

ANαliZA

3-4

2017 letnik 21.

AN_αliZA, časopis za kritično misel, številka 3-4, leto XXI., 2017

Uredila Olga Markič in Toma Strle

Izdajatelj in založnik: Društvo za analitično filozofijo in filozofijo znanosti,
Aškerčeva 2, 1000 Ljubljana

<http://www.drustvo-daf.si/analiza/>

e-naslov: analiza@drustvo-daf.si

Naročila sprejemamo na zgornji naslov.

Na leto izidejo trije zvezki Analize (dve enojni in ena dvojna številka).

Cena zvezka v prosti prodaji je 6,00 EUR, za naročnike 4,00 EUR oz. letna naročnina 12,00 EUR (z vključenim DDV).

Izdajateljski svet:

Bojan Borstner, David Owens, Matjaž Potrč, Marko Uršič,
Matthias Varga von Kibéd, Elizabeth Valentine, Bojan Žalec

Glavna in odgovorna urednica: Maja Malec

Uredniški odbor:

Smiljana Gartner, Božidar Kante, Vojko Strahovnik, Toma Strle,
Borut Trpin, Sebastjan Vörös

Lektor: Borut Cerkovnik

Oblikovanje: K.reg@

Tisk: Demat, d.o.o., Stegne 3, 1000 Ljubljana

Naklada: 200 izvodov

Izid te revije je sofinancirala

JAVNA AGENCIJA ZA KNJIGO REPUBLIKE SLOVENIJE.

Revija Analiza je indeksirana v THE PHILOSOPHER'S INDEX.

ISSN 1408-2969

Kazalo

Filozofski, družbeni in etični vidiki sodobnih kognitivnih tehnologij

Franc Mali Razvoj novih tehnologij kot izziv sodobnim bioetičnim pogledom na človeka in svet	5
Dejan Savič Etični pomisleki o filtriranju vsebin na družbenih omrežjih	19
Meta Kokalj in Anže Mihelič Uvajanje eksperimentalnih kognitivnih tehnologij v terapično in družbeno prakso	35
Tjaša Zapašek Umetna inteligenca in njena pravna ureditev	55
Toma Strle in Olga Markič Neuroznanost in pravo: od redukcionizma do krožnosti nevroprava	77
Zvezdan Pirtošek Možgani, nevrotehnologija in etika	91
Andrej Ule Wittgensteinska kritika zamisli o mislečih strojih	109

Wittgensteinove *Filozofske raziskave*

Urh Vele Forma v službi metode: kako brati Wittgensteinove <i>Filozofske raziskave</i>	123
--	-----

Na kratko o svobodi govora

Borut Ošljaj Svoboda svobode govora	141
---	-----

Franc Mali

Fakulteta za družbene vede Univerze v Ljubljani

Razvoj novih tehnologij kot izziv sodobnim bioetičnim pogledom na človeka in svet

Prispevek se ukvarja z nekaterimi izzivi, ki jih v sodobno bioetiko prinaša razvoj novih tehnologij krepitve človekovih umskih in telesnih zmožnosti. Poudarek je na institucionalnem delu bioetike, ki ga v evropskem prostoru predstavljajo različna bioetična strokovna telesa. V prispevku je predstavljen razvoj nacionalnih etičnih komisij, ki se v svojem delovanju srečujejo z vprašanji razvoja novih tehnologij. Sočasno se v prispevku išče odgovor na vprašanje, zakaj ta institucionalni del bioetike v prostoru Srednje in Vzhodne Evrope, kamor spada tudi Slovenija, ni uspel preseči tradicionalnih pogledov na pomembne (bio)etične izzive današnjega časa. Avtor prispevka ocenjuje, da bi nadaljnjemu utiranju poti v smeri bolj progresivnih bioetičnih pogledov na svet in človeka v tem prostoru najbolj škodile ideološko pogojene kontroverze glede razvoja novih tehnologij, ki so ponekod tudi že prerasle v novo obliko »kulturne vojne«.

Ključne besede: bioetika, nove tehnologije krepitve človekovih umskih in telesnih zmožnosti, človeška kognicija, nacionalni etični sveti, kulturni boj v bioetiki

Nujnost preseganja normativizma in disciplinarnih omejitev v sodobni bioetiki

Živimo v času hitrega razvoja naprednih tehnologij, ki močno vplivajo na življenje ljudi in celotno družbeno okolje. Medsebojno povezovanje nano-, info-, bio- in kognitivnih znanosti vodi k številnim družbenokoristnim uporabam rezultatov znanstvenih raziskovanj. Te koristne uporabe novih naprednih znanosti in tehnologij segajo praktično na vsa področja človekovega in družbenega življenja. Glede na današnji izredno hiter razvoj znanosti in tehnologije je upravičeno pričakovanje, da so vedno večji dinamiki podvrženi tudi pogledi moderne progresivne bioetike na človeka in svet. Treba je reči, da se je bioetika kot posebno področje profesionalnega zanimanja pojavila relativno pozno. Pojem »bioetika« je prvič uporabil šele leta 1970 ameriški biolog in onkolog Van Rensselaar Potter (glej več: ten Have, 2012).

Seveda so vsebinske razprave okrog tem, ki zadevajo bioetiko, prisotne skozi celotno človeško zgodovino. V zadnjih dveh desetletjih se bioetika nahaja, ravno zaradi izrednega napredka novih tehnologij krepitve človekovih umskih in telesnih zmožnosti, pred vrsto novih izzivov. Fumikazu Shibasaki pravi, da

/.../ je v središču problemov, s katerimi se danes sooča bioetika, dejstvo, da napredek tehnologije omogoča neomejeno manipulacijo s človeškim življenjem. V preteklosti so se lahko vsi pojavili, ki so bili povezani s človekovim življenjem, zgolj prepustili silam »narave«. Vendar je bilo skozi čas mogoče z genskimi manipulacijami ustvariti takšne forme življenja, ki, če se tako izrazimo, niso več del izvirnega »programa« narave. (Shibasaki, 2005: 488).

Na področje teh genskih manipulacij lahko uvrstimo številne vrste (tehnoloških) posegov, ki spreminjajo človekovo kognicijo, tako v okviru biomedicinskih zdravljenj kot tudi krepitve človekovih umskih zmožnosti. Raziskovalni projekt »*Brain Research Through Advancing Innovative Neurotechnologies*«, ki ga že od leta 2014 finančno močno podpira ameriška Nacionalna agencija za zdravje (NIH), je definirala tri temeljna področja preučevanja človeških možganov: področje vmesnikov med možgani in računalniki, nevrološke stimulacije ter terapijo z nevrološkimi matičnimi celicami. Etične dileme, ki so povezane s posegi v človeško kognicijo, se v okviru omenjenega projekta zelo poudarjajo, saj se podporniki tega projekta zavedajo, da brez širše javne diskusije družbeno-etičnih vprašanj nadaljnji razvoj kognitivnih znanosti ni mogoč (McCullagh et al., 2014). Številne etične dileme se še zlasti pojavljajo zaradi možnosti izvajanja različnih genskih terapij na področju človeške kognicije. Nahajamo se – če uporabim izraz Petra Singerja – v naraščajoči spirali etičnih dilem, ki se širi na vse strani (od posameznika na družbo, naravo itd.). Najnovejši hit biogenetskega razvoja predstavlja CRISPR/Cas9 tehnika na področju sintezne biologije,¹ ki bi lahko privedla do sprememb genskih značilnosti vseh naslednjih človeških generacij.

V današnjih pogojih delovanja, ki so povezani s silnim napredkom novih tehnologij, je za bioetiko, ki postaja neke vrste globalno »podjetje«, izredno pomembno, da zmore preseči deficit, ki izvira iz njene morebitne prevelike nagnjenosti k normativizmu in disciplinarnim omejitvam. Kar zadeva prvi vidik, gre pri bioetiki za nujnost iskanja večjega ravnovesja med teoretskimi in praktičnimi odgovori na pereče družbene in etične izzive razvoja novih naprednih tehnologij. Kar zadeva drugi vidik, gre za zahtevo po bolj interdisciplinarnem pristopu k bioetičnim vprašanjem. Moderna bioetika se lahko družbeno legitimira edinole kot interdisciplinarno zasnovano področje vedenja. Njena tematska širitev je v zadnjem času

¹ Sintezni biologi so začeli uporabljati sistem CRISPR/Cas9 za spreminjanje človeške genske DNA v različnih celičnih linijah. Na najbolj preprost in poenostavljen način bi sistem CRISPR/Cas9 lahko označili kot neke vrste molekularnih škarij, ki selektivno odstrižejo nezaželen del genoma in ga nadomestijo z novimi DNK-a strukturami. (EASAC, 2017).

povezana z ustanavljanjem novih področij, kot so nevro- in nano-etika. Pred moderno bioetiko se postavljajo zahteve po iskanju odgovorov na vrsto povsem novih družbeno-etičnih izzivov. Ko se je bioetika v sedemdesetih letih prejšnjega stoletja pojavila kot samostojna disciplina, je bila skoraj izključno usmerjena k vprašanjem, ki so zadevala človekovo zaznavanje in osmišljanje zdravja in bolezni. Eden izmed njenih poudarkov je bil reševanje problema dehumanizacije medicine, ki je nastopil zaradi neprimernih zdravniških praks (Coenen et al., 2011). Z razvojem novih tehnologij, zlasti tistih tehnologij, ki se ukvarjajo z vprašanjem krepitve človekovih telesnih in umskih zmožnosti, gre za postopno preusmeritev od bolj osebnih medicinskih tem k širšim družbeno-etičnim vprašanjem. Res je sicer, da so se že tudi v prvih fazah razvoja bioetike kot samostojne discipline pojavile tudi bolj dolgoročne vizije glede možnih etičnih posledic znanstvenega in tehnološkega razvoja. Nemški filozof Hans Jonas je že v sedemdesetih letih prejšnjega stoletja podal poglobljeno refleksijo o možnih katastrofalnih posledicah za človeško družbo, kolikor bi se napredek novih tehnologij dogajal zunaj vsakršne kontrole (Jonas, 1979). Vendar so bili takšni celostni pogledi bolj izjema kot pravilo.

Danes se vprašanja bioetike povezujejo vedno bolj z vprašanji etike znanosti. Moderna bioetika mora preseči meje individualne etike in prevzeti funkcijo družbeno odgovorne etike. Družbeno odgovorna etika znanosti mora vključiti poglobljeno refleksije o prednostih in (vseh možnih) tveganjih bodočega razvoja novih tehnologij in se zavedati svoje lastne dinamike in spremenljivosti. Naj to na kratko ilustriramo s primerom. Nobelovec James Watson, eden od izumiteljev DNK-ja in direktor velikega projekta »*Human Genom Project*«, je bil eden prvih, ki se je leta 2001 neposredno zavzel za nov (bio)etični pristop v okviru velikih projektov biogenetskih znanosti. S tem ko je zahteval, da se del proračuna nameni preučevanju družbenih vidikov razvoja biogenetske znanosti, je odprl prostor – če uporabim mednarodno sprejeto kratico ELSI² – »elsifikaciji« znanosti (Penders et al., 2008). Od takrat naprej so vodilni akterji na področju znanstvene in tehnološke politike, ki skrbijo za institucionalne okvire ter financiranje raziskovanja in razvoja, začeli izražati pričakovanje, da predlogi projektov s področja novih bioznanosti vključujejo vsebine, ki zadevajo etične, pravne in družbene vidike njihovega razvoja (Zwart in Nelis, 2009). Dobro desetletje za vključitvijo ELSI vprašanj v raziskovanje biogenetike sta Paul Rabinow in Gaymon Bennett že pozvala k spremembi »koncepta ELSI«, saj, kot sta trdila, sodobni postgenomski raziskovalni programi, kakršnega predstavlja sintezna biologija, ne morejo več temeljiti na povsem enakem tipu družbenih, etičnih in pravnih razmislekov, ki so veljali še kakšno desetletje nazaj (Rabinow in Bennett, 2012). Četudi ne moremo pričakovati, da bi eksperti, ki delujejo na področju bioetike, prišli do neke »konsenzualne

² Kratica ELSI se v angleščini nanaša na etične, pravne in družbene implikacije razvoja naravoslovnih in tehničnih znanosti.

etične platforme« (Stehr, 2003: 202), se vseeno zdi, da se jih vedno več zaveda, kako pomembno in družbeno odgovorno je spremljanje razvoja novih naprednih tehnologij.

V nadaljevanju naše razprave bomo skušali odgovoriti na vprašanje, zakaj je institucionalni okvir modernega bioetičnega diskurza še posebej povezan z etiko znanosti. Hkrati nas bo zanimalo vprašanje, zakaj ti procesi institucionalizacije ponekod še vedno potekajo zelo počasi. V različnih bioetičnih telesih Vzhodne in Srednje Evrope, kamor spada tudi Slovenija, je še vedno premočan poudarek zgolj na tradicionalnih bioetičnih temah. Premalo je zanimanja za etična in družbena vprašanja razvoja novih tehnologij. Zato se lahko v celoti strinjamo z oceno Angusa Dawsona izpred nekaj let, ki pravi, da se v Vzhodni in Srednji Evropi bioetična vprašanja v glavnem še vedno obravnavajo izključno kot vprašanja medicinske etike, kot da bi pojem bioetike in medicinske etike pomenila eno in isto, četudi »je bioetika dejansko bistveno širša po svojem področju delovanja kot medicinska etika« (Dawson, 2010: 218).

Procesi institucionalizacije modernega bioetičnega diskurza: nacionalni bioetični sveti kot svetovalna telesa političnim odločevalcem in kot (še vedno neuresničeni) prostor dialoga s civilno družbo

Dejali smo, da mora sodobni pojem bioetike razvijati čim širšo družbeno platformo. Družbeni značaj bioetike je vezan na njene filozofske začetke. Idejo o podpori bioetike za človeka je zagovarjal že Kant, idejo o koristnosti so zagovarjali filozofi utilitarizma, idejo o pravičnosti Aristotel (Baron, 2006). Pojem bioetike lahko povezujemo z akademskim značajem te dejavnosti ali pa z njeno praktično naravnostjo. V prvem primeru naj bi šlo bolj za formulacijo etičnih kodeksov, ki se pojavljajo v okviru moralne filozofije. V drugem primeru naj bi šlo bolj za prizadevanje strokovno usposobljenih posameznikov oziroma institucij pri pomoči ljudem, ki se težko odločajo glede vprašanj, ki zadevajo medicinska vprašanja. Predhodne delitve zvenijo nekoliko zastarelo. Uvodoma smo namreč že dejali, da nove etike znanosti, kot sta nevroetika in nanoetika, nastopajo bolj kot področja širitve bioetike, ne pa kot področja, ki se razvijajo neodvisno od bioetike. V polje bioetike danes uvrščamo tudi okoljsko etiko. V okviru različnih klasifikacij je zanimiva klasifikacija Raanana Gillona, ki deli bioetiko na (1) teoretsko bioetiko, ki se ukvarja z idejnimi temelji te vede, (2) klinično bioetiko, ki se ukvarja z vprašanji, ki nastanejo kot posledica odnosov med pacienti in temi, ki skrbijo za njihovo zdravje, (3) »policy« etiko, ki se ukvarja z regulativnimi in pravnimi vidiki, in (4) kulturno bioetiko, ki skuša bioetična vprašanja povezati glede na zgodovinski in kulturni kontekst (Gillon, 2014).

Kljub temu da se včasih pojavljajo pomisleki, češ, da pretirani poudarek na institucionalne okvire delovanja moderne bioetike ogroža subtilnost filozofskih mo-

ralnih refleksij, pa »empirični premik« bioetike in njeni procesi institucionalizacije prispevajo k njenemu razvoju in večji družbeni relevantnosti. Še zlasti, ko govorimo o vlogi bioetike pri (vrednotni) presoji možnih družbenih posledic razvoja novih naprednih tehnologij.

Do bolj utrjenih institucionalnih okvirov bioetičnih razmislekov je prišlo v Evropi z ustanovitvijo različnih vrst strokovnih bioetičnih svetov, komisij in drugih organov na različnih nivojih njihovega delovanja. Od samega začetka oblikovanja prvih bioetičnih teles koncem sedemdesetih let prejšnjega stoletja je obstaja velika heterogenost v njihovem poslanstvu in načinih delovanja. Ta heterogenost je bila v veliki meri odvisna tudi od števila bioetičnih komisij, ki so delovale v posameznih evropskih državah (Druml et al., 2009).

Z vidika bioetičnih razmislekov o razvoju novih naprednih tehnologij, kar nas v naši razpravi posebej zanima, je bila najpomembnejša ustanovitev nacionalnih etičnih svetov (ang. *National Ethical Councils*). V Zahodni Evropi so bile te vrste ekspertnih teles ustanovljene že v začetku osemdesetih let. Od tod so se potem širila tudi v druge dele Evrope (Bogner, 2010; Fuchs, 2005).

Če gledamo na razvoj nacionalnih etičnih svetov iz neke širše zgodovinske perspektive, potem je treba poudariti, da so se v svojih začetnih fazah ustanavljanja pogosto oprli na modele delovanja predhodno ustanovljenih uradov za ocenjevanje tehnološkega razvoja. Slednji so nastopali kot svetovalna telesa parlamentov. Njihovi ustanovitvi je botroval optimizem parlamentarnih politik, da prihodnji razvoj tehnologij lahko predvidimo na relativno enostavnih in natančno izmerljivih parametrih (glej več: Delvenne et al, 2011; Weingart, 2001). V tej prvi fazi ustanavljanja bioetičnih teles na nacionalni ravni so bili njegovi člani predvsem naravoslovci in bioetiki v ožjem pomenu besede. Tem zadnjim je bila podeljena neke vrste pravica oziroma dolžnost, da razpravljajo o fundamentalnih (bio)etičnih vprašanjih. Pomemben premik v prvi fazi oblikovanja nacionalnih etičnih svetov se je zgodil, ko se je spremenil njihov status, tj. njihova sprememba iz začasnih v trajna ekspertna telesa, ki svetujejo vladam, parlamentom in drugim političnim institucijam. Ilustrativen primer je britanski Warnock etični svet. Mary Warnock, ki je spodbudila nastanek omenjenega sveta (po njej ima ta svet sedaj tudi ime), si je od vsega začetka njegove ustanovitve prizadevala, da dobi status stalnega svetovalnega telesa, saj je menila, da je le tako mogoče vzpostaviti kontinuirano spremljanje etičnih dilem razvoja novih naprednih tehnologij (glej več: Wilson, 2011).

Nekateri avtorji, ki so preučevali nastanek in kasnejši razvoj nacionalnih etičnih svetov, so celo trdili, da so bioetična vprašanja postala odločilni element pri družbenem legitimiranju političnih odločitev (Braun in Kropp, 2010). Menim, da še nismo v tej (demokratični) fazi razvoja bioetike, saj bi v tem primeru omenjeni sveti vodili dialog s širšo javnostjo in jo tudi vključevati v procese sprejemanja strokovnih odločitev. Žal se redko srečujemo s takšnimi praksami.

Zavezanost institucionalnega okvira bioetike v prostoru Srednje in Vzhodne Evrope tradicionalnim temam in njegovo (ne)zanimanje za vprašanja razvoja novih tehnologij

Današnji bioetični diskurz v državah Srednje in Vzhodne Evrope, pri čemer Slovenija ni izvzeta, je še vedno preveč okupiran s tradicionalnimi bioetičnimi temami (npr. vprašanje abortusa, evtanazije itd.) in tudi dokaj konzervativnimi pogledi. To se kaže tudi v premajhnem zanimanju za nove družbene in etične izzive razvoja novih naprednih tehnologij. Obstajajo različni razlogi za takšno stanje. Mednje je mogoče še vedno šteti dediščino politične kulture, ki izvira iz polpreteklega obdobja. V času komunistične družbene ureditve so bila bioetična vprašanja obravnavana kot družbeno irelevantna ali pa so bila podrejena pretiranim ideološkim interpretacijam. V večini držav Vzhodne in Srednje Evrope pred političnimi spremembami v začetku devetdesetih let ni obstajala institucionalna mreža bioetičnih komisij. Bioetične teme tudi niso bile sprejete kot družbeno relevantne. Kot ugotavljajo številni avtorji, prevlada paternalizma v biomedicinskih krogih in prevlada vulgarnega scientizma v naravoslovnih znanostih nista ustvarjala pogojev za odpiranje kompleksnih bioetičnih vprašanj (glej več: Glasa, 2002; McKee, 2004; Prodanov, 2001). Hierarhična in paternalistična kultura na področje biomedicine ni priznavala individualnih pravic pacientov, njihove pravice do t. i. obveščenega soglasja³ itd. V teh družbeno političnih okoliščinah je bil ob kolektivno utilitaristični interpretaciji bioetike dan prostor samo še konzervativnim bioetičnim pogledom uradne cerkve. Ravno ta je v času politične tranzicije začela prevzemati ključno vlogo v interpretaciji kompleksnih bioetičnih vprašanj in na ta način predstavlja bistveno večjo oviro za razvoj bolj progresivnega bioetičnega diskurza v tem delu Evrope (Jennings, 2011). Pod progresivnim bioetičnim diskurzom razumem v prvi vrsti takšen pristop bioetičnih teoretikov in praktikov, ki izhaja iz liberalnih vrednot, zavezanosti načelom transparentnega delovanja, poudarku na strokovnosti in ne ideologiji ter prizadevanjem za doseganje socialne pravičnosti, ko gre za uživanje sadov znanstveno-tehnološkega napredka.

Tudi na temelju lastnih socioloških empiričnih raziskovanj, ki so sicer nekoliko starejšega datuma, vendar glede na spoznanja še vedno aktualna, smo prišli do nekaterih zanimivih ugotovitev, ki govorijo v prid tezi, da se v državah Srednje in Vzhodne Evrope srečujemo z relativno večjimi težavami pri institucionalizaciji progresivnega bioetičnega diskurza, kot je to primer v državah Zahodne Evrope. V okviru projekta »*Ethics in public Policy-making: the Case of Human enhancement*«, ki ga je finančno podprla Evropska komisija in je bil izveden še v Sedmem

³ Resnici na ljubo je treba reči, da pravica obveščenega soglasja pacientov predstavlja temo, ki je bila zanemarjena v vseh političnih sistemih, ne glede na njihov ideološki predznak. Šele v zadnjem času v bioetične razmisleke v večji meri vstopa obravnava tega pomembnega načela. S tem vprašanjem se je Evropa začela ukvarjati bolj intenzivno šele leta 1997, v času priprave in sprejema Konvencije o človekovih pravicah in biomedicini, ki je bolj znana pod imenom Oviedo konvencija.

okvirnem programu (2009–2012), smo raziskovalci iz Centra za preučevanje znanosti na Fakulteti za družbene vede Univerze v Ljubljani z različnimi empiričnimi pristopi (on-line anketiranje med predsedniki oziroma tajniki bioetičnih komisij, kvalitativna vsebinska analiza »policy« dokumentov, ki so jih pripravile bioetične komisije itd.) preučevali transparentnost delovanja nacionalnih etičnih svetov v Evropi. Da bi ugotovili, kakšna je javna dostopnost informacij o delovanju nacionalnih etičnih svetov, smo med drugim opravili natančen pregled uradnih spletnih strani teh svetov v 32 evropskih državah (glej več: Mali et al., 2012). Naš pregled spletnih strani se je nanašal na razpoložljive informacije o članstvu teh svetov, opisu njihovega poslanstva, objavljenih in on-line dostopnih gradivih itd. Naša natančna analiza spletnih strani nacionalnih etičnih svetov je pokazala, da se z dokaj izpopolnjenimi informacijami o svojem delovanju ponašajo etični sveti iz Zahodne Evrope, precej bolj netransparentna, kar zadeva javno podobo in javni angažma, je bila situacija pri nacionalnih etičnih svetih v državah Srednje in Vzhodne Evrope. Tudi druge pred tem opravljene empirične študije so prišle do podobnih rezultatov kot naša analiza (glej več: Dranseika et al., 2010; Ahvenharju et al., 2006; Fuchs, 2005). Večjo težavo seveda predstavljajo neustrezni kadrovske sestavi teh svetov (npr. neuravnotežen članstvo različnih profesionalnih in disciplinarnih profilov, neuravnoteženost glede zastopanosti ljudi z različnimi nazor-skimi pogledi itd.) in pogosto prisoten odklonilni odnos do bolj odprtega dialoga z javnostjo, kar vodi v birokratizacijo in netransparentnost delovanja nacionalnih etičnih svetov (Pustovrh in Mali, 2015; Dranseika et al., 2010).

Pričakovali bi, da bodo nacionalne bioetične komisije v državah Srednje in Vzhodne Evrope, ki so postale članice EU, prišle pod bolj neposredni vpliv smernic na področju bioetike, ki jih postavlja Evropska komisija. Vendar ta pričakovanja niso vedno najbolj realna, saj kot ugotavljajo posamezne študije, lahko številne brusselske direktive pogosto zgolj vzbujajo vtis enotnosti (Hedgcock et al., 2006). Glede tega vprašanja svojih pričakovanj žal ne izpolnjuje tudi Evropska komisija za etiko znanosti in nove tehnologije (EGE). EGE je neodvisno in multidisciplinarno strokovno telo, ki svetuje Evropski komisiji v zvezi z etičnimi, družbenimi in pravnimi vidiki razvoja novih naprednih tehnologij (glej več: EGE, 2017). Evropski komisiji svetuje že od leta 1991. Takrat je bilo ustanovljeno strokovno telo, ki se je ukvarjalo z etičnimi posledicami razvoja biotehnologije (GAEIB). Leta 1997 je Evropska komisija sklenila GAEIB zamenjati z Evropsko skupino za etiko v znanosti in novih tehnologijah (EGE) in razširiti mandat skupine, da bi pokrival vsa področja uporabe znanosti in tehnologije.

EGE ima status (bio)etične komisije na evropski ravni, torej deluje predvsem kot svetovalno telo Evropske komisije. V obdobju od 2010 do 2016 je pripravila vrsto dokumentov, ki se ukvarjajo z etičnimi vprašanji razvoja novih tehnologij zdravljenja (EGE, 2015), tehnologij za varnost in nadzor (EGE, 2014), informacijskih in komunikacijskih tehnologij (EGE, 2012) itd. Trenutno pripravlja mnenje o vplivu robotike in drugih naprednih tehnologij na prihodnost delovnih mest.

Vendar v okviru EGE-ja deluje tudi Forum nacionalnih etičnih svetov za znanost (ang. *Forum of National Ethics Committees*), ki ga je Evropska komisija ustanovila leta 2003, njegov primarni namen pa je koordiniranje dela nacionalnih etičnih svetov članic Evropske Unije. Eden izmed pomembnih ciljev tega foruma je izmenjava in širjenje informacij, izkušenj in dobrih praks med članicami EU. Četudi pri uresničevanju tega cilja forum ni vedno najbolj uspešen, vseeno menim, da je njegov obstoj pomemben že na simbolni ravni. Vsaj na načelni ravni spodbuja zavedanje skupne evropske identitete na področju današnjih kompleksnih bioetičnih vprašanj.

Moderni progresivni bioetični diskurz ne potrebuje ideoloških sporov o vprašanih razvoja novih tehnologij

V tranzicijskih državah Srednje in Vzhodne Evrope, za katere ugotavljamo, da zaostajajo v uvajanju bolj progresivnega bioetičnega diskurza, bi bilo družbeno skrajno neproduktivno, če bi se bioetične diskusije še bolj zapletale v ideološko pogojene spore, kot smo jim neredko priča že sedaj. To velja še posebej pri vrednotnih ocenah razvoja novih tehnologij krepitve človekovih umskih in telesnih zmožnosti, saj se, kot so ugotavljali udeleženci enega izmed forumov, ki je potekal na to temo, ideološki spori, največkrat zaradi popolnega ignoriranja empiričnih dejstev, pogosto spremenijo v neke vrste »fantomsko razpravo« (Lucke et al., 2010). Tudi v prostoru Srednje in Vzhodne Evrope potrebujemo takšno bioetiko, ki bo našla ustrezno ravnovesje med »proakcijskimi« (ang. *proactionary*) in »svariteljskimi« (ang. *precautionary*) načeli (Fuller, 2014) in ki ne bo zavračala razvoja novih tehnologij zaradi nekih ideološko pogojenih »tehnofobij«, temveč bo zmožna v razvoju teh tehnologij videti sredstvo za izboljševanje naravnih in družbenih pogojev človekovega življenja (Sorgner, 2016).

Z ideološkimi spori na področju bioetike, ki so prerasli v zagrizen kulturni boj, so se na primer soočali v ZDA v začetku tega stoletja. Takrat je bioetična komisija, ki jo je po nastopu svojega mandata ustanovil ameriški predsednik George Bush mlajši, vstopila v javno razpravo o vprašanju nadaljnjega raziskovanja matičnih celic in tudi razvoja vseh ostalih biogenetskih tehnologij s skrajno konzervativnimi stališči. Članstvo Bushove bioetične komisije so tvorili konzervativci, ki so vztrajali, da se je treba lotiti etičnih presoj o tehnologijah krepitve človekovih telesnih in umskih sposobnosti neposredno, glede na naravo njihovih posegov, ne da bi te razmisleke vstavili v nek širši bioetični okvir (Briggle, 2009; Berger in Moreno, 2010). Takšen pristop je vodil v skrajno ideološka zavračanja napredka na področju biomedicinskih znanosti in še bolj krepil vrednotne spopade med »tehnoprogresivisti in tehnokonzervativci«. (Hughes, 2006: 288)

Allen Buchanan je kritiziral bioetične poglede tehnokonzervativcev na razvoj biomedicinskih tehnologij izboljševanja človekovih umskih in telesnih zmožnosti.

Zanje je značilno, da so namesto racionalnih argumentov uporabljali prazno retoriko. V svoji analizi je prišel do zaključka, da so se ameriški konzervativci na področju bioetike, ki so nastopali zoper napredek biomedicinske znanosti, običajno zatekali k trem tipom ideoloških dogem. Po prvem dogmatskem pogledu je narava človeka po svojem bistvu nespremenljiva, saj je stvaritev boga oziroma božje previdnosti. Po drugem dogmatskem pogledu je človeška narava striktno in trajno omejena glede možnosti, da bi se je dalo izboljševati po družbeni poti (kognitivne pomanjkljivosti naj bi bile najboljši dokaz človekovih neuspešnih prizadevanj za odpravo naših naravnih omejitev). Po tretjem dogmatskem pogledu naj bi v primeru naših prizadevanj po kakršnihkoli spremembah nespremenljive narave človeških bitij zgolj povzročili škodo človeškemu življenju (glej več: Buchanan, 2011: 145-152).

Strinjam se z oceno Allena Buchanana (2011), da takšna dogmatska moraliziranja niso dorasla družbenim izzivom razvoja novih, zlasti biogenetskih, tehnologij. Še posebej na področju biogenetike ni mogoče postavljati apriornih delitev na posege, ki se nanašajo na zdravstveno terapijo, in na posege, ki presegajo meje te terapije. Kakršenkoli poskus, da bi se vnaprej vzpostavilo strogo delitev, je obsojen na neuspeh. Smo v podobnem položaju kot pri problemu razmejevanja naravnega (prirodnega) in privzgojenega (družbenega). Ni mogoče izhajati iz neke toge dihotomije med prirodnim in privzgojenim, saj je ta z napredujočimi spoznanji o kompleksnosti razmerij med genotipi in fenotipi vedno bolj postavljena pod vprašaj (Fuller, 2011). Podobno je pri razmejevanju medicinske terapije in krepitve človekovih telesnih in umskih sposobnosti. Celo tam, kjer smo prepričani, da se vse dogaja zgolj v okviru medicinske terapije, lahko pridemo v situacijo, ko se krepitev človekovih individualnih zmožnosti dogaja fenotipsko, tudi kar zadeva njegove umske in telesne zmožnosti. To lahko ilustriramo s primerom, ki ga naveda John Harris. Po Harrisu (2009) bo terapija ljudi z možganskimi poškodbami, ki temelji na regeneraciji izvornih (možganskih) celic, nedvomno predstavljala obliko medicinske terapije. Če bi to isto terapijo izvajali na ljudeh, ki so normalni in nimajo teh poškodb, bi to lahko povečalo zmožnost možganskih funkcij. Meja med medicinskim zdravljenjem in izboljševanjem človekovih zmožnosti je tu relativno jasna. Vendar, kot nadaljuje, Harris, »če pa bi ta isti poseg uporabljali za sistematično zdravljenje neke starostne bolezni, kar bi vodilo k temu, da ljudje bistveno podaljšajo svojo življenjsko dobo, bi bilo res težko vzpostaviti ločnico med medicinsko terapijo in izboljševanjem človekovih zmožnosti« (Harris, 2009: 142).

Že ta kratek ekskurz v kulturni boj med tehnoprogresivisti in tehnokonzervativci v ZDA nam je lahko v poduk, da prava pot niso ideološke polarizacije pri vprašanjih bioetike, temveč »empirični« obrat in inter- oziroma trans-disciplinarno povezovanje bioetičnih spoznanj, o čemer smo več spregovorili že v uvodu naše razprave. Tega dejstva bi se moral v večji meri zavedati tudi člani različnih bioetičnih komisij, svetov, profesionalnih združenj itd., ki predstavljajo institucionalni

okvir današnjim bioetičnim razpravam v Srednji in Vzhodni Evropi (vključno s Slovenijo) in skozi svojo svetovalno funkcijo, ali pa tudi v širšem dialogu s celotno javnostjo, odpirajo (naj bi odpirala) tudi pomembna etična vprašanja razvoja novih tehnologij.

Zaključek

V prispevku smo skušali opozoriti, da sodobni bioetični pogledi nujno potrebujejo večjo preusmeritev k družbenim, etičnim in pravnim posledicam razvoja novih naprednih tehnologij, vključno s tehnologijami krepitve človekovih umskih in telesnih sposobnosti. To velja tudi oziroma predvsem za tisti del Evrope (in sem spada tudi Slovenija), ki je zaradi različnih kulturno historičnih in družbeno političnih okoliščin nekoliko zaostajal v procesih institucionalizacije bioetičnega diskurza. Zavedati se namreč moramo, da bioetika nasploh nima neke dolge tradicije, saj je kot samostojna disciplina nastopila šele v sedemdesetih letih prejšnjega stoletja. Moderni bioetični diskurz potrebuje več razmisleka in javne diskusije o družbenih, etičnih in pravnih implikacijah razvoja novih naprednih tehnologij, vendar ne v obliki ideoloških ekskluzivizmov in kulturnih bojev, s katerimi se danes neredko še vedno srečujemo. Ideološko pogojeni spori o možnih tveganjih razvoja novih tehnologij, zlasti tehnologij krepitve človekovih umskih in telesnih zmožnosti, so družbeno neproduktivni. V preteklosti smo se ob vsakem revolucionarnem odkritju znanosti in tehnologije ravno tako srečevali s skepto o družbeni koristnosti takšnih znanstvenih odkritij, saj tehnofobije niso zgolj značilnost modernih časov. Tovrstne skepticizme in zavračanja znanstvenih in tehnoloških dosežkov so se zaradi njihove kasnejše uporabnosti hitro izkazale kot neupravičene in nepotrebne. S podobnimi situacijami se srečujemo tudi danes, ko doživljamo izjemen napredek ravno na področju razvoja tehnologij krepitve človekovih telesnih in umskih zmožnosti. Njihov razvoj ponuja številne družbene koristi, zato konzervativno, ideološko pogojeni predsodki in odpori ne bi smeli ustaviti tehnološkega razvoja. Potrebujemo celovito družbeno in etično presojo posledic razvoja teh novih naprednih tehnologij, ne potrebujemo pa apriornih zavračanj. Moderna progresivna bioetika, ki naj bi predstavljala najširšo teoretsko refleksijo o biomedicinskih pogojih bivanja sodobnega človeka, bi se zato morala še posebej posvetiti – naj še enkrat poudarimo, brez vnaprejšnjih ideoloških predsodkov – družbenim in prirodnim posledicam razvoja teh tehnologij.

The Development of New Technologies as a Challenge for Modern Bioethical Views on Human Beings and World

The article deals with bioethical challenges arising from the progress of human enhancement technologies. The focus is on the institutional framework of modern bioethics which is built in Europe from various kinds of bioethical expert bodies. In the article, special emphasis is given to the origin and development of national ethic councils in Europe which deal with the ethical dilemmas of scientific and technological progress. At the same time, the article provides a detailed explanation why the institutional framework of bioethics in East and Central European (CEE) countries, i.e. the new EU Member States, did not succeed in following the shift from traditional to modern bioethical topics. The latter is directly connected with the ethical dilemmas of the tremendous progress of human enhancement technologies. The author concludes that bioethical "cultural wars" pose the biggest threat to a shift towards a more progressive bioethical discourse in this area of Europe.

