

ETIČNE IN POLITIČNE DILEME PROGRAMIRANJA SAMOVOZNIH AVTOMOBILOV ZA ODLOČANJE O NEIZOGIBNI ŠKODI

FRIDERIK KLAMPFER

Univerza v Mariboru, Filozofska fakulteta, Maribor, Slovenija
friderik.klampfer@um.si

Sinopsis Od avtonomnih vozil (AV) si lahko obetamo učinkovitejši promet, zmanjšanje onesnaženosti in odpravo večine prometnih nesreč. Nesrečam se v prihodnosti seveda ne bo dalo povsem izogniti. Vsaj občasno se bodo AV znašla v situacijah, o katerih so v zadnjih štirih desetletjih žolčno razpravljali moralni filozofi: Naj zapeljejo v gručo pešcev ali zavijejo na pločnik in tam poveljijo eno osebo? Ali naj žrtvujejo lastne potnike, da bi obvarovala več drugih udeležencev v prometu? Naj se zdijo takšni scenariji danes še tako neverjetni, z množično rabo AV-jev bodo postali dovolj pogosti, da zahtevajo jasne moralne napotke. Izbiro pravil za odločanje o porazdelitvi neizogibne škode v trkih AV otežujejo širši družbeni in ekonomski vidiki: zakonodajalci bodo morali najti razumen kompromis, ki bo (i) etično opravičljiv, (ii) užival zadostno javno podporo in (iii) ne bo negativno vplival na prodajo AV. Nedavne empirične raziskave kažejo, da bi znala biti to zahtevna naloga – vsak od nas bi hotel, da se drugi vozijo v 'utilitarističnih' vozilih (ki zmanjšujejo skupno škodo), sami pa bi raje uporabljali ne-utilitaristična vozila (ki postavijo potnike na prvo mesto). V članku pokažemo, da je ta specifična različica problema koordinacije načeloma rešljiva, če smo pripravljeni omejiti svobodno potrošniško izbiro.

Ključne besede:
avtonomna vozila,
moralni algoritem,
prometna varnost,
prometna
pravičnost,
potrošniška
svoboda, državni
paternalizem,
dregljanje

ETHICAL AND POLITICAL DILEMMAS OF PROGRAMMING AUTONOMOUS VEHICLES FOR DECISIONS IN CASES OF UNAVOIDABLE HARM

FRIDERIK KLAMPFER

University of Maribor, Faculty of Arts, Maribor, Slovenia
friderik.klampfer@um.si

Abstract Autonomous vehicles (AVs) promise to increase traffic efficiency, reduce pollution, and avoid most traffic accidents. Occasionally, AVs will face the kind of choices that were made prominent in the past four decades by moral philosophers: whether to run over several pedestrians or run down a single bystander, or whether to sacrifice their own passenger(s) to spare a young family in the oncoming car. As unlikely as such scenarios may appear today, they will become sufficiently common with millions of AVs to require clear moral guidance. The choice of decision-rules for distributing unavoidable harm in future AV accidents is further complicated by wider social and economic aspects: lawmakers will need to find a reasonable compromise that will (i) be morally defensible, (ii) generate sufficient public support, and (iii) will not adversely affect the sales of AVs. Recent empirical findings suggest that this may be a rather demanding task –people apparently want to see others driving utilitarian, i.e., harm-minimizing vehicles, but would themselves prefer to use non-utilitarian, i.e., passenger-prioritizing vehicles. In the paper, we show that this particular version of Coordination problem is resolvable, if we are willing to circumvent free consumerist choice.

Keywords:

autonomous
vehicles,
moral algorithm,
traffic safety and
justice,
free consumerist
choice,
state paternalism,
nudging

1 Uvod

Obdobje avtonomnih vozil (AV) s seboj ne prinaša le zmanjšanja potrebe po človeški delovni sili v prometu in s tem povezane izgube zaposlitve za milijone poklicnih voznikov tovornjakov in taksijev, z njihovo uporabo se nam obeta občutno izboljšanje prometne učinkovitosti, zmanjšanje porabe goriva, zmanjšanje onesnaževanja, veliki prihranki in preprečitev, če že ne vseh, pa vsaj večine prometnih nesreč.¹ Slednje z obličja Zemlje seveda ne bodo izginile čez noč. Vsaj občasno bodo AV prisiljeni sprejemati odločitve v situacijah, s kakršnimi so se v zadnjih štirih desetletjih v svojih naslanjačih obsedeno ukvarjali moralni filozofi: Ali naj avto povozi več pešcev na prehodu za pešce ali namesto tega zavije na pločnik in tam ubije enega? Čigava življenja naj ohrani, potnikov v vozilu ali drugih udeležencev v prometu? Tudi če so taki scenariji malo verjetni, se bodo ob vsakodnevnih množični uporabi AV-jev pripetili dovolj pogosto, da zahtevajo jasne moralne napotke. V algoritme, ki upravljajo z AV, bo zato nujno vgraditi moralna načela, ki bodo uravnavala porazdelitev neizogibne škode v teh in podobnih situacijah.

Kar bi utegnilo izbiro pravil za porazdeljevanje neizogibne škode v primeru trčenja avtonomnih vozil otežiti in kar tovrsten premislek razlikuje od zloglasnega miselnega eksperimenta z imenom 'Tramvaj', je njegova širša družbena in ekonomska plat. Zakonodajalci bodo morali najti razumen kompromis, ki bo (i) etično opravičljiv, (ii) užival zadostno podporo v strokovni in splošni javnosti in (iii) kupcev ne bo odvrnil od nakupa tovrstnih vozil. Nedavne empirične raziskave dajejo slutiti, da bi znal biti to kar trd oreh - vsak od nas bi namreč raje videl, da drugi kupujejo in uporabljajo 'utilitaristična' vozila, ki v primeru trčenja kar najbolj zmanjšajo skupno škodo, medtem ko bi hoteli zase pridržati pravico do nakupa in uporabe ne-utilitarističnih vozil, takih torej, ki bodo prednostno zaščitila svoje potnike. V prispevku dokazujem, da je opisana različica problema kolektivnega delovanja v načelu rešljiva, če smo le kot skupnost pripravljeni žrtvovati del toliko opevane, a zlasti v ekonomski znanosti vulgarno razumljene neomejene potrošniške svobode. Natančneje, v prispevku bom iskal odgovore na dve sorodni, a vseeno logično neodvisni vprašanji:

¹ Po mnenju avtorjev poročila skupine A.T. Kearney se bo število prometnih nesreč zmanjšalo za 70 odstotkov. Čeprav se v pričujočem prispevku osredotočam na izboljšanje prometne varnosti in pravičnosti, to ne pomeni, da bi smeli zanemariti druge splošno družbene in individualne koristi od množične proizvodnje in uporabe samovožnih avtomobilov. Celovita moralna in politična presoja bo seveda vključevala tudi okoljske, ekonomske, psihološke, kulturne in socialne vidike, ki jih tu zavestno postavljam ob stran.

- a) moralno: *Z vidika individualnega potrošnika, kateri algoritem bi bilo moralno pravilno vstaviti v svoj AV? in*
- b) politično: *Z vidika politične skupnosti, občega interesa, javnega dobra, družbene pravičnosti itd., proizvodnjo, distribucijo in nakupovanje katerih vrst vozil bi morala vlada predpisati ali vsaj spodbujati in s kakšnimi sredstvi si sme pomagati pri uresničevanju omenjenih ciljev?*

V odgovoru na prvo vprašanje bom dokazoval, da obstajajo prepričljivi razlogi za proizvodnjo oz. nakup utilitarističnih avtonomnih vozil (UAV). (Vendar z določenim pridržkom.) V zvezi z drugim pa, da če niti zakonska prisila niti popolnoma svobodna izbira potrošnikov nista zares moralno in/ali politično privlačni možnosti, bodo vlade morda prisiljene potrošnikom izbiro omejiti na način, ki ne bo sočasno posegel v njihovo avtonomijo, recimo tako, da jih bodo dregnile k nakupu samovoznih avtomobilov.

2 Etika programiranja samovoznih avtomobilov

Vsaka filozofska raziskava temelji na določenih predpostavkah. Kaj vse bom torej sam predpostavil brez dokazovanja? Predvideval bom, da je izbira AV očitno moralno boljša od izbire ne-AV, tudi če voznikom upravljanje slednjih olajša množica sistemov za pomoč in varnejšo vožnjo. Izognil se bom tudi vprašanjem, ali je treba AV zaradi njihove inteligence, odločitvene in delovajske avtonomije ali česa tretjega priznati poseben moralni status, in jih za potrebe ovrednotenja argumentov obravnaval kot stroje brez izvornega oz. prvinskega – tj. nededovanega, neizpeljanega – moralnega statusa, s tem pa se izognil labirintom, v kakršne bi zašli, če bi dopustili konflikte med pravicami in interesi ljudi na eni in strojev na drugi strani. Prav tako se bom iskanja pravilnega moralnega algoritma za AV-je lotil z vidika najboljše človeške in ne morda strojne oz. robotske morale. Čisto mogoče je, da bi morali namesto namestitve moralnega algoritma, ki se zdi po tehtnem premisleku pravilen nam, tj. človeškim bitjem z za ljudi značilnimi pristranskostmi do sebe, potomcev in bližnjikov, ali ki ga kot edino pravilno ravnanje odobrava ta ali ona kultura, nalogo izoblikovanja pravilnih, razumsko utemeljenih načel za porazdelitev neizogibne škode (ali več teh) prepustiti inteligentnim strojem. In končno, pustil bom ob strani (čeprav se bom v resnici kasneje vrnil k njemu) nemara filozofsko najbolj zagatno vprašanje: Kaj se zgodi z moralno odgovornostjo za

tveganja in nastalo škodo, ko moralno presojo in odločanje v prometu enkrat v celoti prepustimo strojem? Nas bo to razbremenilo sleherne moralne in/ali pravne (kazenske in odškodninske) odgovornosti ali pa bo, nasprotno, zahtevalo temeljit premislek o pogojih, v katerih smo odgovornost človeškim akterjem tradicionalno pripisovali, in njihovo posledično revizijo?

