepet játszhatnak.

A „morális gomb” koncepciójával kapcsolatban az egoista beállítások diszincientivizálásának problémája mellett van egy másik, a mi szempontunkból még sürgetőbb kérdés. Az egyéni beállítások az adott személy védelmét attól a kontingens körülménytől tennék függővé, hogy „egoista” vagy „altruista” üzemmódba kapcsolt autóval találkozik-e. Így az, hogy egyes személyek védelmet élveznek, míg mások nem, pusztán szerencse kérdése lenne. Olyanok, akik releváns tulajdonságaikban nem különböznek szerencsésebb társaiktól (tehát ebben az esetben nem azért részesülnének kevesebb védelemben, mert pl. idősebbek vagy rosszabb egészségi állapotban vannak, hanem mert történetesen egoista üzemmódba állított autóval találkoztak) nagyobb kockázatnak és potenciális ártalomnak lennének kitéve.

Azok, akik nem támogatják az egyéni választás eszméjét, gyakran azt javasolják, hogy a szélsőséges esetekre vonatkozó döntéseket az autógyártóknak kellene meghozniuk. Sokak számára ez az alapértelmezett megoldás, és bizonyos

szempontokból valóban kézenfekvőnek látszik, hogy autógyártó vállalatok saját hatáskörben kezeljék a szóban forgó problémákat. Egyrészt a gyártók találkoznak közvetlenül azok tényleges, technikai megoldásának lehetőségeivel, végzik a műszaki tervezést és kivitelezést. Még ha az utas választaná is az üzemmódot, azt, ahogyan az „egoista”, „altruista” vagy „semleges” viselkedés a gyakorlatban megvalósul, nem tudná meghatározni vagy befolyásolni, hanem az a jármű programozásán múlna.

Milyen döntéseket hozhatnak a gyártók a morális beállítások tekintetében? A szabályozások jelenleg meglehetősen általánosak. Csak nagyon tág irányelveknek kell megfelelni, például lehetőség szerint csökkenteni kell a károkat és nem szabad feláldozni nem érintett személyeket (BVI 2017. 11). Az autógyárak számára érthető módon a saját utasok védelme lenne prioritás, amint azt a Mercedes-Benz vezetőjével, Christoph von Hugóval készített interjú (Taylor 2016) is tanúsítja. Ez az egyéni szintű egoista beállítások megfelelője lenne. Amint az interjúra adott reakció mutatta, a közvélemény nem tartja elfogadhatónak a közlekedők ilyen rangsorolását. Amennyiben a beállításokat vállalati szinten határozzák meg, a saját utasok védelmével kapcsolatos nyomás azonban valószínűleg mégis befolyásolni fogja ezeket a döntéseket.

Az utasok rangsorolásával szembeni megfontolások részben a kölcsönöségben alapulhatnak. Az önvezető autók utasai az etikai döntések meghozói és elszenvedői is lehetnek. Ennek a ténynek arra kellene ösztönöznie, hogy a védelmet ne egyenlőtlenül osszák el az utasok és a közlekedés más résztvevői között (Gogoll–Müller 2017). Egy adott gyártó által előállított autó utasa lehet, hogy örül a gyártó által biztosított aszimmetrikus védelemnek, mégsem szeretne egyenlőtlen feltételeket, amikor a gyártó egy másik autójával ütközik. Ez önmagában is ok arra, hogy magasabb szintű, egységes szabályozást szorgalmazzunk.

Egy másik érv is párhuzamos a fentiekben az egyéni beállítások tekintetében megfogalmazott megfontolással: a gyártó döntéseinek kontingens volta miatt az önvezető autók különböző, releváns tulajdonságaikban hasonló úthasználókkal szemben eltérő bánásmódot alkalmazhatnak. Jelenleg fennáll az a lehetőség, hogy mondjuk a Google autófejlesztői deontológiai, míg a Tesláé utilitarista erkölcsi intuíciókat használnak, bár valószínűbb, hogy mindkettő hibrid módon támaszkodik különböző erkölcsi felfogásokra. Az is könnyen előfordulhat, hogy absztrakt elveik konvergálnak, de a szélsőséges események tekintetében eltérően gondolkodnak. Az önvezető járművek etikájára vonatkozó diskurzus többnyire az elfogadható erkölcsi álláspontok pluralizmusával számol: nincsen legjobb választás, nincs az értékek mérlegelésének egyértelműen legjobb módja, és talán még egy legelfogadhatóbb erkölcsi elmélet sem áll a javasolt döntések hátterében (bár a legtöbb elméletalkotó inkább az utilitarista, mintsem például deontológiai vagy erényetikai szempontokra támaszkodik). Bár ezzel egyetérthetünk, kérdés, hogy ezt a helyzetet tükröznie kell-e a szabályozásnak oly módon, hogy a gyártók számára jogilag lehetővé váljon a különböző értékvá-