Keywords: bioethics, human cognition, new human enhancement technologies, human cognition, national ethic councils, cultural war in bioethics

Literatura

- Ahvenharju, S., Halonen, M., Uusitalo, S., Launis, V. in Hjelt, M. (2006). *Comparative analysis of opinions produced by National Ethics Councils*. Final report. Helsinki: Gaia Group Ltd.
- Baron, J. (2006). *Against Bioethics*. Cambridge, Mass.: The MIT Press.
- Berger, S. in Moreno, J. (2010). »Bioethics Progressing«. V Moreno, J. in Berger, S. (ur.), *Progress in Bioethics: Science, Policy, and Politics*, Boston: The MIT Press, str. 3–23.
- Bogner, A. (2010). »Let's disagree! Talking ethics in technology controversies«. *Science and Technology Innovation Studies*, 6(2), str. 183–201.
- Braun, K. in Kropp, C. (2010). »Beyond Speaking Truth? Institutional Responses to Uncertainty in Scientific Governance«. *Science, Technology & Human Values*, 35(6), str. 771–782.
- Briggle, A. (2009). »The US President's Council on Bioethics: modeling a thicker knowledge politics«. *Innovation - The European Journal of Social Science Research*, 22(1), str. 35–51.
- Buchanan, A. (2011). *Beyond Humanity?: The Ethics of Biomedical Enhancement*. Oxford: Oxford University Press.

- Coenen, C., Schuijff, M. in Smits, M. (2011). »The Politics of Human Enhancement and the European Union«. V Savulescu, J. et al. (ur.), *Enhancing Human Capacities*, Oxford: Wiley-Blackwell, str. 486–503.
- Dawson, A. (2010). »The future of bioethics: three dogmas and a cup of hemlock«. *Bioethics*, 24(5), str. 218–225.
- Delvenne, P., Fallon, C. in Brunet S. (2011). »Parliamentary technology assessment institutions as indications of reflexive modernization«. *Technology in Society*, 33(1-2), str. 36–43.
- Dranseika, V., Gefenas, E., Cekanauskaite, A., Hug, K., Mezinska, S., Peicius, E., Silis, V., Soosaar, A. in Strosbe, M. (2010). »20 Years of Human Research Ethics Committees in the Baltic States«. *Developing World Bioethics*, Wiley, 11(1), str. 48–54.
- Druml, C., Wolzt, M., Pleiner, J. in Singer, E. A. (2009). »Research ethics committees in Europe: trials and tribulations«. *Intensive Care Medicine*, 35(9), str. 1636–1640.
- EASAC (2017). *Genome editing: scientific opportunities, public interests and policy options in the European Union*. EASAC Policy Report 31, March 2017. Halle (Saale): German National Academy of Sciences Leopoldina. Dostopno na: www.easac.eu [15.12.2017].
- EGE (2017). *Introduction to The European Group on Ethics in Science and New Technologies*. Dostopno na: <https://ec.europa.eu/research/ege/index.cfm?pg=about> [9. 12. 2017].
- EGE (2015). *The ethical implications of new health technologies and citizens participation. Opinion no. 29 of the European Group on Ethics in Science and New Technologies*. Brussels, 13. October, 2015. Dostopno na: https://ec.europa.eu/research/ege/pdf/opinion-29_ege.pdf [15. 12. 2017].
- EGE (2014). *Ethics of Security and Surveillance Technologies. Opinion no. 28 of the European Group on Ethics in Science and New Technologies*. Brussels, 20. May, 2014. Dostopno na: <https://publications.europa.eu/sl/publication-detail/-/publication/6f1b3ce0-2810-4926-b185-54fc3225c969> [15. 12. 2017].
- EGE (2012). *Ethics of information and communication technologies. Opinion no. 27 of the European Group on Ethics in Science and New Technologies*. Brussels, 22 February 2012. Dostopno na: <https://publications.europa.eu/sl/publication-detail/-/publication/c35a8ab5-a21d-41ff-b654-8cd6d41f6794> [15. 12. 2017].
- Fuchs, M. (2005) *Nationale Ethikräte: Hintergründe, Funktionen und Arbeitsweisen im Vergleich*. Berlin: Nationaler Ethikrat.
- Fuller, S. (2014). *The Proactionary Imperative – A Foundation for Transhumanism*. Hampshire and New York: Palgrave Macmillan.

Fuller, S. (2011). *Humanity 2.0: What it Means to be Human Past, Present and Future*, Hampshire and New York: Palgrave Macmillan.

Gillon, R. (2014). »Defending the four principles approach as a good basis for good medical practice and therefore for good medical ethics«. *Journal of Medical Ethics*, 41(1), str. 111–116.

Glasić, J. (2002). »Establishment and work of ethics committees in Central and Eastern European countries«. *Medicinska etika & Bioetika*, 9(1-2), str. 9–12.

Harris, J. (2009). »Enhancements Are Moral Obligations«. V Savulescu, J. in Bostrom, N. (ur.), *Human Enhancement*, Oxford: Oxford University Press, str. 131–155.

ten Have, H. (2012). »Potter's notion of bioethics«. *Kennedy Institute of Ethics Journal*, 22(1), str. 59–82.

Hedgecoe, A., Carvalho, F., Lobmayer, P. in Raka, F. (2006). »Research Ethics Committees in Europe: Implementing the Directive, Respecting Diversity«. *Journal of Medical Ethics*, 32(8): str. 483–486.

Hughes, J. (2006). »Human Enhancement and the Emergent Technopolitics of the 21st Century«. V Bainbridge, W. in Roco, M. (ur.), *Managing Nano-Bio-Info-Cogno Innovations: Converging Technologies in Society*, Dordrecht: Springer, str. 285–309.

Jennings, B. (2011). »Bioethics Between Two Worlds: The Politics of Ethics in Central Europe«. V Myser, C. (ur.), *Bioethics Around the Globe*, Oxford in New York: Oxford University Press, str. 93–106.

Jonas, H. (1979). *Das Prinzip Verantwortung: Versuch einer Ethik für die technologische Zivilisation*. Frankfurt am Main: Suhrkamp.

Lucke, J., Bell, S., Partridge, B. in Hall, W. (2010). »Weak evidence for large claims contribute to the phantom debate«. *BioSocieties*, 5(4), str. 482–483.

Mali, F., Pustovrh, T., Groboljsek, B. in Coenen, C. (2012). »National Ethics Advisory Bodies in the Emerging Landscape of Responsible Research and Innovation«. *Nanoethics*, 6(3), str. 167–184.

McCullagh, P., Lightbody, G., Zygierevicz, J. in Kernohan, G. (2014). »Ethical Challenges Associated with the Development and Deployment of Brain Computer Interface Technology«. *Neuroethics*, 7(2), str. 109–122.

McKee, M. (2004). »Post-communist transition and health in Europe«. *BMJ*, 329, str. 1355–1356.

Penders, B., Horstman, K. in Vos, R. (2008). »A ferry between cultures: Crafting a new profession at the intersection of science and society«. *EMBO Reports*, 9(8), str. 709–713.

- Prodanov, V. (2001). »Bioethics in Eastern Europe: A Difficult Birth«. *Cambridge Quarterly of Healthcare Ethics*, 10(1), str. 53–61.
- Pustovrh, T. in Mali, F. (2015). »(Bio)ethicists and (Bio)ethical Expertise in National Ethical Advisory Bodies: Roles, Functions and Perceptions«. *Prolegomena*, 14(1), 47–69.
- Rabinow, P. in Bennett, G. (2012). *Designing Human Practices: An Experiment with Synthetic Biology*. Chicago: The University of Chicago Press.
- Shibasaki, F. (2005). »Technology and ethics«. *Philosophy & Social Criticism*, 31(4), str. 487–498.
- Sorgner, S. (2016). *Transhumanismus: »Die gefaehrlichste Idee der Welt«?! Freiburg in Breisgau: Verlag Herder GmbH*.
- Stehr, N. (2003). *Wissenspolitik: Überwachung des Wissens*. Frankfurt am Main: Suhrkamp.
- Weingart, P. (2001). *Die Stunde der Wahrheit?: Zum Verhaeltnis der Wissenschaft zu Politik, Wirtschaft und Medien in der Wissensgesellschaft*. Velbrueck: Wissenschaft Weilerwist.
- Wilson, D. (2011). »Creating the ‘ethics industry’: Mary Warnock, in vitro fertilization and the history of bioethics in Britain«. *BioSocieties*, 6(2), str. 121–141.
- Zwart, H. in Nelis, A. (2009). »What is ELSA genomics?«. *EMBO Reports*, 10(6), str. 540–544.

Dejan Savić

Etični pomisleki o filtriranju vsebin na družbenih omrežjih

Družbena omrežja so postala tako vplivna, da oblikujejo razmišljanje prebivalstva o družbenih in političnih vprašanjih. S pomočjo oglaševanja in narave njihovega delovanja lahko prispevajo k volilnim izidom držav z dolgo demokratično tradicijo, kot smo videli v primeru *Brexita* in pri zadnjih volitvah ameriškega predsednika (2016). Osrednja skrb pričujočega besedila je orisati etično sporne vidike delovanja družbenih omrežij, ki lahko vodijo v oviranje delovanja sodobne demokracije. V besedilu bodo orisani nekateri etično sporni vidiki njihovega delovanja, ki velikemu delu prebivalstva mnogih demokratičnih držav soustvarjajo in ohranjajo sistem dostopanja do informacij in svetovno nazorskih prepričanj. Sporne metode delovanja družbenih omrežij ustvarjajo pogoje za politično manipulacijo volivk in volivcev; pogosto zaradi nepreglednosti onemogočajo demokratični nadzor nad političnimi kampanjami in sredstvi, ki se uporabljajo za prepričevanje volivcev; spodbujajo širitev »lažnih novic« in krepijo drobljenje družb in polarizacijo družbenih skupin, zaradi česar je oviran politični proces reševanja pomembnih družbenih in političnih vprašanj. Soočenje z omenjenimi – proti-demokratičnimi – pojavi zahteva samoregulacijo družbenih omrežij, morda pa tudi spremembe zakonske regulacije njihovega delovanja.

Ključne besede: družbena omrežja, internetna etika, spletni algoritmi, etika tehnologij, politična filozofija

Predgovor

Pričujoče besedilo je za filozofsko razpravo nenavadno gosto prežeto z vsebinami iz konkretnega empiričnega sveta. Izhajajoč iz prepričanja, da bi lahko prihodnje posledice nespremenjenega delovanja *družbenih omrežij* (odslej SDO)¹ imele odločilne posledice na sodobne demokratične družbe, jih je potrebno pozorno spremljati iz vidika (uporabne) etike in politične filozofije in s tem prispevati k obvarovanju ključnih demokratičnih institucij pred novimi oblikami zlorab, ki se

¹ SDO, tj. storitve družbenih omrežij, je izraz, ki ga *Stanford Encyclopedia of Philosophy* uporablja za družbena omrežja. SDO je neposredni prevod angleške kratice *SNS*, ki je okrajšava za *social network services* (Vallor, 2016).

pojavljajo na družbenih omrežjih. Predpogoj dobrega razumevanja filozofskih problemov, ki se odpirajo na področju etike interneta in družbenih omrežij, zahteva pravilno razumevanje materije – družbenih omrežij, orodij, ki so upravljavcem SDO na voljo za oblikovanje stvarnosti družbenih omrežij in zlorab do katerih prihaja v okviru SDO. Za tako razumevanje pa se je potrebno opreti na empirične dokaze o njihovem dejanskem delovanju. Kot podarja Shannon Vallor (2016) so etični pomisleki na področju etike družbenih omrežij vse prej kot »akademski«, zato si niti filozofi niti širša skupnost ljudi ne morejo privoščiti razkošja obravnave teh vprašanj kot zgolj akademskih.

Uvod

Družbena omrežja (SDO) in druge aplikacije pametnih telefonov pomembno zaznamujejo sodobne družbene okoliščine in vplivajo na celo vrsto odnosov med ljudmi. SDO lahko na primer zaznamujejo politične preference prebivalstva in vplivajo na politično stvarnost (Madrigal, 2017). Če se je nekoč kazalo, da je ta vpliv le pozitiven in krepi demokratične težnje prebivalstva, se v zadnjih letih kaže, da imajo lahko socialna omrežja tudi negativen vpliv in lahko ogrožajo demokratično družbeno ureditev.

Shannon Vallor poudarja, da družbena omrežja, kot so Facebook, MySpace, Twitter in YouTube, preoblikujejo družabno in politično delovanje posameznikov in institucij po vsem svetu in njihov dostop do informacij. Tovrstne spremembe kličejo pa odzivu filozofov in etikov, ki se ukvarjajo s tehnologijo in uporabno etiko. Družbena omrežja namreč vplivajo na to, kako ljudje »vzpostavljajo in/ali vzdržujejo praktično vse tipe etično pomembnih družbenih vezi in vlog: prijatelj-prijatelj, starš-otrok, sodelavec-sodelavec, delodajalec-zaposleni, učitelj-študent, sosed-sosed, prodajalec-kupec in zdravnik-pacient« (Vallor, 2016).

V zadnjih letih se kaže, da družbena omrežja bistveno določajo tudi odnos posameznikov do družbeno političnih vprašanj. Obstajajo utemeljeni pomisleki, da nekatera družbena omrežja v imenu maksimiranja oglaševalskega zaslužka uporabljajo sporne metode za prepoznavanje in preoblikovanje posameznikovih družbenih in političnih preferenc.

1. Politične manipulacije v okviru SDO

SDO so postala tako vplivna, da oblikujejo način razmišljanja prebivalstva in lahko prispevajo k volilnim izidom držav z dolgo demokratično tradicijo. Jamie Doward in Alice Gibbs poudarjata, da so v času Brexita in ameriških predsedniških volitev v letu 2016 uporabljali sporne storitve profiliranja uporabnikov (Doward in Gibbs, 2017). Podjetje Cambridge Analytica, ki je sodelovalo v obeh kampanjah, na svoji spletni strani na ameriškem primeru pojasnjuje svoje delova-

nje: »Zberemo do 5000 podatkovnih točk o več kot 220 milijonih Američanov in uporabimo več kot 100 podatkovnih spremenljivk za modeliranje ciljnih skupin in za predvidevanje vedenja podobno mislečih ljudi« (Doward in Gibbs, 2017). Predstavnik omenjenega podjetja Alexander Nix je dejal, da je podjetje »uspešno zelo uspešno okrepiti kampanjo *Leave.EU* na družabnih omrežjih z zagotavljanjem, da na spletu prava sporočila dosega prave volivce« (Nix, 2016).

Alexis C. Madrigal (2017) opozarja, da Facebook kot eno najbolj popularnih SDO (ima več kot 1,6 milijarde uporabnikov po svetu), uporablja sporne algoritme, ki Facebookovim oglaševalcem omogočajo uporabo nepreglednih prijemov za oblikovanje političnih preferenc Facebookovih uporabnikov. Isto družbeno omrežje naj bi v predsedniški kampanji Donalda Trumpa omogočalo vnaprej profilirane posameznike (mlade izobražence, ženske in temnopolte) odvracati od udeležbe na volitvah. Profilirane skupine naj bi namreč v primeru oddaje svojega glasu, le tega namenile Trumpovi politični nasprotnici (Madrigal, 2017). Glavno Facebookovo orodje za poosebljenje uporabniške izkušnje je News Feed,² ki s pomočjo algoritmov ustvarja vsakemu uporabniku posebej prilagojeni digitalni svet novic in drugih objav. Madrigal delovanje NewsFeed-a pojasnjuje takole:

Všečkaj stran, prejmi več objav všečkane strani; všečkaj zgodbo, prejmi več takih zgodb; izmenjaj komunikacijo z osebo, prejmi več novosti od te osebe. Facebook določa pojavnost novic na posameznikovem News Feedu tako, da daje v ospredje zgodbe, za katere obstaja večja verjetnost, da jim bo uporabnik všečkal, komentiral ali delil. Pri tem so delitve novic vredne več kot komentiranje novic, oboje pa je vredno več kot všečkanje, v splošnem pa velja, da večja kot je verjetnost, da se bo uporabnik angažiral z objavo, višje v News Feedu se bo zgodba pojavila. (Madrigal, 2017)

Tovrstno delovanje Facebookovih algoritmov je posledica dolgoletnega učenja in evolucije tega socialnega omrežja. Osnovna gonilo razvoja podjetja Facebook je bilo povečevanje časa, ki ga uporabniki preživijo na tem omrežju, ohranjanje pozornosti uporabnikov za vsebine, ki se tam pojavljajo, in izgradnja ugodne uporabniške izkušnje, ki bo vodila k pogostemu obiskovanju Facebooka. Razpolaganje s časom uporabnikov in podatki, ki kažejo, kaj je tem uporabnikom všeč, ustvari uspešno kombinacijo za oglaševalce. Oglaševanje je namreč glavni vir dohodka tega družbenega omrežja. Podjetje uživa zelo veliko zaupanje vlagateljev, kar se odraža v njegovi tržni vrednosti, ki je v začetku novembra 2017 znašala 520 milijard ameriških dolarjev in je bila štiri krat večja kot leta 2013. Vrednost podjetja Facebook je približno osem krat večja od vrednosti največjega ameriškega proizvajalca avtomobilov *General Motors* (60 milijard ameriških dolarjev). (»Facebook Market Cap (FB)«, 2017)

² Pojem bi lahko prevedel kot *podajalnik novic* ali *dovajalnik novic*, a ker gre za Facebookovo tehnologijo s natančno določenimi lastnostmi, ohranjam originalni angleški izraz.

Ciljno oglaševanje je eden glavnih adutov Facebooka, ko gre za prodajo »oglasnega prostora« političnim kampanjam. Pri tem velja poudariti, da poskus oblikovanja posameznikovih političnih preferenc ni nič novega in samo po sebi nič spornega. Navsezadnje vse volilne kampanje potekajo ravno zato, da bi dosegle zaželeno vedenje volivcev, s čimer bi kampanja oblikovala volilni rezultat po volji določene stranke ali njene kandidatke oziroma kandidata. A vendarle se zdi, da nekatera družbena omrežja uporabljajo nelegitimne metode za oblikovanje volilnih preferenc svojih uporabnikov. Interes podjetij po čim bolj pregledni in dobro profilirani bazi uporabnikov (gledalcev oglasov), ki se jo dobro prodaja naročnikom oglasov, lahko vodi v informacijsko zaprtost³ teh uporabnikov. Kolateralna škoda vztrajnega in dolgotrajnega podajanja točno določenih sporočil določene politične kampanje lahko nekatere uporabnike vodi k zmotnemu dojemanju družbenih razmer, ki se lahko začne oddaljevati od sveta dejstev. Posledično lahko zmanipulirani volivci pripeljejo do zmage politične kandidatke z dvomljivo demokratično legitimnostjo. Kajti če zavedeni in zmanipulirani volivci na podlagi zmotnega in na lažeh utemeljenega prepričanja izvolijo kandidata, ta kandidat nima potrebne demokratične legitimnosti, ampak je bolje reči, da je na oblast prišel s prevaro. Če tako prevaro v zameno za denar pomagajo organizirati SDO, je to moralno sporno. Da bi bolje razumeli moralna vprašanja, ki izvirajo iz tveganja za tovrstne politične manipulacije ljudi s pomočjo SDO, bom v nadaljevanju prispevka orisal nekaj negativnih pojavov povezanih s SDO. Pri tem bom skiciral okvire razprave o morebitnem omejevanju delovanja SDO in vpeljal demokratičnega nadzora nad njihovim delovanjem, ko gre za rabo spornih tehnologij za oblikovanje preferenc državljanek in državljanov.

2. Sporne posledice obstoječega delovanja SDO

SDO posegajo v prost pretok informacij tako, da posamezniku filtrirajo (politične) vsebine do katerih dostopa, in tako, da delovanje algoritmov, ki sledijo popularnosti novic, družbeno omrežje spodbuja k širjenju »lažnih novic«. Posledično uporabniki (državljan) oblikujejo svoje politične reference na podlagi laži in ne na podlagi dejstev.

2.1 Informacijski silos

Algoritmi SDO sledijo vsebinam, ki nas zanimajo, zato da bi nam lahko v prihodnje ponujali več podobnih vsebin. Facebook je v letu 2016 predstavil novo politiko filtriranja vsebin na svoji platformi: »če bi lahko vsakodnevno pogledali tisoče zgodb in izbrali 10, ki vam največ pomenijo, katere bi izbrali? Odgovor bi moral biti vaš News Feed. Je subjektiven, osebni in unikaten – in določa duh tega, za

³ Problem informacijske zaprtosti bo podrobneje obravnavan nekoliko kasneje v razpravi o spornih posledicah obstoječega delovanja SDO.

kar upamo, da bomo dosegli» (Mosseri, 2016). Facebook bi torej rad dosegel popolnoma poosebljen dostop do novic za vsakega posameznika, to pa bo dosegel s pomočjo algoritmov, ki pozorno beležijo klike, komentarje in delitve uporabnikov. Eli Pariser, avtor knjige *Filter Bubble (Filtrirni mehurček)*, poudarja, da so algoritmi družbenih omrežij naravnani tako, da uporabnikom ponujajo tematike, ki so relevantne zanje, glede na izkazano zanimanje v preteklosti (Pariser, 2011). Algoritmi, ki natančno sledijo uporabnikovim klikom in drugim interakcijam z vsebinami na družbenem omrežju, sčasoma ustvarijo filtrirni mehurček – uporabniku se ponujajo »vsebine, ki temeljijo na uporabnikovih znanih preferencah, kar uporabniku ustvari vtis, da je njegov pogled na svet prevladujoč« (»Filter bubble«, 2017).

V tako izgrajenem poosebljenem okolju družbenih omrežij posameznikove objave odmevajo kot v »odmevni komori« (ang. *echo chamber*): uporabnikovo objavo pokažejo le vnaprej profiliranim »prijateljem«, za katere algoritmi izračunajo, da jih zanimajo podobne stvari. Ti »prijatelji« nato všečkajo, komentirajo ali delijo uporabnikovo vsebino, zaradi česar dobi uporabnik občutek zadovoljstva, družbenega priznanja in občutka pomembnosti. Pojav »odmevne komore« označuje okolje »na družbenih omrežjih, kjer velja, da bo katerakoli izjava ali mnenje verjetno sprejeto z odobravanjem, kajti prebrali ga bodo le tisti, ki imajo podobna stališča« (»Echo chamber«, b. d.). Pariser poudarja, da »v poosebljenem svetu [družbenih omrežij], pomembne, a kompleksne ali neprijetne vsebine – na primer rastoča populacija zapornikov ali brezdomstvo – zelo verjetno sploh ne bodo prišle do naše pozornosti« (Pariser, 2011: 15) [op. D. S.].

Cass Sunstein opozarja, da že internet na sploh, SDO pa še veliko bolj, spodbuja ustvarjanje »odmevnih komor« in »filtrirnih mehurčkov«, ki prispevajo k drobljenju javne sfere: ustvarjajo se informacijski silosi podobno mislečih posameznikov, ki se preko interakcije z vsebinami, ki so jim všeč, in z ignoriranjem vsebin, ki jim niso všeč, postopoma ogradijo od izpostavljenosti različnim mnenjem. Informacijski silos je v tem primeru razumljen kot del sistema, ki ne komunicira z drugimi deli sistema, drugih delov ne razume ali z njimi ne sodeluje. Tovrstno zapiranje posameznikov v informacijske silose zbuja skrb, da tovrstna izolacija spodbuja ekstremizem in krepitev slabo utemeljenih stališč, obenem pa preprečuje državljanom v demokratični skupnosti, da bi prepoznali interese in izkušnje, ki so državljanom skupni. (Sunstein, 2008).

2.2 Psihografično ciljno oglaševanje

Dobro profilirani, v filtrirne mehurčke ločeni uporabniki so privlačni za ciljno oglaševanje. Madrigal opozarja, da Facebook omogoča poseben tip ciljnega oglaševanja, tako imenovano temno oglaševanje (ang. *dark ads*) ali psihografično targetiranje, ki določeno vsebino pokaže le točno določeni ciljni skupini in nikomur drugemu (Madrigal, 2017; Hern, 2017). Alex Hern za tovrstno oglaševanje pravi, da gre za »kontroverzno dejavnost, ki skupinam [oglaševalcev] omogoča prilaga-

janje njihovih sporočil tako, da ustrezajo osebnostnim tipom njihovih tarč v politični kampanji» (Hern, 2017) [op. D. S.]. Facebook na svoji spletni strani svoja orodja politikom oglašuje kot »popoln način za 'prepričevanje volivcev' in 'vplivanje' na izide na spletu in zunaj spleta« (»Connect with constituents...«, 2017).

Madrigal poudarja, da novinarji ali kdorkoli drug ne more videti sporočil, ki so se dejansko prikazovala ljudem (Madrigal, 2017). Preko ciljnega oglaševanja, brez javnega nadzora algoritmi SDO uporabnikom izgrajujejo svet, ki se kaže kot resničnost, a je v resnici postavljen po volji oglaševalca, ki ima morda namene, nasprotno interesom uporabnika. Tovrstno nepregledno oglaševanje odpira tveganje zlorab in manipulacijo posameznikove politične volje.

Takih manipulacij v medijih načeloma ne dopuščamo, a ko do njih pride, jih obsojimo ali jih moralno – včasih pa tudi pravno – sankcioniramo. Sporne oglaševalske prijeme načeloma poskušamo nadzorovati in zakonsko regulirati. Če algoritmi SDO zaprejo uporabnikov svet pred demokratičnim nadzorom, se lahko zgodi, da je pomemben del vsebin, ki prihajajo do posameznika in oblikujejo njegove politične preference, spornih ali celo lažnih.

V dnevni pred ameriškimi predsedniškimi volitvami so novinarji odkrili, da 100 spletnih strani, ki promovirajo Donalda Trumpa, izhaja iz majhnega makedonskega kraja s 45000 prebivalci. Kot pravi Madrigal, so neki najstniki ugotovili, kako lahko SDO algoritme izkoristijo za dodaten zaslužek ob ameriških volitvah, in so kar tako postali del informacijske mreže, ki je pomagala Trumpu premagati protikandidatko Clinton (Madrigal, 2017).

Jonathan Albright (2017) iz Columbia Univerze je v analizi ugotovil, da so bile objave šestih, z Rusijo povezanih javnih strani na Facebooku deljene 340 milijon krat. Analiza je bila narejena za šest od 470 strani, za katere je Facebook ugotovil, da so povezane z ruskimi operativci. Madrigal opozarja, da v tem primeru zelo verjetno govorimo o milijardah delitev novic, ki so bil deležne neznano koliko ogledov z neznano določenimi ciljnim sporočili za neznane ciljne skupine (Madrigal, 2017). Trenutni način delovanja SDO torej omogoča celo vrsto zlorab, zato je reforma njihovega delovanja nujna.

3. Prost pretok informacij kot javna dobrina

Demokratična politična ureditev temelji na ideji, da državljanke in državljani posredno upravljajo z javnimi zadevami. Volivke in volivci na svobodnih volitvah političnim predstavnicam in predstavnikom podelijo mandat, da v časovno omejenem obdobju (navadno štiri leta) vladajo v njihovem imenu. Prost pretok informacij je temeljna predpostavka demokracije. Posamezniki kot volivci in kot gradniki javnega mnenja svobodno presojujejo informacije, na podlagi katerih oblikujejo svoje politične preference. Državljanke in državljani potrebujejo dostop do

celovitih informacij o delovanju oblasti in opozicije, zato da lahko sprotno oblikujejo svoje mnenje (javno mnenje) in na naslednjih volitvah odločijo, komu bodo podelili mandat za vodenje države. Odločitev za podelitev mandata bi morala temeljiti na informiranih političnih preferencah (Kymlicka, 2005), kar je predpogoj, da posamezniki odločajo na podlagi dejstev in ne na podlagi neresnic ali laži. Če je prost pretok informacij prekinjen, državljani ne morejo dobro presoјati delovanja oblasti in opozicije. Posledično nastanejo neinformirane politične preference, ki lahko vodijo do volitev, kjer oblast dobi politična opcija, ki politične podpore ljudstva ne bi uživala, če bi bile politične preference državlјank in državlјanov informirane. Filtriranje novic je močno sporno, ker lahko vodi v situacijo, ko pomanjkljivo informirani državljani izvolijo političnega predstavnika, ki ga ne bi volili, če bi imeli neoviran dostop do informacij in bi poznali vsa dejstva o določeni kandidatki ali kandidatu.

4. Vzpostavitev demokratičnega nadzora

Mandat za (pre)oblikovanje družbene ureditve je nujno pogoјen z voljo ljudstva. Ta se zavezujoče preverja na volitvah, v nezavezujoči obliki pa ob meritvah javnega mnenja. Izvoljeni predstavniki ljudstva imajo legitimno pravico izvajati oblast v času trajanja mandata, kolikor uživajo podporo legitimno izražene in legitimno oblikovane volje ljudstva. Če se izkaže, da algoritmi SDO posameznike ovirajo pri oblikovanju informiranih političnih preferenc, potem določeni algoritmi SDO ovirajo normalno delovanje demokracije. Ravnanje, ki je v temeljnem nasprotju z demokratično družbeno ureditvijo, je potrebno omejevati ali prepovedati. Država bi torej morala omejevati rabo SDO algoritmov, ki uporabnike vodijo v informacijsko zaprtost, znotraj katere si je težko oblikovati informirane politične preference in se politično odločati na podlagi informiranih političnih preferenc.

V razpravah o etiki SDO se pojavljajo različne pobude po regulaciji SDO. Potrebno bi bilo določiti obliko demokratičnega nadzora, ki bi legitimnim predstavnikom javnega interesa (npr. informacijskemu pooblaščenцу, raziskovalcem, novinarjem, nevladnim organizacijam) omogočil dostop do orodji, ki jih oglaševalci in politične kampanje uporabljajo v svojem delovanju. Tak nadzor bi moral omogočati izločanje in morebitno sankcioniranje nelegitimnih ravnanj. Pri ločevanju legitimnih oglaševalskih orodij od nelegitimnih bi se lahko zgledovali po veljavni ureditvi (ZMed-D, 2016), ki nadzira spoštovanje demokratičnih standardov v klasičnih medijih (televizija, radio, tisk).

Zdi se, da bi obravnava SDO kot medijev lahko prispevala k omejevanju najbolj negativnih učinkov delovanja samovoljnega filtriranja novic v SDO. Mediji so zavezani uravnoteženemu poročanju, zavezani so resnici, biti morajo transparentni glede virov, podrejeni so medijski zakonodaji in kodeksu novinarske etike.

Omogočajo pravico do popravka in ne nazadnje imajo uredništva odgovorne urednike, ki jamčijo za zakonito in legitimno delovanje določenega medija.

Ker se SDO le razlikujejo od tradicionalnih medijev, zaželenih učinkov ne bomo dosegli z uporabo natanko takšnih orodij, kot veljajo za tradicionalne medije. Potrebno bo razviti posebna orodja, ki so prilagojena posebni naravi SDO. Na tem mestu ni prostora za podrobno razpravo o primernih orodjih za preoblikovanje delovanja SDO, razmislimo pa lahko o prvih smiselni korakih v smer prediranja filtrirnih mehurčkov in osvobajanja uporabnikov iz odmevnih komor. Pariser (2011) meni, da bi morali algoritmi zasledovati ne le vsebine, ki naj bi bile relevantne za posameznika, ampak bi moral prikazovati tudi vsebine, ki jih je pomembno razumeti za razumevanje določene tematike. Uporabnika bi bilo potrebno izpostaviti določenim novicam, tudi če bi bila taka izkušnja zanj neprijetna, ker bi morda pod vprašaj postavljala njegova prepričanja. Uporabniki bi prav tako morali biti redno izpostavljeni vsebinam, ki izhajajo iz drugega svetovno nazor-skega zornega kota (Pariser, 2011).

Pariser (2011) poudarja, da je družba v preteklosti že bila v podobni situaciji, in sicer ob prodoru časopisov v začetku 20. stoletja. Časopisi so bili razpeti med pogosto trivialnimi vsebinami, ki so bile zelo popularne med bralci, in splošno družbeno dolžnostjo pokrivati vsebine, ki so pomembne za dobro delujočo demokracijo. Bralci so prepoznali vrednost časopisov, ki so pokrivali družbeno pomembne vsebine in posledično tudi z naročninami in nakupi podprli resne časopise, da so nadaljevali svoje poslanstvo na področju krepitve delovanja demokratične družbe.

Omeniti je potrebno tudi nekaj možnosti, ki bi bile na voljo za (samo)regulacijo, s katerimi bi razblinili filtrirne mehurčke in uporabnikom omogočili dostop do vsebin uporabnikov, ki imajo drugačna svetovno nazorska stališča. Sunstein (2017) naniza nekaj možnosti, ki bi se lahko hitro uveljavile v SDO.

Svetovnonazorsko profilirane spletne strani, kot so politične stranke, bi lahko na svojih straneh omogočale dostop do vsebin, ki so nasprotno njihovim. Sunstein si ob takih objavah zamišlja naslednji zapis: »Imamo jasno stališče in upamo, da bo več ljudi verjelo v to, kar počnemo. A hkrati smo predani demokratični razpravi in pogovorom med ljudmi, ki razmišljajo drugače. Temu v prid ponujamo povezave do drugih strani, ker bi radi prispevali k iskreni razpravi o teh vprašanjih.« (Sunstein, 2017: 210) Sunstein je prepričan, da bi ob sodelovanju več pomembnejših spletnih strani različnih nazorskih prepričanj pri tovrstnih prizadevanjih lahko skupaj zmanjševali drobljenje javnosti.

Sunstein (2017) predvidi še dve zelo zanimivi rešitvi, prilagojeni za SDO. Prva je vpeljava spletnega gumba »nasprotno stališče« in gumba »naključne vsebine« (ang. *serendipity button*). Gumb »nasprotno stališče« je ideja Geoffreya Fowlerja (2016), za katero sam pravi, da bi omogočila ljudem dostop do neprijateljskih stališč s klikom na gumb. Fowler nas poziva, naj si predstavljamo, da bi le s premikom gumba na Facebooku spremenili vsa konservativna stališča, ki jih vidimo, v

liberalna, in obratno. Zavedali bi se, da so naše novice lahko videti čisto drugače od sosedovih (Fowler, 2016).

Nekaj podobnega v smeri dostopa do raznovrstnih stališč bi prispeval spletni gumb »naključnih vsebin«. Ta bi omogočil, da bi poleg vsebin uporabnikovih prijateljev videli tudi druge stvari: »vaš News Feed bi bil videti kot veliko mesto ali resničen časopis« (Sunstein, 2017: 211-212).

Iz samega dejstva, da SDO, kljub vsem javnim kritikam, še vedno ne uvajajo sprememb dovolj hitro, napeljuje k temu, da obstajajo dobri pomisleki za ohranjanje obstoječega stanja. Zato je na mestu, da v nadaljevanju premislimo nekaj pomislekov proti spremembam politike filtriranja novic SDO.

5. Nekaj ugovorov proti regulaciji SDO

5.1 »SDO niso mediji«

Nekdo bi lahko zagovarjal stališče, da SDO niso mediji, temveč le digitalni prostori, ki so v glavnem namenjeni kratkočassenju, zato SDO ne bi smeli obravnavati in nadzirati kot da so mediji. Zagovorniki državno nereguliranih SDO bi lahko trdili, da posamezniki svoje politične preference še vedno oblikujejo na podlagi klasičnih medijev, zato SDO-jem ni treba nalagati odgovornosti za oblikovanje informiranih političnih preferenc posameznikov.

V odgovor tem pomislekom je potrebno izpostaviti, da v mnogih državah SDO predstavljajo pomemben del medijske krajine, zato bi morali biti zavezani regulacijam, ki skrbijo za zagotavljanje medijskega informiranja državljanov. Vprašanje vloge SDO kot vira informacij za oblikovanje političnih preferenc je empirično vprašanje in raziskave lahko pokažejo, kako pomembni so SDO znotraj posameznih političnih skupnosti. Craig Silverman (2016) ugotavlja za ZDA, da so v zadnjih treh mesecih ameriške predsedniške kampanje najbolj popularne »lažne novice«, povezane z volilno kampanjo, dosegle večjo angažiranost uporabnikov, kot je to uspelo glavnim zgodbam medijskih velikanov, kot so *The New York Times*, *The Washington Post*, *The Huffington Post*, *NBC News*. Silverman opozarja, da je v času predvolilne kampanje Facebookov algoritem, ki uporabnikom kaže novice, ki so v določenem trenutku »trending«, redno promoviral lažne novice (Silverman, 2016). Facebook z 1,6 milijarde aktivnih uporabnikov po vsem svetu (Sunstein, 2017: 211) v določenih državah vpliva na to, kako velik delež prebivalstva dostopa do novic preko družbenega omrežja. SDO so torej zelo pomembni posredniki vsebin do uporabnikov, zato bi morali biti zavezani istim moralnim in zakonskim standardom kot tradicionalni mediji.