2.1 (Neobvezujoča subjektivna) mnenja in (obvezujoče objektivne) norme

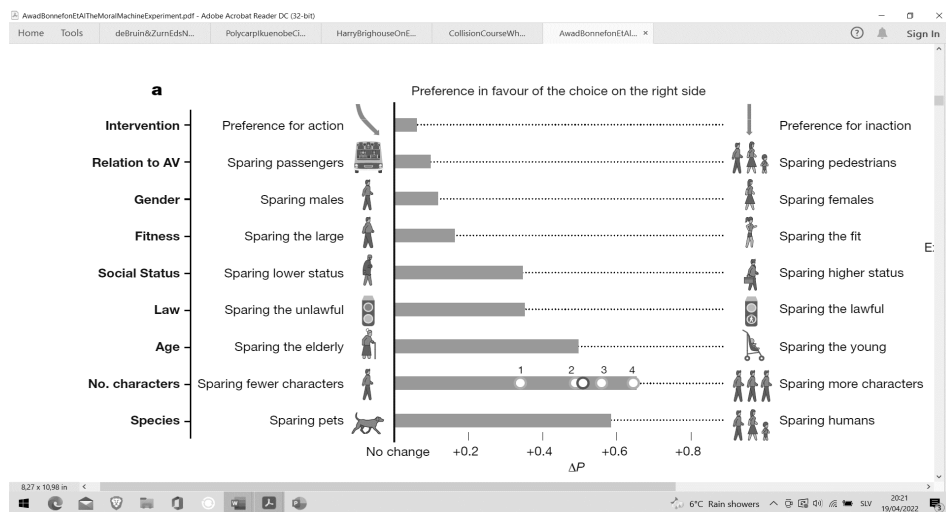
Premišljevanje in odločanje o načelih/navodilih za razreševanje moralnih dilem, ki bi jih veljalo vstaviti v samovozne avtomobile, bi lahko prepustili (a) inženirjem (ki so se s tem problemom spoprijeli prvi), (b) strokovnjakom za etiko oz. poklicnim etikom, (c) slehernikom oz. laični javnosti (kot v projektu *The Moral Machine*) ali pa (d) računalnikom oz. strojem. Naj na kratko predstavim razloge, zakaj je ta naloga najbolj pisana na kožo strokovnjakom za etiko, tj. poznavalcem splošnih etičnih teorij in njihovih praktičnih implikacij.

Začnimo s predlogom, da bi bilo iskanje najboljšega moralnega pravila za porazdelitev neizogibne škode v prometnih nesrečah najbolje pustiti samoučečim se strojem, ki so navsezadnje bolj imuni na pristranost do sebe in predsodke do drugih kot mi, ljudje. Glavno težavo tega predloga vidim v tem, ali ne bomo mogli vedeti, kdaj in če sploh so ti našli pravilne odgovore in rešitve, ali pa njihovih odgovorov in rešitev ne bomo hoteli sprejeti kot pravih, če se bodo slabo ujemali z našimi trdnimi intuicijami. (Rini 2016) Ahmad in kolegi so lepo izrazili ta pomislek, ko so zapisali: »Še nikoli doslej v zgodovini človeštva nismo dovolili, da bi stroji samostojno odločali, kdo naj živi in kdo umre, in to v delčku sekunde in brez nadzora v realnem času«. (Ahmad et al. 2020) Zakaj pa ne bi te zahtevne in nehvaležne naloge zaupali inženirjem, ki so navsezadnje problem tudi prvi prepoznali in se ga lotili? Iz enakega razloga, se bojim, zaradi katerega snovanja kodeksov zdravniške etike ni pametno prepustiti izključno zdravnikom ali sestavljanja kodeksov novinarske etike izključno novinarjem – ker podobno kot prej omenjena poklicna profila tudi inženirji niso posebej usposobljeni za iskanje odgovorov na zagatna moralna vprašanja oz. razreševanje moralnih dilem in ugank. Kot mezdni delavci so za povrh podrejeni vodstvom avtomobilskih koncernov, ki jih pri snovanju, proizvodnji in prodaji avtomobilov vodi prej prizadevanje za čim večji dobiček kot pa skrb za obče dobro ali sledenje moralnemu čutu/svoji vesti.

Zadnji predlog, da bi izbiro po moralni plati optimalnega algoritma prepustili javnemu mnenju oz. jo uskladili z ugotovljenimi večinskimi preferencami ljudi, nas postavi pred zanimivo metodološko dilemo. Snovalci spletišča Moral Machine, kjer so zbrali in obdelali podatke o sodbah skoraj 40 milijonov ljudi iz 233 držav po svetu, o tem, kako naj samovozni avtomobili razrešujejo moralne dileme v primeru neizogibnih prometnih nesreč,² recimo, predpostavljajo, da so podatki o pričakovanjih javnosti in potencialnih kupcev takih vozil za snovalce prometne politike dragoceni in koristni, tudi če bi se uspelo poklicnim etikom čudežno dogovoriti, katera od možnih načel za porazdeljevanje škode v takih situacijah so veljavna in katera ne:

Odločitve o etičnih načelih, ki bodo uravnavala samovozna vozila, ne moremo prepustiti le inženirjem ali etikom. Da bi potrošniki prešli s tradicionalnih avtomobilov, ki jih upravlja človek, na avtonomna vozila, in da bi se širša javnost sprijaznila s cestno prevlado vozil z umetno inteligenco, bosta morali obe skupini razumeti, od kod izvirajo v ta vozila programirana etična načela. Drugače rečeno, tudi če bi se etiki dogovorili, kako naj avtonomna vozila razrešujejo moralne dileme, bo njihov trud zaman, če državljeni s predlagano rešitvijo ne bomo soglašali in se bodo zato spremembe, ki jih avtonomna vozila obljublajo namesto statusa quo, odmaknile v prihodnost. Vsak poskus izoblikovanja etike umetne inteligence je treba uskladiti z javno moralo. Potemtakem je nujno, da spoznamo družbena pričakovanja o tem, kako naj bi avtonomna vozila razreševala moralne dileme. (Awad et al. 2020)

² Če zanemarimo drobne razlike med tremi geografskimi sklopi ali pa se osredotočimo na ti. 'zahodnjaški' sklop prioritete, bi morali, če bi želeli zakonodajno, davčno, zavarovalno ipd. politiko ukrojititi po javnem mnenju, AV-je programirati tako, da bi ti dajali prednost večjemu pred manjšim številom in mlajšim pred starejšimi (močno izraženi preferenci), ženskam pred moškimi, bolj pred manj zdravimi ter direktorjem pred brezdomci (šibkeje izražene preference).



Slika 1: Globalne preference

Vir: zajem zaslona, prirejeno po (Awad, E., Dsouza, S, Kim, R. idr. 2018), 2022

Načelnega odgovora na vprašanje, na katere od zgornjih ugotovitev raziskave naj se opre prometna politika, pa bralci žal ne dočakamo. Individualne razlike (po demografskih kazalnikih, kot so starost, spol, izobrazba, dohodek, politična in verska stališča) so neznatne, in čeprav znajo biti teoretsko pomembne, jih naj politikom ne bi bilo treba upoštevati. Po drugi strani bi po njihovem pri izbiri algoritmov, s katerimi bodo (brez izjeme?) opremljeni (vsi?) samovožni avtomobili, državni uradniki morali, recimo, upoštevati, da je širša javnost naklonjena ideji o prednostni zaščiti otrok pred odraslimi, žensk pred moškimi. A čemu pravzaprav? In kaj naj politika počne s podatkom o kulturno specifičnih sklopih tovrstnih pričakovanj? Jih je dolžna vgraditi v algoritme in tako posredno legitimirati? Čemu le?³ K tem zagatnim vprašanjem se v nadaljevanju še vrnemo.

³ Za pritrdilen odgovor na to vprašanje glej Polzer 2021. Sam se bom razpravi o tem, ali je morala univerzalna ali relativna oz. kulturno specifična, zaradi prostorskih omejitev raje izognil. V pričujočem besedilu zagovarjam vgradnjo univerzalnega moralnega načela, se pa dobro zavedam, da izvira univerzalna privlačnost, če jo sploh premore, iz njene skrajno abstraktne formulacije. Kulturno oz. civilizacijsko pogojene razlike, ki so jih v sodbah o hipotetičnih prometnih scenarijih zaznali Awad in njegovi sodelavci, se bodo pokazale šele na nižji, izvedbeni ravni, ko bo algoritem med ukrepanjem in neukrepanjem, med več in manj žrtvami, med mlajšo in starejšo žrtvijo, prisiljen izbirati v situacijah, kjer je odločitev že na prvi pogled tesna. Ali naj proizvajalci samovožnih avtomobilov te drobne razlike v ovrednotenju izidov upoštevajo in v kolikšni meri, bo odvisno od dvojega – koliko bolj privlačen bo zaradi tega njihov izdelek v očeh potrošnikov in kako razumno utemeljena so posamezna odstopanja.

2.2 Tramvaj ali ne?

Scenariji, ki jih preigrava Moralni stroj, vsaj na prvi pogled ustrezajo tako zgradbi kot tudi izvornemu namenu miselnega eksperimenta, ki se ga je sprva med filozofi, kasneje pa še med psihologi, ki proučujejo dejavnike moralne presoje, oprijelo ime Tramvaj – tako ali drugačno deterministično dogajanje, ki ga ni sprožila namerna človeška dejavnost (gibanje podivjanega vlaka, zemeljski ali snežni plaz, epidemija nalezljive bolezni) ogrozi življenja večjega števila (običajno petih) ljudi in če ne bomo mi, ki za to grožnjo sicer osebno nismo odgovorni, posredovali in jo preusmerili na manjše število ljudi (običajno enega), ki jih bomo žrtvovali namesto njih, bo teh pet ljudi umrlo. Ta filozofski 'minižanr', skratka, preizkuša naše intuicije glede omejitev, ki jih morala nalaga prizadevanjem, da bi proizvedli čim več dobrega oz. preprečili čim več slabega. Natančneje, rešili življenja oz. preprečili smrti čim več ljudi. Za omenjeno moralno dilemo je ključno, da je preživetje enako v interesu vseh vpletenih, da ni nihče od njih izgubil svoje pravice do življenja oz. moralnega varstva pred oškodovanjem in da posledično vsi uživajo polnovreden in enak moralni status – z drugimi besedami, njihove zahteve oz. pričakovanja, da jim s svojim ravnanjem ne bomo škodili, so enako legitimne, četudi je morda res, da bodo imeli nekateri med njimi od preživetja več koristi kot drugi ali da so eni od njih moralno boljši ljudje kot drugi. (Kauppinen 2019: 51)

V zgornjem opisu se že kaže nekaj razpok.⁴ Namišljenih prometnih situacij, ki jih, recimo, preigrava spletna aplikacija Moralni stroj, se že po površnem ogledu ne da brez preostanka preslikati na sto in eno različico Tramvaja. Osnovno podobnost med njima bi bilo seveda neumno zanikati – z obema miselnima pripomočkoma ugotavljamo, ali in kdaj smemo ubiti, žrtvovati življenje enega človeka, da bi jih rešili (bolje, se izognili uboju) pet(ih). Vendar pa se prometne situacije iz aplikacije Moralni stroj od najbolj popularnih različic Tramvaja, kakršne sta si zamislili Philippa Foot in Judith Thomson, bistveno razlikujejo v naslednji ključni podrobnosti: ne stojimo na varnem ob kretnici ali robu cestišča in premlevamo, ali naj podivjani tramvaj