lasztások alapján történő döntés. Az alapvető, kiküszöbölhetetlen nézeteltérés ténye önmagában még nem feltétlenül indokolja a szabályozásra vonatkozó pluralizmust (Overeem–Verhoef 2014).

Hogy nincs egyetlen jó megoldás és hogy minden választáshoz szükségszerűen jelentős erkölcsi veszteség társul, nem jelenti azt, hogy minden megoldás egyformán elfogadható. Nem jelenti azt sem, hogy különböző megoldások szabadon alkalmazhatók különböző esetekre. Ami önmagában megengedhető, nem biztos, hogy más megoldásokkal együtt is az. Az összeegyeztethetetlen nézetek pluralitása nem indokolja, hogy önkényes, esetleges módon átültessük azokat a gyakorlatba.

Itt válik fontossá az ismétlődés, amelyet fentebb már említettünk, mint az olyan helyzetek potenciális jellemzőjét, melyekre vonatkozó döntésekről szó van. Olyan erkölcsi dilemmákkal van dolgunk, melyek kezelése súlyos kockázatot vagy kárt jelent egyes személyek számára. Az egyetlen ok, amiért az ilyen döntések megengedhetőnek tekinthetők, az, hogy a döntés elmulasztása egy ugyancsak megengedhetetlen döntés meghozatalával ér fel. Kényszerű választási helyzetben, amikor nem tagadhatjuk meg a döntést, úgy tűnhet, hogy egy ésszerű tartományon belül bármely döntés megengedhető, ha alternatívái szintén nagy erkölcsi áldozattal járnak.

Emlékezzünk azonban az ún. „villamos probléma” („trolley problem”) forgatókönyvére, amelyre a szakirodalomban a szélsőséges önvezető járműves esetek tárgyalása épül. Az elszabadult villamos esetében a választás a cselekvéstől való tartózkodás – mely esetben a sínen lévő öt embert elgázoljuk – és a kitérés között van, melyben egyetlen, másik sínen lévő embert áldozunk fel, az előttünk lévő sínen tartózkodó ötöt megmentve. Mindkét cselekvési mód (kitérés vagy az eredeti pályán maradás) csak a másik választáshoz képest megengedett.⁵ Ha a mellékvágányon nem lenne egy ember sem, a villamos vezetőjének egyértelmű kötelessége lenne kitérni; ha a kitérés nem mentené meg a többi ember életét, biztosan nem lenne megengedhető egy személy elgázolása. Ezek a normális körülmények között elfogadhatatlan választások csak azért megengedettek, mert nincs olyan választás, amely normális körülmények között elfogadható lenne. Ebből a szempontból hasonlít a fentebb említett egészségügyi vészhelyzetben történő választásra. A választás indoklásának mindkét esetben lényegi eleme, hogy nincsen „jó” választás. Ezek a helyzetek szöges ellentétben állnak azokkal, ahol választási lehetőségek pluralitását, amelyeket szintén összeegyeztethetetlen nézetek támasztanak alá, inkább egyenlő elfogadhatóságuk, mintsem egyenlő elfogadhatatlanságuk indokolja.

⁵ Nem veszem figyelembe azt a lehetséges álláspontot, hogy az utilitarista megoldás egyértelműen jobb, és mindkét cselekvést megengedhetőként, de egyiket sem kötelezőként kezelem.