5.2 »Filtriranje novic je le odraz posameznikovih preferenc«

Zagovorniki posebnih družbenih pravil za SDO bi morda vseeno vztrajali, da filtriranje novic ne bi smelo biti ovirano. Ljudje izberejo, kaj bodo brali, s kom se bodo družili, od katerih virov bodo prejeli novice. Podobno kot ljudje izberejo, da bodo brali politično profilirano revijo (v Sloveniji npr. Mladino ali Reporter), ljudje z nastavitvami svojih preferenc v okviru SDO izbirajo ali bodo sledili tem ali onim ustvarjalcem političnih vsebin. Z vsakim klikom izražajo svoje svetovnonazorske preference, algoritem si le prizadeva zadovoljevati te preference. Skratka, filtrirni mehurček je posledica uporabnikovih preteklih izbir, zato je ožene informacijskega političnega horizonta kot posledica filtrirnega mehurčka, rezultat posameznikovih preferenc in kot tako popolnoma legitimno. Posameznik naj bi torej užival pravico do oblikovanja svojih (političnih) preferenc na podlagi omejene informiranosti v okviru SDO in si tako sam neposredno določil svoj filtrirni mehurček z nastavitvami v okviru SDO, kot tudi s svojo aktivnostjo in interakcijo ob posameznih novicah.

Sunstein (2017) pritrjuje, da je filtriranje novic in zapiranje v nišne skupine vedno obstajalo in da pojav na internetu ni nič novega. Kljub temu pa vztraja, da komunikacijski sistem, ki posamezniku podeli neomejeno moč poosebljanja in filtriranja, spodbuja prekomerno fragmentacijo družbe. »Internet predstavlja veliko nevarnost polarizacije družbenih skupin preprosto zato, ker je na internetu tako enostavno podobno mislečim posameznikom govoriti drug z drugim – in v skrajnosti, se premakniti proti ekstremnim, včasih celo nasilnim, stališčem« (Sunstein, 2017: 236).

Dodatno se je potrebno vprašati, ali lahko uporabnikovo sledenje tem ali onim politikom oziroma ustvarjalcem političnih vsebin res razumemo kot določni znak posameznikovih preferenc. Will Kymlicka (2005) poudarja, da je potrebno razlikovati med splošno izraženimi preferencami in med seznanjenimi preferencami: »Ko nimamo ustreznih informacij ali ko napačno izračunamo stroške in koristi nekega dejanja, se to, kar je za nas dobro, lahko razlikuje od naših tedanjih preferenc« (Kymlicka, 2005: 39). Ronald Dworkin podobno poudarja, da ljudje hočejo »imeti ali narediti, kar je vredno imeti ali narediti, to pa se lahko razlikuje od tega, kar hočejo imeti ali narediti v tistem trenutku« (Dworkin, 1983: 24-30).

Kar ljudje klikajo na socialnih omrežjih, to odraža njihovo takratno željo, ne pa nujno tudi njihovih splošnih preferenc. Posameznikove preference so usmerjene v tisto, čemur bi ljudje dali prednost na podlagi dobrih razlogov in na podlagi presoje o tem, kaj bi resnično izboljšalo njihovo življenje (Kymlicka, 2005: 41). Če je dostop uporabnikov SDO do informacij oviran, le ti niso seznanjeni s celovitim naborom perspektiv, ki so potrebne za oblikovanje seznanjenih preferenc. Kymlicka poudarja vlogo družbenih okoliščin za oblikovanje seznanjenih preferenc, ki se razlikujejo od lažnih preferenc:

Morda res ne moremo ugotoviti, katere posebne preference so popačile napakna prepričanja ali prilagoditvene preference⁴, vendar lahko za to, da bi ljudem zagotovili dostop do ustreznih informacij in/ali možnosti preizkušanja alternativnega življenja in/ali zaščito pred lažnimi, ali popačenimi predstavami, ali propagando, preučimo družbene in kulturne pogoje, v katerih ljudje oblikujejo ali revidirajo svoje preference. Problema lažnih ali prilagoditvenih preferenc nam ni potrebno reševati neposredno z njihovim prepoznavanjem in izločanjem, ampak z odstranjevanjem okoliščin, ki takšne preference generirajo. (Kymlicka, 2005: 46)

Uporabniki SDO najbrž ne bi ponudili »dobrih razlogov« in posebne utemeljitve stališča, da si resnično želijo biti zaprti v filtrirni mehurček. Morda je položaj v odmevni komori kratkotrajno prijeten, a malo je posameznikov, ki bi bili prepričani, da bo algoritmično omejevanje dostopa do informacij resnično izboljšalo njihovo življenje in bi si zato želeli vnaprejšnjega filtriranja informacij.

Če bi uporabnike SDO povabili k utemeljeni in informirani razpravi o tem, kakšno filtriranje informacij je zanje sprejemljivo ali zaželeno, bi si uporabniki vsaj v obravnavi pomembnih informacij najbrž želeli podobno kot Pariser (2011), da bi do njih prišle informacije, ki so pomembne za razumevanje področja, četudi vzbujajo neprijetna občutja, prihajajo iz drugih svetovnonazorskih gledišč ali so izzivalne do uporabnikovih stališč. Kajti uporabnik bi verjetno soglašal, da pestrost stališč odraža pestrost teh v družbi glede določene tematike in da je zato filtriranje, ki pestrost stališč zakrije, enako zakrivanju resnice kot dejanskega stanja stvari. Kljub temu da se vsakdo občasno zateče k eskapizmu, bi si najbrž malokdo – seznanjen z vsemi posledicami tovrstne informacijske zaprtosti – na daljši rok želel živeti v svetu, kjer odmeva le njegov glas in glas podobno mislečih ljudi.

5.3 »SDO so neproblematični, ker prenašajo večinoma trivialne vsebine«

Naslednji ugovor regulaciji SDO bi se lahko skliceval na domnevno družbeno nepomembnost večine vsebin, ki se pretakajo preko SDO. Večine ljudi dnevna politika ne zanima, SDO uporabljajo za spremljanje življenja svojih prijateljev, znancev, sodelavcev itn.

Veliko vsebin preko SDO je najbrž res trivialnih in večina vsebin ni bistvena za oblikovanje posameznikov družbeno-političnih preferenc. A tudi če se le majhen delež prebivalstva močno zanaša na informacije iz SDO, je filtriranje, ki tem ljudem oziroma dostop do pomembnih informacij, moralno sporno. Narava političnega sistema v večini demokratičnih držav je taka, da lahko angažirana večina z minimalno prednostjo v rezultatih glasovanja usodno odloči o pomembnem političnem

⁴ S pojmom »prilagoditvene preference« Kymlicka označi pojav, »ko si ljudje postopoma nehajo želeči, česar ne morejo doseči«. Pojav je znan kot »problem kislega grozdja« iz Ezopove basni o lisici (Kymlicka 2005: 40).

vprašanju celotne skupnosti. *Brexit* – odločitev Velike Britanije, da izstopi iz EU – je bil odločen z 51.89% proti 48.11% (Electoral Commission, 2016). Le 2% volivcev bi drugače volila in usoda Velike Britanije v Evropi bi se pisala popolnoma drugače. Podobno je aktualni predsednik Donald Trump svojo protikandidatko Hillary Clinton na volitvah 2016 premagal z minimalno prednostjo, pri čemer je prejel 2,87 milijona glasov manj od tekmice (Federal Election Commission, 2017). Potrebno je poudariti, da ameriški volilni sistem zmagovalca volitev ne določa glede na največ zbranih glasov posameznih volivcev, ampak je zmagovalec volitev tisti, ki zmaga v dovolj zveznih državah, da vse zvezne zmage skupaj prispevajo k zmagovitem seštevku elektorskih glasov. Vsaka država prinese določeno količino elektorskih glasov, kdor zmaga v tej državi, dobi vse glasove te zvezne države. Državne volitve zmaga, kdor zbere več elektorskih glasov. Trump je na zadnjih volitvah osvojil 306 elektorskih glasov, Clintonova pa 232 (»Presidential results«, 2016).

5.4 »Samoregulacija SDO je dovolj«

Ločevanje uporabnikov na profilirane skupine je glavna privlačnost SDO za oglaševalce. Profiliranje posameznikov v jasno določene skupine je v strateškem interesu SDO in predstavlja ekonomsko podstat njihovega poslovnega modela. Zahteva po samoregulaciji SDO bi lahko postavila njihove lastnike v konflikt interesov. Na eni strani bi morda želeli prispevati k demokratičnim procesom v družbi in delovati proti fragmentaciji in polarizaciji družbenih skupin, na drugi strani pa bi morda ugotovili (to nikakor ni nujno), da vsebine, ki so nasprotne stališčem posameznika, tako poslabšajo uporabniško izkušnjo, da bi bilo uporabnikovo nadaljnje angažiranje v SDO vprašljivo. Pri iskanju ustrezne rešitve za to zagato se je potrebno zavedati, da je možnosti več, a morda je delna rešitev v tem, da pooblaščen javna institucija postavi cilje vsebinske pluralnosti na SDO, SDO-jem pa bi bilo prepuščeno, da razvijejo lastna orodja za uresničevanje zastavljenih ciljev. Zahteva javnega pooblaščenca bi denimo bila, da mora uporabnikom vselej biti na voljo dostop do pomembnih alternativnih stališč o določeni tematiki, SDO pa sami določijo, kako točno bi dostop do teh vsebin zagotovili na lastni spletni platformi.

Kot je bilo zgoraj že omenjeno bi lahko delno rešitev pomenila vzpostavitev gumba »nasprotna stališča« (glej zgoraj), pri čemer se uporabniku prepusti, ali bo dejansko kliknil na gumb ali ne. Večina uporabnikov bi lahko vsaj občasno dostopala do nasprotnih stališč in s tem prebadala »iluzijo« lastnega miselnega mehurčka. Poseben problem nastane, če bi ugotovili, da večina ljudi ne želi slišati za nasprotna stališča. Morda bi njihove seznanjene preference želele vsebinsko pestrost informacij in nazorskih perspektiv, takojšnje preference pa bi morda uporabnika le ohranjale v udobju odmevne komore.

Zaključek

Upravljanje svet komunikacij, informacijskih tokov in medijev je za vsako demokracijo zelo pomembno vprašanje. Internet in družbena omrežja so v zelo kratkem zgodovinskem obdobju zaznamovala naše družbe in vedno znova odkrivamo nove potenciale pa tudi grožnje, ki jih nove tehnologije predstavljajo za družbo in demokracijo. Prejudicirati primerno reformo delovanja SDO bi bilo prenačljeno in bi lahko povzročilo več škode kot koristi. Pomembno pa je, da se prepozna problematičnost obstoječega stanja in se poskuša nadgrajevati internet in SDO v smeri večje skladnosti z idejami *dobro delujoče demokracije*. Za upravljavce SDO kot tudi za javne organe nadzora bi vodilo v iskanju primerne modela delovanja družbenih omrežij lahko našli v Sunsteinovi misli: »v svobodni demokratični republiki državljani stremijo k sistemu, ki zagotavlja širok obseg izkušenj – z ljudmi, vsebinami in idejami –, ki jih sami ne bi posebej vnaprej izbrali« (Sunstein, 2017: 239).

Vse, ki so zadržani do regulacije interneta in SDO, ali se jim predstavljene ideje reform delovanja SDO zdijo slabo domišljene, zgrešene ali kako drugače neprimerne, je potrebno spodbuditi z vprašanjem, ki naj kritike vodi k podajanju boljših rešitev. Vsi skupaj pa bi morali poiskati odgovor na vprašanje, ki ga zastavlja Sunstein: »[Č]e hočemo, da bi trenutne tehnologije služile demokratičnim idealom, kakšne vrste ravnanj bi bile boljše od obstoječih?« (Sunstein, 2017: 213).

Ethical Considerations of Content Filtering on Social Media

Social networks services (SNS) have become so influential that they are now shaping the way people think about social and public issues. Through advertising and the nature of their workings, SNS can influence election results in countries with a long-standing democratic tradition, the cases in point being Brexit and the 2016 American presidential elections. The article's main concern is to outline the ethically questionable aspects of the SNS workings, which can hinder the functioning of contemporary democracy. It will highlight certain ethically controversial aspects of SNS, which determine how a relevant part of the population of many democracies is accessing information and alternative world views. Questionable practices of SNS create a breeding ground for the political manipulation of voters; their lack of transparency often prevents democratic control over political campaigns and the means used to convince voters; they help spreading "fake news" and strengthen the fragmentation and polarization of societies. The latter is hindering the political process needed to confront many important societal and political problems. Confronting the above-mentioned anti-democratic tendencies requires the self-regulation of SNS and, possibly, changes in the legal regulation of the SNS workings.

Keywords: social networks, internet ethics, online algorithms, ethics of technology, political philosophy

Literatura

Albright, J. (2017). »Itemized Posts and Historical Engagement - 6 Now-Closed FB Pages«. *Tableau data*. Dostopno na: <https://public.tableau.com/profile/d1gi#!/vizhome/FB4/TotalReachbyPage> [16. 12. 2017].

»Connect with constituents and voters on Facebook« (2017). *Facebook*. Dostopno na: <https://politics.fb.com/> [5. 11. 2017].

Doward, J. in Gibbs, A. (2017). »Did Cambridge Analytica influence the Brexit vote and the US election?«. *The Guardian*. Dostopno na: <https://www.theguardian.com/politics/2017/mar/04/nigel-oakes-cambridge-analytica-what-role-brexit-trump> [3. 11. 2017].

Dworkin, R. (1983). »In Defense of Equality«. *Social Philosophy and Policy*, 1(1), str. 24–40.

»Echo chamber« (b.d.). V *Collins English Dictionary*. Dostopno na: <https://www.collinsdictionary.com/dictionary/english/echo-chamber> [4. 11. 2017].

Electoral Commission (2016). »EU referendum results«. Dostopno na: <https://www.electoralcommission.org.uk/find-information-by-subject/elections-and-referendums/past-elections-and-referendums/eu-referendum/electorate-and-count-information> [1. 11. 2017].

Federal Election Commission (2017). »Official 2016 Presidential General Election Results«. PDF, 30. januar 2017. Dostopno na: <http://www.fec.gov/pubrec/fe2016/2016presgeresults.pdf> [1. 11. 2017].

»Filter bubble« (2017). V *Collins English Dictionary*. Dostopno na: <https://www.collinsdictionary.com/submission/18679/filter+bubble> [4. 11. 2017].

Fowler, G. A. (2016). »What If Facebook Gave Us an Opposing-Viewpoints Button?«. *Wall Street Journal*, May 18, 2016. Dostopno na: <http://www.wsj.com/articles/what-if-facebook-gave-us-an-opposingviewpoints-button-1463573101> [6. 11. 2017].

Hern, A. (2017). »Facebook 'dark ads' can swing political opinions, research shows«. *The Guardian*. Dostopno na: <https://www.theguardian.com/technology/2017/jul/31/facebook-dark-ads-can-swing-opinions-politics-research-shows> [5. 11. 2017].

Kymlicka, W. (2005). *Sodobna politična filozofija: uvod (2. izdaja)*. Ljubljana: Krtina.

Madrigal, C. A. (2017). »What Facebook Did to American Democracy«. *The Atlantic*. Izdaja oktober. Dostopno na: <https://www.theatlantic.com/technology/archive/2017/10/what-facebook-did/542502/> [14. 10. 2017].

Mosseri, A. (2016). »Building a Better News Feed for You«. *Facebook Newsroom*, 29. junij, 2016. Dostopno na: <https://newsroom.fb.com/news/2016/06/building-a-better-news-feed-for-you/>. [7. 11. 2017].

Nix, A. (2016). »How big data got the better of Donald Trump«. *Campaign*. Dostopno na: <https://www.campaignlive.co.uk/article/big-data-better-donald-trump/1383025> [3. 11. 2017].

Pariser, E. (2011). *The Filter Bubble: What Internet Is Hiding from You*. New York: The Penguin Press.

»Presidential results« (2016). *CNN*. Dostopno na: <http://edition.cnn.com/election/results/president> [13.12. 2017].

Silverman, C. (2016). »Here's Why Facebook's Trending Algorithm Keeps Promoting Fake News«. *Buzz Feed News*. Dostopno na: <https://www.buzzfeed.com/craigsilverman/can-facebook-trending-fight-off-fake-news> [1. 11. 2017].

Sunstein, R. C. (2008). »Democracy and the Internet«. V van den Hoven, J. in Weckert, J. (ur.), *Information Technology and Moral Philosophy*, Cambridge: Cambridge University Press, str. 93–110.

Sunstein, R. C. (2017). *#Republic: Divided Democracy in the Age of Social Media*. Princeton: Princeton University Press.

Vallor, S. (2016). »Social Networking and Ethics«. V Zalta, E. N. (ur.), *The Stanford Encyclopedia of Philosophy*. Dostopno na: <https://plato.stanford.edu/archives/win2016/entries/ethics-social-networking/> [29. 10. 2017].

»Facebook Market Cap (FB)« (2017). *Ycharts*. Dostopno na: https://ycharts.com/companies/FB/market_cap [5. 11. 2017].

Zakon o medijih (ZMed-D) (2016). *Uradni list RS*, št. 39/16 z dne 3. 6. 2016. Dostopno na: <http://www.pisrs.si/Pis.web/pregledPredpisa?id=ZAKO1608> [12. 12. 2017].

Meta Kokalj in Anže Mihelič

doktorska kandidata

Pravna fakulteta Univerze v Ljubljani

Uvajanje eksperimentalnih kognitivnih tehnologij v terapično in družbeno prakso

Življenje posameznikov in družbe kot celote je čedalje bolj prežeto z vse bolj naprednimi tehnološkimi orodji, včasih celo do take mere, da prihaja do spajanja človeško-organskih in strojnih sistemov in ustvarjanja novih tipov entitet. V prispevku sta predstavljena dva primera uvajanja kognitivnih tehnologij (kognitivna biometrija in možganski vmesniki) in poskus izpostavitve nekaterih etičnih, pravnih in praktičnih dilem, ki izhajajo iz njihove uporabe. S tehnološkimi zmožnostmi se neizogibno širi ponudba izvorno nepredvidenih (a v dobi velikega podatkovja ne neuporabnih) uporabniških produktov omenjenih tehnologij, iz njihove uporabe pa vznikajo vedno novi načini poseganj v človekovo osebnost. Vse to zahteva hiter odziv v obliki celovite pravne ureditve problematike, ki bo na domišljen način vgradila ustrezna varovala pred nezakonito in neetično uporabo kognitivnih tehnologij v družbeni praksi.

Ključne besede: kognitivne tehnologije, možganski vmesniki, kognitivna biometrija, odgovornost, tveganja

1 Uvod

Nenehen in izrazito hiter razvoj tehnologije povzroča vse pogostejšo, intenzivnejšo in invazivnejšo vpetost tehnoloških orodij v naša življenja, celo v tolikšni meri, da v določenih primerih prihaja do njihovega stapljanja z uporabnikom in ustvarjanja novih tipov entitet. Tehnologije kognitivnega inženiringa omogočajo zajemanje signalov možganskih aktivnosti uporabnika, kar se vse pogostejše uporablja na mnogih področjih življenja v (post)informacijski družbi; od podpore sistemu avtentikacije do upravljanja s terapevtskim pripomočkom zgolj preko lastnih misli s pomočjo možganskega vmesnika (ang. *brain-computer interface*). S širje-

njem uporabe kognitivnih tehnologij pa se neizogibno povečujejo tudi možnosti za nepredvidene, v mejnih primerih tudi škodne dogodke, pri čemer se zaradi naprednih tehnoloških sistemov zabrisuje jasna vzročnost med sprožilnimi dejavniki in posledicami, ki izidejo iz delovanja sistema.

Človeštvo se je razvijalo hkrati z orodji, s katerimi je izboljševalo svojo produktivnost. Z vgrajevanjem v njegovo telo pa ta orodja čedalje bolj postajajo gradniki človekove identitete (Bockman, 2010: 1332). Artefaktno (tisto, kar je ustvarjeno od človeka) je vedno izraz človekove svobode, pravzaprav je tudi človek sam artefakt oz. stvaritev, v tistem delu, v katerem je skozi lastna dejanja zmožen ustvarjanja samega sebe (Galván, 2005). Danes je v razvitem svetu težko najti človeka, ki v sebi ne bi združeval organskih in anorganskih elementov (pomislimo samo na zobne zalivke in implantate, množično precepljenost, srčne spodbujevalnike, zdravila za izboljšanje počutja, slušne aparate, prostetične roke, noge ipd.). Sofisticirane naprave omogočajo tako ohranjanje življenja kot povečevanje razpona njegovih potencialnih funkcij. Brez teh možnosti delež populacije ne bi mogel preživeti. Druga plat pospešenega razvoja nevroznanosti in robotike pa se izraža skozi spreminjajoč se pogled na človeško telo, krepi se koncept telesa kot stroja, katerega nedelujoče ali nezadostno delujoče dele je mogoče preprosto zamenjati oz. nadgraditi (Bockman, 2010: 1317). Tako kartezijsko razumevanje telesa podpira tudi sodobna zahodna medicina in njen prevladujoči biomedicinski model, ki obravnava človekovo telo povsem mehansko in redukcionistično (Mehta, 2011: 203–208).

Znanost sama kot pomembna družbena institucija nosi največjo odgovornost pri uvajanju novih tehnologij in njenih posledic, ki se kažejo na družbenih podsistemih (Illes in Racine, 2005; Russell, 1960: 61–62; Stirling, 2011: 94–95). Na trenutni stopnji intelektualnega razvoja družbe si je nemogoče zamisliti vse bodoče družbene in etične implikacije (kognitivnih) tehnologij. Dozorel odnos družbe v razmerju do tehnoloških dosežkov bi se moral izkazovati zlasti v ponotranjenju previdnostnega načela, tj. v jasnem razumevanju dejstva, da »odsotnost dokazov« ni »dokaz odsotnosti« tveganja (Stirling, 1999: 33–35). Ne smemo zanemariti niti potencialnih sinergičnih učinkov ob stekanju različnih tehnologij (Stirling, 2011: 94).

Pri načrtovanju prihodnosti nikakor ne smemo prezreti »etične brezpogojnosti človeškega dostojanstva« (Mlinar, 2009: 191), ki je tisto vrednostno izhodišče, iz katerega izvajamo vse temeljne človekove pravice (Moka-Mubelo, 2017: 89–90). V kontekstu tega prispevka so pomembne tiste temeljne človekove pravice, ki se manifestirajo tudi kot osebne pravice (Polajnar-Pavčnik, 1997: 150–154), zlasti – gledano s stališča posegov v duševnost drugega – pravica do duševne celovitosti, pravica do zasebnosti in svoboda mišljenja. Pravice do lastnega notranjega duševnega življenja, doživljanja in predstav, zaščite pred vmešavanji drugih v notranje procese posameznika (Finžgar, 1985: 121, 137) ter zaščite duševnih proce-

sov posameznika, ki tvorijo njegove misli in občutke (Bublitz, 2011: 103–107), ne bi smeli votliti pod pritiski neobvladljivega tehnološkega napredka. Da ne bi degradirali in zanikali svoje narave (Constant, 1815: 26), »moramo ohraniti neko minimalno območje človeškega obstoja«, ki »mora ostati zunaj družbenega nadzora«, to pa je »[t]isto, čemur se človek ne more odpovedati, ne da bi prizadel esenco svoje človeške narave« (Berlin, 1958: 126).

Pričujoči prispevek predstavlja dva primera področij uporabe eksperimentalnih tehnologij, ki izkoriščajo kognitivne podatke, tj. odzive človekovega živčnega sistema ob izpostavitvi nekemu dražljaju: kognitivno biometrijo, kot novejšo vejo sistemov za samodejno biometrično prepoznavanje in ugotavljanje identitete uporabnika, in možganske vmesnike, ki omogočajo upravljanje zunanjih naprav preko uporabnikovih namer, brez pomoči perifernega živčevja in mišic. Poskusili bomo nakazati predvsem nekatere etične, pravne in praktične dileme, s katerimi se že srečujemo ob zaenkrat še sporadični praksi uvajanja tovrstnih tehnologij. S povečevanjem dostopnosti in širjenjem njihove uporabe (tudi v obliki tržnih izpeljank tovrstnih tehnologij) bo treba brez nepotrebnega odlašanja zavzeti celovit pristop do problematike v obliki domišljene pravne ureditve, ki bo znala predvideti ustrezna varovala pred nezakonito in neetično uporabo kognitivnih tehnologij v družbeni praksi.

2 Kognitivna biometrija

Čeprav se biometrijo pogosto povezuje z visokotehnološkimi znanstvenimi doseganji in kompleksno tehnološko infrastrukturo, ideja o identifikaciji s pomočjo individualnih značilnosti oseb ne izhaja iz dobe hitrega razvoja informacijsko-komunikacijskih tehnologij. Biometrija kot zasnovan in predlagan (do neke mere tudi izveden) koncept sega v drugo polovico 19. stoletja, ko je Alphonse Bertillon zasnoval identifikacijski sistem, ki je temeljil na antropometriji, antropološki metodi proučevanja merljivih delov človeškega telesa.

Tako je biometrijo v najširšem smislu treba ločevati od tehnologij za samodejno biometrično prepoznavanje in ugotavljanje identitete uporabnika, slednje pa od krovnega pojma tehnologij za samodejno identifikacijo. Strojna in programska oprema, ki spremljata in v osnovi sploh omogočata (samodejno) identifikacijo oz. avtentikacijo z biometrijo, sta najnovejši odgovor znanosti na vprašanje razvoja samodejne identifikacije, ki za sredstvo prepoznavne uporabljajo izmerljive osebne in/ali osebnostne lastnosti uporabnika (Mordini in Massari, 2008: 488), v nasprotju z ostalimi pristopi, ki temeljijo na tem, kaj uporabnik *ve* (npr. gesla) ali kaj uporabnik *ima* (npr. identifikacijska kartica ali elektronski generatorji gesel) (Renaud in Hoskins, 2015: II).

Predpogoj za katerikoli biometrični postopek je obstoj osebnih značilnosti, ki jih je – ob ustreznem in zakonitem postopku – mogoče nedvoumno pripisati le eni

osebi. Šele tedaj je mogoče to značilnost osebe opredeliti kot biometrični identifikator. Kot taka, da je smiselna in zanesljiva, mora zadostiti štirim najosnovnejšim pogojem (Jain, Ross in Prabhakar, 2004: 1-2): (1) pridobljivost (lastnost je mogoče izolirati in izmeriti); (2) splošnost (lastnost je mogoče najti pri vseh osebah); (3) enkratnost (lastnost mora biti lastna le eni osebi in neponovljiva) in (4) trajnost (izmerljive značilnosti morajo ostati nespremenjene, neodvisno od staranja osebe). Med vsemi osebnimi lastnostmi je vsem pogojem zadostil le odtis papirnih linij (Mordini in Massari, 2008: 494).

Razvoj znanosti je prispeval k merljivosti človeškega telesa in njegovih kognitivnih značilnosti na način in v obsegu, ki je bil še pred desetletji nesluten. Kognitivna biometrija, kot nov pristop k identifikaciji in/ali avtentikaciji uporabnika, za samodejno prepoznavanje uporablja odziv človekovega živčnega sistema, kot neke vrste kognitivni podatek (Revett, 2012: 1). Taka naj bi v kombinaciji z drugimi biometričnimi tehnikami povečala stopnjo varnosti na mnogih področjih informacijsko-družbenega življenja – od dostopanja in upravljanja z osebnimi finančnimi sredstvi do preprečevanja operativnih odklonov v zdravstvu. Potrditev pristnosti uporabnika temelji na ujemanju njegovega odziva na sproženi dražljaj z vnaprejšnjim posnetkom možganov, ki služi kot referenčna točka (Palaniappan in Revett, 2014: 96). Ti sistemi uporabljajo osebne značilnosti, bodisi za identificiranje njihovega nosilca bodisi za preverjanje pripadnosti teh značilnosti osebi, ki je bila pooblaščen za uporabo tega sistema (Sprokkereef in de Hert, 2009: 6). Ker omogoča zbiranje edinstvenih bioloških ali vedenjskih značilnosti posameznika, lahko tovrstna tehnologija na različne načine povzroči nova zasebnostna in varnostna tveganja (Sprokkereef in de Hert, 2009: 15), na kar nakazuje zgodovina skoraj slehernega identifikacijskega sistema (Agamben, 2004; Davies, 1994: 44).

Odklon v uporabi (angl. *function creep*) v obliki postopne spremembe namembnosti ali sistema razširitve uporabe podatkov, procesov ali sistemov v namene, za katere je primarno vzpostavljen (Mordini in Massari, 2008: 490), ni nov in spremljač pojav razvoja tehnologij in »informacijske avtoceste«, temveč ljudi spremlja že stoletja. Zaradi postopnosti uvajanja sprememb namembnosti omenjenih podatkov, procesov ali sistemov, povprečni uporabniki pogosto sprememb ne zaznajo ali pa se na njih ne odzivajo. Nesporen in nam najbližji primer je nedvomno slovenska davčna številka, ki je bila vzpostavljena in oblikovana z namenom katalogizacije zavezancev za davek in naj bi se uporabljala izključno v zvezi z vsemi davki, kasneje pa se je njena *de facto* uporaba, nedvomno zaradi pragmatičnosti in univerzalnosti številčnega označevanja, razširila na mnoga druga, z davki (tudi popolnoma) nepovezana področja. Obstoječi sistem, ki zagotavlja visoko integriteto in je pridobil (vsaj relativno) javno zaupanje, nedvomno ustvarja neustavljivo skušnjavo za široko uporabo in medsebojno povezavo številnih do sedaj ločenih zbirk osebnih podatkov (Davies, 1994: 44).

Informacijska tehnologija je v kombinaciji z nevroznanostjo začela razkrivati ozadja atributov osebnosti in značaja, odstirati vpogled v skrite aktivatorje človekovih odločitev, s tem pa še dodatno prispeva k človekovemu inherentnemu nelagodju v obliki nadaljevanega občutka nadzora in negotovosti glede tega, koliko dejstev o sebi je znanih nepoklicanim osebam (Alpert, 2003: 302–303; Illes in Racine, 2005: 6, 14; Murphy, Illes in Reiner, 2008: 294). Biometrija v procesu pridobivanja primerjalnega vzorca oziroma predloge¹ digitalizira fizične lastnosti ali vedenja uporabnikov z namenom hranjenja in obdelovanja pridobljenih informacij. V kontekstu biometričnih tehnologij merjene značilnosti izhajajo iz nezavednega dela posameznika in s tem posegajo v posameznikovo najintimnejšo sfero. Naj zato poleg *širjenja uporabe* informacij o uporabniku preko izvirno zamišljenih okvirov dodamo še (zlo)rabo biometričnih tehnologij za *zbiranje in hranjenje* informacij, ki za postopek biometrične avtentikacije *niso neizogibno potrebne*, vendar jih je mogoče pridobiti v postopku ustvarjanja biometričnih predlog brez dodatnih sredstev, na vsakem koraku oziroma z vsakim modulom, uporabljenim v biometričnem postopku. Ker napredna nevrotehnoška orodja omogočajo dostop in nadzor nad informacijami o posamezniku (Van den Berg, 2011: 421, 423, 428; Bublitz, 2011: 110–114), je težko zagotoviti varovanje skritosti vsega tistega, »česar posameznik ne želi, da se odkrije ali ima interes nad tem, da se ne odkrije« (Finžgar, 1985: 121), saj ni nujno, da se bo posameznik sploh zavedal, kdaj implicitno izdaja kakšna pomembna, a za biometrični postopek ne neizogibno potrebna dejstva o sebi.

V situacijah, ko je posameznik pred izbiro: ali privolitev v razkritje ali izguba določenih pravic, prednosti, privilegijev, koristi, ki imajo lahko v konkretni situaciji za konkretno osebo velik pomen, ne moremo govoriti o prostovoljnosti razkrivanja informacij o sebi (Burkert, 1997: 128; Nabhan, 1998: 876–877). Italijanski filozof Giorgio Agamben v znak protesta ni vstopil v Združene države Amerike, ker je bil za to potreben vnos prstnih odtisov in fotografije v njihovo podatkovno bazo. Biometrijo je opisal kot označevanje, podobno nacističnemu tetoviranju² med drugo svetovno vojno, saj naj bi to človeka osiromašilo golega življenja³ (Agamben, 2004). Subverzivnosti postopkov zbiranja, hranjenja in obdelovanja

¹ Biometrični postopek avtentikacije uporabnika vključuje tri osnovne faze: (1) ustvarjanje predloge in njeno umeščanje v podatkovno bazo; (2) odčitavanje določene telesne, vedenjske ali kognitivne značilnosti in (3) primerjavo s predlogo, ki obstaja v podatkovni bazi. Odločitev o uspešnosti avtentikacije v primeru biometrije ni binarna (kot denimo pri geslih), pač pa temelji na intervalu zaupanja, odstotku, ki razlaga, v kolikšni meri se pridobljene informacije ujemajo s predlogo v podatkovni bazi.

² Ideja o vsesplošnem označevanju oseb vsekakor ni nova. Za potrebe identifikacije pred oblastmi je splošno tetoviranje (imen na zapestja) predlagal že Jeremy Bentham v 18. stoletju.

³ Tehnologizacija telesa prispeva k vse bolj izrazitemu razlikovanju med *zoe* in *bios*, med prvinsko golim življenjem in določeno obliko življenja, kot ju je iz grške filozofije povzel in razvil Giorgio Agamben. *Zoe* je tisto življenje, ki je skupno živalim, ljudem, bogovom in predstavlja le življenje v najožjem smislu, *bios* pa je življenje, ki je lastno le ljudem, saj ga živimo v človeškem kontekstu s pomenom in smislom (Mordini in Massari, 2008: 493).

informacij sicer še zdaleč ne gre pripisovati le biometričnim tehnologijam, vendar velja opozoriti na možnost širjenja uporabe biometričnih sistemov, kar bi (upali bi si trditi, da *bo*) lahko rezultiralo v ekstenzivnih podatkovnih bazah, napolnjenih z najintimnejšimi podatki o uporabnikih, ki bi jih bilo domala nemogoče pridobiti s kakršnimkoli drugimi sredstvi ali kjerkoli drugje.

S članstvom v skupnosti smo implicitno privolili v to, da so splošna dejstva o nas znana navzven, še vedno pa gojimo pričakovanje, da dejstva iz ožjega kroga zasebnosti ostanejo zgolj v naši sferi. Stopnja informacijske zasebnosti, ki jo nekdo uživa, se presoja po tem, koliko ima posameznik možnosti pri omejevanju dostopa do svojih osebnih informacij ali omejevanju količine informacij, ki jih mora – najsibo prostovoljno ali ne – dati tretjim. Informacijska zasebnost je delno utemeljena v posameznikovi želji po ohranjanju lastne anonimnosti in omejevanju fizičnega vdora in opazovalnega nadzora, da ohrani prostor, kjer se mu ni treba prilagajati zahtevam »sprejemljivega« oz. »normalnega« vedenja (Alpert, 2003: 303). Brez nadzora nad informacijami o sebi posameznik ne more vzdrževati osebnih odnosov in podobe o sebi, ki jo želi sporočiti drugim, niti ne uresničevati lastne avtonomije in sprejemati smiselnih odločitev o sebi (Alpert, 1993: 18; Alpert, 2003: 303–304).

3 Vmesniki med možgani in računalniki

Nenehni dosežki na področju nanobiotehnologije, električnega inženiringa, računalništva, socialne robotike, nevrobiologije itd. generirajo pomembne inovacije na področju raziskav interakcije med človekom in robotom, in eksperimentalne tehnologije vedno bolj prenašajo tudi v klinično in industrijsko prakso (Nijboer, Clausen, Allison in Haselager, 2013: 541). Živahno področje umetne inteligence učvrščuje razvoj v smeri novih tipov entitet »učečih se« strojev z vgrajeno sposobnostjo za strojno učenje, kjer se zgodi preskok od le avtomatičnega proti avtonomičnemu delovanju robota in prevzemanja »lastnega kibernetičnega življenja (Allen in Widdison:, 1996: 25–52; Clausen, 2009).

Tako tehnologija možganskih vsadkov oz. vmesnikov med možgani in računalniki (v nadaljevanju BCI – *brain-computer interface*), ki sodi na področje kognitivnega inženiringa, omogoča vgraditev in uporabo neposrednega komunikacijskega kanala za upravljanje zunanjih naprav (računalnika, invalidskega vozička, nevroproteze ipd.) zgolj preko uporabnikovih namer, brez pomoči periferne živčevja in mišic (Lee in Jang, 2013: 72).