⁴ Zakaj je sploh pomembno, ali gre pri prometnih scenarijih, ki jih preigrava spletna aplikacija Moralni stroj, za različice Tramvaja ali ne? Po eni strani zato, ker če so, potem so filozofi morda že zadovoljivo odgovorili na vprašanja, ki bi se utegnili v vsakodnevnih prometnih situacijah zastavljati avtonomnim vozilom. Po drugi pa zato, ker bi jih znalo to spoznavno razvrednotiti – če gre namreč verjeti standardnim ugovorom proti taki metodologiji filozofskega raziskovanja, imajo spontane sodbe, ki jih v odgovor na take scenarije oblikujemo bodisi poklicni filozofi bodisi filozofsko neuki ljudje, bolj skromno, če sploh kakšno, spoznavno vrednost. Za zadržke do uporabe Tramvaja za modeliranje prometnih situacij, v kakršnih se bodo morali znajti v AV-jih vgrajeni algoritmi, glej Furey in Hill 2020.

preusmerimo na tir, kjer bo namesto petih povozil eno osebo, temveč sedimo v tramvaju in prelevamo, ali naj podivjani tramvaj preusmerimo na tir, kjer bo trčil v zid in bomo umrli. Še vedno je to intuitivno vodeno miselno eksperimentiranje s ciljem, da najdemo taka splošna moralna pravila ali načela, ki bodo naše posamične intuitivne sodbe med seboj uskladila oz. poenotila (in posredno zmanjšala njihovo poljubnost),⁵ le da postane sicer abstrakten moralni izračun zdaj nenadoma preklemansko oseben, kajti na tehtnici je naše lastno življenje.

2.3 Katero načelo za porazdeljevanje neizogibne škode?

Potem ko smo za silo utemeljili, čemu bi veljalo premišljevanje o načelih za porazdeljevanje škode v prometnih nesrečah prepustiti strokovnjakom za etiko, je zdaj čas za odgovor na bolj vsebinsko vprašanje, katero od ponujenih načel bi se po tehtnem premisleku izkazalo za edino veljavno. Osnovna izbira je jasna – ali naj algoritem prednostno zaščiti potnike v samovoznem avtomobilu ne glede na ceno, ki jo bodo morali za njihovo preživetje plačati drugi udeleženci v prometu, ali pa naj poskrbi, da bo celokupna škoda kar najmanjša ne glede na to, kdo jo bo utrpel. Imenujmo prvi algoritem neutilitaristični in drugi utilitaristični in si najprej na kratko oglejmo tri nezdrave argumente v prid vgradnji oz. izbiri neutilitarističnega algoritma.

- (i) Če je reševanje potnika v avtomobilu edini gotovo dober izid, medtem ko so vsi drugi izidi, vključno s tistimi, ki bi preprečili več škode, bolj ali manj negotovi, potem bi morali izbrati edino pot, ki nas bo zagotovo pripeljala do dobrega rezultata;
- (ii) avto je podaljšek voznikovega (ali, v primeru samovoznih avtomobilov, potnikovega) sebstva in kot tak upravičen do samoobrambe v primeru neposrednega trčenja;
- (iii) ker „morali bi rešiti več življenj“ ni veljavno moralno načelo, mora biti dopustno, da damo prednost svojemu življenju (in življenju svojih bližnjih) pred življenjem številnih drugih, če seveda drži, da nima nihče od vpletenih

⁵ Ugotoviti poskušamo, kaj je ljudem, ki se morajo v okoliščinah prisilne izbire odločiti, kdo naj preživi in kdo naj umre, bolj in kaj manj moralno pomembno – koliko ljudi bo v primeru take ali drugačne reakcije avtonomnega vozila preživelo in koliko umrlo, njihova identiteta, koristnost, družbeni status, starost in/ali spol, ali so se v nevarnosti znašli po svoji krivdi ali nekrivdno, tj. zaradi kršenja prometnih predpisov ali kljub njihovem spoštovanju itd.

močnejše pravice do življenja/preživetja kot kdorkoli drug (t. i. problem moralne relevantnosti števil Johna Taureka)

Filozofsko najbolj izzivalen je zadnji, zato mu bomo namenili nekaj več pozornosti.

2.3.1 (Nezdrav) argument za ne-UAV v duhu Johna Taureka

Da je lahko moralno sprejemljiv tudi algoritem, ki daje pri zaščiti pred oškodovanjem potnikom v avtu prednost pred drugimi udeleženci v prometu, čeravno je slednjih neprimerno več in bomo za preživetje potnikov plačali neprimerno višjo ceno, bi lahko zagovorniki neutilitarističnih algoritmov dokazovali s pomočjo naslednjega argumenta:

- (1) Ni slabše (ne za prizadete in ne iz brezosebne perspektive), če umre več ljudi, kot če umre manj ljudi.
- (2) Zato ne more biti narobe, če se moj AV namesto življenj petih pešcev na pločniku raje odloči rešiti moje življenje (zaradi zgornjega razloga).
- (3) Nobenih drugih razlogov ni, zaradi katerih bi bila lahko odločitev algoritma, da reši raje enega človeka (potnika v avtu) kot pet ljudi (pešcev na pločniku), napačna.
- (4) V primeru, ko so vsi življenjsko ogroženi enako (močno ali šibko) upravičeni do preživetja, nam pravičnost nalaga, da priskrbimo vsem enako in pošteno možnost za preživetje.

Zato,

- (5) kadar smo soočeni z izbiro med rešitvijo enega in rešitvijo več življenj tujcev, bi veljalo odločitev prepustiti metu kovanca.
- (6) Nakup in raba ne-UAV-ja namesto UAV-ja nista moralno problematična (niti ne razkrivata potrošnikovega moralno pomanjkljivega značaja).

Zato lahko sklenemo, da

- (7) bo algoritem, ki daje odločilno prednost varnosti oz. preživetju (manjšega števila) potnikov v AV-jih pred varnostjo oz. preživetjem (večjega števila) drugih udeležencev v prometu, brez posebnih težav opravi minimalni moralni preizkus.

Zgornji argument zveni prepričljivo, toda temeljita analiza hitro pokaže usodne slabosti. Kje mu torej spodrsne? Parfit (1978) prepričljivo ovrže njegovo prvo premiso. Kamm (1985) in Scanlon (1998) naredita enako za tretjo premiso. Njegova poglobljena slabost pa je, da premisi (4) in (5) stojita in padeta z domnevo, da so vsa ogrožena življenja moralno enakovredna oz. da imajo vsi, ki so vpleteni v prometno nesrečo, enako pravico do preživetja oz. do zaščite pred smrtjo/oškodovanjem kot kdorkoli drug. V prenekateri prometni situaciji, tudi teh, ki jih preigrava Moralni stroj, pa pogoj moralne simetrije ni izpolnjen, kajti medtem ko so eni posamezniki v njih udeleženi po lastni krivdi (ali pa so vsaj, četudi brez krivde, vzročno prispevali k njenemu nastanku), so se drugi v stiski znašli nekrivdno – kdor življenja ali zdravja drugih ljudi ne ogroža, pa bi moral v primerjavi s tistimi, ki to ali po svoji krivdi ali pa nekrivdno počnejo, uživati prednostno zaščito pred resnim in nepopravljivim oškodovanjem. K implikacijam, ki jih ima ta moralna asimetrija za vprašanje pravičnosti v porazdelitvi neizogibne škode, se še vrnemo.

2.3.2 ... in nekaj prepričljivih moralnih razlogov za proizvodnjo utilitarističnih AV-jev

Nakup in rabo UAV podpirajo najmanj tri vrste prepričljivih moralnih razlogov:

Konsekvenencialistični razlog: Moralno gledano je bolje, če v prometnih nesrečah umre manj ljudi, kot če umre več ljudi, če je vse ostalo enako (*ceteris paribus*).

Argument iz pravičnosti v izidu: Poleg krivice, ker vsem, ki so se znašli v smrtni nevarnosti, nismo v enaki meri zagotovili preživetja, obstaja še krivica, ki se nam zgodi, ker nas niso rešili, pa bi nas morali – in krivic te druge vrste bo več v primerih, ko namesto rešitve čim večjega števila življenj odločitev o tem, koga bomo rešili, manjšo ali večjo skupino ljudi, prepustimo naključju (npr. metu kovanca).

Pogodbeniški: Ne pripišemo enakega pomena življenju vsakega posameznika, ki mu grozi oškodovanje, če se kljub dodajanju vedno novih ljudi v skupino, rešenih sprva, uravnotežena tehtnica sčasoma ne nagne na stran številčnejše skupine.

2.4 Kakšen algoritem torej?

Tri zgoraj navedene vrste legitimnih premislekov, prvič, da naj algoritem poskrbi, da bo v prometni nesreči nastalo kar najmanj škode, drugič, da naj bo porazdelitev škode oz. tveganja za oškodovanje pravična, in tretjič, da naj bodo primerljivi interesi vseh vpletenih – prima facie, tj. v odsotnosti tehtnih nasprotnih razlogov – upoštevani v enaki meri, nas pripeljejo do prvega približka etičnega načela za porazdeljevanje neizogibne škode v primerih, ko bo samovozni avtomobil zaradi tehnične okvare ogrozil življenja udeležencev v prometu.

Načelo zmanjševanja celokupne škode, prilagojeno pravičnosti (NZŠP):
»Poskrbi, da bo škoda zaradi prometne nesreče, ki se ji ni dalo izogniti, čim manjša in med udeležence porazdeljena karseda pravično oz. pošteno«.