Az önvezető járművek esetében – a villamos gondolatkísérlettől és az egészségügyi esettől eltérően – nem egyetlen eseményre vonatkozóan hozunk vészhelyzeti döntéseket, hanem iterálható, reflektáltan jóváhagyott megoldást kell találnunk. Képzeljük el, hogy az elszabadult villamos forgatókönyvét a következőképpen egészítjük ki: van egy rendszerszintű hiba, amely miatt a villamos probléma egy adott helyen ismétlődik. Minden mást változatlanul hagyunk: nincs kiút semmilyen technikai korrekcióval, a sofőr nem követhet el öngyilkosságot és nem hagyhatja el a villamost. Vajon az eset egymást követő előfordulásai során az egyes járművezetők továbbra is azt a döntést hozhatnák, amelyik jobbnak tűnik számukra, pl. az egyik utilitarista, a másik deontológiai alapon? Nyilvánvalónak látszik, hogy ilyen körülmények között ki kellene dolgozni valamiféle szabályokat, amelyet a járművezetők kötelesek betartani. Ha az egyik járművezető kitér, a következő pedig nem, a második baleset áldozatainak hozzátartozói okkal sérelmezhetnek ezt.

Ami egyetlen váratlan helyzetben megengedett, nem biztos, hogy a tervezés során, pontenciálisan ismétlődő helyzetekre hozott kényszerű döntések esetében is megengedhető. Itt az erkölcsösség, illetve méltányosság kritériumainak nem csupán egyetlen eseten belül, hanem eseteken átívelő módon is érvényesülniük kell. Az adott helyzetben adott módon kezelt személyekhez hasonló helyzetben lévő, releváns tulajdonságaikban hasonló személyeket (amennyiben nem pusztán személy voltukra vagyunk tekintettel) hasonló módon kell kezelni. A preskriptív egyenlőség eszményének kell érvényesülnie:

Ugyanabban a helyzetben lévő személyek jogosultak azonos bánásmódra pusztán azért, mert azonos helyzetben vannak. Másképpen fogalmazva, a preskriptív egyenlőség valójában az az elv, hogy a pusztán tény, hogy egy személyt egy bizonyos módon kezelték, önmagában indokolja, hogy egy másik, azonos helyzetben lévő személyt ugyanígy kezeljenek. (Peters 1997. 1223.)

Hogyan lehetne elkerülni azonban, hogy egy egységes szabályozás azt a benyomást keltse, hogy az egyetlen megengedhető megoldást alkalmazzák, és alternatív erkölcsi álláspontok mégiscsak elfogadhatatlanok? Nem úgy, hogy kiválasztunk egy bizonyos nézetet és azt a szabályozásban az alternatívák fölé helyezzük, hanem úgy, hogy hosszú, minden érintett fél (képviselőinek) bevonásával zajló diszkurzív folyamat során gyakorlati megoldásokhoz, kompromisszumokhoz jutunk. A cselekvési döntések amúgy sem feleltethetők meg az erkölcsi álláspontoknak közvetlen módon, mivel ez utóbbiak aluldeterminálják a konkrét helyzetekben meghozható tényleges döntéseket. Az absztrakt utilitarista vagy deontológiai álláspont általában nem fordítható le egyszerűen konkrét döntésekre. (Vegyük újra a villamos gondolatkísérletet: Mi lenne az „utilitarista” álláspont, ha választani kellene aközött, hogy 5 idős embert ütünk el a főpályán vagy 3 kisgyereket a mellékpályán? Vajon a „deontológiai” álláspont ugyanilyen egyértelmű maradna-e, ha a főpályán 100 személyt vagy a mellékvá-

gányon 1 személyt kellene elűtni?) Mivel egyik erkölcsi álláspont sem egyedül elfogadható, mindegyiknek jelentős erkölcsi ára van, és természetesen maga az erkölcsi nézőpont sem kizárólagos (minimálisan technikai és jogi megfontolások is szerepet játszanak), ezért ésszerű morális elveket is szem előtt tartó, meg-egyezős, pragmatikus megoldásokra törekedni. Ezek nem szüntetik meg az értékorientációk közötti konfliktusokat: az erkölcsi nézeteltérések nem oldódnak meg, hanem félreteszik azokat a gyakorlati, erkölcsi elemeket is magukban foglaló megoldások javára, amelyeket a felek el tudnak fogadni.