BCI naprava najprej preko elektromagnetnega valovanja meri možgansko aktivnost, ki naj bi reprezentirala uporabnikove namere, ki jih nato prevede v kontrolna navodila za zunanje naprave. To prevajanje signalov v ukaze in dejanja vključuje zaporedje računalniškega procesiranja signalov in tehnik za prepoznavanje vzorcev. V prvem koraku je treba zajeti možganske signale uporabnika s snemanjem

njihove električne aktivnosti bodisi preko snemalnih naprav v obliki elektrod na skalpu ali na površini možganov bodisi preko elektrod, ki so vsajene globoko v možgane, neposredno v korteks uporabnika (Hammock, 2010; Lee in Jang, 2013: 72; Tamburrini, 2009: 139). Glede na stopnjo poseganja v biološko tkivo se tehnologija BCI loči na invazivno in neinvazivno oz. polinvazivno (Lebedev, 2014: 99). Že s samo biološko ne/združljivostjo komponent možganskih vsadkov s tkivom preiskovanca in izrabo uporabljenih materialov so povezana določena tveganja. Zaradi vstavitve elektrod v možgansko tkivo lahko pride do lokalnih mehaničnih poškodb tkiva, v določenih primerih pa celo do odpornega odziva organizma na tujek (brazgotinjenje tkiva), zaradi česar sčasoma upade kvaliteta zaznavanja signalov (Klein in Ojemann, 2016: 2). Sistem za merjenje signalov je povezan z računalnikom, ki podatke obdeluje v realnem času (Hammock, 2010). Merjenju oz. snemanju signalov sledi izločanje možganskih signalov od spremljajočih motečih zvokov in njihovo razvrščanje na podlagi klasifikacijskih modelov glede na lastnost signala, tu že preidemo v fazo oblikovanja ukazov in upravljanja z zunanjo napravo oz. umetno okončino in posredovanja povratne informacije uporabniku (Lee in Jang, 2013: 72). Učenje upravljanja z napravo zahteva določeno obdobje treninga, ko se uporabnik nauči izzvati ustrezne možganske signale za svojo namero, da bo lahko z mislimi premikal kurzor in tako posredoval svoja sporočila navzven (Starc, 2010).

Še posebej ključne so možnosti terapevtske uporabe možganskih vmesnikov pri osebah s prizadetostjo senzoričnih ali motoričnih sposobnosti, kjer je zaradi nezmožnosti hotenih gibov onemogočeno komuniciranje med možgani in telesom (Lee in Jang, 2013: 72). V klinični praksi je najbolj izpostavljena uporaba BCI tehnologij pri pacientih s t. i. sindromom zaklenjenosti (ang. *locked-in syndrome*) (Hammock, 2010), ki se opisuje kot redka nevrološka motnja s popolno paralizo voljnih oz. skeletnih mišic na vseh predelih telesa, razen tistih, ki nadzirajo premikanje oči.⁴ Take osebe obdržijo polno zavest in se lahko sporazumevajo z zunanjim svetom preko mežikanja z očmi, saj živčevje, ki povezuje možgane in očne mišice, ni poškodovano (prav tam).

Če računalnik in napravo preko misli usmerja uporabnik (Podgorelec, 2010), se samo po sebi ponuja sklepanje, da gre pri delovanju naprave še vedno za delovanje v sferi uporabnika, za uresničevanje njegovih namer, ki izidejo v želeno oz. zahtevano delovanje. Vendar je ob uporabi možganskih vsadkov uporabnik lahko zaveden z iluzijo delovanja, ko v resnici ne deluje, ali pa je, obratno, dejaven, ko misli, da ne deluje. Vsiljuje se prispodoba avtomatizma, ko se zdi, da se je terapevtski pripomoček osamosvojil od uporabnika in začel delovati v skladu s svojo lastno logiko (Haselager, 2013: 410, 414).

⁴ Gre za povzetek definicije sindroma zaklenjenosti s strani Nacionalnega inštituta za nevrološke motnje in kap (URL: <http://www.ninds.nih.gov/disorders/lockedinsyndrome/lockedinsyndrome.htm> [20. 11. 2017]).

Protokoli delovanja vmesnikov med možgani in računalnikom zahtevajo prenos dela nadzora nad zunanjimi napravami od uporabnika na računalniški sistem. Gre za prenos prepoznavanja uporabnikovih namer, njihovo pretvorbo v ukaz in nadzor nad podrobno izvedbo ukaza (Tamburrini, 2009: 141). Naprava lahko implicitno prevzame nadzor nad procesom, ko ne pride do eksplicitne intervencije uporabnika oz. ko za kratek čas nadomesti voljo uporabnika. Do tega lahko pride v času nezmožnosti zaznavanja možganskih signalov zaradi hrupa, nezmožnosti ustvarjanja ustreznih duševnih stanj uporabnika zaradi motenj pozornosti uporabnika, utrujenosti, tesnobe, ali pa tudi spričo nezavednih odzivov na zaznavne dražljaje (Haselager, 2013: 405). Treba se je namreč zavedati, da naše vedenje vsaj deloma uravnavajo samodejni procesi spontanega sledenja različnim (ne nujno zavedno zaznavnim) sporočilom iz okolja (Camerer, Loewenstein in Prelec, 2005: 11, 37–38). Četudi bi uporabnik npr. želel premakniti napravo v določeno smer, se lahko zgodi, da ne bo uspel izzvati potrebnih duševnih stanj, ki se izrazijo v ustreznih možganskih signalih, za izvedbo take naloge, ali pa bo kljub ustreznemu duševnemu stanju naprava napačno prebrala signale.

V takih situacijah bo naprava usmerjala delovanje brez zavedanja uporabnika, na podlagi strojnega interpretiranja situacije in protokolov predvidevanja uporabnikove namere iz preteklih izkustev. V takem kontekstu posredovanja naprave je prišlo do motnje v verigi ustvarjanja zelenega delovanja, brez zavedanja uporabnika in na način, ki za uporabnika ni bil predvidljiv (Haselager, 2013: 405).

Pri tako kompleksnem nizu interakcij med uporabnikom, robotom in računalnikom vedno obstaja možnost napačnega zaznavanja ali prevajanja možganskih signalov v prognozo uporabnikovega namena za oblikovanje izhodnega ukaza za usmerjanje zunanje naprave (Clausen, 2009: 6. odst). Problematična so že razhajanja med semantičnimi konteksti uporabnika in strojnega sistema, saj se predstave uporabnika ne ujemajo nujno s tistimi zaznavnimi stanji, ki so kot vhodni podatki vprogramirani v robota (Tamburrini, 2009: 143). Upravljanje zunanje naprave lahko zmoti že okoljski dejavnik, denimo hrup, ki lahko kljub pravilni interpretaciji uporabnikove namere pripelje do delovanja, ki ne bo ustrezalo dejanski situaciji oz. hotenemu rezultatu (prav tam). Tekom procesa lahko pride tudi do nezavednih in nehotenih gibov z glavo, kognitivnih procesov (kot npr. miselne asociacije, blag stres), ki lahko zmotijo merjenje in dekodiranje signalov in bistveno spremenijo izhodni podatek, ki se prevede v ukaz za napravo. Uporabnik pa bo zaznal rezultat iz prejšnjega cikla oblikovanja ukaza šele naknadno, ko bo že začel naslednji cikel z novo duševno aktivnostjo (Haselager, 2013: 407).

Tako se lahko zgodi, da kljub najbolj skrbnemu načrtovanju, implementaciji in upravljanju z napravo pride do škodnega dogodka, ko ne bo mogoče vzpostavljati krivdne pravne odgovornosti za dejanje niti programerja, aplikacijskega strokovnjaka ipd. niti uporabnika oz. lastnika naprave, pač pa bo treba razmišljanje prenesti v konceptualni kontekst objektivne odgovornosti, tj. odgovarjanje ne glede

na krivdo za protipravno ravnanje ali opustitev, ali vsaj v sistem objektivne odgovornosti z elementi krivde odgovornosti, kolikor jo bo v konkretnem primeru v določenem deležu sploh mogoče pripisati odgovorni osebi ali osebam. (Clausen, 2009; Tamburrini, 2009: 143) Ker je tehnološki razvoj vgrajen v samo kulturo družbe, industrijske izpeljave tehnologij pa olajšujejo življenje v družbi, ki tudi ustvarja implicitne pritiske po uporabi tovrstnih pripomočkov, bi odgovornost za nenadzorovano in za uporabnika nepredvidljivo delovanje naprav lahko pripisali le družbi kolektivno, ob tem pa bi naleteli na problem, tj. vprašanje, ali gre za nedeljivo kategorijo ali pa je odgovornost mogoče razporediti med posamezne člane te skupnosti in če slednje, še kako in v kolikšni meri (O'Brolchain in Gordijn, 2014: 177, 179). Vendar takšno abstraktno izhodišče ne bi bilo kaj dosti uporabno v primerih konkretnih škodnih dogodkov, ki bi izhajali iz uporabe stvari ali dejavnosti, ki zaradi svojih lastnosti nepredvidljivosti in neobvladljivosti pomeni nevarnost za nastanek škode (analogija s pravnima standardoma nevarne stvari in nevarne dejavnosti v odškodninskem pravu) (Jadek Pensa, 2003: 850–851). V vsakem posameznem primeru bi morali ugotavljati, kdo je tisti, ki poseduje in/ali uporablja tako nevarno stvar ali izvaja dejavnost, kdo ima od nje koristi oz. kdo jo je uvedel v določeno okolje in s tem povzročil potencialno nevarnost za življenje ljudi in okolje (prav tam: 848). Ugotavljanje objektivnih okoliščin na strani oškodovalca, ki opredeljujejo odgovornost za vzrok za škodo (in ne za ravnanje), bi lahko pripeljalo tudi do več deljeno objektivno odgovornih oseb za isti škodni dogodek (praviloma uporabnik take terapevtske naprave ne bo hkrati tudi lastnik naprave ipd.).

Če bi splošni sistem objektivne odgovornosti dopolnjevali z elementi odgovornosti, ki temelji na krivdi povzročitelja za ravnanje ali opustitev, bi si morali, če pogledamo na primeru uporabnika, postaviti naslednja vprašanja. Ali je rezultirajoče dejanje izvirno uporabnikovo, če je do njega prišlo brez intervencije njegovega senzornega in gibalnega sistema, včasih pa celo brez zavestne namere? Lahko rečemo, da uporabnik navzven še vedno deluje kot notranje koherentna enota – avtonomen subjekt v skladu z intencionalno držo (prepričanju in hotenju kot vzroki teh dejanj)? Šele ko bo tehnologija omogočala tudi sprotno beleženje in shranjevanje podatkov o dogajanju na nevralni ravni, bo omogočeno ugotavljanje za nazaj, ali je uporabnikova misel zares sprožila ukaz za nevoljno dejanje naprave oz. neetično posledico (O'Brolchain in Gordijn, 2014: 178–180). Še pred spraševanjem o subjektivnem odnosu uporabnika do posledice (krivda), bi bilo tako treba ugotoviti, ali je imel uporabnik tovrstnih terapevtskih pripomočkov še v oblasti svoje gibe (primer bioničnih okončin), ravnanje in vedenje in kdaj bi ga lahko štel zgolj za *orodje* nekoga/nečesa (npr. načrtovalca programske opreme, po katere navodilih oz. protokolu naprava uresničuje svoj predločen namen) (Bavcon in Šelih, 1996: 240, 142). Gre za vprašanje vzpostavljanja razmerja vzroka in posledice med namero uporabnika, ki pomeni navodilo za napravo, in nastalo prepovedano posledico (prav tam: 147). Te vzročnosti ne bi bilo mogoče priznati v prime-

rih, ko bi se naprava le odzvala na podzavestne vzgibe uporabnika ali njegove napačne zaznave (prisluihi, vidne iluzije), ki se jih uporabnik ne bi niti zavedal. Ne moremo se strinjati z glediščem, po katerem naj bi uporabnik že s tem, ko se je »odločil« (praviloma niti nima prave »izbire«, saj gre za omogočanje njegovih vitalnih življenjskih funkcij ob pomoči take naprave) za vgradnjo možganskih vmesnikov, implicitno sprejel tudi svojo odgovornost za celotno prihodnje delovanje mehanizma (O'Brolchain in Gordijn, 2014: 178–179). Takšno vnaprejšnje pripisovanje eventualnega naklepa uporabniku za vsa *njegova* bodoča ravnanja (torej da se je uporabnik že s sprejetjem terapijskega pripomočka zavedal, »da lahko zaradi njegovega ravnanja nastane prepovedana posledica (oz. posledice)« in »privolil«, da posledica (oz. posledice) nastane (Bavcon in Šelih, 1996: 225)), ne more biti dopustno. V primeru, da bi do neželenega dogodka prišlo tudi zaradi pomanjkljivih ali ne dovolj jasnih navodil aplikacijskega strokovnjaka uporabniku glede rokovanja z napravo, bi del krivde za svoj prispevek k škodi, poleg uporabnika, moral nositi tudi tak strokovnjak.

Spajanje človeško organskega in mehničnega ustvarja novo hibridno entiteto, pri čemer ne moremo izključiti možnosti, da bi vgrajena programska oprema nasprotovala notranji logiki delovanja uporabnika (Allhoff, 2010: 14). Poleg tega so možni tudi zelo dramatični stranski učinki za uporabnika, kot so motnje živčnega sistema, psihiatrične motnje, sprememba vzorcev vedenja, sistema preferenc, potencialni nehoteni učinki na kvaliteto življenja, odnosov z bližnjimi, soočimo pa se tudi s pomisleki spreminjanja človekove osebnosti (Clausen, 2009: 7. in 9. odst.; Fuchs, 2006: 600, 603). Možnost povezave možganov z računalnikom bi lahko v preveliki meri obremenila spominske zmožnosti bioloških možganov, kako asimilirati vse možne vhodne podatke v konsistentno sliko. Prav tako je mogoče, da s tehnološko optimizacijo določene človeške sposobnosti/funkcije posežemo v drugo, s prvo subtilno povezano sposobnost/funkcijo (Wallach, 2011: 204). S tem se slabi oz. ruši koncept duševne celovitosti uporabnika in njegova zmožnost sprejemanja koherentnih odločitev.

Kljub nespornemu priznanju dosežkov pri olajšanju življenja funkcionalno oviranih ljudi, ni odveč etični pomislek, da gre pri tem vsaj deloma za »tiranijo normalnega«, saj se s tovrstno tehnologijo prizadeta oseba prilagaja svetu, ki je oblikovan za »normalne« ljudi, s tem pa tudi sama prispeva k družbenemu dojemanju prizadetih oseb kot inferiornih (Clausen, 2009: 15. odst.). Vir zadreg je tudi dejstvo, da tovrstna tehnologija z računalniško spodbudo presega tisto, kar šteje za le normalno funkcioniranje (Koops in Leenes, 2011: 195–196; Merkel, 2007: 290–292; Rosahl, 2007: 136–137). Bistvo tovrstnih izboljšav bi moralo biti prizadevanje – preko omogočanja normalnega funkcioniranja posameznikov – po izboljšanju življenja članov skupnosti, ne pa ustvarjanje izboljšanih verzij posameznikov (kodiranje zaželenih telesnih funkcij, sposobnosti, delovanja, kognitivnega instrumentarija ipd. (Brownsword, 2006: 10–11).

3.1. Izzivi pridobivanja soglasja za terapevtski poseg

V skladu z načelom avtonomije ima vsakdo pravico sprejemati odločitve o samem sebi (*Voluntas aegroti suprema lex esto*).⁵ Pred vsakim medicinskim postopkom, raziskavo ali zdravljenjem se tako, v uresničevanju njegove pravice do samoodločbe, zahteva soglasje pacienta oz. preiskovanca za tak poseg, nobenega posega v njegovo telo mu ni mogoče vsiliti (razen v izjemnih zakonsko določenih primerih).⁶ Ustavna prepoved prisilnega zdravljenja⁷ je podrobneje izpeljana v Zakonu o pacientovih pravicah (ZpacP).

Posamezniku, ki sprejema odločitev, mora biti na njemu razumljiv način predstavljena njegova zdravstvena situacija, možni postopki zdravljenja in potencialne koristi ter tveganja (Klein in Ojemann, 2016: 1). Privolitvena sposobnost se presoja v vsakem posamičnem primeru, navezuje pa se na stopnjo njegove razsodnosti, tj. sposobnosti razumeti pomen in posledice lastne odločitve. Odločitev za poseg mora biti sprejeta v neprisilnem okolju in kontekstu, torej popolnoma prosta vplivov (prav tam). V postopku pridobivanja soglasja je treba preiskovancu razložiti raziskovalno naravo postopka, se zadrževati pred zbujanjem nerealnih pričakovanj pri preiskovancu in poudariti njegovo pravico, da ima vsak hip možnost umakniti svoje soglasje (Klein, 2016: 1313).

A kako pridobiti veljavno soglasje, če so možgani, kot organ privolitve, poškodovani? Pri preiskovancu, ki je nezmožen ali omejeno zmožen komuniciranja, bo ob pridobivanju soglasja za poseg treba tehtati med njegovo pravico do najvišjega dosegljivega fizičnega in mentalnega zdravja in potencialnimi fiziološkimi in emocionalnimi poškodbami, povezanimi s tehnologijo možganskih vsadkov. V primeru, da bi bila vstavev možganskega vmesnika edini možen način za vzpostavitev vsaj delne komunikacije preiskovanca s svetom, bi bilo tu mogoče soglasje prizadete osebe domnevati do vzpostavitve vsaj minimalne komunikacijske odzivnosti, ko bo preiskovanec sposoben sam izraziti voljo po nadaljevanju ali prekinitvi preiskave (Klein, 2016: 1306)? Družinski člani oz. skrbniki, ki v določenih primerih nadomestijo voljo za soglasje pacienta k posegu, imajo interes po povečanju samostojnosti preiskovanca v skrbi zase, saj imajo sami od tega posredne koristi, izhajati pa bi bilo treba le iz največje koristi za preiskovanca (Klein, 2016: 1308).

V takih primerih bi lahko bila v veliko pomoč funkcijsko-slikovna orodja proučevanja možganov, ki prinašajo tudi nove metode razumevanja in razlikovanja med različnimi stopnjami motenj zavesti in s tem možnost natančnejše in objektivnejše

⁵ Latinsko za: Pacientova volja je vrhovni zakon.

⁶ Omejitve pravice do samoodločanja za pacienta določa Zakon o pacientovih pravicah v 4. odstavku 4. člena in sicer, kadar je to potrebno zaradi zagotavljanja javne varnosti, varovanja javnega zdravja ali/in varovanja pravic drugih oseb (Zakon o pacientovih pravicah (ZpacP), Ur. l. RS, št. 15/08.).

⁷ »Nikogar ni mogoče prisiliti k zdravljenju, razen v primerih, ki jih določa zakon.« (3. odst. 51. člena URS).

ocene pacientovega duševnega stanja, kar je lahko ključno pri objektivnem ugotavljanju, kdaj gre za primere z realno možnostjo zdravljenja (Fins, 2011: 336; Habermas, 2007: IV; Kabraji, Naylor in Wood, 2008: 9). Ker taka tehnološka orodja lahko pravočasno prepoznajo prehod iz trajno vegetativnega stanja v stanje minimalnega zavedanja, ko je že mogoče zaznati minimalne, a določne znake zavedanja samega sebe in okolice, bi lahko v nadaljevanju take paciente z vsaj minimalno zmožnostjo odzivnosti s pomočjo BCI naprav pritegnili v odločanje o sebi ter tako zdravnika vsaj nekoliko razbremenili težkih etičnih dilem (Habermas, 2007: IV; Kabraji, Naylor in Wood, 2008: 9; Owen et al., 2006: 1402; Pleterski-Rigler in Trontelj, 2001: 477–80). Paciente, pri katerih bodo z diagnostiko preslikavanja možganov ugotovili ohranjeno ali na novo pridobljeno sposobnost komuniciranja, bo mogoče s pomočjo sporazumevanja preko živčnega signala z vgraditvijo možganskih vsadkov tudi aktivneje vključevati v proces odločanja o njihovem nadaljnjem zdravljenju in jih postopoma tudi ponovno vključevati v družbo (Fins, 2015: 124–125; Kabraji, Naylor in Wood, 2008: 10).

Tudi po že veljavno vzpostavljeni začetni volji za poseg lahko tekom izvajanja preiskave oz. terapije pride do problematičnih situacij. Slabitev kognitivnih zmožnosti preiskovanca tekom izvajanja preiskave oz. terapije lahko postavi pod vprašaj veljavnost začetnega soglasja glede nadaljevanja izvajanja postopka, ker preiskovanec iz različnih razlogov ne izpolnjuje več pogoja razsodnosti za privolitveno sposobnost. V primeru kakršnegakoli suma v zmanjšanje zmožnosti za sprejem odločitve za poseg, je treba obstoj soglasja za nadaljevanje preiskave preveriti. To ni le etična, ampak tudi varnostna zahteva, saj že sama tehnologija možganskih vsadkov ob poslabšanju kognitivnih zmožnosti uporabnika v primerjavi z izhodiščnimi izgubi velik del svoje zanesljivosti, kar lahko vodi tudi do tveganih situacij (Klein, 2016: 1305).

4 Razprava

Zdi se, da imamo ljudje inherentno potrebo po primerjanju sebe z drugimi (Davis, 1995: 24). V družbi vlada hegemonija normalnosti, ki se uveljavlja skozi t. i. proces normalizacije, tj. stalnega procesiranja, primerjanja, oblikovanja podob »normalnosti« in temu kontrastne »nenormalnosti« oz. prizadetosti (prav tam: 44). Vzdrževanje te binarnosti je družbeni konstrukt, katerega vsebino zapolnjujejo potrebe in zahteve določenega časa in prostora (prav tam: 2–5, 9–10). S tržno dostopnostjo aplikativnih tehnologij in terapevtskih postopkov ter pripomočkov nam družba vsiljuje »dolžnost« uporabiti jih brez oprijemljive možnosti umika tej zapovedi družbenega konteksta, ki ga s svojo poslušnostjo dejansko soustvarjamo in vzdržujemo in v katerem koreninijo naše predstave o sebi in svetu (Iannitti, 2004: 34–35). Avtonomija se na tem področju ne more udejanjiti (prav tam: 24, 32–33, 63), saj bi odklonitev uporabe določene tehnološke aplikacije oz. postopka, ki bi bil na voljo, in ki bi končno pomenila zgolj popolno izpeljavo etičnega načela av-

tonomije izbire in odločanja (možnost izbrati tok akcije med več razpoložljivimi možnostmi) (O'Connor, 2016: I) – pravica sprejeti postopek in pravica zavrniti ga –, sprožila mehanizem družbenega sankcioniranja. Vsako človeško življenje pa je – ne glede na morebitne pomanjkljivosti, drugačnosti in nezmožnosti uživanja materialnih in socialnih koristi sodobne družbe – izvorni in končni nosilec dostojanstva in ga je nedopustno prisiljevati v prilagajanje merilom družbeno pripisane identitete (Iannitti, 2004: 61-72). Široka dostopnost tehnologij bi lahko v svoji skrajni posledici odprla možnost sprejemanja povsem potrebam posameznika prilagojenih evgeničnih odločitev (Van Camp, 2012: 36) glede osebno ali družbeno zaželenih ali nezaželenih lastnosti in nagnjenj. Stopnjevanje patosa nad vsezmožnostjo znanosti pa lahko vodi do momenta nadčloveškega dojemljanja življenja in zavračanja slehernega odstopanja v lastni biološki zasnovi od predpostavljeno »normalnega« kot nevrednega obstoja. Popolna izpeljava avtonomije pri upravljanju z lastnim telesom – ki je, protislovno, z uklanjanjem posameznika diktatu sicer vedno izmuzljive popolnosti, hkrati vse bližje izvirnosti pojma avtonomije (prav tam: 53) – odpira tveganja, da človek v svojem samoustvarjalnem prizadevanju nenamerno postane pretirano izboljššan, postčloveški, odtujen od svoje samodoločevalne naravnosti (Academy of Medical Sciences, 2012: 45).

S tem ko se ustvarja vtis, da je z uporabo tehnologije mogoče odpraviti sleherni odklon od »statistično značilnega funkcioniranja vrste« (Kingma, 2007: 128–131) oz. tistega, kar naj bi bilo »normalno«, se spodkopava solidarnostni temelj družbe: občutek odgovornosti do šibkejših članov družbe (do tistih z manj sreče z naravno loterijo) se temu ustrezno zmanjša, na ta način pa se družbi odvzame priložnost za kultiviranje občutka odprtosti do izrazov drugačnosti (Caldera, 2008; Sandel, 2004: 25–62 odst.). Sčasoma bi lahko v družbi, ki bi vsako duševno abnormalnost, neprilagojeno osebnostno lastnost ali funkcioniranje telesa, ki odstopa od pričakovano značilnega funkcioniranja, dojemala kot napako, ki jo je moč odpraviti, prišlo do izrazitega preoblikovanja pogleda na (telesno ali duševno) prizadetost, čemur bi sledila zmanjšana skrb in strpnost do prizadetih, saj bi bili družbeno sprejemljivi le popolni⁸ ljudje (Academy of Medical Sciences, 2012: 6, 23–24, 37, 45–46). Na tem mestu je treba opozoriti, da gre proces v smer približevanja, mestoma pa celo preseganja običajnih človeških funkcij, zato je marsikje že močno zabrisana meja med terapevtskimi posegi⁹ (obnovitev okvarjene, oslabele ali nedelujoče naravne funkcije) in posegi izboljševanja človekove biološke

⁸ Popolnost je v tem kontekstu treba razumeti kot popolno doseganje tistega, kar v družbi šteje za normalno. Ko je normalnost družbena norma in je vsako odstopanje od normalnega nepopolno, je tisto, kar naj bi bilo normalno, popolno.

⁹ Težavna pa je že sama opredelitev zdravja in bolezni, ki za sabo potegne določitev meje med izboljšavo in terapijo (Clausen, 2009: 14. odst.; Koops in Leenes, 2011: 203). Terapija naj bi bila namenjena obravnavi patologij, ki zmanjšujejo zdravje ali nivo funkcioniranja pod mejo, ki je tipična za vrsto ali ki je statistično normalna, vse, kar pomeni razširjanje človeških sposobnosti nad to mejo, pa se tako šteje za nadgrajevanje oz. krepitev (Allhoff, 2010; Bockman, 2010: 1317, 1333).

osnove (Koops in Leenes, 2011: 202–204, 210–211). Nekateri avtorji (npr. Goertzel in Bugaj, 2006; Laughlin, 1997) celo menijo, da smo priča procesu »kiborgizacije«, ¹⁰ v smislu procesov nadaljevanega tehnološkega poseganja v človeško telo, ki vključuje nadomeščanje ali razširjanje struktur, ki so odgovorne za različne fizične in duševne attribute, ki jih običajno štejemo za človeku lastne (Saniotis, 2009: 6).

Ena od ključnih nalog države je, da vsem državljanom zagotovi pravičen dostop do dosežkov kolektivnega napredka, kamor zagotovo spadajo tudi napredna tehnološka (zlasti terapijska) orodja, v nasprotnem bi se morali »mali« ljudje zadovoljiti s konvencionalno zdravniško obravnavo, privilegirani pa bi imeli boljše možnosti zdravljenja in ohranjanja oz. krepitev svojih duševnih/telesnih funkcij. Človekovo blagostanje pa ni zgolj rezultat funkcionalne odličnosti njegovega organizma, ampak preplet izpolnjenosti določenih bioloških, izkustvenih in družbenih pogojev, kot okvir, v katerem lahko deluje človek kot integrirana celota (Saniotis, 2009: 8).

Introducing Experimental Cognitive Technologies into Therapeutic and Social Practice

Increasingly advanced technological tools are becoming progressively integrated into the lives of individuals and society. The process of integration can even go to the extent that a technological tool merges with its user, which can result in the creation of new types of entities. The paper presents two examples of cognitive technologies (cognitive biometrics and brain-computer interfaces) in an attempt to highlight some of the ethical, legal and practical dilemmas which arise from their use. Increasing accessibility to these originally-unforeseen technological products and the unavoidable expansion of their use germinate new ways of interference with human personality. Therefore, a comprehensive legal framework should be drafted without undue delay in order to introduce reasonable and adequate safeguards against the unlawful and unethical use of cognitive technologies in social practice.

Keywords: cognitive technologies, brain-computer interface, cognitive biometrics, accountability, risks

¹⁰ »Kiborg« je izraz za entiteto, ki združuje organske in mehanske elemente in informatiko. Prva resna znanstvena pobuda in utemeljitev kiborga je prišla s strani enega najbolj znanih in kontroverznih britanskih znanstvenikov – Johna Desmonda Bernala – v 20. letih prejšnjega stoletja.

Literatura

- Academy of Medical Sciences (2012). *Human Enhancement and the Future of Work*. Dostopno na: https://www.britac.ac.uk/sites/default/files/12308_academy_HE_report_2012_web.pdf. [18. 11. 2017].
- Agamben, G. (2004). *No to Bio-Political Tattooing*. Dostopno na: <https://ratical.org/ratville/CAH/totalControl.html> [20. 11. 2017].
- Allen, T. in Widdison, R. (1996). »Can Computers Make Contracts?«. *Harvard Journal of Law in Technology*, 9(1), str. 25–52.
- Allhoff, F. (2010). »Ethics of Human Enhancement: 25 Questions and Answers«. *Studies in Ethics, Law, and Technology*, 4(1), str. 1–39.
- Alpert, S. A. (1993). »Smart Cards, Smarter Policy: Medical Records, Privacy, and Health Care Reform«. *The Hastings Center Report*, 23(6), str. 13–23.
- Alpert, S. A. (2003). »Protecting Privacy in the Age of Genetic Information«. *Journal of Social Issues*, 59(2), str. 301–322.
- Bavcon, L. in Šelih, A. (1996). *Kazensko pravo. Splošni del*. Ljubljana: Uradni list Republike Slovenije.
- Berlin, I. (1958). »Two Concepts of Liberty«. V Berlin, I. (ur.), *Four Essays on Liberty*, Oxford: Oxford University Press, str. 118–172.
- Bockman, C. R. (2010). »Cybernetic-Enhancement Technology and the Future of Disability Law«. *Iowa Law Review*, 95, str. 1315–1340.
- Brownsword, R. (2006). »Neither East Nor West, Is Mid-West Best?«. *A Journal of Law, Technology in Society*, 3(1), str. 15–33.
- Bublitz, J. C. (2011). »If man's true palace is his mind, what is its adequate protection? On a right to mental self-determination and limits of interventions into other minds«. V Van den Berg, B. in Klaming, L. (ur.), *Technologies on the Stand: Legal and Ethical Questions in Neuroscience and Robotics*, Nijmegen: Wolf Legal Publishers, str. 95–121.
- Burkert, H. (1997). »Privacy-Enhancing Technologies: Typology, Critique, Vision«. V *Technology and Privacy: The New Landscape*, London: The MIT Press, str. 125–146.
- Caldera, E. O. (2008). »Cognitive Enhancement and Theories of Justice: Contemplating the Malleability of Nature and Self«. *Journal of Evolution in Technology*, 18(1), str. 1–8.

- Camerer, C. F., Loewenstein, G. F. in Prelec, D. (2005). »Neuroeconomics: How Neuroscience Can Inform Economics«. *Journal of Economic Literature*, 43(1), str. 9–64.
- Clausen, J. (2009). »Man, Machine, and in Between«. *Project Syndicate* (spletna stran). Dostopno na: <http://www.project-syndicate.org/commentary/man--machine--and-in-between> [15. 11. 2017].
- Constant, B. M. (1815). *Principes de politique applicables à tous les gouvernements représentatifs et particulièrement à la Constitution actuelle de la France*. Paris: Alexis Eymery.
- Davies, S. G. (1994). »Touching Big Brother: How Biometric Technology Will Fuse Flesh and Machine«. *Information Technology in People*, 7(4), str. 38–47.
- Davis, L. J. (1995). *Enforcing normalcy: disability, deafness, and the body*. London, New York: Verso.
- Fins, J. J. (2011). »Neuroethics, Neuroimaging, and Disorders of Consciousness: Promise or Peril?«. *Transactions of the American Clinical and Climatological Association*, 122, str. 336–346.
- Fins, J. J. (2015). *Rights Come to Mind: Brain Injury, Ethics, and the Struggle for Consciousness*. New York: Cambridge University Press.
- Finžgar, A. (1985). *Osebnostne pravice*. Ljubljana: Slovenska akademija znanosti in umetnosti.
- Fuchs, T. (2006). »Ethical Issues in Neuroscience«. *Current Opinion in Psychiatry*, 6(19), str. 600–607.
- Galván, J. M. (2005). »Techno-ethical and anthropological aspects of the Synergy of Natural and Artificial Hardware«. *CRA 2005, IEEE International Conference on Robotics and Automation - Workshop on Robo-Ethics*, Barcelona, 18. april 2005. Dostopno na: <http://www.roboethics.org/icra2005/galvan.htm> [24. 11. 2017].
- Goertzel, B. in Bugaj, S. V. (2006). *The Path to Posthumanity: 21st Century Technology and Its Radical Implications for Mind, Society and Reality*. Bethesda: Academica Press, LLC.
- Haberfeld, E. (2007). *Neuroimaging: Visualizing Brain Structure and Function – Foundation Text*. Dostopno na: <http://ccnmtl.columbia.edu/projects/neuroethics/module1/foundationtext> [15. 11. 2017].
- Hammock, A. (2010). »The future of brain-controlled devices«. *CNN* (spletna novice), 4. januar 2010. Dostopno na: <http://edition.cnn.com/2009/TECH/12/30/brain.controlled.computers/index.html> [7. 11. 2017].

Haselager, P. (2013). »Did I do that? Brain-computer interfacing and the sense of agency«. *Minds and Machines*, 23(3), str. 405–418.

Iannitti, V. E. (2004). *Manufacturing eugenics: the role of scientific norms and liberal values in the debate over prenatal genetic technologies*. University of British Columbia.

Illes, J. in Racine, E. (2005). »Imaging or imagining? A neuroethics challenge informed by genetics«. *American Journal of Bioethics*, 5(2), str. 5–18.

Jadek Pensa, D. (2003). »Komentar k 133. - 154. členu OZ«. V Juhart, M. in Plavšak, N. (ur.), *Obligacijski zakonik s komentarjem (splošni del): 1. knjiga (1. do 189. člen)*. Ljubljana: GV Založba, str. 760–889.

Jain, A. K., Ross, A. in Prabhakar, S. (2004). »An Introduction to Biometric Recognition«. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1), str. 4–20. Dostopno na: <https://doi.org/10.1109/TCSVT.2003.818349> [20. 11. 2017].

Kabraji, S., Naylor, E. in Wood, D. (2008). »Reading minds? Ethical implications of recent advances in neuroimaging«. *Penn Bioethics Journal*, 4(2), str. 9–11.

Kingma, E. (2007). »What is to be healthy?«. *Analysis*, 67(2), str. 128–133.

Klein, E. (2016). »Informed Consent in Implantable BCI Research: Identifying Risks and Exploring Meaning«. *Science and Engineering Ethics*, 22(5), str. 1299–1317.

Klein, E., in Ojemann, J. (2016). »Informed consent in implantable BCI research: identification of research risks and recommendations for development of best practices«. *Journal of Neural Engineering*, 13(4), str. 43001.

Koops, B.-J. in Leenes, R. (2011). »Cheating with implants: Implications of the hidden information advantage of bionic ears and eyes«. V Van den Berg, B. in Klaming, L. (ur.), *Technologies on the Stand: Legal and Ethical Questions in Neuroscience and Robotic*, Nijmegen: Wolf Legal Publishers, str. 195–214.

Laughlin, C. D. (1997). »The Evolution of Cyborg Consciousness«. *Anthropology of Consciousness*, 8, str. 144–159.

Lebedev, M. (2014). »Brain-machine interfaces: an overview«. *Translational Neuroscience*, 5(1), str. 99–110.

Lee, K.-Y. in Jang, D. (2013). »Ethical and social issues behind brain-computer interface«. *2013 International Winter Workshop on Brain-Computer Interface (BCI)*, str. 72–75.

Mehta, N. (2011). »Mind-body Dualism: A critique from a Health Perspective«. *Mens Sana Monographs*, 9(1), str. 202–209.