V tej abstraktni formulaciji bi moral biti NZŠP sprejemljiv tako za pristaše konsekvencializma kot tudi za deontologe. Za prve, ker lahko vlogo, ki jo nekoliko netipično odmerja premislekom o pravičnosti in poštenosti, upravičimo s sklicevanjem na pozitivne dolgoročne posledice take politike – s tem, ko zanje zmanjšamo tveganje, da bodo v prometu staknili poškodbo in/ali na cesti celo umrli, nagradimo tiste, ki spoštujejo prometne predpise,⁶ na ta način pa v voznikih avtomobilov in drugih udeležencih prometa posredno okrepimo previdnost in zatreto objestnost. Za druge pa, ker se bo z zmanjševanjem števila vseh žrtev prometnih nesreč samodejno zmanjšalo tudi absolutno število teh, ki v prometnih nesrečah škodo (v skrajnem primeru izgubo življenja) utrpijo po krivici.⁷

Kljub svoji teoretični privlačnosti pa vzbuja tak predlog najmanj dve vrsti upravičenih pomislekov (zaenkrat pozabimo na tehnično, kako sploh tako abstraktno pravilo narediti opravilno). Prvič, posamezniki, ki skrbijo le zase, za svoje življenje, zdravje in premoženje, medtem ko jim je za pravičnost in poštenost bolj

⁶ Pravičnost nam navsezadnje, vsaj intuitivno, nalaga (in bo enako zahtevala od algoritma v AV-ju), da v primeru, ko se trčenju ni več mogoče izogniti, lahko pa izbiramo, ali bomo trčili v avto s pripetim voznikom ali pa v takega z neprijetim voznikom, avto usmerimo v slednjega, čeravno bodo posledice predvidljivo hujše, kot če bi izbrali prvo možnost. (Furey in Hill 2020: 152)

⁷ Dodatno bi lahko opredelili in razvrščali vrste škode (s človeškimi smrtmi na vrhu, ki jim sledijo težke in trajne poškodbe, nato smrti nečloveških bitij, za njimi lažje poškodbe in na koncu izključno materialna škoda), zaradi česar bi postajal algoritem čedalje bolj zapleten. Bo zaradi tega prej ali slej postal celo prezapleten, da bi še bil praktično uporaben? Ne, načeloma bi moral delovati enako kot računalniški šahovski programi, ki vsak položaj na šahovnici analizirajo in vrednotijo skozi prizmo kombinacij več izbranih strategij in le redko ene same.

malo mar ali pa se znajo prepričati, da ima njihovo preživetje upravičeno prednost pred preživetjem vseh drugih tudi z vidika morale ali pravičnosti, bodo tako pravilo odločno zavrnil. Drugič, čeprav bi marsikdo to z veseljem podprl kot meta-, drugostopenjsko načelo za uravnavanje algoritemskih izbir, lahko pričakujemo pogosta nesoglasja o tem, ali algoritemska ‚rešitev‘ vsakokratne prometne dileme izpolnjuje dva pogoja ali ne (oz. ju izpolnjuje bolje kot kaka druga rešitev) – na primer, ali bomo, če bomo zavili in trčili v dve starejši ženski, ki prečkata cesto pri zeleni luči, ravnali bolj krivično, kot pa če bi vztrajali v prvotni smeri in do smrti povozili dva mladostnika, ki sta cesto prečkala pri rdeči luči.

Pristranskost do sebe oz. določena stopnja sebične preudarnosti, skrbi za lastne interese (za svoje preživetje, zaščito telesne integritete itd.) – in z njo tudi dajanja prednosti le-tem pred interesi drugih – je na prvi pogled morda moralno nepriljubljiva, a tako kot je nesporno psihološko dejstvo, je tudi ena od osrednjih moralnih vrlin. Z moralo je seveda nezdržljivo, da bi zaščiti oz. uveljavljanju lastnih interesov dajali absolutno in/ali neulovljivo prednost, določena mera sebične skrbi pa je, četudi morda ne dobrodošla, vsaj moralno sprejemljiva. Tak uvid nas pripelje do prve in zadnje revizije algoritemskega načela za porazdeljevanje neizogibne škode:

Načelo zmanjševanja celokupne škode, prilagojeno pravičnosti in rahli pristranskosti do sebe (NZŠPS): Vedno zmanjšaj celokupno škodo (z rahlo pristranskostjo do sebe) pod pogojem, da so izpolnjene zahteve pravičnosti (»poskrbi, da se nikomur ne bo zgodila krivica oz. da se jih bo zgodilo kar najmanj«) in poštenosti (»poskrbi, da bodo vsi obravnavani kot enaki, tj. da se njihovim interesom v moralni enačbi prizna enaka, ne večja in ne manjša izhodiščna teža kot vsem drugim«).

Pot do rešitve druge uganke je nekoliko bolj vijugasta, a vseeno prevozna. O tem, katera nasilna smrt je bolj krivična, mladostnika ali starostnika, ženske ali moškega, revnega ali premožnega, človeka ali živali, se predstave razlikujejo ne le med posamezniki, temveč tudi med kulturami. Vseh teh mnenjskih razlik se v algoritmu ne bo dalo vgraditi. In se jih tudi ne bi smelo, ker v nekaterih od teh spontanih moralnih sodb evidentno odzvanjajo starizem, seksizem, rasizem, etnocentrizem in drugi iracionalni in moralno arbitrarni predsodki, ki bi jih s tem neupravičeno legitimirali. Enako nesporno je, da izvirajo po drugi strani nekatere razlike v sodbah o tem, katera porazdelitev neizogibne škode je v danih okoliščinah bolj in katera

manj pravična iz različnih pojmovanj pravičnosti, ki so enako dobro razumsko utemeljena in so izraz globokih in nepomirljivih vrednostnih razhajanj. V tej luči bo smiselno, če se namesto iskanja ene same, edino pravilne algoritemske rešitve za vsako prometno moralno enačbo, sprijaznimo z razponom možnih rešitev, ki so še znotraj polja razumno podprtega in/ali moralno sprejemljivega.

3 Politika programiranja samovoznih avtomobilov

Predpostavimo, za potrebe ovrednotenja argumenta, da nam je uspelo pravilno odgovoriti na prvo vprašanje – če je vse ostalo enako (*ceteris paribus*), bi bilo v primeru nakupa osebnega avtomobila edino moralno pravilno, da potrošnik izbere samovozni avtomobil, ki je opremljen z utilitarističnim (v ohlapnem pomenu besede) algoritmom. Naj država tovrstno odločitev prepusti ljudem in se sprijazni s tveganjem, da bodo ti množično dajali prednost opciji, ki maksimira njihove osebne koristi (tj. preudarnosti), pred edino moralno pravilno opcijo, ali pa jim naj za ceno posega v osebno svobodo edino pravilno izbiro vsili, tj. predpiše nakup utilitarističnih samovoznih vozil?⁸ Proti drugemu predlogu govorita dva premisleka – da bi država/oblast s tem močno preseгла svoje pristojnosti in pa da bi znal biti tak ukrep v luči nedavnih raziskav o odnosu ljudi do samovoznih avtomobilov kontraproduktiven.

3.1 Liberalna oblast in individualna svoboda

Kaj vse sploh velja za legitimen razlog, zaradi katerega sme oblast/vlada posameznikom omejiti svobodo odločanja in delovanja? Mnenja o tem so med političnimi filozofi in filozofinjami deljena. Moralni legalisti, recimo, vidijo eno od funkcij oblasti/države tudi v zaščiti javne morale. Zanje oblastno vsiljevanje moralno pravilne odločitve v primeru nakupa samodejnih vozil potemtakem ne bo nujno sporno. Sam bom namesto tega raje izhajal iz liberalnega filozofsko-pravnega okvirja. Znotraj le-tega je mogoče zakonske prepovedi in zapovedi upravičiti le, če je to nujno za zaščito posameznikov pred oškodovanjem in/ali krivicami, ki bi ga

⁸ V pričujočem prispevku se bomo pretvarjali, da sta pojem in vrednost osebne svobode enoznačna in neproblematična. Resnica je seveda ravno nasprotna – z izrazom 'svoboda' označujemo zelo različne stvari, od odsotnosti preprek ali prisile preko zmožnosti za nekaj do odsotnosti nadvlade (angl. *domination*) in nekatere od teh stvari so že na prvi pogled bolj – in iz drugačnih razlogov – dragocene kot druge. Potrošniška svoboda, ki bi jo ogrozila morebitna zakonska prepoved prodaje in nakupa ne-UAV-jev, je vsaj na prvi pogled bolj trivialne vrste. Zaradi prostorskih omejitev se v to, sicer filozofsko, nadvse zanimivo razpravo ne bom spuščal. Lep pregled ponuja Carter (2021).

oz. jih ti utrpeli zaradi izbir in ravnanja drugih ljudi, in če bodo (edino) taki zakonski ukrepi dovolj učinkovito opravili zadano nalogo (medtem ko je blažji posegi v osebno svobodo posameznikov ne bi). Liberalno pojmovanje o nalogah/funkcijah kazenskega in civilnega prava je seveda kompleksnejše, kot pa daje slutiti zgornja poenostavitvev. Liberalno-demokratska oblast sme svobodo odločanja in delovanja posameznikov z zakoni in predpisi omejiti, s političnimi in ekonomskimi ukrepi pa preusmeriti v družbeno zaželeno smer iz cele vrste razlogov: da bi ene ljudi zavarovala pred škodo, ki bi jo utegnili utrpeti zaradi odločitev ali dejanj drugih ljudi; da bi preprečila krivično (angl. *wrongful*), čeravno morda neškodljivo ravnanje oz. ljudi odvrnila od takega ravnanja; da bi poskrbela, da bodo ti, ki drugim delajo škodo ali krivico, za svoje ravnanje kazensko ali odškodninsko odgovarjali; da bi oškodovancem in žrtvam krivic zagotovila določene koristi iz naslova odškodnin in poprave krivic; in končno, da bi pravičnejše porazdelila breme družbeno neželenega ravnanja. (Simpson 2012)

3.2 Legitimni cilj in sredstva prometne politike

Zakonska določba, da morajo biti vsi samovožni avtomobili opremljeni z utilitarističnim algoritmom, se lepo umešča v kategorijo regulatornih javnih politik, ki določajo pogoje oz. zarisujejo omejitve ravnanju posameznikov, organizacij in korporacij iz legitimnih razlogov, namreč zaradi izboljšanja prometne varnosti in večje prometne pravičnosti. Ne-UAV-ji bi svojim morebitnim voznikom in potnikom resda zagotovili boljšo zaščito v prometu v primerjavi z UAV-ji, a ker bi si boljšo varnost zagotovili na račun zvišanega tveganja za vse druge udeležence v prometu, bi tak aranžma ustvaril to, čemur pravijo ekonomisti negativne eksternalije – koristi za ene (delno ali v celoti) na račun stroškov za druge, ki niso imeli pri razmisleku, ali se jim bo tak vložek obrestoval, nobene besede. Če bi torej država posameznikom dovolila nakup in uporabo ne-UAV-jev, bi si težko opredeljiv, a najbrž ne nezanemarljiv delež potrošnikov (v taksonomiji/terminologiji Bergmanna et al. vsi sebičneži, ki jih je približno za petino populacije, in vsaj nekateri preklopniki) na ta način zagotovil otipljive koristi, stroške zanje pa – v obliki povečanega tveganja za v prometu povzročene poškodbe in smrt – prevalil na pleča drugih in posredno, v obliki zvišanih stroškov zdravstvene oskrbe in dodatnih smrti, na celotno skupnost. Politične skupnosti upravičeno skrbi, kadar eni ljudje s svojimi – tudi potrošniškimi – odločitvami in dejanji zvišujejo tveganja za zdravje in življenje

drugih ljudi, in jih smejo od tega odvrniti, če ne gre drugače, tudi z zakonskimi prepovedmi in omejitvami.⁹ (Wolff 2011: 86)