Az önvezető járművek működésének szabályozásában az értékkonfliktusok nem feltétlenül oldhatók fel. A mérlegelés szempontjainak és eredményeinek átláthatóság itt alapvető fontosságú. Ennek jelentőségét olyan példákön láthatjuk be, mint pl. a bizonyos tulajdonságokkal rendelkező autók védelme az utakon. Ha pl. a kisebb autók egységesen védelmet élveznek, autóvásárlásnál szem előtt lehet tartani ezt a szempontot, amennyiben ez az információ a potenciális vásárló számára elérhető.

Miközben az erkölcsi pluralizmus szabályozási pluralizmusra való átültetése nem segít az alapvető nehézségen, vagyis azon, hogy lehetetlen elkerülni az erkölcsileg kifogásolható és talán igazságtalan döntéseket, egy további, ugyanilyen súlyos nehézséget is magában hordoz: az egyenlő bánásmód hiányát. Az utolsó szakaszban amellet érvelek, hogy az egyenlőtlen vagy önkényes bánásmód kulturális és gyakorlati szempontból is komoly problémát jelentene.

IV. EGYENLŐ BÁNÁSMÓD KÜLÖNBÖZŐ AUTÓK UTASAIVAL SZEMBEN

Ha az egyenlő bánásmód elvét további absztrakt erkölcsi értékekre való hivatkozással akarjuk igazolni, a legkézenfekvőbb a méltányosság elve (ahogyan azt a villamos probléma iterált esetében már megfigyelhettük). Ha egy személy jogsult a károkozással szembeni védelemre, akkor egy másik, hasonló helyzetben lévő, hasonló tulajdonságokkal rendelkező személyt (amennyiben tulajdonságait egyáltalán számításba vesszük) ugyanúgy jogosultnak kell tekinteni erre.

Az egységesség ebben a tekintetben az önvédelem lehetőségeinek szempontjából is fontos. Egészségügyi és biztonsági kérdésekben az érintetteknek esélyt kell adni arra, hogy saját kockázataikat csökkentsék (ez a célja pl. a betegségmegelőzési programoknak). Az autonóm vezetéssel összefüggésben megemlíthetjük a bukósisak nélküli motorkerékpáros jól ismert példáját: Ütközés esetén egy járműnek az egyiket vagy a másikat kell inkább védenie? A sisakot viselőnek jobbak a túlélési esélyei, ami a másik védelmét indokolná; ugyanakkor nem tűnik ésszerűnek a közlekedési szabályokat be nem tartó magatartás jutalmazása és támogatása. Ha van egységes szabályozás a sisakosok védelmére, a motoros választhatja a kockázat csökkentését megfelelő magatartással. Ha egyének vagy gyártók különböző döntései alapján eltérő beállítások lehetségesek, ez megfosztja az utasokat attól a

lehetőségtől, hogy csökkentse saját kockázatukat. Másik (fentebb már említett) példa, hogy ha a kisebb méretű autók védettek (mivel veszélyeztetett úthasználók), ez szempont lehet az autó vásárlásakor. Különböző beállítások mellett a többi közlekedő nem lesz képes felmérni saját kockázatait közlekedési helyzetekben és ezekre megfelelően cselekedni. A közlekedők akkor tudnak megalapozott döntéseket hozni önmaguk védelmére, ha mások kiszámítható módon viselkednek. Ehhez egységes szabályozásra van szükség.⁶

Ha fenntartjuk, hogy minden egyénnek joga van a minimális egészségügyi ellátáshoz való hozzáférésre (például létfontosságú szervei működésének helyreállítására sürgős esetekben), akkor azt is fenn kell tartanunk, hogy joguk van az – egészségnél is alapvetőbb – élet egyetemes védelmére is. Ez a követelmény, mint már említettük, a nyugati társadalmakban alkotmányosan elismert. Amikor az étellel és a biztonsággal kapcsolatos zéróösszegű játszmaról van szó (valakinek a biztonságát, illetve életét szükségszerűen veszélyeztetni kell egy vagy több másik személy védelmében), akkor legalább olyan pártatlan és erkölcsileg releváns indokokra van szükség a döntések meghozatalánál, mint az egészség esetében. Nem lenne szabad pl. pénzügyi megfontolások alapján (melyre példa fentebb a 'biztosítás' kiváltása volt) mentesíteni személyeket.