- Merkel, R. (2007). *Intervening in the Brain: Changing Psyche and Society*. Berlin, Heidelberg: Springer-Verlag.
- Mlinar, A. (2009). »Bioetika, človekove pravice in trajnostni razvoj«. V Škof, L. (ur.), *Človekove pravice, Poligrafi: revija za religiologijo, mitologijo in filozofijo*, 14(53/56), Ljubljana: Nova revija, str. 161–220.
- Moka-Mubelo, W. (2017). *Reconciling Law and Morality in Human Rights Discourse: Philosophy and Politics - Critical Explorations, vol 3*. Cham: Springer.
- Mordini, E. in Massari, S. (2008). »Body, biometrics and identity«. *Bioethics*, 22(9), str. 488–498.
- Murphy, E. R., Illes, J. in Reiner, P. B. (2008). »Neuroethics of Neuromarketing«. *Journal of Consumer Behaviour*, 7(4-5), str. 293–302.
- Nabhan, A. (1998). »Review: Technology and Privacy: The New Landscape«. *Harvard Journal of Law in Technology*, 11(3), str. 871–880.
- Nijboer, F., Clausen, J., Allison, B. Z. in Haselager, P. (2013). »The asilomar survey: Stakeholders' opinions on ethical issues related to brain-computer interfacing«. *Neuroethics*, 6(3), str. 541–578.
- O'Brolchain, F. in Gordijn, B. (2014). »Brain-Computer Interfaces and User Responsibility«. V Grübler, G. in Hildt, E. (ur.), *Brain-Computer-Interfaces in their ethical, social and cultural contexts*, Dordrecht, Heidelberg, New York, London: Springer, str. 163–182.
- O'Connor, T. (2016). »Free Will«. V Zalta, E. N. (ur.), *The Stanford Encyclopedia of Philosophy* (izdaja poletje 2016). Dostopno na: <http://plato.stanford.edu/archives/sum2016/entries/freewill/> [19. 11. 2017].
- Owen, A. M. et al. (2006). »Detecting awareness in the vegetative state«. *Science*, 313(5792), str. 1402.
- Palaniappan, R. in Revett, K. (2014). »Pin generation using EEG: A stability study«. *International Journal of Biometrics*, 6(2), str. 95–105.
- Pleterski-Rigler, D. in Trontelj, J. (2001). »Ko bolnik ne sodeluje ali ne more sodelovati«. *Zdravniški vestnik*, 70(9), str. 477–480.
- Podgorelec, D. (2010). *Uporabniški vmesniki, I. poglavje, Uvod HCI in UI: predavanja*. Univerza v Mariboru (UM), Fakulteta za elektrotehniko, računalništvo in informatiko (FERI). Dostopno na: <http://sciget.com/Predogled/3856/08d739ee4a7a00c67b3cc18cc65d79f20ff39e94> [12. 11. 2017].
- Polajnar-Pavčnik, A. (1997). »Temeljne pravice kot osebnostne pravice«. V Pavčnik, M., Polajnar-Pavčnik, A. in Wedam-Lukić, D. (ur.), *Temeljne pravice*, Ljubljana: Cankarjeva založba (Zbirka Pravna obzorja), str. 150–172.

Renaud, K. in Hoskins, A. (2015). »Biometric Identification: Are We Ethically Ready?«. *Information Security for South Africa*.

Revett, K. (2012). »Cognitive biometrics: a novel approach to person authentication«. *International Journal of Cognitive Biometrics*, 1(1), 1–9. Dostopno na: <https://doi.org/10.1504/IJCB.2012.046516> [15. 11. 2017].

Rosahl, S. K. (2007). »Medicine and society Neuroprosthetics and neuroenhancement: can we draw a line?«. *Virtual Mentor - American Medical Association Journal of Ethics*, 9(2), str. 132–139.

Russell, B. (1960). »The Social Responsibilities of Scientists«. *Science*, 131(3398), str. 391–392.

Sandel, M. J. (2004). »The Case against Perfection«. *Atlantic Monthly* (spletna verzija), april 2004. Dostopno na: <http://www.theatlantic.com/magazine/archive/2004/04/the-case-against-perfection/302927/> [19. 11. 2017].

Saniotis, A. (2009). »Future Brains: An Exploration of Human Evolution in the 21st Century and beyond«. *World Future Review*, 1(3), str. 5–11.

Sprokkereef, A. in de Hert, P. (2009). *The Use of Privacy Enhancing Aspects of Biometrics: Biometrics as PET (privacy enhancing technology) in the Dutch private and semi-public domain*. Tilburg: Tilburg Institute for Law, Technology and Society.

Starc, M. (2010). »Možganski vmesniki«. *Predstavitveni poster na dogodku »Teden možganov«* (v okviru mednarodne izobraževalne akcije »Brain Awareness Week«), Sinapsa, 15.-21. marec 2010. Dostopno na: <http://www.sinapsa.org/tm/media/-pripone/a187-Mozganski%20vmesniki.pdf> [15. 11. 2017].

Stirling, A. (1999). *On science and precaution in the management of technological risk: volume I - a synthesis report of case studies*. Institute for Prospective Technology Studies. Dostopno na: <http://ftp.jrc.es/EURdoc/eur19056en.pdf> [20. 11. 2017].

Stirling, A. (2011). »Governance of neuroscience: challenges and responses«. V *Brain Waves Module 1: Neuroscience, society and policy*, London: The Royal Society, str. 87–97.

Tamburrini, G. (2009). »Brain to computer communication: Ethical perspectives on interaction models«. *Neuroethics*, 2(3), str. 137–149.

Van Camp, N. (2012). »Heidegger and the Question Concerning Biotechnology«. *Journal of Philosophy of Life*, 2(1), str. 32–54. Dostopno na: <http://www.philosophyoflife.org/jpl201203.pdf> [11. 11. 2017].

Van den Berg, B. (2011). »Techno-elicitation: Regulating behaviour through the design of robots«. V Van den Berg, B. in Klaming, L. (ur.), *Technologies on the stand: Legal and ethical questions in neuroscience and robotics*, Nijmegen: Wolf Legal Publishers, str. 417–436.

Wallach, W. (2011). »From Robots to Techno Sapiens: Ethics, Law and Public Policy in the Development of Robotics and Neurotechnologies«. *Law, Innovation and Technology*, 3(2), str. 185–207.

Pravni viri

Zakon o pacientovih pravicah (ZPacP), Ur. l. RS, št. 15/08.

Ustava Republike Slovenije (URS), Ur. l. RS, št. 33/91I, 42/97 – UZS68, 66/00 – UZ80, 24/03 – UZ3a, 47, 68, 69/04 – UZ50, 69/04 – UZ43, 69/04 – UZ14, 68/06 – UZ121,140,143, 47/13 – UZ90,97,99, 47/13 – UZ148.

Tjaša Zapušek

Pravna fakulteta Univerze v Kopenhagnu

Umetna inteligenca in njena pravna ureditev

Prisotnost umetne inteligence je mogoče zaznati na praktično vseh področjih našega življenja. V zadnjih desetletjih je začela uspešno prodirati tudi na trg dela. V nadaljevanju bodo na primeru medicinskih robotov predstavljene različne »razvojne stopnje robotov«, ki zahtevajo različne pravne pristope. Trenutno se namreč soočamo s pomanjkanjem ustrezne zakonodaje, ki bi urejala delovanje robotov, opremljenih z algoritmi, ki ne vsebujejo vnaprej predvidenih vzorcev reševanja naloge, temveč jih med reševanjem naloge glede na zaznave iz okolja ustvarjajo algoritmi sami. Kljub temu da naj bi roboti sčasoma izboljšali učinkovitost delovnega procesa, pri tem ne smemo pozabiti na ključnost pravne ureditve, ki bo predvidela osebo, odgovorno za škodo, ki jo bodo povzročili takšni roboti. Članek poleg ocene različnih možnosti glede pravne ureditve vsebuje tudi zamisli o preureditvi delovnega razmerja. Ta bi nas lahko pripeljala do zelenega cilja – do izdelave zakonodaje, ki bi zavarovala osebe v primeru škode, ki bi jo povzročili roboti.

Ključne besede: umetna inteligenca, pravna ureditev, medicinski robot, algoritem, delovno razmerje, razvojne stopnje robota

Uvod¹

»Imagination is more important than knowledge. Knowledge is limited. Imagination encircles the world.«²

Albert Einstein

¹ Zakonodajni predlogi, ki so predstavljeni v tem prispevku, so del širšega raziskovalnega projekta na Univerzi v Princetonu v Ameriki, kjer sem skupaj s petimi profesorji napisala poglobljeno pravno analizo. Ta je bila posredovana institucijam, ki se trenutno osredotočajo na kreiranje pravne ureditve umetne inteligence. Znotraj omenjenega projekta sem razdelila robote (glede na njihovo uporabnost) na dve skupini: robote, namenjene zasebni uporabi (avtomatizirana prevozna sredstva, roboti kot asistenti za starejše ljudi ipd.), in robote, vključene v delovne procese. V tem prispevku sem se osredotočila zgolj na slednje. Osnovna ideja je bila že predstavljena v *Asia Pacific Journal of Health Law and Ethics, Volume 11* ter na mednarodni konferenci *Data driven Health Care* v Seoulu, Južni Koreji.

² Slovenski prevod: »Domišljija je bolj pomembna kot znanje. Znanje je omejeno. Domišljija pa obkroža svet.«

Albert Einstein je leta 1929 med intervjujem za revijo *Post* poudaril, da domišljija dominira nad znanjem. Znanje je zgolj omejeno, domišljija pa obkroža cel svet (J. Nilsson, 2010). Nič čudnega, da sta ustvarjalnost in domišljija lastnosti človeške narave, ki ju je najtežje avtomatizirati. Ustvarjalnost posameznika je plod njegove domišljije, saj skozi čustveno dožemanje okolice in njenih lastnosti presega že pridobljeno znanje in svoje ustaljene navade. Računalniški sistemi pa za svojo ustvarjalnost ne potrebujejo domišljije, potrebujejo samo znanje (podatke), s pomočjo katerega ustvarijo nek rezultat.

V začetku prispevka bo prikazan proces približevanja umetne in človeške inteligence, ki je rezultat hitrega razvoja na področju algoritmov in dejstva, da znanstveniki z dopuščanjem krepitve avtonomnosti algoritmov omogočajo osvajanje večšin človeške inteligence. V nadaljevanju pa bo predstavljena pravna analiza ter pravni predlogi glede ureditve različnih stopenj algoritmov.

1. Razvoj zamisli o umetni inteligenci

Pomen besede »umetna inteligenca« je zaradi odsotnosti definicije podvržen številnim predstavam in interpretacijam, hkrati pa je bila ravno njena odsotnost tista, ki je pripomogla k širitvi in pospešitvi njenega raziskovanja. Odsotnost definicije seveda ne pomeni, da raziskovalci s tega področja niso poskušali začrtati pomen-skih okvirov omenjenega pojma. Nils J. Nilsson je umetno inteligenco interpretiral z naslednjimi besedami: »Umetna inteligenca je aktivnost, katere namen je opremiti strojno napravo z inteligenco. Inteligenca pa je kvaliteta, ki napravi omogoči primerno delovanje in predvidevanje situacij v okolju.« (N. J. Nilsson, 2010, v Stone et al., 2016: 12).

Gledano z Nilssonove perspektive, karakteristike umetne inteligence temeljijo na umetno ustvarjeni programske in strojne opremi, katere cilj je doseči primerno delovanje naprave in njenega predvidevanja situacij (prav tam).

Pojem inteligenca ostaja kompleksen fenomen, ki je že dolga leta deležen velike pozornosti strokovnjakov bioloških, nevroznanstvenih, filozofskih, pravnih in drugih znanosti. Je inteligenca zgolj razum oziroma uporaba uma ali je to sinonim za obstoj duševnih zmožnosti? Howard Gardner pravi, da je oseba, ki je bila sposobna uporabljati svoje duševne zmožnosti, že od nekdaj izstopala in bila deležna spoštovanja okolice. To je tudi eden izmed razlogov, da se je povečalo zanimanje za razvoj novih programov, ki poskušajo človeško inteligenco še dodatno okrepiti. Preden se osredotočimo na učenje in sposobnosti umetne inteligence, je smiselno omeniti različne oblike inteligence pri človeku in s tem začrtati mejo med umetno in človeško inteligenco.

Psiholog Howard Gardner je človeško inteligenco razčlenil na sedem podskupin (ta splošna znana razdelitev je povzeta po Howard Gardner's Theory of Multiple Intelligences, 2009):

Jezikovna inteligenca je povezana z interpretacijo in razlago idej, prenesenih pomenov ter informacij, pridobljenih preko jezikovnega sporazumevanja. Gardner pri tej vrsti inteligence poudarja ključnost razumevanja razmerja med komunikacijo in njenim pomenom.

Logično-matematična inteligenca se nanaša na spretno analiziranje problemov, odkrivanje (novih) vzorcev, sposobnost opravljanja matematičnih kalkulacij, sposobnost logičnega sklepanja ter, ne nazadnje, sposobnost razumevanja razmerja med vzrokom in posledico. Tovrstna inteligenca se je nedvomno že izrazila na področju umetne inteligence in na določenih področjih že presegla tovrstno človeško inteligenco, kar dobro prikazuje video, naslovljen »Can a robot pass a university entrance exam«³ (Arai, 2017).

Glasbena inteligenca izpostavlja pomembnost zavedanja in uživanja ob uporabi zvoka, hkrati pa Gardner poudarja pomembnost sposobnosti zaznave tonskih in ritmičnih vzorcev, ki jih je moč zaslediti ob poslušanju glasbe. Ob tem pa izpostavlja ključnost razumevanja razmerja med zvokom in občutki, ki se porajajo ob njihovem poslušanju.

Telesno-gibalna inteligenca je osredotočena predvsem na nadzor gibanja telesa, kot so očesno-telesna koordinacija, ročne spretnosti, gibljivost telesa in ohranjanje ravnotežja. Tovrstna oblika inteligence se je že dodobra izrazila na področju umetne inteligence. Eden izmed novejših znanstvenih dosežkov na področju robotske motorike je predstavljen v videu What's new, Atlas?⁴ (2017).

Prostorsko-vizualna inteligenca je sposobnost vizualiziranja in interpretiranja zamišljenih slik, sposobnost slikovnega izražanja, hkrati pa tudi razumevanje razmerja med prostorom in učinkom ter sliko in pomenom, ki ga le ta sporoča.

Interpersonalna inteligenca poudarja sposobnosti poistovetenja s sočlovekom in vživetja v njegovo situacijo. Pri interpersonalni inteligenci je zelo pomembna interpretacija obnašanja posameznika in njegove komunikacije z zunanjim svetom, obenem pa tudi razumevanje razmerja med ljudmi in situacijami, s katerimi se soočajo.

Intrapersonalna inteligenca se v primerjavi z interpersonalno inteligenco osredotoča na razumevanje samega sebe, na samozavedanje. Intrapersonalno inteligenten posameznik je sposoben samozavedanja in razumevanja samega sebe ter svo-

³ Slovensko: »A bi lahko robot uspešno opravil sprejemni izpit za vpis na fakulteto?«

⁴ Slovensko: »Kaj je novega, Atlas?«

jih lastnih potreb. Pri tem je ključnega pomena, kako se oseba prilagaja na spremembe, stresne situacije ter kakšen je njen osebni odnos do zunanjega sveta.

Kot že rečeno, je razvoj in vključevanje umetne inteligence v sodobno družbeno okolje nedvomno rezultat sinteze vseh zgoraj omenjenih znanosti (bioloških, nevroznanstvenih, filozofskih, pravnih in drugih znanosti). Kot primer nam lahko služi umetna nevrnska mreža, ki predstavlja srce velike večine novih dognanj, inovacij in rešitev, pridobljenih s pomočjo umetne inteligence. Potovanje živčnega signala in prenos informacij med biološkimi nevroni so znanstvenikom služili kot inspiracija in vodilo pri oblikovanju njene strukture in načina delovanja. Trenutno se preučevanje umetne inteligence osredotoča predvsem na algoritme, ki so zaradi sposobnosti samoučenja v določeni meri nepredvidljivi in kjer se določene vrste človeške inteligence ne morejo izraziti (Hildebrandt, 2016: 24–25). Tukaj govorimo o tistih vrstah inteligence, ki zahtevajo poleg interpretacije že pridobljenega znanja tudi čustveno doživljanje okolice (prav tam). A kot bomo videli v nadaljevanju, bi bil lahko ta scenarij s pomočjo programske in strojne opreme, na katero so bile predhodno prekopirane funkcionalne strukture biološkega živčnega sistema (ang. *whole brain emulation* – WBE), v prihodnosti tudi izvedljiv (Sandberg, 2013: 8).

Kljub temu, da začetke proučevanja umetne inteligence povezujemo z razvojem naprednih tehnologij, njeni idejni začetki segajo daleč nazaj v zgodovino (Ploj, 2013: 4).

Prve zametke lahko zasledimo že v grških mitih in legendah, ki nam razkrivajo, kakšne so bile prve predstave o umetni inteligenci. Ena izmed zgodb govori o tem, kako se je Bog Zevs zaljubil v lepo Evropo in ji podaril Kreto, ki naj bi služila kot njeno novo domovanje. V skrbi za njeno varnost je zaprosil Hephaestusa, naj iz medenine ustvari velikana Talosa, ki naj bi jo varoval pred nevarnostmi. Vsakdo, ki je uspel priti na otok, se je kmalu znašel v Talosovem objemu, pri čemer je bil stisk njegovih medeninastih rok tako močan, da so te sčasoma začele žareti in nepovabljene goste spremenile v kup pepela. (McCaughrean, 1997: 61)

Eno bolj znanih literarnih del, ki se dotika tega tematskega področja, je roman *Frankenstein*, ki ga je leta 1818 objavila avtorica Mary Shelley. Njegov namen je bil posvariti prebivalstvo pred človeško predrznostjo in opozoriti na negativne posledice izbruha takratne industrijske revolucije. Roman pripoveduje zgodbo o švicarskem znanstveniku Victorju Frankensteinu, ki s pomočjo kombinacije kemije, alkimije in električnega toka iz mrtvih trupel ustvari prvega umetnega človeka. Slednji se kmalu prelevi v velikega nasilneža, k čemur pa botruje znanstvenikovo nerazumevanje njegovih potreb. Victor se namreč počuti krivega, da je dal življenje pošasti, ne da bi ob tem pomislil, kdo bo pravzaprav skrbel zanj. Umeten človek je kljub svoji surovosti in nečlovečnosti predstavljen kot trpeče, osamljeno in čustveno bitje – kot žrtev znanosti. Avtorica pa v vlogi žrtve ne vidi zgolj umetnega človeka, temveč tudi znanstvenika, ki zaradi vsakodnevnega boja s hudimi

duševnimi bolečinami in osamljenostjo začne izgubljati svojo človečnost. (Shelley, 1818)

Ne dolgo zatem je ideja umetnega človeka dobila tudi svoje ime – robot. Češki pisatelj Karel Čapek je leta 1921 v svojem scenariju za gledališko igro R.U.R. prvič uporabil to besedo (Gural, 2014). Zgodba govori o tovarni robotov, ki so jo odprli zato, da bi dosegli pocenitev človeške delovne sile, ki pa bi jo sčasoma v celoti izključili iz delovnih procesov, tj. nadomestili z roboti. S tem bi dosegli enakost med prebivalstvom in na dolgi rok izkoreninili revščino. Ideja, ki je v začetku obetala pozitivne spremembe v življenju prebivalstva, je kmalu dobila svoj negativen epilog. Rodnost prebivalstva je zaradi dejstva, da roboti opravljajo večino človeških nalog, močno upadla. K poslabšanju situacije je še dodatno doprinesel upor robotov, ki so se odločili, da ne bodo več ubogali navodil in bodo uničili človeštvo. (Gural, 2014)

Leta 1924 pa je ime robot dobilo tudi svojo vizualno podobo. Zamisel se je porodila nemškemu filmskemu režiserju Fritzju Langu, ko se je med sprehajanjem po ulicah New Yorka zastrmel v silhueto visokega nebotičnika. Poln vtisov in idej je na podlagi scenarija, ki ga je napisala njegova žena, zrežiral film z naslovom *Metropolis*. Film ne ponuja zgolj odlične upodobitve umetnega človeka, temveč predstavlja prototip same tematike (vzpon robotov spremlja kaos), ki je dolga leta služila kot inspiracija režiserjem znanstvenofantastičnih filmov. (Gural, 2014)

Ideja o robotih pa ni ostala zgolj upodobljena na filmskih platnih, temveč roboti danes predstavljajo nepogrešljiv del delovnih procesov. V nadaljevanju bom na primeru robotov, ki se uporabljajo za medicinske namene, poskušala predstaviti različne »razvojne stopnje robotov«, ki zahtevajo različne pravne pristope.

2. Razvoj umetne inteligence v medicini

Ob tehnoloških inovacijah na področju umetne inteligence in njihovem vključevanju v delovne procese si zastavljamo številna vprašanja o tem, kako ob ustvarjanju/preoblikovanju pravnih pravil, ki bodo urejala popolnoma novo dimenzijo/umetno inteligenco, ohraniti moralno noto in premagati moralno-etične ovire današnjega časa.

Hipokratova zaprisega je skupaj s svojo moralno dediščino postavila temelje sodobni medicini in začrtala moralno-etične vzorce delovanja zdravnikov in preostalega zdravstvenega osebja (Jotterand, 2005: 110). Prisega predstavlja enega prvih moralnih pečatov, ki se po njeni zaslugi danes odraža v vseh razmerjih, ki so sestavni del medicinske prakse. Ime je dobila po očetu medicine, grškem zdravniku Hipokratu (Ackerknecht, 1982: 55). Uvodni del prisege je namenjen izražanju pripadnosti bogovom in boginjam, kar nas privede do zaključka, da je posameznikova sposobnost zdravljenja že od nekdaj veljala kot nekaj zelo veličastnega,

če ne celo nadnaravnega: »Prisegam na Apolona zdravnika in Asklepija, na Higiejo in Panakejo, na vse bogove in vse boginje, ki jih jemljem za priče, da bom po svoji moči in sodbi izpolnjeval tole prisego in tole pismo pogodbo.«

Z vključevanjem umetne inteligence v medicino pa postavimo pod vprašaj filozofijo zaupnih razmerij, ki so sestavni del medicinske prakse. Medicina ima v samem bistvu izjemno občutljivo in moralno noto, saj predstavlja združitev zelo osebnih in zaupnih razmerij. Z vsako medicinsko obravnavo je pacientovo zdravje zavestno položeno v roke zdravnikov, torej ljudi, ki pacienta ne poznajo. Edina stvar, na katero se pacient lahko zanese, je v prvi vrsti zaupanje in vera, da zdravnikove odločitve ne bo krojilo zgolj njegovo medicinsko znanje, temveč tudi njegova moralna presoja.

Vse večja avtonomnost robotov pri njihovem delovanju bi privedla do situacije, ko ti, zaradi svojih sposobnosti in samostojnih reakcij pri reševanju nalog, ne bi več ustrezali moralni ter pravni zasnovi že obstoječih norm, ki obravnavajo stroje zgolj kot pripomočke za lažjo izvedbo neke naloge. Dolga leta je bila vključenost robotov na področju medicine omejena na asistiranje zdravnikom pri posegih (ang. *robot-assisted surgeries* – RASs), kar je pripomoglo h kvalitetnejšemu in hitrejšemu zdravniškemu obravnavanju. Napredek tehnologije na področju umetne inteligence pa odpira nove možnosti in predstavlja nove načine njihovega vključevanja v delovne procese znotraj medicine.

V nadaljevanju bosta predstavljena dva načina uporabe robotov znotraj medicine, in sicer uporaba robotov asistentov in robotov, ki samostojno opravljajo terapije/posege.

2.1 Roboti asistenti

V letu 2016 je bilo s pomočjo robotov asistentov po svetu opravljenih približno 750.000 operacij. Pri tem je vredno poudariti, da roboti, ki sodijo v to kategorijo, posegov ne opravljajo samostojno, temveč pod nadzorom in v sodelovanju z zdravnikom. Sodelovali so predvsem pri kirurških posegih na prostati in maternici, v nekoliko manjšem obsegu pa tudi pri operacijah ledvic, debelega črevesa, srca in drugih organov. (New surgical robots, 2017: 71–72) Podjetje, ki je zaslužno za izdelavo večine robotov asistentov, je registrirano v Kaliforniji, pod imenom *Intuitive Surgical, Inc.* Monopol na trgu si je prislužilo z lansiranjem robota da Vinci, danes pa se mu s svojimi inovacijami nevarno približujejo tudi druga podjetja. Po podatkih *Intuitive Surgical, Inc.* vsakih 60 sekund zdravstvena ustanova, nekje na svetu, uporablja pri operaciji njihovega robota (*Intuitive Surgical, b. d.*). Vključevanje inovativnih produktov v medicini se je pokazalo kot izjemno učinkovito in predstavlja velik korak k uspešnejšemu in hitrejšemu zdravljenju bolezni. Z njihovo pomočjo zdravniki posredno in bolj precizno vihtijo kirurške instrumente, kot bi jih, če bi jih držali v lastnih rokah. Robot da Vinci je opremljen s štirimi rokami, pritrjenimi na ploščo. Tri izmed njih držijo kirurške instrumente,

četrti pa je opremljen s kamero, ki zdravniku omogoča vpogled in posledično nadzor nad celotno situacijo. Ukaze izvršuje s pomočjo konzole, opremljene s krmilno palico in pedali, ki med operacijo avtomatsko izničijo tresenje zdravniškove roke (Intuitive Surgical, b. d.). Prednost tovrstnih kirurških posegov (ang. *keyhole surgery*) je, da zdravnik preko robota, na katerega so pritrjeni kirurški instrumenti, poseže v pacientovo telo skozi majhne luknje, s čimer se zmanjšajo tveganja posega in zato pride do hitrejšega okrevanja (New surgical robots, 2017: 71–72). Primeri resnejših in življenjsko nevarnih zapletov, ki podaljšajo zdravljenje pacienta, terjajo dolgotrajno hospitalizacijo in celo ponovno operacijo, so poškodbe tkiv in organov, krvavenje ter infekcije (Intuitive Surgical Inc.: »Instrumentation«). Za preprečitev zgoraj omenjenih rizikov, ki so posledica bodisi zdravnikove napake oziroma nepravilnega delovanja robota, je podjetje oblikovalo dokument »Surgical Risk Document« (b. d.), ki opisuje možne zaplete in vzpostavi določene varnostne mehanizme za njihovo preprečitev.

V letu 2017 je britansko podjetje *Cambridge Medical Robotics* predstavilo svoj najnovejši produkt, robota Versius, ki se bo uporabljal predvsem pri operacijah trebušne votline in hrbtenice. V primerjavi z robotom da Vinci njegove roke niso pritrjene na isto ploščo, temveč je vsaka samostojna, opremljena s samostojno bazo. Versiusove roke so za razliko od rok njegovega predhodnika grajene podobno strukturi človeških rok, imajo torej tri sklepe – ramo, komolec in zapestje (Technology, b. d.). Prednost tovrstne izdelave je po besedah izvršnega direktorja *Cambridge Medical Robotics*, Martina Frosta, ta, da bodo zdravniki med operacijo lahko opravljali enake gibe, kot bi jih, če bi operacijo izvajali neposredno, brez pomoči robota (New surgical robots, 2017: 71–72).

Italijansko podjetje *Medical Microinstruments* (MMI) pa se je osredotočilo na izboljšave pri robotih, ki opravljajo mikrokirurške posege. Eden izmed njegovih ustanoviteljev v videu na njihovi spletni strani pove: »Robotika predstavlja revolucijo medicinskih operacij. Spretno manevriranje z najmanjšimi anatomskimi strukturami odpira vrata na nova področja medicine. MMI želi omogočiti, da bi vsak kirurg bil sposoben opravljati mikrokirurške operacije, tudi tiste, ki danes štejejo za neizvedljive.« (MMI, b. d.) Glavni inženir podjetja, Massimiliano Simi, je dodal: »Zaradi naprednih materialov in novih mikroproizvodnih procesov smo ustvarili najmanjše artikulirano zapestje, ki omogoča opravljanje mikrokirurških operacij z robotsko preciznostjo« (MMI, b. d.). S pomočjo tovrstnih robotov lahko zdravniki natančneje in uspešneje operirajo poškodovane žile in živce, pri čemer jim je samo stanje poškodbe predstavljeno v taki velikosti, kot če bi gledali skozi mikroskop (New surgical robots, 2017: 71–72).

Četrto podjetje je *Auris Robotics*, ki za svoj prihajajoči produkt napoveduje, da bo sposoben odstranjevati pljučne tumorje. Po statistiki je pljučni rak najbolj smrtonosen, zaradi te bolezni na letni ravni izgubi življenje 1,7 milijona ljudi (New surgical robots, 2017: 71–72). Kje tiči vzrok za njegovo smrtonosnost ter kakšna

je prednost robotskega posega v primerjavi z običajnim kirurškim posegom? Pri klasični operaciji je za tovrsten kirurški poseg običajno, da se pacientu odpre prsni koš in nato odstrani bolno tkivo. Ker operacijo spremlja relativno visoko tveganje, se zdravniki mnogokrat odločijo, da v primeru manjšega tumorja odložijo operacijo in spremljajo njegov razvoj. V primeru, da se tkivo začne povečevati in ni pravočasno odstranjeno (je torej puščeno *in situ*), je bolezen v večini primerov smrtonosna. Težave, povezane z odstranitvijo bolnega tkiva v kasnejših stadijih bolezni, so povezane s povečanimi riziki in dejstvom, da je večje tkivo veliko težje odstraniti. Prednost posega z Auris robotom pa je, da doseže bolno tkivo preko ust in sapnika, kar zniža tveganje in težavnost odstranitve tkiva (New surgical robots, 2017: 71–72).

V bližnji prihodnosti bo svojega robota predstavilo tudi podjetje *Medtronic*, ki slovi kot največji proizvajalec medicinske opreme na svetu. Robota so že patentirali ter mu nadeli ime MIRO. Ta bo omogočil zdravnikom, da se med operacijo zanesejo tudi na otip, saj bodo MIROvi instrumenti opremljeni z naprednimi senzorji, ki bodo pošiljali signale v krmilno palico in posledično v zdravnikove roke. Prej omenjeni roboti niso tako tehnološko dovršeni, da bi zdravnik preko krmilne palice lahko otipal, začutil mehko človeškega tkiva, kar pa je eden izmed ključnih faktorjev pri oceni stanja oz. postavitvi diagnoze (New surgical robots, 2017: 71–72).

Kot predstavljeno v razdelku o pravni analizi, ta vrsta robotov s pravnega vidika ni problematična, saj se obravnava kot pripomoček v rokah zdravnika in ni opremljena z algoritmi, ki bi v času samostojnega reševanja problema oblikovali lastne vzorce delovanja in posledično onemogočali predvidljivost.

2.2 Roboti, ki samostojno opravljajo terapije/posege

Vključevanje robotov v medicinske posege je bilo dolga leta omejeno zgolj na robotsko asistiranje (Hiyaguha, 2016). Trenutni znanstveni dosežki pa nam predstavljajo popolnoma drugačno sliko, saj nas soočajo z različnimi oblikami naprednih robotov, ki so sposobni avtonomno ter v celoti sami opraviti določen medicinski poseg. Robot STAR, ki je sposoben posnemati najboljše medicinske operacije, je s pomočjo svojega kompleksnega računalniškega programa (avtonomnega algoritma) oblikoval načrt za izvedbo operacije poškodovanega mehkega tkiva. Znanstveniki so po opravljenem poskusu primerjali končne rezultate robota, zdravnika z robotsko asistenco ter zdravnika, ki je samostojno opravil laparoskopsko operacijo. Pri primerjavi so se osredotočili na enakomernost in gostoto šivov, število napak pri samem šivanju ter čas opravljanja medicinskega posega na mehkem tkivu. Primerjava je pokazala, da je končen rezultat pri posegu, ki ga je opravil robot, bistveno bolj precizen in dovršen v primerjavi s končnim rezultatom zdravnika z robotsko asistenco ter zgolj zdravnika (Shademan et al., 2016: 3). Pozitivni rezultati nakazujejo, da bi avtonomni roboti lahko imeli pomembno, samostojno vlogo pri opravljanju posegov, pri čemer se ne bi izboljšala zgolj kva-

liteta tovrstnih posegov, temveč tudi konsistentnost in dostopnost najnovejših kirurških tehnik.

Napredni roboti niso dosegli odličnih rezultatov zgolj pri operacijah, temveč so se izkazali za zelo uspešne tudi, ko so opravljali terapije s pacienti, ki se soočajo z različnimi oblikami duševnih bolezni. Znanstveniki v Singapurju so 7. 3. 2016 svetu prvič predstavili Nadine, emocionalno-inteligentnega humanoidnega robota. Raziskovalci na Nanyang tehnični univerzi v Singapurju so jo opremili s sposobnostmi priklica preteklih pogovorov s pacienti in s tem omogočili prilagoditev njene terapije potrebam vsakega posameznika. Nadia Thalmann, direktorica Inštituta za medijske inovacije, je poudarila: »Ker se države po svetu soočajo z izzivi staranja populacije, socialni roboti lahko predstavljajo eno izmed rešitev pomanjkanja delovne sile, lahko postanejo osebni spremljevalci otrok in starostnikov ter lahko celo služijo kot platforma za zdravstvene storitve v prihodnosti« (Emotionally intelligent robot, 2016). Thalmannova predvideva, da bodo Nadinine sposobnosti svojevrstne interakcije z ljudmi pomagale posameznikom, ki trpijo za posledicami demence, drugimi duševnimi boleznimi in avtizmom (Emotionally intelligent robot, 2016). Za slednje je značilno atipično obnašanje in nespretnost pri socialni komunikaciji.

V centru za nadzor in preprečevanje bolezni (*The center for disease control and prevention* – CDC) so razkrili, da je glede na opažanja eden izmed 88 otrok in približno eden izmed 54 dečkov v ZDA diagnosticiran z avtizmom. To predstavlja 78-odstotni porast bolezni v primerjavi z zadnjim CDC poročilom v letu 2009 (Bekele et al., 2013: 289). Ta porast ni zanemarljiv, saj v tem trenutku veliko številu družin ni finančno sposobnih zagotoviti svojim otrokom optimalno medicinsko obravnavo. Slednja običajno vključuje dolgotrajno angažiranje usposobljenih terapevtov, katerih storitve niso dostopne večini avtistične populacije. Kot odziv na pomanjkljivo dostopnost optimalnih terapevtskih pristopov so znanstveniki izoblikovali robotske terapije, ki omogočajo ponavljajoče vaje za krepitev socialne interakcije otrok z avtizmom. Znanstveniki so ugotovili, da ljudje z avtizmom lažje sodelujejo z roboti terapevti kot pa z zdravniki. Ugotovili so, da pacienti gledajo robota terapevta 27,65 odstotkov dlje kot pa človeškega terapevta, kar predstavlja izrazito preferenčno obnašanje med robotskimi terapijami (Bekele et al., 2013: 295).

Kot bomo videli v nadaljevanju, tovrstni avtonomni roboti, ki način zdravljenja prilagajajo lastnim preteklim zaznavam okolja, ustvarjajo občutek nepredvidljivosti, kar je sporno z vidika pravnega reguliranja.

3. Raznovrstnost algoritmov kot glavni kriterij za pravno ureditev umetne inteligence

V sklopu delovnega procesa bo robot soočen z različnimi nalogami, ki bodo rešene preko dobro definiranih zaporednih faz, ki so v matematičnem svetu opisane kot algoritem. V tem razdelku bodo predstavljene razlike med štirimi različnimi procesi odločanja, ki so rezultat štirih različnih vrst algoritmov. S tem bo prikazano, zakaj potrebujemo različne pravne pristope za različne vrste algoritmov, s katerimi so opremljeni roboti.

3.1 Determinističen algoritem

Determinističen algoritem je najbolj osnovna oblika algoritma, saj je njegov potek v celoti predvidljiv. Obnašanje naprave, opremljene z determinističnim algoritmom, je s tabelo in vhodnim podatkom natančno določeno. Naprava po sprejemu vhodnega podatka vedno poišče identično rešitev ter sledi enakemu postopku oz. zaporedju (Žerovnik, 1989: 60). Ta algoritem preprečuje robotu avtonomno odločanje, oblikovanje novih vzorcev in posledično ustvarjanje nepredvidljivih izidov (Hildebrandt, 2016: 23). Ti roboti so zgolj pripomoček za lažjo in bolj precizno izvedbo naloge, pri tem pa sledijo človekovim ukazom, kar je bilo prikazano na primeu robotov asistentov. V primeru determinističnih algoritmov bomo vedno našli odgovorno osebo, ki bo odgovarjala za morebitne napake robota. Kot bomo videli v poglavju o pravni analizi, je ravno koncept predvidljivosti tisti, ki nam omogoča uporabo pravil o producerski odgovornosti in pravil o objektivni odgovornosti delodajalca za nevarno stvar.