Ukrepe, ki uravnavajo načrtovanje, oblikovanje, proizvodnjo, prodajo ali uporabo avtomobilov iz tovrstnih razlogov, ni težko najti, od zakonskih določb o maksimalnih izpustih do varnostnih standardov za rezervne dele in dodatke, rednih tehničnih pregledov in obveznega avtomobilskega zavarovanja pa vse do zaprtja mestnih središč za avtomobilski promet, omejitev hitrosti in predpisane uporabe varnostnih pasov ali čelad za motoriste. Zadnji v tej vrsti je odlok, s katerim je Evropska komisija proizvajalcem avtomobilov določila zgornjo letno mejo povprečnih emisij na proizveden avto in predpisala kazni za njeno prekoračitev. Predpis, ki bi proizvajalcem samovoznih avtomobilov nalagal vgradnjo utilitarističnih algoritmov oz. otežil ali onemogočil prodajo neutilitarističnih različic, torej ne bi v ničemer odstopal ne od siceršnje regulative in ne od ciljev, ki jih države z vrsto pravnih, fiskalnih, administrativnih in političnih ukrepov zasledujejo na področju avtomobilske industrije, cestnega prometa, splošne mobilnosti in, zakaj pa ne, tudi sprememb življenjskih navad. Sklicevanje na individualne pravice in svoboščine – da ima vsak od nas pravico izbrati avto po svojem okusu, če le ta izpolnjuje osnovne tehnične in varnostne zahteve, je v tej luči kratkovidno in naivno.¹⁰

Kritik vladnega vmešavanja v poslovni odnos med proizvajalcem, ki je pripravljen ne-UAV izdelati, trgovcem, ki ga je pripravljen prodajati, in potrošnikom, ki ga želi kupiti, bi lahko vztrajal, da bi bila zakonska prepoved tovrstnih transakcij še posebej moralno ali politično problematična zato, ker bi država s tem posameznikom proti njihovi volji naprtila določena tveganja, ki jih sami po tehtnem premisleku morda ne bi bili pripravljeni prevzeti nase. Toda nekaj podobnega bi lahko očitali tudi obvezni rabi varnostnih pasov ali obveznemu cepljenju zoper nevarne in hudo nalezljive bolezni – z obema ukrepoma nam država nalaga določena, resda minimalna tveganja

⁹ S tem nočem reči, da so eksternalije vedno dovolj dober razlog za oblastne posege v svoboščine posameznikov ne glede na višino stroškov ali resnost tveganj, ki jih ti ustvarjajo za druge brez njihovega soglasja. Prodaja glomaznih SUV-jev je, denimo, pri nas in v večini držav po svetu dovoljena, čeprav so ti za druge udeležence v prometu dokazano bolj nevarni, praviloma pa tudi bolj potratni in okolju neprijazni kot vozila manjših dimenzij. Še pa vsa omenjena dejstva o SUV-jih sestavijo vsaj v prima facie prepričljiv razlog za ukrepe, s katerimi bi država posameznike odvrnila od nakupa tovrstnih vozil oz. jim zagrenila tako odločitev.

¹⁰ V daljši verziji tega članka kritično obravnavam argument za svobodo delovanja in poslovanja, ki se navdihuje pri Miltonu Friedmanu (Friedman 1970) in njegovi ideji, da je skrb za varnost – tudi prometno – stvar avtonomnih posameznikov in ne vsiljive, pokroviteljske oblasti.

in nas istočasno oropa za priložnost, da jih samostojno ocenimo in kot sprejemljiva odobrimo oz. kot nesprejemljiva zavrnemmo (Giubilini in Savulescu 2019). Morebitna vladna prepoved prodaje in nakupa neutilitarističnih samovožnih avtomobilov ne bi bila nič bolj sporna, kot sta zaradi posegov v individualno avtonomijo – v izhodišču, ne pa tudi po tehtnem premisleku – sporna obvezna uporaba varnostnih pasov ali obvezno cepljenje.

Doslej smo možne utemeljitve za zakonsko regulacijo trga AV-jev iskali izključno znotraj klasičnega liberalnega okvirja. Če ta okvir razširimo in državni oblasti priznamo širši nabor funkcij in nalog, s tem pa podelimo tudi širša pooblastila, se sorazmerno s tem poveča tudi nabor legitimnih razlogov za tovrstno regulacijo. Od nje si lahko, recimo, realno obetamo tudi določene moralne koristi. Če bi namreč država proizvodnjo in prodajo AV-jev z vgrajenim ne-utilitarističnim algoritmom dovolila, bi umetna inteligenca/stroji prevzeli vlogo spodbujevalca (angl. *enabler*) neetičnega obnašanja, ki je sicer brez dvoma družbeno nezaželeno, vendar pa posameznikom omogoča, da žanjejo sadove neetičnega obnašanja in se izognejo plačilu stroškov zanj – res da so na račun nekoga drugega preživeli prometno nesrečo, a ker je to zanje koristno, četudi moralno sporno odločitev sprejel nekdo drug in ne oni sami, si nimajo česa očitati. Z morebitno zakonsko prepovedjo vgradnje neutilitarističnih algoritmov v AV-je bi tovrstnemu moralnemu parazitstvu naredili konec.¹¹

Če povzamem. Država/oblast ne bi preseгла svojih pristojnosti, če bi se zaradi izboljšanja splošne prometne varnosti, pa tudi v imenu prometne pravičnosti, odločila kupcem avtonomnih vozil omejiti možnost izbire oz. vsiliti svojo voljo. Oba cilja sta sama po sebi legitimna in ker je prizadevanje zanju učinkovito le, če je podprto z ustreznimi sistemskimi ukrepi, tudi v pristojnosti države, tovrsten poseg pa niti približno ne bi bil ne prvi in ne edini te vrste – država je v imenu teh istih (obvezna uporaba varnostnih pasov) ali podobnih ciljev (obvezno cepljenje proti nalezljivim boleznim) že sprejemala in še vedno sprejema pravne, politične,

¹¹ Tezo o moralnih koristih, ki si jih lahko obetamo od vsiljevanja utilitarističnih algoritmov, lahko razumemo na dva načina, v močnejšem smislu kot tezo, da bodo ti izboljšali človeški značaj, ali v šibkejšem smislu kot tezo o izboljšanju človeškega obnašanja. Za obvezno vgradnjo utilitarističnih algoritmov bi se utegnilo izkazati, kar Giubilini in Savulescu 2019 ugotavljata o predpisani uporabi varnostnega pasu – namreč da se je kljub začetnemu odporu v zgolj nekaj letih iz eksternega zakonskega predpisa spremenila v ponotranjeno družbeno normo. V tem primeru bi bile moralne koristi od obvezne vgradnje utilitarističnih algoritmov globoke in trajne. A tudi če bo omenjeni ukrep ljudem pomagal zgolj bolje izpolniti njihove moralne dolžnosti, moralne koristi od njega ne bodo zanemarljive. Več o moralnem dometu zakonodajnih ukrepov glej Thomson (1986) in Hunt (2014).

ekonomske itd. ukrepe, ki v imenu javnega dobrega ožijo prostor individualne svobode.¹² Pa bi bila taka odločitev tudi zares pametna, preudarna? Rezultati nedavnih raziskav dajejo slutiti, da bi se znala vrniti kot bumerang.

3.3 Ko bi le bili vsi ljudje nesebični ... razen mene!

Bonnefon et al. so izvedli vrsto eksperimentov, da bi proučili odnos ljudi do AV-jev obeh vrst, utilitaristične in neutilitaristične. Pri tem so se dokopali do nekaterih nadvse zanimivih ugotovitev. V vseh študijah je večina udeležencev izrazila moralno preferenco za AV-je, ki žrtvujejo svoje potnike, da bi rešili večje število pešcev. Ta moralna preferenca je bila močna v okoliščinah, ko so si udeleženci zamišljali sebe v AV-ju v družbi sodelavca, družinskega člana ali lastnega otroka. V tej luči je presenetljivo, da udeleženci niso izrazili primerljive preference tudi za nakup utilitarističnega AV-ja, še zlasti ne, ko so si morali v vlogi sopotnikov zamisliti družinske člane. Za povrh udeleženci niso bili naklonjeni ideji, da bi oblasti predpisale utilitaristične algoritme za AV-je, in so dali vedeti, da se najverjetneje ne bi odločili za nakup AV-ja, če bi bili vsi tovrstni avtomobili brez izjeme opremljeni s takimi algoritmi.

3.4 Resnična ali zgolj navidezna družbena dilema?

Tabela 1: Individualna odločitvena matrika (problem koordinacije)

	Večina drugih ljudi (se) vozi (v) neutilitaristične/ih samovozne/ih avtomobile/ih	Večina drugih ljudi (se) vozi (v) utilitaristične/ih samovozne/ih avtomobile/ih
jaz (se) vozim (v) neutilitaristični/em samovozni/em avtomobil/u	(1) Zame tretja najboljša možnost (z vidika splošne prometne varnosti pa najslabša možnost)	(2) Zame najboljša možnost
Jaz (se) vozim (v) utilitaristični/em samovozni/em avtomobil/u	(3) Zame najslabša možnost	(4) Zame druga najboljša možnost (z vidika prometne varnosti najboljša možnost)

¹² Da bi šlo pri zakonski prepovedi prodaje in nakupa ne-UAV-jev za razmeroma blag poseg v svobodo delovanja in poslovanja posameznikov in organizacij, nakazuje še en premislek. Nekateri teoretiki verjamejo, da država bolj drastično poseže v svobodo posameznikov, če iz obstoječega nabora možnih izbir eno (recimo nakup in uživanje tobačnih izdelkov) odstrani, kot pa če prepreči, da bi ta ali oni ponudnik izdelkov ali storitev obstoječemu naboru možnih potrošniških izbir eno dodal. Prikrajšanje za določeno opcijo naj bi bilo, skratka, vedno lažje opravičiti kot odvzete taiste. Takšno stališče ni neproblematično, kot v svoji kritiki lepo pokaže Schmidt (2016), a če ga pogojno sprejmemo, iz njega izhaja, da bo v primeru zakonske regulacije algoritmov v AV-jih letvica za upravičenje državne intervencije postavljena razmeroma nizko, ker pač potrošnikom s tem možnost izbire ne-UAV-ja ne bo odvzeta (angl. *withdrawn*), temveč bodo zanjo zgolj prikrajšani (angl. *withheld*).