Nem arról van szó, hogy minden úthasználót ugyanolyan mértékben kellene védeni vagy mentesíteni a kockázat alól ütközés esetén – a szóban forgó szélsőséges forgatókönyvek éppen arra a feltételezésre épülnek, hogy ez lehetetlen. Állításunk az, hogy *különböző úthasználókat azonos alapon kell védelemben részesíteni, és a különbségek csak akkor elfogadhatóak, ha erkölcsileg jelentősek.*

A méltányosság követelménye mellett kulturális és pragmatikus érvek is szólnak amellett, hogy az önvezető autók működése során az egyenlő bánásmód elve érvényesüljön. Ahogy Erwin Chemerinsky alkotmányjogász írja, „egyetlen érték sem gyökerezik olyan mélyen a nyugati kultúrában, mint az egyenlőség fogalma”, amely „a nyugati kultúra főáramában ma már minimális normaként elfogadott” (Chemerinsky 1983). A személyek erkölcsi egyenlőségének gondolata alátámasztja a hasonló helyzetben lévő személyekkel szembeni egyenlő bánásmód követelményét. Ahogy Kent Greenawalt megjegyzi:

[...] legalábbis a mi kultúránkban, és valószínűleg a nyugati kultúrában általában, meglehetősen mély és erős az érzület, melynek értelmében nem részesülhetünk rosszabb bánásmódban, mint más, egyenrangúnak vélt személyek [...]. Ha ez a feltételezés helytálló, akkor ésszerűen elvárható, hogy a jó döntéshozók bizonyos mértékig figyelembe vegyék ezt az érzületet. (Greenawalt 2016. 215.)

⁶ Hangsúlyozni kell, hogy a kockázat csökkentése itt nem azt jelenti, hogy a saját kockázatomat baleset esetén másokra hárítom, hanem azt, hogy kihasználom a *bárki* számára elérhető lehetőséget, hogy a nyilvános információk felhasználásával óvintézkedéseket tegyek az ellen, hogy baleset áldozatává váljak.

Az egyenlő bánásmód értékének támogatottsága nem csak absztrakt jelentőséggel bír. Az önvezető járművek működésének morális szabályozását a politikai közösség erkölcsi attitűdjeihez kívánják igazítani, amint azt olyan felmérések is jelzik, mint a Moral Machine kísérlet és a Pew Research Center felmérése (Smith–Anderson 2017), amelyekkel a közvélemény számára elfogadható erkölcsi döntéshozatali elveket szándékozták meghatározni. Ahhoz, hogy a közvélemény elfogadja és használja az önvezető autókat, bizalomra van szükség:

Úgy véljük, a gyors technológiai változások közepette a bizalom továbbra is a társadalmak, közösségek, gazdaságok és a fenntartható fejlődés alapköve marad. Ezért alapvetően fontos, hogy a mesterséges intelligencia megbízható legyen, mivel emberek és közösségek csak akkor tudnak majd bízni a technológia fejlődésében és alkalmazásaiban, ha világos és átfogó keretrendszer áll rendelkezésre a megbízhatóság eléréséhez. (HLEG 2019. 4.)

Az elmúlt évtized technológiai fejlődésével, és különösen a mesterséges intelligencia széles körű alkalmazásával, amely várhatóan alapvető módon fogja megváltoztatni életünket, munkánkat és társadalmi rutinjainkat, és különböző területeken döntéseket hoz helyettünk, sokan érzik úgy, hogy rosszul járhatnak az automatizálás elterjedésével. Felmérések szerint „több az aggodalom, mint a lelkesedés” (Smith–Anderson 2017. 2) a várható fejleményekkel kapcsolatban. Például „az amerikaiak 76%-a arra számít, hogy a gazdasági egyenlőtlenség sokkal nagyobb lesz, ha a robotok és a számítógépek képesek lesznek sok olyan munkát elvégezni, melyet jelenleg emberek végeznek” (uo. 3). Annak biztosítása, hogy az olyan tényezők, mint az anyagi helyzet, ne legyenek tényezők az utakon való biztonságos közlekedés szempontjából, hogy a szabályok hasonló bánásmódot garantáljanak a relevánsan hasonló tulajdonságokkal rendelkező egyéneknek, akik egy átlátható és egységes szabályozásra támaszkodva járulhatnak hozzá saját biztonságukhoz, feltehetően növelné a bizalmat és az elfogadottságot.