3.2 Nadzorovan učni algoritem

Nadzorovan učni algoritem je naprednejša oblika algoritma, ki omogoča strojno učenje (Hildebrandt, 2016: 24–25). Eden izmed pomembnejših procesov, ki se odvija med časom učenja, je generalizacija, z drugimi besedami posploševanje znanja. Predstavlja situacijo, ko posameznik iz podatkov, pridobljenih v procesu učenja, izlušči določene splošne zakonitosti. Globlje razumevanje teh zakonitosti pripomore k temu, da je posameznik sposoben razrešiti nek primer, ki se sicer glede na dane podatke razlikuje od primera, iz katerega je izluščil splošne zakonitosti, vendar je pristop reševanja problema enak (Ploj, 2013: 6). V procesu učenja robot prestane različne faze treninga, ki mu ga pripravi programer. Treninki so sestavljeni iz posameznih podatkov in vnaprej definiranih vzorcev, ki dajo robotu »občutek«, kakšen izid/rešitev je zaželen. Roboti, opremljeni s tovrstnimi algoritmi, lahko pridejo do podobne ali celo boljše rešitve problema od predvidene (Hildebrandt, 2016: 24–25).

Nadzorovano učenje – ta vrsta strojnega učenja opremi stroj z učnimi podatki, ki vsebujejo vhodne vrednosti ter obenem tudi želene izhodne vrednosti (Ploj, 2013: 7). Primer strojnega učenja je razvrščanje fotografij ljudi glede na določene krite-

rije (barva las, barva kože, telesna masa, spol, velikost ipd.), pri čemer pa ne moremo vedeti, za kateri od ponujenih kriterijev se bo odločila posamezna naprava.

3.3 Nenadzorovan učni algoritem

Druga oblika naprednejšega algoritma je nenadzorovan učni algoritem, ki v primerjavi z nadzorovanim učnim algoritmom vzpostavlja korelacije med pridobljenimi podatki brez predhodne postavitve hipoteze (Hildebrandt, 2016: 23). Brez zagotovitve treninga se robot ne more seznaniti z informacijo, kaj predstavlja želen izid.

Nenadzorovano učenje – za ta način strojnega učenja je značilno, da je stroj pri svojem delovanju opremljen zgolj z vhodnimi vzorci, ne pa tudi z idejo, kakšen točno naj bo končni rezultat. Naprava, ki je v postopku nenadzorovanega učenja, sama poišče podobne vhodne vzorce in jih po lastnih kriterijih razvrsti v razrede (Ploj, 2013: 6). V znanstvenih člankih se pogosto omenja, da je rezultat nenadzorovanega učenja nepredvidljiv, kar stigmatizira to vrsto strojnega učenja kot nezaželeno. Ob tem bi rada poudarila razliko med opredelitvijo končnega cilja in opredelitvijo, kakšen rezultat šteje kot najbolj zaželen. Ljudje mnogokrat napačno sklepamo, da odsotnost natančno določenih vzorcev reševanja problema predstavlja popolno avtonomnost robota, na kar nimamo kot posamezniki nobenega vpliva, kar seveda ni res. Robot ves čas, v procesu svojega delovanja in optimalnega reševanja problema, teži k zadovoljitvi cilja, ki mu ga je zastavil človek. Pri razvrščanju fotografij ljudi tako ne vemo, katere kriterije za njihovo razvrstitev bo izbrala posamezna naprava. Pri tem pa ne moremo trditi, da gre za odsotnost predvidljivosti končnega cilja, želeli smo imeti razvrščene slike in sedaj imamo pred seboj razvrščene slike. Neopredelitev želenega rezultata ne pomeni hkrati tudi odsotnosti cilja.

3.4 Umetna nevronska mreža

Umetna nevronska mreža je model za obdelavo informacij, ki je bil narejen po vzoru delovanja biološkega živčnega sistema (človeških možganov). Človeške možgane sestavlja zelo veliko število prepletajočih se živčnih celic oziroma nevronov in njihovih podpornih celic. Nevron navadno sprejme signal iz sosednjega nevrona preko goste mreže izrastkov, ki jih imenujemo dendriti (Mramor, 2007: 4). Signal nato z dendritov potuje do jedra nevronske celice, ki pošlje impulz naprej po dolgi, tanki, nitasti strukturi, imenovani akson ali nevrit, ki se lahko razcepi na tisoče vej. Na koncu vsake veje ali izrastka je struktura, imenovana sinapsa, ki signal iz aksona preko nevrottransmiterjev prenese na sosednjo živčno celico, kjer se sproži električni dražljaj. Slednji lahko vzbudi ali zavira dražljaje, ki se prenesejo na sosednji nevron. Ko nevron prejme dražljaj, ki je dovolj velik v primerjavi z vhodnim pragom (v sinapsi je izločenih dovolj nevrottransmiterjev, da le

ti presežejo prag za sprožitev dražljaja), pošlje dražljaj naprej po aksonu.⁵ »Možgani uravnavajo učinkovitost prenosa signala preko sinaps, kot odgovor na ugodne in neugodne dražljaje iz okolja, tako da se organizem čim bolj prilagodi svoji okolici.« (Mramor, 2007: 4).

Cilj posnemanja delovanja možganov je kreacija sistemov, ki imitirajo resnične biološke organizme v smislu njihovega obnašanja in notranjih vzročnih struktur. Roboti, opremljeni s takimi možgani, bi bili lahko sposobni čutiti bolečino in čustveno doživljati svet okoli sebe (Sandberg, 2013: 8). Sandberg omenja, da bi s posnemanjem možganskega delovanja lahko popolnoma opustili poskuse na živih organizmih (živalih) in jih začeli izvajati na umetnih. Kantova teorija o pomanjkanju moralne avtonomije pri živalih argumentira, da čeprav živali niso sposobne moralnega zaznavanja, je zelo pomembno, kako ravnamo z njimi. Naše dolžnosti do živali se izražajo kot posredne dolžnosti do humanosti. Krutost do živali tako posledično poškoduje našo lastno humanost. (Gruen, 2017) Torej, če se šteje kot kruto dejanje to, da biološki miški odrežemo rep, se enak impulz človeške krutosti izrazi pri tem, ko odrežemo rep miški, ki je opremljena s programsko in strojno opremo, na katero so kopirani človeški možgani (Sandberg, 2013: 10). Uspešno prekopirani možgani bi omogočili robotu, da se bi na situacijo odzval na podoben način, kot bi se biološko bitje. Eden izmed najbolj znanih primerov je projekt *Cyber Child* – simuliran otrok, ki ga je ustvaril Rodney Cotterill in je deloval na podlagi biološke inspiracije živčnega sistema (Sandberg, 2013: 10). Njegovo telo je imelo nekaj simuliranih mišic in bioloških stanj, kot je na primer raven sladkorja v krvi, mleko v želodcu in urin v mehurju. Simuliran glas in gibanje sta mu omogočala interakcijo z uporabnikom, preko katere je želel sporočiti, da je lačen in potrebuje mleko. V tem projektu je bilo uporabljenih presenetljivo malo nevronske mreže (20 nevronov na območje), a kljub temu ta simulacija predstavlja poskus izpeljave modela samozavedanja. Začetnik projekta trdi, da ni nobenih utemeljenih razlogov, zakaj se sposobnost zavedanja ne bi sčasoma vzpostavila v takšnih umetnih možganih (Sandberg, 2013: 10).

Umetna nevronska mreža na podlagi proučevanja primerov iz preteklosti pridobiva nova znanja, kar pa lahko, zaradi njene samostojnosti pridobivanja/kreiranja, privede do situacije, ko robotovo znanje preseže znanje učitelja. Neredko se zgodi, da so algoritmi iz resničnega življenja zelo zapleteni in jih ljudje ne zmoremo dojeti v vseh podrobnostih. Klasičen pristop k obdelovanju podatkov nas omejuje na reševanje problemov, ki jih že dobro poznamo. Eden izmed ciljev nevronske mreže pa je, da bi rešili tudi tiste probleme, ki jih ljudje še ne razumemo dobro in za njihovo razrešitev ne znamo spisati ustreznih algoritmov. (Ploj, 2013: 8-10) Kot rečeno, pri nevronske mreže samo-reševanje poteka nekoliko drugače, saj med učenjem mreže same poiščejo rešitev določenega problema. Do rezultata

⁵ Za pomoč pri besedilu, ki se nanaša na delovanje biološkega živčnega sistema (človeških možganov), se zahvaljujem Janiju Zapušku, dr. med.

pridejo tako, da znotraj različnih učnih podatkov poiščejo splošna pravila. Če želimo doseči uspešno učenje, morajo biti učni podatki natančno izbrani in morajo vsebovati veliko število reprezentativnih primerov. Tovrsten postopek se imenuje strojno učenje, pri katerem ne sodelujejo programerji. Uporaba nevronske mreže nam omogoča rešitev problema tudi, kadar programer algoritma, potrebnega za izračun končnega rezultata/rešitve problema, ne pozna. S tovrstnim načinom dosežemo, da na neki točki znanje nevronske mreže preseže celo znanje strokovnjakov. (Ploj, 2013: 8–10)

4. Pravna ureditev umetne inteligence

Kot vidimo, se svet, ne zgolj na področju medicine, temveč (predvsem) tudi na drugih področjih, sooča z vključevanjem robotov v delovni proces, pri čemer se delodajalci ne omejujejo zgolj na tiste robote, katerih delovanje je popolnoma predvidljivo. Vprašanje, ki se pojavi, je: Kako pravno ovrednotiti dejanja robotov in potencialno škodo, ki izhaja iz njihovega delovanja, ter, ne nazadnje, kako jih pravno umestiti v obstoječ sistem?

4.1 Dodelitev pravne osebnosti

Ideja dodelitve pravne osebnosti je bila prvič predstavljena v okviru predloga Evropskega parlamenta in se je nanašala na dodelitev posebnega pravnega statusa robotom – statusa »elektronske osebe«.

Pravni sistemi niso sposobni samouresničitve, njihov *spiritus agens* (gonilna sila) je oseba, nosilka pravnih akcij, sprememb in pravnih posledic (Pavčnik, 2011: 136). Oseba je v pravnem sistemu predstavljena kot pravni igralec, kateremu pravni sistem priznava neko določeno pravno vlogo. Pravna vloga, katero igramo, je sestavljena iz pravic in obveznosti in nam omogoča, da vstopamo v pravna razmerja. Če povzamemo, pravni igralec je glede na naravo vloge, ki se mu pripisuje, lahko zgolj in samo človek. Slednji igra pravno vlogo bodisi kot fizična oseba bodisi kot član pravne osebe, ki jo opisujemo kot umetno tvorbo oziroma s pravnim aktom ustvarjen pravni subjekt, ki dejavnosti ne more opravljati sam, brez (ne)delovanja fizične osebe. V obeh primerih je človek kreator končne odločitve, le manifestacija le-te se razlikuje. »Kreator« se pri tem zaveda pravih posledic, ki bi jih imele morebitne napačne odločitve, in dobro ve, da se bodo vse odločitve, skupaj s pravicami in obveznostmi, vrnila nazaj k njemu in ga zavezovale – *Hominum causa omne ius constitutum* (Pavčnik, 2011: 137). Po besedah Marijana Pavčnika je glavni razlog za pripis pravne subjektivitete pravnemu subjektu »vrednostna odločitev, da se določene pravice in dolžnosti pripisujejo (prištevajo) tem subjektom« (prav tam). S pripisom take kvalitete se prizna (na novo ustvarjenemu) pravnemu akterju, ki se zaveda pravnih posledic svojih dejanj, sposobnost kreiranja in pravnega manifestiranja lastnih želja in odločitev. Torej, če

se pravni sistem odloči za razširitev vrst subjektov/pravnih akterjev, mora med drugim vzeti v obzir, da je subjekt človek oz. da se skozi na novo ustvarjen pravni subjekt manifestira človeška volja. Vzemimo za primer pravno osebo. Zakonodajalec ji ne bi mogel nikoli pripisati pravne osebnosti, če za dejanji, ki jih le-ta manifestira v svet, ne bi stali človeški posamezniki. Enako lahko trdimo za robote. Pripis pravne osebnosti robotom nove generacije, ki so opremljeni z algoritmi, kjer njihovih dejanj in odločitev ne moremo v celoti predvideti, bi bil skrajno nesmiseln in bi spodkopal glavne temelje, na katerih je osnovana ideja pravne osebnosti.

Kot je razvidno iz argumenta, osebno nisem najbolj naklonjena dodelitvi pravne osebnosti robotom, dopuščam pa možnost, da bi bila ta kvaliteta nekoč (ko bo znanost še napredovala) dodeljena *robotom, ki so opremljeni s programsko opremo, na katero so bili predhodno preslikani človeški možgani*. Po mnenju mnogih tu ni izključen postopen pojav samozavedanja in čustvovanja (Sandberg, 2013: 10).

4.2 Producerska odgovornost

Bistvo producerske odgovornosti je, da mora proizvajalec odgovarjati za posledice, ki jih utrpijo kupci oziroma uporabniki produkta. Na območju Evropske unije je producerska odgovornost urejena z Direktivo Evropskega parlamenta 85/374/EEC (v nadaljevanju Direktiva). Skladno z besedilom Direktive produkt predstavlja materialno stvar, ki jo poganja elektrika. Produkt se smatra kot okvarjen, ko med njegovim delovanjem varnost proizvoda ni takšna, kakor jo lahko oseba ob upoštevanju vseh okoliščin upravičeno pričakuje. Pri tem mora biti seznanjena s predstavitvijo proizvoda, predvidljivo uporabo proizvoda ter kdaj je bil proizvod dan v promet (Evropski parlament, Direktiva 85/374/EEC: člen 6). Producent se skladno s pogoji iz sedmega člena omenjene Direktive lahko oprosti odgovornosti za škodo, povzročeno s strani okvarjenega proizvoda, če dokaže, da okvara, ki je bila vzrok za povzročeno škodo, ni obstajala v trenutku, ko je bil produkt dan v obtok. Če bi želeli uporabiti pravila producerske odgovornosti za poškodbe s strani robotov, opremljenih z algoritmi, ki omogočajo robotu opravljanje nepredvidljivih potez, bi kmalu ugotovili, da ta pravila niso ustrezna. Tovrstni roboti v procesu samoučenja kreirajo nove vzorce reševanja nekega problema, s čimer se njihove sposobnosti konstantno povečujejo. Če je bil na začetku robot sposoben opraviti samo določen tip nalog, se v procesu samoučenja ta okvir znatno poveča, pri čemer pa se ustvarja robot, ki se razlikuje od svoje prvotne verzije, kar pomeni, da je producentu odvzeta možnost predvidevanja – to pa skladno s sedmim členom omenjene Direktive izključi odgovornost producenta (Evropski parlament, Direktiva 85/374/EEC: člen 7).

Tovrstna oblika odgovornosti bi se lahko uporabila zgolj za robote, opremljene s takšnimi *algoritmi, ki so v celoti predvidljivi* in med reševanjem naloge niso sposobni kreirati novih vzorcev in tako spreminjati/krepiti svojih zmožnosti. Nepred-

vidljivost je vsekakor ena izmed ključnih težav, ki jih prinaša umetna inteligenca, tako na pravnem področju kot tudi v samem postopku zagotavljanja varnosti, ki je zaradi nepredvidljivosti oteženo.

4.3 Re-regulacija delovnega razmerja

V prvih dveh razdelkih je bila ovrednotena dodelitev pravne subjektivitete in naložitev producerske odgovornosti. Moja ideja o re-regulaciji delovnega razmerja bi odpravila dileme glede pravne ureditve še zadnjih dveh pravno neurejenih sklopov nepredvidljivih robotov, ki uporabljajo *nenadzorovan učni algoritem* in *nadzorovan učni algoritem*. Vprašanje, ki se pojavlja je: Kako pravno ovrednotiti dejanja in kako z opredelitvijo osebe, ki bo odgovorna za škodo povzročeno s strani tovrstnih robotov, zavarovati poškodovano osebo?

Delodajalec je tisti, ki organizira in koordinira delovne procese, ki se odvijajo znotraj njegove primarne dejavnosti. Organizacija delovnega procesa zahteva od delodajalca, da poleg razporeditve dela njegovih zaposlenih zagotovi tudi primer-no delovno okolje, stroje in pogoje za ohranitev dobrega in zdravega delovnega okolja. Narava njegovega dela mu daje veliko mero avtonomije, saj je način delo-vanja njegovih delavcev odvisen od njegovega vodenja in navodil. Oblikovanje slednjih pa, ob težnji po optimalnem delovnem procesu, temelji na tehtanju raz-ličnih možnosti in morebitnih rizikov njihove izvedbe (Plavšak et al., 2009: 578). Vključitev naprednih robotov v delovni proces predstavlja za delodajalca nepred-vidljivost glede morebitnih rizikov, kar pogosto povzroča težave pri sami koordi-naciji. Po drugi strani pa njihova vključitev in pravna ocenitev posledic njihovih ravnanj predstavlja pravno praznino, saj pravo ne predvideva osebe, ki bi bila od-govorna za njihova dejanja, ki nastanejo kot posledica samostojnih korelacij v al-goritmu.

Ko se vprašamo, kdo je odgovoren za škodo, povzročeno *med* oziroma v *povezavi* z delovnim procesom, moramo vedeti, da bo delodajalec v razmerju do tre-tje/poškodovane osebe objektivno odgovoren (odgovornost ne glede na krivdo) za ravnanja njegovega personalnega substrata – njegovih delavcev.

Obstoječe pravo ureja situacijo, ko je škoda povzročena s strani stroja, na način, da pripiše delodajalcu objektivno odgovornost, saj se domneva, da je stroj med njegovim delovanjem uporabljal eden izmed delodajalčevih zaposlenih – perso-nalni substrat (Plavšak et al., 2009: 568). Na tej točki je treba poudariti, da pravo predvideva objektivno odgovornost zgolj za tiste stroje/robote, ki se uporabljajo *zgolj kot pripomoček* in niso opremljeni z algoritmi, ki povzročajo nepredvidljivi-vost. Pravo torej trenutno ureja situacije, ko se domneva, da za dejanji stroja stoji človek, tj. če ta da določen ukaz, hkrati tudi natančno ve, kaj pričakovati. Deloda-jalec se lahko izogne svoji objektivni odgovornosti za škodo, povzročeno s strani stroja, ki se uporablja v delovnem procesu, če dokaže, da je dejanski razlog za povzročeno škodo dejanje ali dogodek, ki je izven njegove sfere (Plavšak et al.,

2009: 580). Če želi to dokazati, mora kumulativno zadovoljiti dvema pogojema. Prvi pogoj zahteva, da se škodi, povzročeni zaradi delovanja stroja, ni bilo mogoče izogniti in tudi ne preprečiti njenih posledic. Drugi pogoj zahteva, da je dejanje nepredvidljivo.

Kot že omenjeno, trenutno pravo ureja zgolj stroje, opremljene z algoritmi, ki v postopku samostojnega reševanja naloge ne morejo kreirati novih vzorcev. Za naprednejše algoritme (samostojno kreiranje vzorcev na podlagi učenja iz preteklih izkušenj) tovrstna zakonodaja enostavno ne bo zadostovala, delodajalec bo namreč vedno oproščen objektivne odgovornosti za škodo, saj le te resnično ne bo mogel pričakovati niti preprečiti (Plavšak et al., 2009: 581).

Predlagam, da izraz »delovna sila« razdelimo na dve podskupini, prvo bo sestavljala človeška delovna sila, drugo pa avtomatizirana delovna sila. Pripis besede »avtomatizirana« k besedni zvezi »delovna sila« omogoči *prilagoditev* (diskriminiranje: npr. roboti ne potrebujejo bolniške ipd.) obstoječih pravnih pravil, ki veljajo za človeško delovno silo, in njihovo uporabo za avtomatizirano delovno silo, ki je opremljena s sposobnostmi za samostojno opravljanje nalog, ki so jih včasih opravljali samo ljudje. Naj povzamem, cilj je vključitev avtomatizirane delovne sile v delodajalčev *substrat*, pri čemer želimo z njeno vključitvijo ustvariti pogoje za uporabo pravil objektivne odgovornosti delodajalca v primeru škodnih dejanj s strani robota. Glavni namen vključitve je obvarovati oškodovano osebo pred situacijo, ko pravo ne predvidi osebe, ki bi odgovarjala za poškodbe s strani naprednejših robotov.

Poglejmo, kako bi tovrsten predlog deloval v praksi. Ob predpostavki, da je robot, opremljen z naprednim algoritmom, med svojim delovanjem napravil škodo in pri tem niso sodelovale druge osebe.

4.3.1 Človeška delovna sila

Ko delavec (človeška delovna sila) povzroči škodo tretji osebi, najprej pogledamo predpogoje za naprtitev objektivne odgovornosti delodajalcu. Predpogoja, ki morata biti izpolnjena, da se škodno ravnanje, povzročeno s strani delavca, smatra kot ravnanje delodajalca, sta: delavec mora biti del personalnega substrata delodajalca; delavec, kot del personalnega substrata delodajalca, povzroči škodo med opravljanjem dela ali v zvezi z delom (Plavšak et al., 2009: 568).

Pri tem je treba preveriti, če lahko uporabimo pravila o krivdni odgovornosti. Ko delavec namenoma povzroči škodo tretji osebi, je slednja upravičena zahtevati povračilo škode bodisi od delavca bodisi od delodajalca (solidarna odgovornost – *liability in solido*) (Plavšak et al., 2009: 570). Če oškodovanec zahteva povračilo od delodajalca in mu delodajalec povrne znesek, s tem ta pridobi zahtevek proti delavcu za povračilo plačanega zneska. Subjektivna oziroma krivdna odškodninska odgovornost je opredeljena v prvem odstavku 131. člena Obligacijskega zakonika, ki določa, da kdor povzroči škodo drugemu, jo je dolžen povrniti, če ne

dokaže, da je škoda nastala brez njegove krivde. Krivda pa je podana, kadar oškodovalec povzroči škodo namenoma ali iz malomarnosti⁶ (Sodba Višjega delovnega in socialnega sodišča: 146/2009). Posebnost odškodninske odgovornosti, ki izhaja iz delovnega razmerja, je predvsem ta, da je delavec odgovoren le za škodo, ki jo namenoma ali iz hude malomarnosti povzroči na delu (torej, pri opravljanju svojih nalog) ali v zvezi z delom (Zakon o delovnih razmerjih: člen 177).

4.3.2 Avtomatizirana delovna sila

Ko robot (avtomatizirana delovna sila) povzroči škodo tretji osebi, najprej pogledamo predpogoje za naprtitev objektivne odgovornosti delodajalcu. Predpogoja, ki morata biti izpolnjena, da se škodno ravnanje, povzročeno s strani avtomatizirane delovne sile, smatra kot ravnanje delodajalca, sta: avtomatizirana delovna sila mora biti del substrata delodajalca; delavec, kot del substrata delodajalca, povzroči škodo med opravljanjem dela ali v zvezi z delom.

Možnost uporabe pravil o krivdni odgovornosti je v tem primeru izključena, saj avtomatizirana delavna sila zaradi svoje narave ni sposobna povzročiti škode namenoma. Torej, v primeru povzročitve škode s strani avtomatizirane delavne sile se delodajalec ne bo mogel otresti objektivne odgovornosti za dejanja, povzročena s strani njegovega substrata.

⁶ Za uspešno uveljavljanje odškodninskega zahtevka morajo biti kumulativno izpolnjene naslednje predpostavke: nastanek škode, da le-ta izvira iz protipravnega ravnanja, vzročna zveza med nastalo škodo in protipravnim ravnanjem ter odškodninska odgovornost povzročitelja škode.

Shematski prikaz:

Človeška delovna sila		Avtomatizirana delovna sila
Objektivna odgovornost - odgovornost neglede na krivdo	Pogoja za uporabo pravil objektivne odgovornosti delodajalca: 1. delavec mora biti del delodajalčevega personalnega substrata 2. škoda mora biti povzročena na delu ali v zvezi z delom (prvi odstavek 147. člena OZ)	Pogoja za uporabo pravil objektivne odgovornosti delodajalca: 1. delavec mora biti del delodajalčevega personalnega substrata 2. škoda mora biti povzročena na delu ali v zvezi z delom (prvi odstavek 147. člena OZ)
Subjektivna odgovornost - krivdna odgovornost	1. Če je delavec škodo povzročil namenoma: Oškodovanec ima pravico zahtevati povrnitev škode TUDI neposredno od delavca (torej, ali od delodajalca ali od delavca) (drugi odstavek 147. člena OZ) 2. Če se povrnitev škode, ki je bila povzročena namenoma ali iz hude malomarnosti zahteva od delodajalca: Delodajalec ima pravico zahtevati povračilo plačanega zneska od delavca, če je ta povzročil škodo namenoma ali iz hude malomarnosti. (tretji odstavek 147. člena OZ)	Uporaba pravil o subjektivni odgovornosti je v tem primeru izključena, saj avtomatizirana delavna sila zaradi svoje narave ni sposobna povzročiti škode namenoma.

Zaključek

Razvoj umetne inteligence zaradi odsotnosti njene pravne opredelitve razkriva pomanjkljivosti trenutne zakonodaje. V članku so bili na primeru medicinskih robotov predstavljeni izjemni dosežki umetne inteligence, vključene v proces zdravljenja. Odsotnost pravne ureditve in standardizacije robotov, opremljenih z naprednejšimi algoritmi, lahko privede do situacije, da jih kljub njihovim izjemnim sposobnostim ne bo mogoče uporabiti v praksi. V poglavju o pravni analizi so bila predstavljena že obstoječa pravna pravila, čemur je sledila krajša analiza njihove uporabe v primeru škode, povzročene s strani robotov, opremljenih z naprednejšimi algoritmi. Prispevek pa se zaključi s predlogom glede preureditve pravnih pravil, ki se nanašajo na delovna razmerja, ki bi po mnenju avtorice razrešil marsikatero pravno dilemo, ki se nanaša na odgovornost glede delovanja robotov.

Artificial Intelligence and Its Legal Regulation

Currently, artificial intelligence impacts virtually each area of our everyday lives. Over the last few decades, it has successfully infiltrated the labour market as well. The article will, by describing different types of healthcare robots, try to answer the key question: Why do different “development phases of robotics” require different legal approaches? At the moment, we are confronted with legal gaps concerning liability for actions performed by robots equipped with algorithms which are not provided with a specific pattern when solving the task, but create their own based on environmental inputs. Even though such robots might enhance work efficiency, the law needs to respond and provide new legal provisions that would determine the person responsible for the harm done by such robots. In the article, I evaluate a few different options for such regulation and propose ideas to re-regulate the employment relationship, which could potentially lead us to the desirable aim, namely to create legislation for the protection of individuals in case of harm done by robots.

Keywords: artificial intelligence, legal regulation, healthcare robot, algorithm, employment relationship, robot’s development phases

Literatura

- Ackerknecht, E. (1982). *A Short History of Medicine*. Baltimore: The Johns Hopkins University Press.
- Arai, N. (2017). »Can a robot pass a university entrance exam« [video]. *Dogodek TED 2017*. Dostopno na: https://www.ted.com/talks/noriko_arai_can_a_robot_pass_a_university_entrance_exam#t-805036.
- Bekele, E. T., Lahiri, U., Swanson, A. R., Crittendon, J.A., Warren, Z. E. in Sarkar, N. (2013). »A Step towards Developing Adaptive Robot-mediated Intervention Architecture (ARIA) for Children with Autism«. *IEEE Transactions on neural systems and rehabilitation engineering*, 21, str. 289–295.
- European Parliament (1985). *Council Directive 85/374/EEC of 25 July 1985* (on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products).
- »Howard Gardner's Theory of Multiple Intelligences« (2009). Northern Illinois University, Faculty Development and Instructional Design Center, str. 1 – 9.
- Gruen, L. (2017). »The moral status of animals«. V Zalta, E. N. (ur.), *Stanford Encyclopedia of Philosophy*. Dostopno na: <https://plato.stanford.edu/entries/moral-animal/>.
- Gural, U. V. (2014). »The Science Fiction Cinema«. V Gural, U. V., *Film History*, BookBaby, poglavje 12 (neoštevilčene strani).
- Hildebrandt, M. (2016). *Smart Technologies and the End(s) of Law*. Edward Elgar Publishing.
- Hiyaguha, C. (2016). »Robots Replacing surgeons?«. *Neutral Health Blog*. Dostopno na: <https://jonbarron.org/doctors-and-drugs/robots-replacing-surgeons>.
- Intuitive Surgical, Inc. (b. d.). [spletna stran]. Dostopno na: <https://www.intuitivesurgical.com>.
- Jotterand, F. (2005). »The Hippocratic Oath and Contemporary Medicine: Dialectic Between Past Ideals and Present Reality«. *Journal of Medicine and Philosophy*, 30, str. 107–128.
- McCaughrean, G. (1997). *Greek Gods and Goddesses*. Margaret K. Mcelderry Books.
- MMI (Medical Microinstruments). (b. d.). [Video predstavitev]. Dostopno na: <http://www.mmimicro.com>.
- Mramor, K. (2007). »Nevronske mreže«. Ljubljana: Fakulteta za matematiko in fiziko, Univerza v Ljubljani.

»New surgical robots are about to enter the operating theatre«. *The Economist: Science and technology* (16.11.2017), str. 71–72.

Nilsson, J. (2010). »Albert Einstein: Imagination Is More Important Than Knowledge«. *The Saturday Evening Post* (10. 3. 2010). Dostopno na: <http://www.saturdayeveningpost.com/2010/03/20/history/post-perspective/imagination-important-knowledge.html>.

Nilsson, N. J. (2010). *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge: Cambridge University Press.

Obligacijski zakonik (OZ) (2007). *Uradni list RS*, št. 97/07 – uradno prečiščeno besedilo in 64/16 – odl. US.

Odločba VDSS sodba Pdp 146/2009. Republika Slovenija Višje delovno in socialno sodišče, Oddelek za individualne in kolektivne delovne spore.

Pavčnik, M. (2011). *Teorija prava*. Ljubljana: GV Založba.

Plavšak, N., Juhart, M. in Vrenčur, R. (2009). *Obligacijsko pravo*. Ljubljana: GV Založba, str. 568 – 581.

Ploj, B. (2013). *Metoda mejnih parov za učenje umetnih nevronske mreže* (Doktorska dizertacija). Maribor: Fakulteta za elektrotehniko, računalništvo in informatiko.

Sandberg, A. (2013). »Ethics of brain emulations«. [draft]. Oxford University, Future of Humanity Institute.

Shademan, A., Decker, R., Opfermann, J., Leonard, S., Krieger, A. in Kim, P. (2016). »Supervised autonomous robotic soft tissue surgery«. *Science Translational Medicine*, 8(337).

Shelley, M. (1818). *Frankenstein*. Lackington, Hughes, Harding, Mavor & Jones. Dostopno na: <https://www.cliffsnotes.com/literature/f/frankenstein/book-summary>.

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J., Tambe, M. in Teller, A. (2016). »Artificial Intelligence and Life in 2030«. *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 (Study Panel)*, Stanford University. Dostopno na: <http://ai100.stanford.edu/2016-report>.

»Surgical Risk Document« (b. d.). *Intuitive Surgical, Inc* [spletna stran]. Dostopno na: <https://www.intuitivesurgical.com/safety/>.

»Technology« (b. d.). *Cambridge Medical Robotics* [spletna stran]. Dostopno na: <http://cmedrobotics.com/versius/technology/>.

»Emotionally intelligent robot comes to life 7 March 2016« (2016). [YouTube video] 2:17, posted by: "tNEWS," 7.3.2016. Dostopno na: https://www.youtube.com/watch?v=pXg33S3U_Oc&t=4s.

»What's new, Atlas?« (2017). [Youtube video]. Boston Dynamics. Dostopno na: <https://www.youtube.com/watch?v=fRj34o4hN4I&app=desktop>.

Zakon o delovnih razmerjih (ZDR-1) (Uradni list RS, št. 21/13, 78/13 – popr., 47/15 – ZZSDT, 33/16 – PZ-F, 52/16 in 15/17 – odl. US).

Žerovnik, J. (1989). »Verjetnostni model računanja: Deskriptorji: Računanje, Verjetnostni model, Algoritmi hevristični«. *Informatica*, 13, str. 58–66.

Toma Strle

Pedagoška fakulteta Univerze v Ljubljani

Olga Markič

Filozofska fakulteta Univerze v Ljubljani

Nevroznanost in pravo: od redukcionizma do krožnosti nevroprava

V prispevku najprej predstaviva porast vključevanja nevroznanosti v pravo (npr. uporabe nevroznanstvenih dokazov na sodišču) in na nekaterih primerih prikaževa zaupanje, ki ga vzbujajo nevroznanstveni dokazi, razlage in opisi. Nato kritično ovrednotiva redukcionističen in mehanicističen pogled nevroznanosti na človeško naravo. Zagovarjama stališče, da moramo biti pri združevanju objektivistične in redukcionistično usmerjene nevroznanosti ter prava, ki temelji na vrednotah in odločitvah pravnih odločevalcev in javnosti, previdni. Namreč, nevroznanost s svojim vedno bolj sprejetim redukcionistično-mehanicističnim pogledom na človeško naravo spreminja naša prepričanja in intuicije o pojavih, ki so relevantni za pravo (npr. osebnost, volicija, samokontrola, odločanje, ipd.), ter s tem vpliva prav na same našete in sorodne pojave. To drži še posebej v kontekstu naraščajočega trenda integracije nevroznanosti v javno domeno, kot je pravo, ki nadalje spreminja naša prepričanja in intuicije o preučevanih pojavih in tako potencialno spreminja mišljenje ter delovanje ljudi. Pogled nevroznanosti na človeško naravo in integracija le tega v pravne kontekste tako vodita do krožnosti nevroprava, kjer nevroznanost in nevropravo pravzaprav vplivata na same (prihodnje) izsledke nevroznanosti. To pa pomeni, da nevroznanosti ne moremo jemati kot objektivne znanosti o človeški duševnosti in vedenju, na podlagi katere bi pravo končno lahko sprejemalo pravilne in/ali nepristrane odločitve.

Ključne besede: moralna odgovornost, naravne in človeške vrste, nevropravo, prepričanja, redukcionizem, samokontrola, svoboda

1. Trend vključevanja nevroznanosti v pravne kontekste: od dokazov do zaupanja v nevroznanstvene razlage

Uporaba dokazov in izsledkov nevroznanosti v kontekstu prava z namenom bolj-
ših in bolj »objektivnih« pravnih odločitev – npr. odločitev o zmanjšanju ali vrsti
kazni obsojenih ali odločitev o bolj splošnih zadevah, ki se tičejo kazenske odgo-
vornosti – je v zadnjem desetletju in pol v močnem porastu. Odvetniki tistih, ki so
obtoženi kriminalnih dejanj, kot dokaze uporabljajo tako slike, ki kažejo možgan-
ske poškodbe (npr. z magnetno resonanco), kot tudi slike, ki naj bi pojasnjevale
aktivnost možganov in vedenje, dozdevno povezano s to aktivnostjo (npr. s funk-
cijsko magnetno resonanco (fMR) ali elektroencefalografijo (EEG)). Takšne vrste
dokazi naj bi na primer kazali, da obdolženci z dozdevno »nenormalno« delujo-
čimi možgani pač naj ne bili ali pa naj bi bili vsaj manj odgovorni za svoja deja-
nja.

Leta 2005 je na primer odvetnik Gradyja Nelsona, ki je brutalno umoril svojo že-
no, poskušal zagovarjati trditev, da si Grady Nelson zasluži zmanjšano kazen. To
je trdil predvsem na podlagi njegove dozdevno nenormalne možganske aktivnosti,
ki naj bi bila odgovorna za njegova dejanja (iz Jones et al., 2013b). Leta 2007 je
obramba Braunsteina, ki mu je zaradi ugrabitve, spolne zlorabe in ropa grozila
dosmrtna zaporna kazen, predstavila MR slike njegovih možganov in trdila, da le-
te kažejo na Braunsteinovo shizofrenijo, ki naj bi mu preprečila ustrezen nadzor
nad njegovim nagnjenjem k nasilju (iz Church, 2012; glej Church, 2012 in Catley,
2016 za nadaljnje primere).