Nakup in uporaba neutilitarističnega samovoznega avtomobila je, skratka, z vidika posameznika enoznačno dominantna in v toliko najbolj preudarna strategija. Pa vendar, če bo vsak poskušal povečati osebno korist na račun drugih – in upravičeno pričakujemo, da se bomo tako obnašali vsi ali vsaj večina med nami – bomo na koncu dobili kolektivno/družbeno neoptimalen izid. Vsaj na prvi pogled smo se torej ujeli v začaran krog:

- (i) Nakup in raba UAV je najbolj zaželena izbira tako z moralnega kot z družbenega vidika (in velika večina ljudi se s to oceno celo strinja).
- (ii) Kljub temu bi večina ljudi, če bi to odločitev prepustili njim, namesto tega zase raje kupila in uporabljala ne-UAV.
- (iii) Če bi država ponudbo AV-jev z zakonom slučajno omejila zgolj na UAV-je oz. če bi dovolila zgolj prodajo AV-jev z utilitarističnim algoritmom, pa bi ljudje celo raje izbrali kaj drugega, recimo ne-AV.

Izsledki Bonnefonove študije snovalce prometne politike vsaj na prvi pogled postavljajo pred nerešljivo družbeno dilemo: naj žrtvujemo svobodo, da bi izboljšali varnost, ali naj raje žrtvujemo varnost, da bi ohranili svobodo? Toda ali imamo pri programiranju samovožnih avtomobilov in njihovi promociji res opraviti z zagatnim problemom koordinacije, kjer bi vsak posameznik najraje videl, da vsi drugi storijo, kar je prav in družbeno koristno, tj. optimalno z vidika seštevka interesov vseh vpletenih, zase bi hotel pa izpogajati izjemo od tega pravila? Izsledki nekaterih drugih raziskav o pripravljenosti ljudi, da se zaradi preživetja večjega števila drugih ljudi sprijaznijo z lastnimi poškodbami ali celo smrtjo, so nekoliko bolj optimistični.¹³ Raziskava, ki so jo izvedli Bergmann in sodelavci, je, recimo, z uporabo podobnih prometnih scenarijev med udeleženci odkrila »presenetljivo visoko stopnjo pripravljenosti, da žrtvujejo sebe, da bi rešili druge«. Natančneje:

/.../ udeleženci so delovali bolj altruistično od pričakovanj, čeprav so rezultati v dani situaciji odvisni od števila potencialnih žrtev. V več kot polovici poskusov so bili udeleženci pripravljene rešiti celo samo dve osebi. S povečevanjem števila ogroženih ljudi raste tudi delež tistih, ki so se pripravljene žrtvovati zanje. Medtem ko so bili za rešitev dveh oseb pripravljene umreti v 52 odstotkih primerov, je pri treh rešenih življenjih ta

¹³ In sploh niso edini, ki Bonnefonu očitajo preveč pesimistično interpretacijo zbranih podatkov. Glej npr. Greene 2016 in Hübner in White 2018.

delež narasel na 57 in pri štirih na 63 odstotkov. Pri skupinah petih, šestih in sedmih oseb so bili rezultati precej podobni, za toliko ljudi se je bilo pripravljenih žrtvovati približno 70 odstotkov anketirancev. Rezultati podpirajo zaključek o obstoju različnih skupin odločevalcev, ki sledijo vsaka svoji strategiji. Prva skupina se je dosledno odločala za lastno preživetje na račun drugih in jo lahko zato označimo za (moralne) egoiste. Druga skupina, recimo ji altruisti, je, nasprotno, dosledno izbirala možnost samožrtvovanja. Tretji skupini bi lahko rekli preklopniki, ker so člani svoje odločitve uravnavali po velikosti skupine, pri čemer so se odločili za samohranitev v preizkušnjah z manj in za samožrtvovanje v preizkušnjah z več ogroženimi ljudmi. (Bergmann et al. 2018)

Da lahko ljudi po pripravljenosti, da se žrtvujejo za druge, razvrstimo v eno od treh skupin, med sebičneže, nesebičneže in tiste vmes, ne preseneča. Bolj so v luči Bonnefonove pesimistične študije presenetljivi – in razveseljivi – sorazmerni deleži vsake od njih. Če bi jih z veliko mero poenostavljanja neposredno prevedli v nakupne odločitve (pa jih seveda ne moremo), bi si lahko obetali, da se bo nekje med polovico in dvema tretjinama kupcev odločila za nakup AV-ja z utilitarističnim algoritmom. Celo ob upoštevanju dejstva, da se marsikatera v anketi zabeležena sodba o tem, kateri razplet je bolj in kateri manj moralno sprejemljiv in/ali pravičen, nazadnje ne bo prevedla v ustrezno potrošniško odločitev, najbrž drži, da so Bonnefon in kolegi močno podcenili delež tistih, ki bodo izbrali samovozni avtomobil z utilitarističnim algoritmom prostovoljno, se pravi brez oblastne palice in korenčka.

3.5 Kontraproduktivna prometna politika?

Kaj pa če je Bonnefonova skupina vendarle natančneje izmerila (brez)srčni utrip javnosti/ljudstva kot Bergmannova? Za oblikovalce utilitarističnih politik bo v tem primeru najpametneje, da se odrečejo vsiljevanju utilitarističnih samovoznih avtomobilov. Kar je paradokсно. Vendar pa je tudi ta paradoks bolj navidezen kot resničen – zavajajoče je namreč imenovati avtomobile, ki zmanjšajo število smrtnih žrtev/škodo v posameznih primerih, vendar pa zaradi svoje nepriljubljenosti pri potrošnikih ne bodo zmanjšali splošne škode zaradi prometnih nesreč, 'utilitaristični'. V ugotovitvi, da je politika, za katero bi pričakovali, da bo po utilitarističnih kriterijih moralno optimalna, namreč spodbujanje nakupa in rabe

izključno utilitarističnih AV-jev, potem ko seštejemo vse, tudi dolgoročne učinke tovrstne politike, šele na drugem mestu za nakupom in uporabo AV-jev, ki sicer zmanjšujejo škodo, a ne za ceno ogrožanja varnosti potnikov, ni v resnici nič paradoksnega. Z utilitarističnega vidika bo v takem primeru bolje, da nekateri ljudje kupujejo in uporabljajo avtomobile, ki bi jih bilo z utilitarističnega vidika bolje ne uporabljati. Če se bo izpolnila napoved, da bodo avtonomna vozila močno oklestila število nesreč in posredno smrtnih žrtev na cestah, hkrati pa bodo utilitaristični samovozni avtomobili med potrošniki izrazito nepriljubljeni, bi lahko vlada z njihovim agresivnim vsiljevanjem zamaknila uvedbo na splošno varnejših avtonomnih vozil. Zato se bomo morda prisiljeni iz utilitarističnih razlogov – da bi kar se da hitro in občutno zmanjšali krvni davek na naših cestah – odpovedati vsiljevanju 'utilitarističnih' AV-jev in sprijazniti z drugo najboljšo možnostjo, nakupom in rabo/prevlado neutilitarističnih, a v primerjavi z neavtonomnimi vozili še vedno veliko varnejših AV-jev.

3.6 Družbena trilema?

Prvoten vtis o paradoksu, v katerega naj bi se neizogibno zapletla na utilitarističnem kalkulu temelječa politika, je zmoten še iz enega razloga. Opisani družbeni in psihološki parametri namreč ne tvorijo dileme med a) svobodo potrošniške izbire in b) prometno varnostjo, temveč trilemo med (a) svobodo izbire, (b) (prometno) pravičnostjo in (c) (prometno) varnostjo/dobrobitjo. Ta spominja na trilemo, ki nas je pestila med pandemijo COVID-19, ko smo bili prisiljeni izbirati med tremi vrednotami: (a) svobodo/svobodščinami, (b) enakostjo/pravičnostjo in (c) varnostjo/dobrobitjo, in smo lahko v najboljšem primeru ohranili dve od navedenih treh. Glede na to, kar smo vedeli o virusu in bolezni, ki jo je ta povzročal, so nam bile namreč na voljo naslednje osnovne opcije: (i) življenje brez slehernih omejitev, vendar za ceno nesprejemljivo visoke stopnje okužb, bolezni in smrti (torej svobodo za vse in enako obravnavo vseh, a za ceno zmanjšane varnosti oz. dobrobiti), (ii) življenje s selektivnimi, ciljanimi ukrepi, ki bi nesorazmerno prizadeli določen del populacije, recimo ljudi z večjim tveganjem za okužbo in/ali bolezen oz. tiste, ki v večji meri ogrožajo druge ljudi (razmeroma dobra zaščita zdravja/dobrobiti vseh, a za ceno neenake obravnave, tj. svobode za večino in nesvobode za manjšino), in (iii) vsesplošno in vseobsegajoče zaprtje/lockdown, ki vse ljudi razmeroma dobro zaščiti pred okužbo, boleznijo in smrtjo in kjer nas oblast vse obravnava enako (slabo, bi

pripomnil cinik), a za varnost in enakopravnost plačujemo visoko ceno v obliki vsesplošnega suspenza pravic in svoboščin. (Pugh et al. 2021)

Podobno kot za pandemijo covid-19 bi se lahko tudi za prihajajočo avtomatizacijo prometa zdelo, da nas namesto z družbeno dilemo sooča z družbeno trilemo, tj. s prisilno izbiro med tremi konkurenčnimi/konfliktnimi (skupinami) vrednot oz. načel: (a) osebno svobodo;¹⁴ (b) prometno varnostjo in (c) prometno pravičnostjo. Na voljo imamo vrsto rešitev, ki ohranjajo po dve od teh treh vrednot, a med njimi ni niti ene, ki bi nam pomagala ohraniti vse tri hkrati – izbiramo lahko med (i) prometno varnostjo in pravičnostjo, a za ceno žrtvovanja svobodne potrošniške izbire; (ii) svobodno potrošniško izbiro in prometno varnostjo, a na račun okrnjene in/ali zamaknjene prometne pravičnosti; ter (iii) svobodno potrošniško izbiro in prometno pravičnostjo, a za ceno poslabšanja prometne varnosti.¹⁵

Na srečo je tudi opisana trilema bolj namišljena kot resnična. Množična raba AV-jev, od katere si – upravičeno ali ne – obetamo pomembno izboljšanje prometne varnosti, namreč ni nujno navzkriž s prometno pravičnostjo. V kombinaciji s potrošniško svobodo bo, če ima Bonnefon prav in se Bergmann moti, res zmanjšala in zakasnila prometno pravičnost (v njenem maksimalnem možnem obsegu), a če drži, da bo ob množični rabi AV-jev žrtev prometnih nesreč in škode zaradi njih neprimerno manj kot sedaj, bo po matematični nujnosti neprimerno manj tudi žrtev prometnih nesreč, ki jih bodo smrt, pohabljenje, težje poškodbe in podobne oblike oškodovanja doletele po krivici. Kar bomo prisiljeni žrtvovati, če prepustimo izbiro algoritma, s katerim bo opremljen njihov AV, samim kupcem, ni toliko prometna pravičnost nasploh kot bolj en – težko izmerljiv – del le-te. V določenem smislu torej drži, da bi bilo mogoče hkrati zagotoviti vse trije: udeležencem v prometu varnost, vpletenim v prometne nesreče pravičnost in potrošnikom svobodno izbiro.¹⁶

¹⁴ Če bi jo hoteli trivializirati, bi rekli ‚svobodo potrošniške izbire‘, a moralni nazori, ob katere bi znal trčiti utilitaristični algoritem in zaradi katerih ga bo marsikdo odklanjal, tvorijo trdno jedro vsakega svetovnega nazora in so kot taki seveda vse prej kot trivialni.