IRODALOM

- Bhargava, Vikram – Wan, Kim Tae 2017. Autonomous vehicles and moral uncertainty. In Patrick Lin, Ryan Jenkins, Keith Abney (szerk.) *Robot Ethics 2.0. From Autonomous cars to Artificial Intelligence*. Oxford, Oxford University Press. 5–19.
- Bonnefon, Jean – Francois – Shariff, Azim – Rahwan, Iyad 2016. The Social Dilemma of Autonomous Vehicles. *Science*. 352/6293. 1573–1576.
- [Bundesministerium für Verkehr und digitale Infrastruktur] 2017. *Ethics Commission on Automated and Connected Driving Report*.
https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile

- Chemerinsky, Erwin 1983. In Defense of Equality: A Reply to Professor Westen. *Michigan Law Review*. 81/3. 575–599.
- Contissa Giuseppe – Lagioia, Francesca – Sartor, Galileo 2017. The ethical knob: ethically-customisable automated vehicles and the law. *Artificial Intelligence and Law*. 25/3. 365–378.
- [European Commission Higher-Level Group, HLEG] 2019. *Ethics guidelines for trustworthy AI*. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>
- Gogoll, Jan – Müller, Julian F. 2017. Autonomous Cars: In Favor of a Mandatory Ethics Setting. *Science and Engineering Ethics*. 23/3. 681–700.
- Greenawalt, Kent 1997. Prescriptive Equality: Two Steps Forward. *Harvard Law Review*. 110/6. 1265–1290.
- Leben, Derek 2018. *Ethics for Robots: How to Design a Moral Algorithm*. Abingdon/NY, Routledge.
- Lin, Patrick 2014. Here's a Terrible Idea: Robot Cars with Adjustable Ethics Settings. *Wired*. 18 Aug. <http://www.wired.com/2014/08/heres-a-terrible-idea-robot-cars-with-adjustable-ethics-settings/>.
- Lin, Patrick 2016. Why Ethics Matters for Autonomous Cars. In M. Maurer et al. (szerk.): *Autonomous Driving: Technical, Legal and Social Aspects*. Cham, Springer. 69–85.
- Marchant, Gary – Lindor, Rachel A. 2012. The Coming Collision between Autonomous Vehicles and the Liability System. *Santa Clara Law Review*. 52/4. 1321–1340.
- Millar, Jason 2014a. An Ethical Dilemma: When Robot Cars must Kill, Who should Pick the Victim? *Robohub.org*. <http://robohub.org/an-ethical-dilemma-when-robot-cars-must-kill-who-should-pick-the-victim/>
- Millar, Jason 2014b. You Should Have a Say in Your Robot Car's Code of Ethics. *Wired*. 2. Sept. <http://www.wired.com/2014/09/set-the-ethics-robot-car>.
- Overeem, Patrick – Verhoef, Jelle 2014. Moral Dilemmas, Theoretical Confusion: Value Pluralism and its Supposed Implications for Public Administration. *Administration and Society*. 46/8. 986–1009.
- Peters, Christopher J. 1997. Equality Revisited. *Harvard Law Review*. 110 /6. 1211–1264.
- Sandberg, Anders – Bradshaw-Martin, H. 2013. What Do Cars Think of Trolley Problems: Ethics for Autonomous Cars? In Jan Romportl et al. (szerk.) *Beyond AI: Artificial Golem Intelligence*, Conference Proceedings. https://www.beyondai.zcu.cz/files/BAI2013_proceedings.pdf:12
- Smilansky, Saul 2019. A Hostage Situation. *Journal of Philosophy*. 116/8. 447–466.
- Smith, Aaron – Monica Anderson 2017. Automation in Everyday Life. Pew Research Center, https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2017/10/PI_2017.10.04_Automation_FINAL.pdf
- Taylor, Michael 2016. Self-Driving Mercedes-Benzes Will Prioritize Occupant Safety over Pedestrians. *Car and Driver* 7. Oct. <https://www.caranddriver.com/news/a15344706/self-driving-mercedes-will-prioritize-occupant-safety-over-pedestrians/>
- Yang Xianhui – Xu, Ming – Cheng, Yao 2010. Risk Analysis for the Modification in Automatic Train Control Systems. In *Proceedings of the 2010 International Symposium on Computer Communication Control and Automation (3CA)*. Tainan, IEEE. 100–103.