Na porast uporabe nevroznanstvenih dokazov na sodiščih kažejo tudi mnoge so-
dobne analize. Analiza Farahanyjeve (2016) na primer razkriva močno porast
uporabe nevrobioloških dokazov na ameriških sodiščih med letoma 2005 in 2012.
Farahanyjeva na primer ugotavlja, da je leta 2012, v primerjavi z letom 2007, več
kot dvakrat več pravnih mnenj (približno 250) navajalo nevrobiološke dokaze z
namenom pokazati, da so možgani tisti, ki so obtoženega (vsaj delno) »prisilili« v
dejanje X (Farahany, 2016), obtoženi pa tako ni oz. je manj odgovoren za svoje
kriminalno dejanje. Farahanyjeva (2016) v svoji analizi nadalje ugotavlja, da ne-
vroznanstveni dokazi niso uporabljeni le kot strategija zagovora obsojenih na naj-
hujše zločine (kot je npr. umor ali posilstvo), ampak se vedno bolj uporabljajo kot
orodje za zagovor zločincev obsojenih ropov, prevar in preprodaje mamil. Tak-
šnih primerov je glede na njeno analizo približno 60 %.

Analiza Farahanyjeve (2016) je sicer omejena na ZDA, vendar je porast uporabe
nevroznanstvenih dokazov na sodiščih zaslediti tudi zunaj ZDA in ne le v državah
z anglosaksonskim oz. običajnim pravom¹ (ang. *Common law*): npr. na Nizozem-

¹ Anglosaksonsko oz. običajno pravo temelji na precedenčnih primerih in, vsaj v ZDA, v pravne odločitve bistveno vključuje porote.

skem, v Singapurju, Kanadi, Veliki Britaniji, Walesu, Indiji in Italiji (glej Church, 2012; Catley in Claydon, 2015; Catley, 2016). Catley in Claydon (2015) sta na primer identificirala 204 primere uporabe nevroznanstvenih dokazov v Angliji in Walesu med letoma 2005 in 2012. Med letoma 2005 in 2008 so bili nevroznanstveni dokazi v povprečju uporabljeni v 17-ih primerih na leto, med letoma 2009 in 2012 v približno 34-ih primerih.

Nadalje, dokazi in izsledki nevroznanosti so bili že uporabljeni kot osnova za pravne odločitve, ki se nanašajo na bolj splošne zadeve, povezane z moralno odgovornostjo. V primeru *Graham versus Florida* leta 2012 je vrhovno sodišče ZDA tudi na podlagi nevroznanstvenih dokazov presodilo, da bo kategorično prepovedalo obsodbo mladoletnih oseb na dosmrtno kazen brez možnosti pogojnega izpusta v primerih, ko ne gre za umor (iz Morse, 2015). Ne nazadnje, nevroznanstveni dokazi so bili že uporabljeni tudi kot osnova za obsodbo. Leta 2008 je Indija postala prva država, kjer so na sodišču dopustili uporabo tehnologije možganskih prstnih odtisov² (ang. *brain fingerprinting*) in na podlagi le-te Sharmo ter čez šest mesecev še dve osebi obsodili na smrtno kazen (iz Church, 2012). In čeprav je znanstvena skupnost relativno skeptična do trenutnih zmožnosti te tehnologije za detekcijo prikritih spominov in do uporabe funkcijske magnetne resonance za detekcijo laži (glej npr. Farah et al., 2014), pa je v znanstveni sferi kot tudi v poročilih sodnikov zaslediti pogled, da bo prihodnji razvoj omogočil uporabo takšnih tehnologij na sodiščih (glej npr. Church, 2012).

Porast vključevanja nevroznanosti v pravo pa je vidna tudi izven neposredne uporabe nevroznanstvenih dokazov in izsledkov na sodiščih. Vedno več pravnih odločevalcev (tako znanstvenikov kot pravnih ekspertov) se vključuje v diskusije in raziskovanje nevroloških osnov sodnih odločitev in pristranosti, ki jih spremljajo: na primer v diskusije in raziskovanje mehanizmov presojanja in odločanja porot, sodnikov, prič, odločitev o pogojnih izpustih ipd. Prav tako so v porastu raziskave in diskusije o možnostih uporabe fMR ali EEG tehnik za detekcijo laži, o možnostih ustvarjanja direktnih možganskih intervencijskih terapij za zločince ipd. Na skrajnem spektru uporabe nevroznanosti v pravo je zaslediti trend diskusij in raziskav o t. i. nevro-osnovani napovedi (ang. *neuro-based prediction*), katere namen je na podlagi podatkov o možganih storilcev kaznivih dejanj vnaprej predvideti potencialno možnost kriminalnih dejanj (za nedavno raziskavo glej Aharoni et al., 2013), tudi npr. v kontekstu odločitev o vrsti kazni, pogojnem izpustu, varščini ipd. (Jones et al., 2013a; glej tudi Jones et al., 2013b; 2014 za razpravo in relativno optimističen pogled na prihodnje možne uporabe nevroznanosti v pravo). V tem kontekstu si ni težko zamisliti nekakšne distopične prihodnosti, kjer bi bila nevroznanost uporabljena kot orodje za kaznovanje »zločina misli«, kot to lepo opiše Orwell v svojem distopičnem romanu *1984*.

² To je EEG tehnologija, ki naj bi omogočala razkritje prikritih informacij (npr. spominov) v možganih s pomočjo detekcije specifičnih možganskih odzivov na določene dražljaje (npr. slike) (Farwell, 2012).

Ne nazadnje se zdi, da nevroznanstveni dokazi in razlage v prejemnikih vzbujajo nekakšno posebno zaupanje, ki je najverjetneje delno odgovorno za hiter porast vključevanja nevroznanosti v pravne kontekste. Takšno posebno zaupanje je lepo vidno v nekaterih sodnih obravnavah in odločitvah porotnikov. Leta 2005 je bil Grady Nelson obtožen brutalnega umora svoje žene. Njegov odvetnik je zagovarjal zmanjšanje kazni na podlagi QEEG podatkov, ki naj bi kazali na relevantne nenormalnosti v delovanju njegovih možganov, ki naj bi bile – ne pa on sam³ – odgovorne za njegovo kriminalno dejanje (primer iz Jones et al., 2013b). Pomenljivo je, da sta v sodnem procesu dva od šestih porotnikov, ki so glasovali proti smrtni kazni (vseh porotnikov je bilo dvanajst), naknadno poročala, da sta spremenila svoje prvotno mnenje zaradi QEEG dokazov (sprva sta bila za smrtno kazen; iz Jones et al., 2013b). Številne študije prav tako kažejo, da nevroznanstveni dokazi in opisi dejanj, v nasprotju s psihološkimi, vodijo v pripisovanje manjše krivde, moralne odgovornosti in svobode volje. Raziskava Gurleyja in Marcusa (2008) na primer kaže, da so sodniki in porote bolj nagnjeni k temu, da sprejmejo zagovor na podlagi neprištevnosti, ko so predstavljeni nevroznanstveni dokazi. Študija Monterossa, Royzman in Schwartza (2005) kaže, da ljudje vedenje drugih presojajo manj ostro, ko je zanj podana fiziološka razlaga, in bolj ostro, ko je podana psihološka razlaga. Nahmias, Coates in Kvaran (2007) v svoji raziskavi med drugim kažejo, da nevroznanstveni opisi oseb in dejanj zmanjšajo pripisovanje svobode volje in moralne odgovornosti.

Vedno boljše in natančnejše razumevanje delovanja možganov, hiter razvoj neinvazivnih tehnik slikanja možganov in vedno večje zaupanje v nevroznanost in njen redukcionistično-mehanicistični pogled na človeško duševnost skratka kaže, da bo nevroznanost v bližnji prihodnosti najverjetneje postala bistven sestavni del prava in njegovih praks, njene razlage duševnosti in vedenja pa bodo čedalje bolj sprejete kot tiste prave, objektivne, najboljše itd. – znotraj in izven prava.

2. Redukcionističen in mehanicističen pogled nevroznanosti na človeško naravo

Vse do sedaj povedano nakazuje, da nevroznanost ni le na poti k temu, da postane neločljivo povezana s pravom, ampak počasi spreminja – in bo najverjetneje v prihodnje še močnejše – naše intuicije, poglede in prepričanja o tem, kaj pomeni biti oseba, kaj je in kako poteka odločanje, naš pogled na prostovoljno dejavnost in odgovornost – koncepti, ki so za pravo bistvenega pomena. Kot to lepo izrazi Farah:

³ Seveda je v kontekstu nevropa pomembno, a povsem nereflektirano vprašanje, kdo sploh naj bi bil Grady Nelson oz. kdo sploh je ta »jaz« ali dejavnik (ang. *agent*), ki naj bi bil entiteta, odgovorna za (svoja?) dejanja. S tem pa področje nevropa – predvsem tisti del, ki ni kritičen do nevroznanosti – bolj ali manj sprejema predpostavko, da so možgani pravzaprav tisti bistveni (če že ne edini) dejavnik, ki »upravlja« dejanja človeka in nad katerimi ta pravzaprav nima prav veliko zavestnega nadzora. Človek tako ni, vsaj ne v veliki meri, odgovoren za (lastna) dejanja (glej tudi nadaljevanje članka).

Nevroznanost ponuja alternativno perspektivo, iz katere lahko človeško vedenje razumemo tudi kot posledico fizičnih vzrokov. /.../ [U]porabe nevroznanosti /.../ bodo ponudile veliko opomnikov, da je naša duševnost v svojem izvoru in bistvu fizični mehanizem. S tem ko nevroznanost ljudi postavlja v in kot del mehaničnega univerzuma, prevprašuje mnoge predpostavke o moralnosti in pojmu osebe. /.../ Sicer ne trdimo: »Ni imel izbire, saj so ga prisilili zakoni fizike.« A vendar bo deterministični pogled, glede na to, da nevroznanost osebnosti, odločanja in nadzora impulzov začena ponujati bolj podrobne razlage fizičnih procesov, ki vodijo do neodgovornega ali kriminalnega vedenja, vseeno najverjetneje močnejše prevzel naše intuicije. (Farah, 2012: 585–586)

In čeprav je nevroznanstvena skupnost *relativno* skeptična do *trenutne* uporabe nevroznanosti na sodiščih in do *trenutnega* dosega nevroznanosti (ne pa toliko prihodnjega; glej npr. Jones et al., 2014), v splošnem sprejema prepričanje, da bo nevroznanost v bližnji prihodnosti pojasnila »dejansko« delovanje človeške duševnosti na bolj ali manj mehanicističen in redukcionističen način ter da bo prikrbela najboljšo (če že ne popolno) razlago človeškega vedenja (glej na primer Gazzaniga et al., 2008; Greene in Cohen, 2004; Jones et al., 2013a; 2013b; 2014, za takšen implicitno ali eksplicitno izražen pogled). Kot je napovedal že Francis Crick v svoji odmevni knjigi *The Astonishing Hypothesis* (1994), naj bi naše veselje in žalost, naši spomini in ambicije, naš občutek osebne identitete in svobodne volje bili zgolj vedenje velikih mrež živčnih celic in z njimi povezanih molekul. Po njegovem, kot temu sam pravi, znanstvenem prepričanju, bo človeški um – vedenje naših možganov – mogoče razložiti z interakcijo med nevroni (in drugimi celicami) ter z njimi povezanimi molekulami. (Crick, 1994: 3–7)

Kot ugotavlja Ted Rockwell (2007), takšno stališče temelji na empirični evidenci, da kadarkoli se dogodi duševni dogodek, se nekaj zgodi tudi v možganih. In prav tako, ko se nekaj zgodi v možganih, to zelo pogosto sproža posledice na duševne dogodke osebe, ki ima te možgane. Te vzajemne vzročne povezave po njegovem vodijo do domneve o izenačitvi duševnosti (uma) in možganov (Rockwell, 2007: 54). Še bolj radikalno stališče zavzemajo eliminativisti, ki vidijo nevroznanost kot edini ustrezeni znanstveni pristop za razlago vedenja in menijo, da je naše običajno razmišljanje o duševnosti povsem zmotno. Postavke tako imenovane zdravorazumske psihologije, na primer prepričanja, želje, strahovi, bo po njihovem mnenju doletela podobna usoda, kot je v kemiji po odkritju kisika doletela flogiston (Churchland, 1988).

Vendar se postavlja vprašanje, ali imata radikalno redukcionistično stališče in eliminativizem dejansko podporo v nevroznanstvenem raziskovanju (glej Markič, 2013). Ian Gold in Daniel Stoljar (1999) sta v svoji analizi ugotovila, da pogosto ni jasno, kaj je mišljeno s tezo »uspešna znanstvena teorija razlage duševnosti bo nevroznanstvena«. Po njunem mnenju gre za dvoumnost med dvema pogledoma.

Po eni strani gre za *trivialno nevronske teorije*, po kateri je duševnost biološki fenomen, ki ga bo potencialno mogoče razložiti z znanostjo. Vendar pri tem ni rečeno, da bo razlaga podana samo v bioloških pojmi, po vsej verjetnosti bodo potrebni tudi psihološki in poljudnopsihološki pojmi. (Gold in Stoljar, 1999: 813). Na drugi strani pa je *radikalna nevronska teorija*, po kateri naj bi nevrobiologija sama priskrbela nujen pojmovni aparat za razlago duševnosti: »Uspešna teorija duševnosti bo teorija o možganih, izražena v izrazih temeljnih strukturalnih in funkcionalnih lastnosti nevronov, povezanih skupin ali struktur« (Gold in Stoljar, 1999: 814) Avtorja podata prepričljive argumente proti radikalni nevronske teoriji, tako da kot bolj verjetno razumevanje ostaja trivialna nevronska teorija, ki pa ne vodi do radikalnih filozofskih posledic. Večjo previdnost pri interpretaciji zagovarja tudi John Bickle (1998; 2003), ki meni, da je treba počakati na nadaljnji razvoj znanstvene psihologije in nevroznanosti in šele potem, ko bosta obe discipline »dovolj zreli«, raziskovati medteoretske relacije med teorijami in morebitno redukcijo in eliminacijo.

Nevroznanstveniki ali, še večkrat in bolj natančno, interpreti nevroznanstvenih raziskovanj, dajejo »razlage«, na primer pomnjenja, strahu, ljubezni, tako da jih umestijo v določena področja možganov, nevronske mreže ali molekularnih procesov (Rose, 2012). Na ta način pripisujejo psihološke predikate možganom ali celo delom možganov. Toda, kot poudarjata Bennet in Hacker (2003), lahko zgolj za človeka ali žival rečemo, da misli, se veseli, vidi, sliši, vonja, okuša. Ljudje, ne pa njihovi možgani ali deli možganov, sprejemajo odločitve ali ostajajo neodločeni. Če torej nadomestimo osebo z možgani ali celo delom možganov, zagrešimo mereološko napako (Dennet, 1991; Bennet in Hacker, 2003).

Razloge za pristop, ki ga omenjeni avtorji označujejo kot mereološko napako v nevroznanosti, lahko najdemo v *modularni hipotezi* (glej Vörös in Markič, 2014), ki je implicitno ali odkrito prisotna v precejšnjem delu sodobnega raziskovanja. Gre za idejo, da duševnost sestavlja množica različnih kognitivnih funkcij ali modulov, ki so udeleženi v diskretnih predelih možganov. Če predpostavljamo, da so določeni predeli možganov specializirani za določene duševne funkcije, potem se zdi, da slednje lahko razložimo z vzročnimi lastnostmi teh predelov možganov. Oziroma, z drugimi besedami, tako modularno pojmovanje sloni na postavki, da so lastnosti določenega duševnega pojava zgolj vsota vseh lastnosti določenega predela možganov, za katerega se je izkazalo, da spremlja to izkustvo. Sodobne slikovne metode (fMR, PET, EEG) tako javnosti ponujajo relativno enostavne vidne podobe, ki lahko zakrijejo kompleksnost procesov, ki vodijo do teh podob. Prav te poenostavitve pa potem lahko pripeljejo do neupravičenega navdušenja nad sodobnimi tehnologijami in do nekritične uporabe tudi na področju prava, kot ugotavlja v tem članku.

Razloge za take prehitre in nekritične interpretacije je nevroetik Eric Racine opredelil s tremi pojmovanji: *nevroesencializem*, *nevrorealizem* in *nevropolitika* (Racine, 2010: 101-107). Pojem nevroesencializma identificira interpretacije, po katerih so možgani samo-opredeljujoče se bistvo osebe, nekakšna sekularna vzporednica duše. Možgani na ta način postanejo nekakšna bližnjica do pojmov, kot je sebstvo, in do značilnosti posameznika, ki jih sicer ne bi pripisovali možganom. Nevroesencializem je pogosto povezan z biološkim redukcionizmom (ibid.: 101). Pojem nevrorealizma namiguje, da raziskovanje s slikovnimi metodami prinaša neposredne podatke o možganskih funkcijah. Na primer, pogost je pogled, da so slikovne metode nekakšno »branje misli« (ibid.: 106). S tretji pojmom, nevropolitika, Racine opisuje poskuse, ko nekateri z rezultati slikovnih metod skušajo promovirati svoje osebne ali politične cilje (ibid.: 107).

Redukcionističen in mehanističen pogled je prinesel poglobljeno razumevanje nevroloških mehanizmov in se izkazal kot pomembno in uporabno eksperimentalno orodje naravoslovnih znanosti in s tem tudi nevroznanosti. Vendar, ko za interpretacije opisov in razlag procesov, ki potekajo na pod-osebni ravni, uporabimo psihološke predikate, smo naredili korak, ki ni samoumeven in temelji na določenih predpostavkah. Dotika se enega temeljnih filozofskih problemov, ki se sprašuje o odnosih med fizičnim in duševnim. Tudi če se dandanes skoraj vsi strinjamo, da možgani igrajo pomembno vlogo pri čustvovanju, kognitivnih procesih in voliciji, še vedno obstajajo različni pristopi, kako razumeti duševnost, ki se navezujejo na različne rešitve problema. Redukcionizem in eliminativizem sta le dve od možnih rešitev. Kritike, namenjene redukcionističnemu in mehanicističnemu pogledu na človeka, po najinem mnenju kažejo, da je treba kritično prevprašati predpostavke takih interpretacij in vzeti v obzir pristope, ki bolj poudarjajo prepletenost možganov, telesa in okolja (tako naravnega kot kulturnega) (glej Strle, 2013; Markič, 2013; Uršič, Markič in Ule, 2012).

3. Krožnosti nevroprava

V preostanku prispevka se bova zadržala pri stvari, ki je izpuščena iz večine diskusij nevroprava in širše. Trdila bova, da vedno bolj popularen in sprejet objektivističen in redukcionističen pogled nevroznanosti na človeka pravzaprav spreminja naša prepričanja in intuicije o pojmi in pojavih, kot so oseba, svoboda, samokontrola in odgovornost, in tako vpliva na same pojave, ki naj bi jih nevroznanost raziskovala objektivno, ločeno od naših prepričanj, teorij in intuicij o teh pojavih. Nenavadna krožnost nevroznanosti, ki je izpuščena iz večine diskusij o njenem vključevanju v pravne kontekste. Slednje drži posebej, če vzamemo v zakup porast sistematične integracije nevroznanosti in njenih predstav o duševnosti v javno domeno, kot je pravo, ki še nadalje utrjuje in krepi nevroznanstven pogled na človeško naravo ter s tem nadalje spreminja same pojave, ki jih nevroznanost

raziskuje dozdevno objektivno, ločeno od naših prepričanj, teorij in intuicij o teh pojavih.

Naj sedaj na kratko predstaviva nekaj študij, ki kažejo, kako spreminjanje človeških prepričanj o zadevah, povezanih s človeško dejavnostjo,⁴ kot sta na primer svoboda volje ali samokontrola, v smer deterministične, mehanicistične in/ali redukcionistične koncepcije teh, vpliva na samo mišljenje, delovanje in možganske procese, povezane s temi pojavi.

Seminalna raziskava Vohsa in Schoolerja (2008) na primer razkriva, da zmanjšanje prepričanj v svobodo volje zmanjšuje moralnost dejanj: vodi v povečanje dejanskega goljufanja udeležencev. Podobne študije nadalje kažejo, da je zmanjšanje prepričanj v svobodo volje povezano z zmanjšanjem pripravljenosti pomagati sočloveku in s povečano agresivnostjo (Baumeister, Masicampo in DeWall 2009); s povečano nagnjenostjo k impulzivnosti in asocialnosti (Rigoni et al., 2012), kar avtorji interpretirajo kot zmanjšanje samokontrole; Rigoni et al. (2011) pa so celo pokazali, da zmanjšanje prepričanj v svobodo volje vpliva na možganske procese, ki so povezani s pripravo voljnih motoričnih dejanj (zmanjšuje potencial pripravljenosti). Implikacije takšnih in podobnih študij lepo izrazijo Rigoni et al. (2011): »Če povzamemo, /.../ to nakazuje, da ima morda abstraktni sistem prepričanj veliko bolj fundamentalen učinek kot bi pričakovala večina ljudi« (str. 617).

Podobno kažejo raziskave t. i. učinka izčrpanosti ega (ang. *ego-depletion effect*), ki so močno relevantne za pravo, saj poskušajo razumeti mehanizme samokontrole. Učinek izčrpanosti ega naj bi kazal, da je moč volje (ang. *willpower*) oziroma zmožnost samokontrole omejena (glej na primer Baumeister, 2002), na primer na način, da uporaba samokontrole za opravljanje mentalnih nalog vodi v slabše reševanje nadaljnjih mentalnih nalog oz. zmanjšuje nadaljnjo zmožnost samokontrole. Po drugi strani novejša raziskava jasno kažejo, da sta učinek izčrpanosti ega in tako samokontrola močno odvisna od implicitnih in eksplicitnih prepričanj ljudi o samokontroli in naravi »moči« volje. Savani in Job (2017) sta na primer pokazala, da je učinek izčrpanosti ega močno odvisen od kulturnih prepričanj ljudi o uporabi in naravi samokontrole. V svoji raziskavi sta pokazala, da Indijci – ki so v nasprotju z Američani v splošnem prepričani, da naprežanje volje vodi v povečanje (mentalne) energije (to je bil prvi rezultat njune študije) – izkazujejo ravno nasproten učinek od učinka izčrpanosti ega. Namreč, uporaba samokontrole Indijcev je, v nasprotju z Američani, izboljšala njihovo nadaljnje reševanje nalog, ki so prav tako zahtevale mentalni napor. Takšen učinek so izkazali tudi Američani, ki so imeli podobna prepričanja o naravi samokontrole kot Indijci.

⁴ Prevod angleškega izraza *agency*.

Omenjene raziskave skratka kažejo, da pojavi, kot so samokontrola, volicija, odločanje, prostovoljna dejavnost itd., nikakor niso zgolj biološke narave (oz. jih ni možno razložiti zgolj z biološkimi in/ali nevrološkimi mehanizmi), ampak so pojavi, ki so neločljivo povezani in odvisni od naših konceptualizacij, opisov in prepričanj o njih. So nadalje pojavi, ki so neločljivo povezani s kulturnim in družbenim prostorom, v katera smo neodtujljivo umeščeni. To pa pomeni, da posameznikova prepričanja in koncepcije o teh pojavih pravzaprav so-konstituirajo same pojave, ki naj bi jih nevroznanost preučevala objektivno, ločeno od kulturnih in individualnih prepričanj, intuicij in teorij o njih. Posledično nevroznanost s svojimi objektivističnimi in redukcionističnimi metodološkimi usmeritvami in spremljajočim pogledom na človeško naravo – ki ga, predvsem preko porasta integracije nevroznanosti v javno domeno, kot je pravo, vedno bolj sprejema tudi javnost – pravzaprav spreminja pojave, ki jih preučuje in tako potencialno vpliva na lastne (prihodnje) empirične izsledke.

Če misel izrazimo v jeziku Hackinga (1995), omenjeni pojavi, ki jih preučuje nevroznanost, izsledke o njih pa vedno bolj zvesto uporablja pravo, niso *naravne vrste* (kar bi si želela in kar predpostavlja nevroznanost), ampak *človeške vrste*, ki so bistveno odvisne od klasifikacij, opisov in človeških dejanj, ki se jih tičejo, ter socialnosti, v katero so umeščene, in ki kot take »lahko vplivajo same nase« (po Brinkmann, 2005). Kot je to provokativno izrazil Alasdair MacIntyre (1985; iz Brinkmann 2005): »Vodi je vseeno, kako jo imenujemo, a za človeške vrste je to [bistvena] razlika. Konec koncev, /.../ molekule ne berejo učbenikov kemije, medtem ko ljudje berejo knjige o psihologiji, ki vplivajo na njihovo samorazumevanje« (Brinkmann, 2005: 775), in treba je dodati, tako tudi na njihovo mišljenje in (moralna) dejanja.

V tem smislu ne moremo trditi, da so redukcionistične in/ali mehanicistične razlage človeške narave, ki jih podaja nevroznanost, ustrezne ali zadostne, tako v kontekstu prava in izven. Nadalje to pomeni, da nevroznanosti ne moremo jemati kot objektivne znanosti o človeški duševnosti in vedenju, na podlagi katere bi pravo *končno* lahko sklepalo pravilne in/ali nepristrane odločitve.

4. Zaključek

Naj povzameva. Tem bolj ko nevroznanost sprejema in spodbuja mehanicistično-redukcionističen pogled na človeško naravo in bolj in bolj ko je ta sprejet s strani javnost in pravnih odločevalcev – in tako nevroznanost postaja vedno bolj sistematično integrirana v pravne prakse in sisteme –, čedalje bolj mišljenje in dejanja ljudi odražajo takšno koncepcijo človeške narave. S tem pa, ko bo (bi) opisani proces krožno napredoval in se razvijal, se bo (bi) vedno bolj zdelo, da ima nevroznanost prav v trditvi, da imamo v »resnici« zelo malo (če sploh) vpliva na lastno mišljenje in dejanja (manj, kot bi ga lahko imeli, če bi imeli drugačna pre-

pričanja o človeški naravi) in da smo tako veliko manj (manj, kot če bi imeli drugačna prepričanja o človeški naravi) ali pa sploh ne(?) odgovorni za lastna kriminalna ali običajna dejanja.

Če privzamemo opisani krožnosti nevroznanosti in nevproprava, nevroznanosti ne moremo jemati kot tiste prave objektivne znanosti o človeški duševnosti in vedenju, ki bo pravu končno prinesla tako dolgo iskani nespremenljivi »temeljni kamen modrosti«, na podlagi katerega bo končno lahko sklepala pravilne in nepristrane odločitve. Pri združevanju objektivistične in redukcioniistično usmerjene nevroznanosti ter prava, ki temelji na vrednotah in odločitvah pravnih odločevalcev in javnosti, moramo biti tako bolj previdni, kot nakazuje prikazani trend integracije nevroznanosti v pravne kontekste.*

Neuroscience and Law: From the Reductionism to the Looping Effects of Neurolaw

In the article, we first present the growing trend of using neuroscientific evidence and findings in legal contexts. Second, we briefly present with a few examples the trust neuroscientific explanations and descriptions seem to prompt in laypersons and legal decision-makers. Third, we critically elucidate neuroscience's reductionist and mechanistic view of the human nature. We argue that we need to be cautious in combining objectivist and reductionist neuroscience with the law based on the values and decisions of legal decision-makers and the public given that neuroscience's reductionist-mechanistic view of human agency alters our beliefs and intuitions about phenomena, which are important to the law (such as personhood, volition, self-control, decision-making), thus influencing these same beliefs and related phenomena. We suggest that this holds even more so in the context of an increasing integration of neuroscience into the public domain, such as the law, which further alters our beliefs and intuitions on the studied phenomena, thus possibly changing human thinking and activity. Neuroscience's view of human nature and its integration into legal contexts thus lead to the looping effects of the law where neuroscience and neurolaw in fact influence the (future) findings of neuroscience. This in turn means that neuroscience is not an objective science of human nature and behavior on which the law could finally base its correct and/or unbiased decisions.

Keywords: moral responsibility, natural and human kinds, neurolaw, beliefs, reductionism, self-control, freedom

* Del prispevka temelji na delu, ki je bilo predstavljeno na 20. mednarodni multikonferenci Informacijska družba (podkonferenca Kognitivna znanost) in objavljeno v obliki razširjenega povzetka (Strle, 2017).

Literatura

Aharoni, E., Vincent, G. M., Harenski, C. L., Calhoun, V. D., Sinnott-Armstrong, W., Gazzaniga, M. S. in Kiehl, K. A. (2013). »Neuroprediction of Future Rearrest.« *Proceedings of the National Academy of Sciences of the United States of America*, 110(15), str. 6223–6228.

Baumeister, R. F. (2002). »Ego Depletion and Self-Control Failure: An Energy Model of the Self's Executive Function«. *Self and Identity*, 1(2), str. 129–136.

Baumeister, R. F., Masicampo, E. J. in DeWall, C. N. (2009). »Prosocial Benefits of Feeling Free: Disbelief in Free Will Increases Aggression and Reduces Helpfulness«. *Personality and Social Psychology Bulletin*, 35(2), str. 260–268.

Bennett, M. in Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Oxford: Wiley-Blackwell.

Bickle, J. (1998). *Psychoneural Reduction: The New Wave*. Cambridge, Mass.: The MIT Press.

Bickle, J. (2003). *Philosophy and Neuroscience: A Ruthlessly Reductive Account*. Dordrecht: Kluwer Academic Publishers.

Brinkmann, S. (2005). »Human Kinds and Looping Effects in Psychology«. *Theory & Psychology*, 15(6), str. 769–791.

Catley, P. (2016). »The Future of Neurolaw«. *European Journal of Current Legal Issues*, 22(2), spletna publikacija. Dostopno na: <http://webjcli.org/article/view/487/651> [6. 12. 2017].

Catley, P. in Claydon, L. (2015). »The Use of Neuroscientific Evidence in the Courtroom by Those Accused of Criminal Offenses in England and Wales«. *Journal of Law and the Biosciences*, 2(3), str. 510–549.

Church, D. J. (2012). »Neuroscience in the Courtroom: An International Concern«. *William & Mary Law Review*, 53(5), str. 1825–1854.

Churchland, P. (1988). *Matter and Consciousness*. Cambridge, Mass.: The MIT Press.

Crick, F. (1994). *The Astonishing Hypothesis: The scientific search for the soul*. London: Touchstone Books.

Dennett, D. (1991). *Consciousness Explained*. Boston: Little, Brown and Company.

Farah, M. J. (2012). »Neuroethics: the ethical, legal and societal impact of neuroscience«. *Annual Review of Psychology*, 63, str. 571–591.

Farah, M. J., Hutchinson, J. B., Phelps, E. A. in Wagner, A. D. (2014). »Functional MRI-based lie detection: scientific and societal challenges«. *Nature Reviews Neuroscience*, 15(2), str. 123–131.

- Farahany, N. A. (2016). »Neuroscience and Behavioral Genetics in US Criminal Law: An Empirical Analysis«. *Journal of Law and the Biosciences*, 2(3), str. 485–509.
- Farwell, L. A. (2012). »Brain Fingerprinting: A Comprehensive Tutorial Review of Detection of Concealed Information with Event-Related Brain Potentials«. *Cognitive Neurodynamics*, 6(2), str. 115–154.
- Gazzaniga, M. S., Keil, F. C., Goodstein, J., Rawson, E., Gray, J. R. in Passingham, R. E. (2008). »The Law and Neuroscience«. *Neuron*, 60(3), str. 412–415.
- Gold, I. in Stojlar, D. (1999). »A neuron doctrine in the philosophy of neuroscience«. *Behavioral and brain sciences*, 22(5), str. 809–869.
- Greene, J. in Cohen J. (2004). »For the Law, Neuroscience Changes Nothing and Everything«. *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences*, 359(1451), str. 1775–1785.
- Gurley, J. R. in Marcus, D. K. (2008). »The Effects of Neuroimaging and Brain Injury on Insanity Defenses«. *Behavioral Sciences & the Law*, 26(1), str. 85–97.
- Hacking, I. (1995). »The looping effect of human kinds«. V Sperber, D., Premack, D. in Premack, A. J. (ur.), *Causal cognition: A multidisciplinary debate*, New York: Oxford University Press, str. 351–383.
- Jones, O. D., Marois, R., Farah, M. J. in Greely, H. T. (2013a). »Law and Neuroscience«. *The Journal of Neuroscience*, 33(45), str. 17624–17630.
- Jones, O. D., Wagner, A. D., Faigman, D. L. in Raichle, M. E. (2013b). »Neuroscientists in court«. *Nature Reviews Neuroscience*, 14, str. 730–736.
- Jones, O. D., Bonnie, R. J., Casey, B. J., Davis, A., Faigman, D. L., Hoffman, M., Montague, R. et al. (2014). »Law and Neuroscience: Recommendations Submitted to the President’s Bioethics Commission«. *Journal of Law and the Biosciences*, 1(2), str. 224–236.
- Markič, O. (2013). »The philosophical framework for understanding neuroscientific research«. *Interdisciplinary description of complex systems*, 11(4), str. 351–362.
- Monterosso, J., Royzman, E. B. in Schwartz, B. (2005). »Explaining Away Responsibility: Effects of Scientific Explanation on Perceived Culpability«. *Ethics & Behavior*, 15(2), str. 139–158.
- Morse, S. J. (2015). »Criminal Law and Common Sense: An Essay on the Perils and Promise of Neuroscience«. *Marquette Law Review*, 99, str. 39–72.
- Nahmias, E., Coates, D. J. in Kvaran, T. (2007). »Free will, moral responsibility, and mechanism: experiments on folk intuitions.« *Midwest Studies in Philosophy*, 31(1), str. 214–242.

- Racine, E. (2010). *Pragmatic Neuroethics*. Cambridge, Mass.: The MIT Press.
- Rigoni, D., Kühn, S., Gaudino, G., Sartori, G. in Brass, M. (2012). »Reducing Self-Control by Weakening Belief in Free Will«. *Consciousness and Cognition*, 21(3), str. 1482–1490.
- Rigoni, D., Kühn, S., Sartori, G. in Brass, M. (2011). »Inducing Disbelief in Free Will Alters Brain Correlates of Preconscious Motor Preparation«. *Psychological Science*, 22(5), str. 613–618.
- Rockwell, T. (2007). *Neither Brain nor Ghost*. Cambridge, Mass.: The MIT Press.
- Rose, S. (2012). »The Need for a Critical Neuroscience: From Neuroideology to Neurotechnology.« V Slaby, J. in Choudhury, S., (ur.), *Critical Neuroscience: A Handbook of the Social and Cultural Contexts of Neuroscience*, Malden, Oxford: Wiley-Blackwell, str. 53–66.
- Savani, K. in Job, V. (2017). »Reverse Ego-Depletion: Acts of Self-Control Can Improve Subsequent Performance in Indian Cultural Contexts«. *Journal of Personality and Social Psychology*, 113(4), str. 589–607.
- Strle, T. (2017). »Problematicnost integracije nevroznanosti v pravne kontekste«. V Markič, O. et al. (ur.), *Kognitivna znanost: zbornik 20. mednarodne multikonference Informacijska družba - IS 2017, zvezek B*, Ljubljana: Institut »Jožef Stefan«, str. 56–59.
- Strle, T. (2013). »Uganka izkustva: vloga prvo- in drugoosebnih metod v kognitivni znanosti«. *Analiza*, 17(1/2), str. 41–60.
- Uršič, M., Markič, O. in Ule, A. (2012). *Mind in nature: from science to philosophy*. New York: Nova Science Publishers.
- Vohs, K. D. in Schooler, J. W. (2008). »The Value of Believing in Free Will: Encouraging a Belief in Determinism Increases Cheating«. *Psychological Science*, 19(1), str. 49–54.
- Vörös, S. in Markič, O. (2014). »Photograph of the Soul: Towards a Critical Neuroscience«. *Teorija in praksa*, 51(5), str. 811–831.

Zvezdan Pirtošek

Medicinska fakulteta Univerze v Ljubljani

Možgani, nevrotehnologija in etika

V članku opisujemo genetsko-evolucijske in civilizacijske mehanizme, ki ves čas, a na različnih časovnih skalah, spreminjajo zgradbo in delovanje možganov. V zadnjih tisočletjih postaja nevroplastičnost vse pomembnejši mehanizem hitrejših adaptacij, zlasti ob velikih tehnoloških in kulturnih prebojih (iznajdba pisave, internet), in vodi do reorganizacije možganskih struktur in novega uma. Uvedba novih nevrotehnologij, s katerimi lahko še bolj dramatično in hipno spreminjamo možganske funkcije v bolezni in zdravju in ki nam že omogočajo interakcijo med možgani in računalnikom, pa človeka postavlja pred velika vprašanja bistva in nadaljnjega razvoja in usode človeka. Prav zato se kot imperativ postavlja potreba po novi etiki, ki mora (pro)aktivno sooblikovati dostikrat nejasno mejo med znanjem možnega in modro uporabo le-tega, morda v duhu *recta ratio speculabilium*, *recta ratio agibilium*.