¹⁵ Opcija (i) je med vsemi najmanj sporna, ustreza namreč popolni zakonski prepovedi nakupa ne-UAV. Kaj pa drugi dve? Za kombinacijo (ii) smo že ugotavljali, da ji ustreza scenarij pospešene proizvodnje, prodaje in rabe obeh vrst AV-jev, utilitarističnih in neutilitarističnih, ob predpostavki, da ti neizogibne škode še ne bodo sposobni odmerjati in razporejati na podlagi natančno odmerjene odgovornosti zanj. Opcija (iii) se zdi med vsemi najbolj neverjetna – težko si je namreč zamisliti okoliščine, ko bi prosta prodaja in množična uporaba neutilitarističnih AV-jev izboljšala prometno pravičnost na račun poslabšanja prometne varnosti oz. povzročila rast skupnega števila žrtev prometnih nesreč, ne da bi se sočasno povečalo število tistih, ki jo v prometnih nesrečah skupijo po krivici.

¹⁶ Kar ponujam tukaj, je bolj oris argumenta kot njegova dokončna podoba. Za oceno vpliva, ki ga bodo posamezni ukrepi imeli na prometno pravičnost, bi potrebovali vsaj približno oceno, v koliko prometnih situacijah neizogibne

3.7 Pravična porazdelitev neizogibne škode?

Do zdaj smo bore malo rekli o merilih za presojo, ali bo določena porazdelitev neizogibne škode oz. tveganj zanj v prometnih nesrečah pravična ali ne. Je sploh smiselno trditi in dokazovati, da vodijo ene reakcije AV-jev v tem ali onem od trinajstih tipskih scenarijev Moralnega stroja do pravičnejšega izida kot druge? In vztrajati, da moralni enačbi brez te spremenljivke nekaj manjka in je moralni izračun zato napačen? Nikomur od nas seveda ni tuj pravičniški bes ali nejevolja, ki nas tipično obide ob prebiranju novinarskih poročil o prometnih nesrečah, v katerih jo vinjeni povzročitelji prometnih nesreč odnesejo brez prask, ne krivi in ne dolžni pešci ali kolesarji ali potniki v vozilu pa njihovo objestnost plačajo z življenjem. Toda v tipičnih prometnih scenarijih, ki jih preigrava Moralni stroj in za potrebe katerih se bo programirala avtonomna vozila, prometno nesrečo brez izjeme zakrivi okvara stroja in ne objestnost ali kaka druga človeška hiba. Se je v takih okoliščinah sploh smiselno spraševati, kdo od vpletenih si bolj in kdo manj zasluži preživeti oz. umreti ali kdo med njimi je morda upravičen do prednostne zaščite pred škodo, ki se ji ni mogoče izogniti? In če je pri trkih avtonomnih vozil za nesrečo nesmiselno kriviti ljudi, na kaj drugega bi se lahko oprla sodba, da bo, recimo, smrt sopotnika v takem vozilu manj krivična kot smrt pešca, v katerega je to vozilo trčilo na pločniku? Na tem mestu se pri stranskih vratih v razpravo vrača vprašanje odgovornosti, ki smo ga v uvodu postavili na stranski tir v upanju, da bo to moralno presojo maksimalno poenostavilo. Kakorkoli kompleksna in zamudna že presoja o pravični porazdelitvi škode je, očitno se ji ne da zares izogniti.

Na tem mestu lahko ponudimo kvečjemu oris take teorije. Pri tem izhajamo iz predpostavke, da morala v izhodišču vse ljudi v enaki meri varuje pred oškodovanjem ali tveganjem zanj, da pa se ta izvorna moralna simetrija podre, kadar je kdo od vpletenih odgovoren za nastanek tveganih okoliščin oz. neizogibne škode, in da ta pogoj izpolnimo, kadar, recimo, prostovoljno sodelujemo v dejavnostih, za

škode bo pravičnost potegnila ta kratko – bodisi zato ker bo neutilitaristični AV, katerega nakup smo potrošnikom omogočili v imenu svobode, absolutno prednostno zavaroval potnike na račun oškodovanja – večjega števila – drugih udeležencev (potnikov v drugem avtu, kolesarjev, motoristov, pešcev, itd.) ali pa ker, če si sposodim Kaupinnenov pomislek, v AV-je vgrajeni algoritmi ne bodo znali pravilno odmeriti odgovornosti za škodo (angl. *liability to harm*) za v nesrečo vpletene posameznike in v skladu z njo mednje porazdeliti neizogibne škode. Ta drugi potencialni vzrok za ohranjanje prometnih krivic nas po pravici povedano niti pretirano ne skrbi, ker skrb za pravično porazdelitev škode med vse v nesrečo vpletene najbrž že zdaj ni ravno med poglavitnimi dejavniki, ki bi usmerjali reakcije voznikov v kočljivih situacijah, in je zato malo verjetno, da bi AV-ji zaradi algoritmov, ki bodo v prvi fazi odgovornost za oškodovanje ali smrt in z njo neizogibno škodo odmerjali nenatančno in nepravično, ta rezultat bistveno poslabšali.

katere bi lahko predvideli (ali bi celo morali predvideti), da bodo ustvarila taka tveganja. Če smo sebe in druge spravili v nevarnost zavestno ali z brezobzirnim ravnanjem, bi morali sami nositi poglaviten delež tako nastalega tveganja oz. škode. Glede tega ni razhajanj. Kaj pa če prometno nesrečo povzroči AV, ki smo ga kupili in uporabljali dobro vedoč, da mu lahko odpovedo zavore ali krmilni mehanizem in da bo v tem primeru ogrozil življenja drugih udeležencev v prometu – smo odgovorni za to, da ta naš AV ogroža življenja drugih ljudi, in smo posledično izgubili pravico do varstva pred oškodovanjem, taisto pravico, ki so jo vse druge potencialne žrtve v teh okoliščinah ohranile? Ter se nam zato ne bi zgodila krivica, če bi se v taki prometni nesreči poškodovali ali umrli, bi pa bilo krivično, če bi v njej namesto nas izgubil življenje kdo drug?

To so vse prej kot lahka vprašanja. Še zlasti ker med filozofi_njami ni soglasja o pogojih, ki morajo biti izpolnjeni za brezprizivno oškodovanje oz. uboj (angl. *liability to being harmed/killed*). Najmanj, kar lahko rečemo, je, da potniki v samovoznem avtomobilu, ki je zaradi okvare povzročilo prometno nesrečo, v takem primeru definitivno ne bodo uživali absolutne prednosti pri zaščiti pred nastalo škodo. In da bi bilo zato krivično, če bi jih algoritem ne glede na druge okoliščine vedno in brez izjeme prednostno zaščitil pred njo. Kaupinnen (2021) celo verjame, da bi bili v vrsti kandidatov za oškodovance na prvem mestu. Razlog za to je preprost – čeravno ne bodo neposredno *krivi* za okvaro AV-ja in posledično nesrečo, ki jo bo le-ta povzročil, bo vendarle res, da so kupili in uporabljali tak avto prostovoljno in ob polni zavesti, da se kaj takega lahko pripeti, kljub temu da bi bili lahko ravnali drugače (tj. kupili kak drug avto ali pa se nakupu avtomobila odrekli), zaradi česar bodo za nesrečo *minimalno moralno odgovorni* in se posledično ne bodo mogli sklicevati na enako pravico do varstva pred oškodovanjem kot drugi, naključni udeleženci. Omenjeni Kaupinnenov sklep seveda ni neproblematičen, sloni namreč na dveh vse prej ko samorazvidnih domnevah: a) da za brezprizivnost oškodovanja zadošča že prostovoljna udeležba v aktivnosti, za katero vemo, da bi lahko prispevala k situaciji neizogibne škode in b) da temu opisu ustrežata že sam nakup in raba AV-ja. A ne glede na to, ali pravičnost v tovrstnih okoliščinah res *zahteva* prednostno žrtvovanje potnikov pred drugimi udeleženci v prometu ali ne, prej naštetih razlogi gotovo podpirajo vsaj blažji sklep, namreč da bi bila brezpogojna prednostna zaščita, kakršno bi svojim potnikom zagotavljal neutilitaristični algoritem, krivična.

Avtomatizacija prometa in prometnih sredstev bo zmanjšala število prometnih nesreč in žrtev le-teh, ne bo pa nujno poskrbela, da bodo vedno rešena tista življenja, ki bi jim moralno gledano morali dati prednost. Če tehnologija ne bo kos nalogi, da odmeri odgovornost posameznih udeležencev v prometu, bo občutnemu deležu udeležencev v prometu še naprej kratena pravica do učinkovitega varstva pred oškodovanjem. Zato je nujno najti razumen kompromis med zmanjševanjem škode in njenim bolj pravičnim porazdeljevanjem. To bo za prihodnjo avtomatizacijo prometa verjetno resnejši izziv kot ta, ki so ga identificirali Boneffon in njegova ekipa.

3.8 Nazaj k družbeni dilemi ... in orisu njene rešitve

Ko grobo skico, ki jo ponujajo Bonnefon et al., dopolnimo s podrobnostmi, se prične sprva oster kontrast med realnostjo in idealom/pričakovani mehčati. Kot prvo, Bonnefon in njegova ekipa so v orisu družbene dileme dvojno pretiravali – moralno junaštvo so zamešali z moralno obveznostjo (rahla pristranost do sebe je namreč v situacijah neizogibne škode moralno dopustna), obenem pa so podcenili pripravljenost ljudi na samožrtvovanje, ko jim tako obveznost po lastni presoji nalaga morala. Ljudje bodo bolj pripravljeni kupovati in uporabljati zmerno utilitaristične AV-je, torej vozila, ki zagotavljajo nekoliko več varnosti svojim potnikom, a ne na račun pravičnosti in/ali bistveno višjega tveganja za druge udeležence v prometu. Razumen kompromis med preudarnostjo oz. sebično skrbjo na eni ter moralnostjo oz. nesebičnostjo na drugi strani se glasi: »brez absolutne prioritete potnikov«, a tudi »brez absolutne prioritete za zmanjševanje škode«. Razkorak med privlačnostjo UAV-jev in ne-UAV-jev bi se dalo, skratka, dodatno zmanjšati s priznanjem legitimnosti skrbi zase in svoje bližnje ali ustrezno prilagoditvijo algoritma razumno utemeljenim pričakovanjem in sodbam ljudi. Vrzel med vozilom, ki bi mikalo večino potrošnikov, in vozilom, ki je politično in moralno gledano najboljša opcija, bi lahko dodatno premostili z vključevanjem takih kulturno specifičnih pomislekov in prednostnih nalog v algoritem, ki jih je mogoče vsaj v grobem razumno utemeljiti (glej Pözlner 2021). Družbena dilema se tako ali sploh ne bi pojavila ali pa bi bila razmeroma enostavno rešljiva.

Najresnejši očitke Bonnefonu in njegovi ekipi pa smo prihranili za konec – izvorno zastavljena dilema med zakonsko prepovedjo in svobodno potrošniško izbiro je ... lažna. Omenjeni so namreč spregledali ali zavestno zamolčali celo vrsto alternativ brezpogojni zakonski prepovedi. Države se lahko v prizadevanju po izboljšanju prometne varnosti in/ali pravičnosti zatečejo k celi paleti pravnih in političnih ukrepov, ki so blažji in zato moralno in politično manj problematični kot zakonska prepoved, ki potrošnike oropa svobodne izbire: od državnih subvencij za kupce UAV-jev do nižjega cestninskega davka. Za vlade liberalno-demokratskih držav že dolgo velja, da se družbeno nezaželenega obnašanja (kajenja in drugih škodljivih razvad, pornografije, sovražnega govora, ipd.) lotevajo s kombinacijo zakonskih, političnih, fiskalnih in izobraževalnih ukrepov. Nekateri od teh sodijo v kategorijo t. i. dregljajev (angl. *nudges*). Dregljaji so intervencije, s katerimi poskušamo na predvidljiv način vplivati na odločitve in ravnanja drugih ljudi, ne da bi bistveno omejili njihovo možnost izbire ali pa jim izbiro družbeno nezaželenih opcij zagrenili tako, da bi jo obremenili s pretiranimi stroški in tako naredili nepriljubljeno, temveč namesto tega s spremembo 'arhitekture izbire' – uokvirjenjem (angl. *framing*) ponujenih opcij, prikrajanjem njihovega vrstnega reda, določitvijo osnovne in dodatnih opcij, ipd.¹⁷ (Thaler in Sunstein 2008) V zadnjem času se tovrstne intervencije množijo v vladnih kampanjah za krepitev javnega zdravja in spodbujanje zdravega življenjskega sloga, čedalje pogosteje pa jih je mogoče zaslediti tudi v prizadevanjih po dvigu prometne varnosti (merilci hitrosti ob cestah, zvočni signal, ki nas v avtu opozarja, da nismo pripeti, razni dodatki za varno vožnjo, ipd.). Država lahko, skratka, potrošnike, ki so do nakupa UAV-jev morda zadržani, k tej družbeno (in navsezadnje tudi individualno) zaželeni odločitvi spodbudi na različne načine, od moralnega uokvirjenja izbire med UAV-ji in ne-UAV-ji preko izbire UAV-ja kot osnovne (in ne-UAV-jev kot dodatne) nastavitve do obremenitve nakupa in uporabe ne-UAV-jev z materialnimi in nematerialnimi stroški, med slednjimi tudi, zakaj pa ne, z družbeno stigmo.

Ker dregljaji, kadar seveda delujejo, svojo učinkovitost dolgujejo znanim človeškim iracionalnostim (pristranskostim, heuristikam, nevednosti, lenobi ipd.), bi se država na ta način izpostavila očitku, da s tovrstnimi pozitivnimi spodbudami za nakup UAV-jev in negativnimi spodbudami proti nakupu ne-UAV-jev nedopustno manipulira s svojimi državljani. A takega dobronamernega državnega paternalizma

¹⁷ Locus classicus je seveda knjiga Richarda Thalerja in Cassa Sunsteina *Dregljaji: kako izboljšati naše odločitve o zdravlju, premoženju in sreči* iz l. 2008.

v celoti gledano ne bo težko moralno opravičiti, če bodo le dregljaji a) transparentni, ne prikriti, b) naslovljeni primarno na sistem 2 (zavestne procese tehtanja in premlevanja razlogov) in šele sekundarno, če sploh, na sistem 1 (avtomatske, podzavestne procese), ter končno c) namenjeni opolnomočenju posameznikov in ne podtikanju tuje jim volje. Spodbujanje nakupa in/ali rabe UAV-jev – oz. sočasno odvrčanje od nakupa in/ali rabe ne-UAV-jev – bi tako postalo lep vzgled za moralno opravičljivo dreganje, saj država ne bi na moralno sporen način zaobšla ali spodkopala potrošniške svobode in/ali avtonomije; nasprotno, s tem, ko bi prilagodili potrošniško ponudbo njihovim pristnim (v raziskavah razkritim) moralnim vrednotam, bi po svoje zavarovali njihovo ogroženo avtonomijo.

4 Namesto zaključka

V pričujočem prispevku sem dokazoval, da obstajajo tehtni razlogi, da v prihodnje AV-je vgradimo zmerno – pravičnosti in rahli pristranskosti do sebe prilagojeno – utilitaristično načelo za porazdelitev neizogibne škode. Če imajo Bonnefon in njegova ekipa prav, pa država takih algoritmov kupcem AV-jev ne bo smela vsiljevati z zakonom, temveč bo z vidika izboljšanja prometne varnosti in pravičnosti smotrnejše, da proizvodnjo in prodajo UAV spodbuja z mehkejšimi prijemi, od nakupom utilitarističnih AV-jev naklonjene davčne, cenovne in zavarovalne politike preko oglaševalskih kampanj do blagih dreglajev oz. premišljene, a hkrati transparentne arhitekture potrošniških izbir.

Viri in literatura

- Awad, E., Dsouza, S., Kim, R. et al. (2018). »The Moral Machine experiment«. *Nature*, 563, str. 59–64.
- Bergmann L.T. et al. (2018). »Autonomous Vehicles Require Socio-Political Acceptance—An Empirical and Philosophical Perspective on the Problem of Moral Decision Making«. *Frontiers in Behavioral Neuroscience*, 12, 31.
- Bonnefon, J., Shariff, A. in Rahwan, I. (2016). »The Social Dilemma of Autonomous Vehicles«. *Science*, 352(6293), str. 1573–1576.
- Carter, I. (2021). »Positive and negative liberty«. V Zalta, E. N. (ur.), *Stanford Encyclopedia of Philosophy* (izdaja pomlad 2022). URL = <https://plato.stanford.edu/archives/spr2022/entries/liberty-positive-negative/>.
- Friedman, M. (1970). »The social responsibility of business is to increase its profits«. *The New York Times*, 13. september 1970.
- Furey, H. in Hill, S. (2021). »MIT's moral machine project is a psychological roadblock to self-driving cars«. *AI Ethics*, 1, str. 151–155.
- Giubilini, A in Savulescu, J. (2019). »Vaccination, risks, and freedom. The seat belt analogy«. *Public Health Ethics*, 12(3), str. 237–249.

- Greene, J. (2016). »Our driverless dilemma«. *Science*, 352(6293), str. 1514–1515.
- Hübner, D. in White, L. (2018). »Crash Algorithms for Autonomous Cars: How the Trolley Problem Can Move Us Beyond Harm Minimisation«. *Ethical Theory and Moral Practice*, 21, str. 685–698.
- Hunt, L. H. (2014). »On improving people by political means«. V: LaFollette, H. (ur.), *Ethics in Practice. An Anthology* (4. izdaja). Oxford: Wiley Blackwell, str. 299–308.
- Kamm, F. M. (1985). »Equal treatment and equal chances«. *Philosophy & Public Affairs*, 14(2), str. 177–194.
- Kauppinen, A. (2019). »Who's afraid of Trolleys?«. V Suikkanen, J. in Kauppinen, A. (urd.), *Methodology and Moral Philosophy*. London: Routledge, str. 49–71.
- Kauppinen, A. (2021). »Who should bear the risk when self-driving vehicles crash?«. *Journal of Applied Philosophy*, 38(4), str. 630–645.
- Keeling, G. (2020). »Why Trolley Problems Matter for the Ethics of Automated Vehicles«. *Science and Engineering Ethics*, 26(1), str. 293–307.
- Köbis, N., Bonnefon, JF. in Rahwan, I. (2021). »Bad machines corrupt good morals«. *Natural Human Behavior*, 5, str. 679–685.
- Lin, P. (2013). »The ethics of autonomous cars«. *The Atlantic*, 8. oktober 2013. URL = <https://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/>.
- Parfit, D. (1978). »Innumerate ethics«. *Philosophy & Public Affairs*, 7(4), str. 285–301.
- Poročilo skupine A.T. Kearney. n.d. »How automakers can survive the self-driving era«. URL: <https://www.es.kearney.com/automotive/article?/a/how-automakers-can-survive-the-self-driving-era>.
- Pözlner, T. (2021). »The Relativistic Car: Applying Metaethics to the Debate about Self-Driving Vehicles«. *Ethical Theory and Moral Practice*, 24, str. 833–850.
- Pugh, J. et al. (2021). »Learning to live with Covid-19 – the tough choices ahead«. *The Conversation*, 16. april 2021. URL = <https://theconversation.com/learning-to-live-with-covid-the-tough-choices-ahead-158992>.
- Rinni, R. (2017). »Raising good robots«. *AEON*, 19. april 2017. URL = <https://aeon.co/essays/creatingrobotscapableofmoralreasoningislikeparenting>.
- Scanlon, T. (1998). *What Do We Owe To Each Other*. Cambridge, Mass.: Harvard University Press.
- Schmidt, A. (2016). »Withdrawing vs. withholding freedoms: the case of tobacco and nudging«. *The American Journal of Bioethics*, 16(7), str. 3–14.
- Taurek, J. (1977). »Should the numbers count?«. *Philosophy & Public Affairs*, 6(4), str. 293–316.
- Thaler, R. in Sunstein, C. (2008). *Nudge. Improving Decisions About Health, Wealth and Happiness*. New Haven in London: Yale University Press.
- Thomson, J. (1986). »Some questions about government regulation of behaviour«. V *Rights, Restitution & Risk. Essays in Moral Theory*. Cambridge, Mass. in London: Harvard University Press; str. 154–172.
- Wolff, J. (2011). *Ethics and Public Policy. A Philosophical Inquiry*. London in New York: Routledge.