Ključne besede: možgani, nevroplastičnost, nevrotehnologija, internet, etika

Nevroplastičnost med empiricismom in racionalizmom

Človeško iskanje resnice in znanja je tisočletja zaznamovala dihotomija med empiričnim in racionalnim. Dileme se je na svojstven način dotaknila in jo delno razrešila sodobna nevroznanost. Nevrologija je večji del dvajsetega stoletja možgane razlagala kot okamenel organ, kjer nove možganske celice – kakor je trdil ugledni nevropatolog, prejemnik Nobelove nagrade, Ramon y Cajal – ne nastajajo. Nekateri znanstveniki so se temu sicer upirali in poudarjali plastičnost možganov (William James je v svojem monumentalnem delu *Principles of Psychology* (1890) poudarjal, da je živčevje izrazito plastično, in mladi nevrolog Sigmund Freud je v svojem nikoli objavljenem delu *Načrt za znanstveno psihologijo* menil, da se povezave med nevroni lahko spreminjajo glede na izkušnjo), a šele konec stoletja so z raziskavami prišli namigi (Paul in sodelavci, 1972) in nato nedvoumni dokazi (Kandel, 2006), da se možganske celice porajajo celo življenje in stiki med njimi vzpostavljajo ob novih izkušnjah.

Na ta način je nevrologija vzpostavila komplementarnost med empiricismom in racionalizmom: genska zasnova, osnovna struktura možganov ter grobe povezave med možganskimi predeli so kantovsko vrojene, sinaptične povezave med nevroni pa se rojevajo in ugašajo z izkušnjami in lockovsko oblikujejo in preoblikujejo

um in možgane. Nevroplastičnost ni prisotna le med sinapsami utrjenih sistemov, ki jih zavestno aktiviramo (npr. spomin v hipokampusu). Je avtomatična in podzavedna, po potrebi se vzpostavi tudi med sicer oddaljenimi in različnimi možganskimi sistemi: pri človeku, ki oslepi, se v asociacijske vidne predele možganov vrastejo slušni nevroni, ali – ko se nauči Braillove pisave – nevroni za dotik. Pri oglušelih se aktivira tisti del možganov, ki procesira dražljaje z obrob-nih predelov vidnega polja in omogoči, da ne-slišano zajamejo z vidom. Pri londonskih taksistih so našli relativno povečan zadnji del hipokampusa (pomemben za prostorsko reprezentacijo), morda na račun sprednjega dela (Maguire in sodelavci, 2000). Nevroplastičnost je prisotna ves čas in povsod, vznikne lahko v urah ali dnevih. Nekatere raziskave descartovsko nakazujejo, da je za nevroplastične spremembe v možganih dovolj že sama predstava aktivnosti ali misel na motorično akcijo (npr. igranje klavirja; gl. Pascual-Leone in sodelavci, 1995). Študija iz leta 2009 je pokazala, da se ob branju zgodbe aktivirajo vsebini pripovedi ustrezni deli možganov – včasih kognitivni, drugič motorični ali senzorični predeli (Speer in sodelavci, 2009).

A nevroplastičnost v evoluciji človeku prinaša dobro in zlo. Osvobaja ga spon genetskega in strukturnega determinizma in odpira vrata svobodni volji, omogoča mu, da se vse življenje uči, popravlja napake, postavlja hipoteze, opaža, si pomembno zapisuje v spomin in po potrebi ta spomin briše. Omogoča razvoj novih farmakoloških in nefarmakoloških terapevtskih pristopov za najrazličnejše bolezni. Po drugi strani pa lahko – s ponavljanimi in vse močnejšimi vzpostavitvami sinaptičnih povezav – prej novo aktivnost spremeni v rigidno navado, potrebo, prisilno misel (obsesijo), dejanje (kompulzijo) ali odvisnost.

Možgani med geni, evolucijo in nevroplastičnostjo

Učinek nevroplastičnih sprememb v zgradbi in delovanju možganov pa lahko opazujemo tudi skozi prizmo razvoja civilizacije, ne samo našega individualnega življenja. To velja še posebej v obdobjih velikih tehnoloških in kognitivnih prebojev. Milijone let je možgane naših prednikov oblikovala evolucija – postopne genetske spremembe kot prilagoditev na nove izzive v okolju. V zadnjih treh milijonih let so se možgani hitro povečevali, vsaj delno sta za to odgovorna gena *microcefalina* in *ASPM*, ki regulirata rast možganov in ki še vedno usmerjata evolucijo tega organa. A ti vse večji možgani niso bili zmožni večjega kognitivnega preskoka. Milijone let je naš prednik hodil po afriških planjavah, vse bolj vzravnano, s preprostim kamnitim orodjem, ki ga je v stotisočih letih malenkostno izboljševal, in v komunikacijsko vse bolj povezani skupini. V zadnjih nekaj sto tisoč letih je verjetno prišlo do genetskih sprememb, ki so v možgane naših prednikov vgradile sposobnost razumevanja načina razmišljanja soljudi, razumevanja njihovih vokalizacij (v temporalnem režnju) in motorično izražanje, vokaliziranje lastnih

misli (v frontalnem režnju). Kdaj natančno je človek ali njegov prednik spregovoril, ne vemo.

A do pravega rojstva uma je prišlo šele pred približno štirideset tisoč leti. Takrat se je pojavila nova varianta gena mikrocefalina, a natančnih vzrokov tega nenadnega kognitivnega »velikega poka« ne poznamo; takrat človek prične na stene paleolitskih jam odtiskovati svoje z okro premazane dlani; risati in iz gline oblikovati podobe živali in ljudi; z noto dostojanstva, reminiscence in vere pokopavati in častiti umrle. Takrat nekako se ob komaj opaznih evolucijskih genetskih spremembah začnejo v možganih z vrtozglavo hitrostjo razvijati mehanizmi, ki omogočajo hitre spremembe, opazne skozi nekaj generacij ali pa že v življenju posameznika. Nov preskok se s prvimi civilizacijami pojavi pred približno 7000 leti (v času, ko se pojavi tudi nova varianta gena ASPM), ko vzniknejo sposobnosti zapisovanja, računanja, branja, merjenja časa, kartografskih predstav. Vse te pridobitve so bile odkrite v zadnjih nekaj tisoč letih in vsak otrok se jih mora naučiti znova; za razliko od govora, ki se je strukturiral in oblikoval skozi dolga obdobja evolucije, verjetno s postopnimi mutacijami genov in spremembami možganske strukture, in ki v otroku samo čaka, da bo – s poslušanjem – nekega dne spontano odkrit. Cerebralna osnova teh novih, relativno hitrih sprememb, je v veliki meri nevroplastičnost.

A tudi te civilizacijske (v nekem smislu »neevolucijske«) sposobnosti zapisovanja, računanja, branja, merjenja časa, kartografskih predstav se razvijajo skozi več stopenj, in na vsaki stopnji je potrebna prilagoditev možganov; včasih zgolj okrepitev obstoječih sinaps, včasih vzpostavitev novih; drugič spet povezave prej slabo ali nepovezanih predelov možganov. Grobo strukturno v zadnjih nekaj tisoč letih možgani ostajajo sicer v glavnem enaki (so se pa, začuda, v zadnjih nekaj deset tisoč letih zmanjšali za približno 10%), a opisane prilagoditve lahko dokažemo z novejšimi, bolj natančnimi slikovnimi in elektrofiziološkimi metodami: možgani nepismenih se tako razlikujejo od možganov pismenih ljudi (Ostrosky-Solís in sodelavci, 2004); angleško govorečih od italijansko govorečih (Paulesu in sodelavci, 2001).

Prostorska predstava sveta se v zori civilizacij začne s primitivno, perceptivno konkretno risbo tega, kar vidimo; brez perspektive, objektivnih razdalj in razmerij. Tisoči let bodo minili, da bo zemljevid z uporabo kompasov in drugih zapletenih merilnih naprav postal natančna, a abstraktna podoba pokrajine. *Občutenje časa* se prične kot linearen ali ciklični tok, ki ga je bilo moč opazovati s potjo planetov in s sončnimi, peščenimi ali vodnimi urami. Šele ko Cerkev v poznem srednjem veku zapove čaščenje boga ob točno določenih časih, ob vsaki uri zadonijo cerkveni zvonovi, v stolpih mestnih hiš pa ob vrtenju mehaničnih kazalcev zakikirikajo petelini in se zavrtijo svetniki. *Pisava* se rodi v rodovitnem predelu med Evfratom in Tigrisom pred več kot pet tisoč leti in pomeni eno največjih revolucij v razvoju civilizacije. Znakovno se je razvijala skozi več stopenj: od pre-

prostih znakov, vtisnjenih v glinasto ploščico, preko logogramov in hieroglifov na papirusu, ki so jih razumeli le izbranci, do feničanskih in grških fonetičnih znakov, sprva zapisovanih brez presledkov (*scriptura continua*).

Prehod od oralnega k pismenem je radikalno spremenil način razmišljanja in s tem strukturo in delovanje možganov. Mentalni svet oralne komunikacije, zasidrane v individualnem in kolektivnem spominu, ritmične in metrično elokventne, recitativne, emocionalno poudarjene in v skupnost zazrte, se je z razvojem pismenosti razvil v novo zavest, manj odvisno od spomina, bolj racionalno in premišljeno, razčlenjeno, a tudi bolj zasebno in individualno. Ure in merjenje časa pričnejo urejevati nov ritem dela, zabave in počitka, v um pa vnašati potrebo po natančnosti, odmerjanju, metodičnosti; ritmi hipotalamusa se usklajujejo s potrebami prefrontalnega režnja. Rojeva se bolj znanstveno-metodičen um. Čitanje zemljevidov okrepi sivo substanco v zadnjem delu hipokampusu. Računanje okrepi povezave med predeli, s katerimi zaznavamo, ločujemo in poimenujemo prste (*angularni girus*) s predeli v frontalnem režnju, ki omogočajo abstraktno razmišljanje. Zapis prvega semantičnega znaka na glinasto ploščico že vzpostavlja bolj trdno povezavo med vidno in semantično skorjo, odgovorno za konceptualno procesiranje. Uvajanje na stotine hieroglifov zahteva dodatno sodelovanje predelov za spomin, prostorsko analizo in odločanje, saj je pomen znaka odvisen od konteksta. Pogled na ideogram, piktogram, črko, hieroglif pa zadnjih 5000 let avtomatično sproži aktivnost predela možganov med okcipitalnim in temporalnim režnjem, nekoč v evoluciji namenjenega za prepoznavanje linij, (pre)križanj, kotov, tistih vidnih elementov, na katere smo v okolju morali biti posebej pozorni. Ta predel možganov se aktivira ne glede na obliko, velikost, font, in znake prepoznamo kot zapisan jezikovni znak.

Na splošno je s časom opazen premik aktivnosti iz predelov možganov, ki svet dojemajo konkretno, senzorično, skupnostno, proti tistim predelom, ki so povezani z abstraktnim razmišljanjem, višjo kognicijo, virtualnostjo, individualnostjo. A potrebno je poudariti, da se včasih zgodi tudi obratno. Tak primer je prehod od kompleksnih logogramov in hieroglifov na zgolj 24 fonetičnih znakov grškega alfabeta, ki vodi do določene »razbremenitve«, zahteva manj spominskega napora in uvaja bolj preprosto senzorično procesiranje. Taka senzorična razbremenitev pa lahko omogoči večjo učinkovitost drugih možganskih procesov – npr. usmerjene pozornosti. Usmerjena in poglobljena pozornost je namreč energetsko posebej zahtevna, saj je za človeka bolj prirodna in evolucijsko razumljiva razpršena pozornost, s katero nadzoruje predvsem za preživetje pomembne spremembe v okoljskih senzoričnih vzorcih.

Vse spremembe, zlasti pa in posebej tehnološke revolucije in dosežki, preko mehanizmov nevroplastičnosti spreminjajo delovanje možganov in naše razumevanje sveta. A vendar te spremembe ne vodijo v popolnoma novo, niso izključujoče, nepovratne in absolutne. Večinoma gre prej za dopolnjevanje, za novo zmožnost

ob ohranjanju ali modificiranju starejše. S prvimi zapisi si je človek razprl nov svet in novo zavest – a s staro oralno tradicijo ne prekine. Tudi v sodobnem svetu ostaja in uživa v svetu razgovorov, posluša predavanja, pridige, hodi v gledališče. Ob zapisovanju svojih spominov na kamen ali papirus – v nasprotju s tem, kar je napovedoval Sokrat v Fajdrosu – ni izgubil svojih spominskih zmožnosti. Ob vzniku poglobljene, usmerjene pozornosti ni izgubil zmožnosti razpršene, hitro se pretakajoče pozornosti in jo je kaj hitro reaktiviral ob vzniku interneta.

Ta sposobnost dodajanja novih zmožnosti ob ohranjanju starih je del evolucije in fiziologije možganov. Nove zmožnosti pogosto nastajajo tako, da evolucija ohrani starejši del možganov in tudi njegovo funkcijo, a nadenj v milijonih let naloži nov sloj možganovine, ki prinaša nove zmožnosti. Včasih – zlasti ob pritisku tehnoloških in civilizacijskih sprememb – pa ne gradi novih možganskih delov, ampak za novost uporabi starejši del možganov, zlasti, če je nova funkcija na nek način povezana s starejšo (primer je prej opisani predel za avtomatično prepoznavo črk v možganski skorji, ki sicer prepozna pomembne linije, obris, nagibe, kote).

Možgani in digitalni mediji

Nekaj tisoč let po iznajdbi zapisovanja črk in števil se soočamo z novim tehnološkim preskokom, ki že v veliki meri spreminja in dopolnjuje naše mentalne funkcije. Gre za razvoj digitalnih medijev, zlasti interneta. Brskanje po internetu nas zasipa z obiljem delno paralelnih in delno izmenjujočih in hitro spreminjajočih se vidnih in slušnih dražljajev, klikanje, uporaba tipkovnic in mišk pa ustvarja drugačne somatosenzorične in proprioceptivne vzorce. Internet aktivira predvsem deljeno, razpršeno pozornost, izvršilne funkcije, procese odločanja in izbire. Reorganizirajo se spomin in procesi učenja. Manj je poudarka na osvajanju in zadrževanju dejstev, manj je poudarka na učenju na pamet. Semantične in epizodične spomine ter znanje o dejstvih vse bolj predajamo »zunanjim izvajalcem«, bolj pa postaja pomembna sposobnost iskanja informacij in fluidnega vzdrževanja informacij (delovni spomin), ki nam omogoča, da v svetu instantnega dostopa do nešteti informacij plujemo k tistemu, kar iščemo. Internet krepi predvsem bolj površno, razpršeno pozornost, ki je sicer pri človeku bolj »prirodno« stanje, saj skozi evolucijo omogoča hitre preskoke pozornosti na tiste senzorične dražljaje, ki se spreminjajo in so morebitno pomembni za beg ali boj. Senzoričnih dražljajev, ki se ne spreminjajo, sčasoma skoraj ne opazimo več (Kandel, 2006). Vendar nam kot vrsti tudi vzdrževana pozornost ni tuja. Med lovom je včasih najbolj koristno tiho, vztrajno in poglobljeno opazovanje.

Z internetom si izgrajujemo možnosti drugega sveta. Ob še vedno ne izgubljenem starem svetu poglobljenih doživetij smo vse bolj sposobni tudi sicer bolj površinskega, a hitrega in izjemno širokega pridobivanja informacij.

Raziskave so jasno pokazale, da uporaba interneta že v nekaj dneh spremeni delovanje možganov. Small in sodelavci (2009) so slikali možgane preiskovancev, ki so redno brskali po internetu, in tistih, ki interneta niso uporabljali. Pri uporabnikih interneta so se aktivirali širši predeli možganov, zlasti pa evolucijsko najmlajši predel možganov, dorzolateralni prefrontalni korteks leve poloble, povezan s procesi odločanja, kratkoročnega delovnega spomina, vrednotenja in reševanja problemov. V drugem delu študije so preiskovance, ki niso imeli izkušenj z internetom, šest dni po eno uro na dan uvajali v internet. V tem kratkem času se je tudi pri tej skupini pojavila izrazita aktivnost levega dorzolateralnega predela.

Te študije dokazujejo, kar intuitivno že vemo: internet, nov digitalni medij, razvija poudarek na nekoliko drugačnih mentalnih funkcijah kot starejši mediji, npr. časopis ali knjiga. Vprašanje pa je, če s tem nepovratno izgubljam tiste sposobnosti, ki so tako značilne za čitanje knjig – sposobnost dolgoročnega spomina, pomnjenja dejstev, usmerjeno pozornost in poglobljeno razumevanje. Zaenkrat se zdi, da temu ni tako. V skladu z biologijo nevropastičnosti se lahko kaj hitro vzpostavijo starejše nevrološke poti, ko in če jih potrebujemo. Če smo preplavljeni z dražljaji, jih lahko zavestno filtriramo (v veliki meri jih sicer filtrira že fiziologija receptorjev). Ni znakov, da bi zaradi uporabe interneta, kot bi lahko sklepali iz naslova popularne knjige (Spitzer, 2012), postajali dementni. Tudi *multitasking* fiziološko ni mogoč: možgani ne morejo sočasno procesirati večjega števila aktivnosti, lahko le zelo hitro menjajo pozornost med eno in drugo dejavnostjo – velikokrat na račun učinkovitosti. Obstaja pa – bolj kot doslej – nevarnost odvisnosti od interneta.

Tudi danes se z vznikom novih nevrotehnologij nahajamo med dvema tehnološkima svetovoma. Izkušnje nas učijo, da je skozi zgodovino človek bil sposoben zadržati najboljše od starega, sprejeti dobro novega in zavreči ali omejiti slabo novega. Res pa je tudi, da je izziv nove digitalne tehnologije drugačen in močnejši kot v preteklosti. Digitalni svet, v katerega smo vstopili, je izrazito dvosmeren: daje nam informacije, a z informacijami ga oblikujemo tudi mi sami in preko njega spletno socialne mreže. Širi se z vrtoglavo hitrostjo in človek preživi vse več časa v interakciji z digitalnimi mediji. Ob vsaki tehnološki revoluciji se nekoliko spremeni in dopolni tudi etika. Dostikrat ta etična dimenzija v svojem času ostaja nevidna in neizrečena; drugič je spet preveč apokaliptična (novost nas bo pokopala) ali pretirano optimistična (novost služi človeku in nanj ne bo vplivala); po pravilu pa je kot zelo pomembna prepoznana (še) v vzratnem zrcalu zgodovine.

Redkokdaj se je v zgodovini človeštva pokazala tako akutna potreba po odgovornosti in po etiki, kot smo ji priča danes, v dobi mogočne nevroznanosti in mogočnih nevrotehnologij. Po eni strani je to zato, ker je stopnja našega (samo)zavedanja višja kot kadarkoli prej v zgodovini, in po drugi, ker ravno digital-

na tehnologija in njene neslutene možnosti odpirajo temeljna vprašanja naše bodočnosti kot človeške vrste.

Možgani v dobi visokih tehnologij

A če je v zadnjih milijonih let možgane in sposobnosti človeka in njegovih prednikov relativno počasi spreminjala evolucija, klimatske in genetske spremembe, v zadnjih nekaj tisoč letih pa preko nevroplastičnosti tudi hitro razvijajoče se civilizacijske pridobitve, smo danes soočeni s spremembo druge vrste in večjim izzivom: nesluten razvoj nevroznanosti, računalniških nevrotehnologij in umetne inteligence omogoča dramatične, včasih izjemno hitre spremembe in razvoj novih, doslej neslutelih zmožnosti. Človekova usoda je iz spon genetike in evolucije v veliki meri prešla v njegove lastne roke.

Hiter vzpon nevroznanosti, nevrofarmakologije in nevrotehnologij odraža delno silovit razmah računalništva in elektronike v zadnji tretjini prejšnjega stoletja, delno pa zavestno spoznanje politikov, da bo nova veda o možganih v veliki meri določila razvoj človeka in družbe. Tako je predsednik Obama leta 2013 lansiral velik in drag projekt *U.S. BRAIN Initiative (Brain Research through Advancing Innovative Neurotechnologies)*, Evropska komisija pa istega leta *Human Brain Project*. Oba sta vzpodbudila raziskave možganov in uvedbo novih nevrotehnologij, ki so sicer v veliki meri izhajale iz potreb medicine. Danes lahko bolj zanesljivo in bolj zgodaj odkrijemo pravo diagnozo ter identificiramo biološke označevalce (biomarkerje), s katerimi lahko zasledimo zasevke bolezni, še preden bolnik ali zdravnik opazita prve klinične znake ali simptome. Odpirajo se neslutene možnosti za zdravljenje nevroloških in psihiatričnih bolezni. Z drobnimi elektrodami v možganih ali na površini lobanje lahko merimo možgansko aktivnost, jo spreminjamo in preko te aktivnosti upravljamo s telesom in okoljem. Sodobne slikovne metode omogočajo študij podrobne strukture možganov in zarisovanje možganskega zemljevida za posamezne funkcije.

Obenem pa nevrotehnologije vse bolj široko odpirajo vrata tudi v razumevanje bolj temeljnih mehanizmov delovanja možganov in vedenjskih značilnosti. Raziskave na tem področju se s področja medicine hitro širijo na druga področja: vojsko, ekonomijo, načrtovanje okolja in pametnih mest. Nevroznanstveniki odkrivajo možne odgovore na vprašanja, o katerih je znanost tisočletja lahko le ugibala, saj so možgani zaradi svoje nepristopnosti ostajali najmanj raziskan organ. V zadnjih desetih letih smo v veliki meri pričeli razumeti, kaj je spomin, kaj pozornost, kako se generira hoteni ali avtomatični gib, kaj je fiziološka in anatomska osnova čustev, raziskovali smo možne mehanizme svobodne volje ali njeno morebitno odsotnost. In s temi spoznanji je prišla možnost, da popravimo tudi nekatere manjše, še fiziološke pomanjkljivosti telesa in uma in da okrepimo nekatere sicer normalne funkcije (npr. koncentracijo ali učenje). Z izrazom »kognitivno

žlahtnjenje« ali »kognitivno krepitev« opisujemo katerokoli od številnih farmakoloških ali instrumentalnih metod, ki izostri, okrepi, pospeši katero od možganskih funkcij.

Nove možnosti (podobno kot ob rojstvu pisave ali interneta) seveda spreminjajo in preoblikujejo naše možganske funkcije, naš um, a tudi interakcijo z drugimi ljudmi in s fizičnim okoljem. Povzročajo tektonske spremembe v družbi 21. stoletja.

Med boleznijo, žlahtnjenjem in »novim normalnim«: farmakologija

Že od samih začetkov nevroznanosti smo skušali znanje o možganih uporabiti predvsem kot pomoč ljudem v bolezni. V minulih stoletjih so bile v ospredju farmakološke intervencije, v zadnjih desetletjih neinvazivne in invazivne stimulacije možganov, v zadnjih letih pa vzpostavljanje vmesnikov med možgani in računalniki.

V 19. stoletju je bila razširjena uporaba kokaina, ki so ga v Prusiji načrtno dajali svojim vojakom, v 20. stoletju uporaba kofeina, nikotina, metilfenidata, propranolola, modafinila, v zadnjem času pa potekajo raziskave z oksitocinom in nizkimi odmerki halucinogenov. K farmakološkemu žlahtnjenju možganskih funkcij se zateka veliko ljudi in že leta 2008 je neformalna anketa med bralci revije *Nature* odkrila, da jih 20% uporablja kognitivna krepitvena zdravila, in njihova uporaba vsako leto v razvitih deželah narašča (Maher, 2008).

Modafinil v medicini uporabljamo pri motnjah s pretirano ali nepričakovano zaspanostjo (narkolepsija, hipersomnije). Natančen mehanizem delovanja substance ni znan, delno pa ga pripisujemo zvišani koncentraciji dopamina. Zdravilo poglubi ter podaljša budno stanje, lahko tudi do 60 ali 80 ur. Pri sicer zdravih preiskovancih modafinil lahko izboljša izvršilne funkcije, kognitivno fleksibilnost, načrtovanje in v kombinaciji s kognitivnim treningom za približno 30% izboljšuje proces učenja (Gilleen in sodelavci, 2014).

Metilfenidat je psihostimulant, ki zvišuje raven dopamina in norepinefrina. Pogosto ga predpisujemo pri narkolepsiji in pri motnji z moteno pozornostjo in hiperaktivnostjo (sindrom ADHD).

Propranolol je beta blokator, zdravilo proti visokemu krvnemu pritisku, ki pa zmanjšuje tudi anksioznost, otopi afekt in slabi spomin. Zato ga v medicini uporabljajo za prekinitev začaranega kroga pogojenih fobij, kjer že sam spomin ali misel na stresen dogodek ali objekt (npr. strah pred letenjem, pajki) vzbudi enako burno reakcijo kot dogodek sam. To zdravilo selektivno oslabi konsolidacijo spomina povezanega s strahom, ne pa običajnega deklarativnega spomina (Kindt in sodelavci, 2009), in ga zato uporabljamo tudi za zdravljenje postravmatske stresne motnje. Mnogi sicer zdravi ljudje jemljejo propranolol za pomiritev in

ublažitev napetosti in treme. Simon Tedeschi je v zanimivem zapisu »priznal« na-
črtno uživanje tega »dopinga« pred koncerti in opisal zelo razširjeno uporabo
propranolola med izvajalci klasične glasbe.

Oxitocin je hormon, ki se izloča v možganih ob in po porodu, zlasti v obdobju do-
jenja in vzpostavljanja tesnega stika med otrokom in materjo. Povezan je z večjo
zaupljivostjo, prosocialnostjo in njegov učinek preskušajo v stanjih kot sta avti-
zem in socialna anksioznost. Raziskovalci so z aplikacijo oksitocina v preisko-
vancih vzbudili večje zaupanje do tujcev in dosegli, da so dobrodelnim organiza-
cijam donirali višje vsote (van Ijzendoorn in sodelavci, 2011).

Sicer pa tudi pri zdravlilih, ki jih ljudje ne uporabljajo za krepitev kognitivnih
funkcij, odkrivamo do zdaj manj opazne spremljevalne učinke, kot npr. blago
spremenjeno moralno kognicijo ali občutek za empatijo. Pri bolnikih, ki jemljejo
antidepresiv (selektivni zaviralec privzema serotonina), in tistih s Parkinsonovo
boleznijo, ki jemljejo prekursor dopamina (levodopo), se kaže po uvedbi terapije
motnja moralne kognicije: pri prvih se nekoliko zniža in pri drugih zviša stopnja
egoizma. Tako ugotovljen selektivni učinek serotonina in dopamina bo potencial-
no moč uporabiti pri socialnih disfunkcijah, ki so opisane kot ena od značilnosti
in tudi kot dejavnik tveganja pri mnogih psihiatričnih boleznih (Crockett in sode-
lavci, 2015).

Med boleznijo, žlahtnjenjem in »novim normalnim«: možganska stimulacija

Trenutno veliko moč novih nevrotehnologij poosebljajo učinki draženja možga-
nov, ki v zadnjem desetletju postaja rutinska metoda pri zdravljenju mnogih bole-
zni. Že stari Egipčani so z električnimi ribami zdravili epilepsijo in nekatere du-
ševne motnje, iz starega Rima pa je znana zgodba, ko je zdravnik cesarju Klavdi-
ju neznosen glavobol skušal olajšati tako, da mu je na čelo položil ohlajeno ribo.
Riba je Klavdiju čudežno pomagala; pa ne zaradi učinka hladu na čelu, pač pa za-
radi učinka električnega šoka ribe *Torpedo torpedo*.

V prejšnjem stoletju uvedena in kontroverzna elektrokonvulzivna terapija (EKT,
»elektrošok«) je danes sprejeto in učinkovito zdravljenje za hude in farmakološko
rezistentne oblike depresij. Mehanizem delovanja ni znan, pri tretjini bolnikov se
lahko pojavijo spominske motnje.

Obetajoče in manj kontroveržno je draženje z bolj sodobnimi neinvazivnimi
transkranialnimi (»čez oz. skozi lobanjo«) metodami indukcije električnih tokov v
možganovini, kot so transkranična magnetna stimulacija (TMS), transkranična
stimulacija z enosmernim električnim tokom (ang. *transcranial direct current
stimulation*, tDCS) in transkranična stimulacija z izmeničnim tokom (ang. *trans-
cranial alternating current stimulation*, tACDS). Mehanizem sicer še ni popol-
noma razjasnjen, verjetno pa s stimulacijo induciramo procese podobne dolgo-

trajni krepitvi (ang. *long-term potentiation*, LTP) ali dolgotrajnemu zaviranju (ang. *long-term depression*, LDP), ki potekajo v hipokampusu pri ponavljanem vzpostavljanju sinaptičnih povezav. Stimulacija spreminja tudi koncentracije nevrottransmiterjev, tako inhibitornih gabaergičnih kot ekscitatornih glutamatnih, in vzpodbuja nastanek nevrotropnih faktorjev, ki so ključni za nevroplastične spremembe pri učenju in spominu.

Transkranialna magnetna stimulacija

Pri transkranialni magnetni stimulaciji izrabljamo visokonapetostno magnetno polje, ki nastane v bakreni tuljavi, po kateri teče električni tok v kratkih močnih sunkih in preko lobanje v možganovini inducira tokove, ki so dovolj močni, da povzročijo depolarizacijo živčnih celic do globine 2 cm. Čeprav tudi pri TMS ne poznamo mehanizma, pa se v primerjavi z ECT ta metoda zdi varnejša. Obilno jo uporabljamo za raziskovanje delovanja možganov, saj na neinvaziven način aktivira nevrone pod tuljavo: aktivacija motoričnega predela bo npr. povzročila (ali pa zaustavila) gib, aktivacija vidnega predela vizualne občutke, aktivacija Brocovega jezikovnega področja pa lahko prekine govor. Raziskovalci so ugotovili, da TMS lahko za krajši čas spremeni moralno presojo (Young in sodelavci, 2010). Preiskovance so postavili v situacijo moralnega odločanja – npr., ali naj fant vzpodbudi dekle, da se napoti po nevarnem mostu, s katerega bi lahko padla. Kontrolni preiskovanci so menili, da je to moralno nesprejemljivo, ne glede na to, ali bo dekle padlo ali ne. Za preiskovance, katerih desni temporoparietalni predel (tesno povezan z moralno obarvanimi odločitvami) je bil stimuliran s TMS, pa je bil odločilen izid dogodka (ang. »*no harm, no foul*«). Po draženju desnega TPP so preiskovanci manj upoštevali akterjevo miselno stanje, zlasti v pogojih, ko je načrtovano moralno sprevrženo dejanje (npr. nekomu ponuditi zastrupljen »sladkor«), a je izhod nepričakovano ugoden (v sladkorju ni bilo strupa); moralna perspektiva se torej premakne od deontološke k bolj utilitaristični.

Transkranialna stimulacija z električnim tokom

Transkranialna stimulacija z enosmernim električnim tokom (tDCS) in transkranialna stimulacija z izmeničnim električnim tokom (tACS) pa uporabljata nizke napetosti in majhne tokove.

S tDCS možgansko skorjo dražimo s šibkim polariziranim električnim tokom, ki spreminja električno ravnotežje ionov membrane in mirovni membranski potencial živčne celice. V vrsti študij je tDCS pokazal zmožnost žlahtenja (Meinzer in sodelavci, 2014; Jacobson in sodelavci, 2012), a včasih tudi poslabšanja nekaterih kognitivnih funkcij. Tako je stimulacija posteriornega parietalnega korteksa facilitirala numerično učenje, a okvarila hitrost in zvečala napor (avtomatizem) pri učenju; nasprotno pa je stimulacija dorzolateralnega prefrontalnega korteksa poslabšala učenje, a izboljšala avtomatizem. Taka dvojna disociacija kaže, da tDCS lahko eno kognitivno funkcijo okrepi na račun drugih (Iuculano in Kadosh,

2013). Tehnologija električnega transkranialnega draženja možganov postaja široko dostopna; v namene večje uspešnosti pri video igrar jo je mogoče naročiti tudi po internetu (npr. <http://foc.us/me/#/shop>). Ker pa je zelo poceni, se – z nakupom 9-voltne baterije in stimulacijskih elektrod – uporaba hitro širi v domove tehničnih in nevro entuziastov in laikov. Uporabniki v blogih poročajo, da so uspešnejši pri vrsti dejavnosti, npr. pri sudokuju, raziskovalci pa glede mogočih bolj dolgotrajnih posledicah tovrstnega draženja možganov nimajo niti podatkov niti prepričljivih napovedi.

Invazivne stimulacijske tehnike

Ob neinvazivnih pa so že desetletja v medicini prisotne tudi invazivne stimulacijske tehnike, ki zahtevajo operativno vstavev elektrod na površino ali v globino možganov.

S stimulatorji, vgrajenimi globoko v možgane (globoka možganska stimulacija, GMS), zdravimo Parkinsonovo bolezen, Tourettovo bolezen, tremorje, distonijo, epilepsijo; indikacije za GMS se hitro širijo. Z umetnim kohlearnim implantatom pošiljamo impulze v slušno skorjo gluhih, slepim ljudem pa se vgrajujejo prvi elektronski čipi, ki vzbujajo aktivnost v vidnih poteh živčevja in omogočajo vid. Znanstveniki z univerze Berkeley so poročali o prvih poskusih z izjemno majhnimi milimetrskimi senzorji v velikosti zrnca peska (nevalni prah), ki so jih vgradili v mišice in periferne živce podgan in ki omogočajo brezžično ultrazvočno polnjenje in nadziranje mišične in živčne aktivnosti (Seo in sodelavci, 2016).

Med boleznijo in opazovanjem sanj: slikovne metode

Možgani so ostajali skrivnostni organ tudi zato, ker tehnologija dolgo ni omogočala slikanja tega organa. Šele ob koncu 20. stoletja so se pojavile tehnike, s katerimi lahko dobro prikažemo zgradbo (CT, MRT) ali delovanje (PET) možganov oz. oboje (kombinacija MRT in PET, funkcijska MRT). Tudi tu smo hitro doživeli pomik od pretežno medicinske diagnostične uporabe k poglobljenim raziskavam patofizioloških mehanizmov in k načinom, kako prikazati slikovne korelate možganskih aktivnosti, kot so spomin, pozornost, čitanje, razumevanje in eksekucija govora.

Danes je že mogoče s (sicer še relativno grobo) vizualno rekonstrukcijo ugotoviti, o čem človek razmišlja. V raziskavah so preiskovancem predvajali filme in slikali vzorce njihove možganske aktivnosti. S posebnim algoritmom so nato vzorce teh aktivnosti rekonstruirali v vidno podobo in na ta način odkrili, kaj je preiskovanec gledal (Nishimoto in sodelavci, 2011; Huth in sodelavci, 2016). Japonski znanstveniki so s 60% natančnostjo rekonstruirali kategorije sanjskih podob, ki so jih preiskovanci videli, ko so vstopili v zgodnje stadije spanja (Horikawa in sodelavci, 2013). Wager in sodelavci (2013) so s pomočjo fMRT objektivno prikazali

prisotnost in stopnjo fizične bolečine v *a priori* lokaliziranem bolečinskem omrežju, ki so ga poimenovali NPS (ang. *neurologic pain signature*). V drugi študiji (Woo in sodelavci, 2015) so preiskovance zaprosili, naj poskušajo na bolečino vplivati z mislijo. Preiskovanci so res poročali manj izrazito bolečino, raziskovalci pa opazili aktivacijo povezav med amigdalo in ventromedialnim prefrontalnim korteksom. To je skladno s hipotezo, da na zaznavanje in poročanje o bolečini vplivata dva možganska sistema: nociceptivno/afektivni v NPS in evaluativno-funkcionalni v amigdalo-prefrontalnih povezavah.

Slikovne metode so lahko v veliko pomoč pri bolnikih, ki zaradi kapi ali poškodbe v strateškem delu možganskega debla ne morejo komunicirati z okoljem, in ob sicer normalnih ciklih spanja dajejo vtis popolnega in perzistentnega vegetiranja, pa vendar imajo višje miselne funkcije ohranjene (sindrom vklejenosti, ang. *locked-in syndrome*); takemu bolniku bomo naročili, naj si npr. zamisli obraz svojega zakonca in to sliko bo moč iz možganske aktivnosti rekonstruirati.

Med boleznijo, opazovanjem misli in »novim normalnim«: jutri

Že danes so razvoj in zmožnosti nevrotehnologij osupljivi, čeprav se učinkov ne kot posamezniki ne kot družba ne zavedamo povsem. V naslednjih letih lahko pričakujemo nov kognitivni veliki pok, ki ga lahko zaslutimo in napovemo.

Z razvojem slikovnih tehnik bomo lahko spremljali in tudi shranjevali misli, sanje, predstave, ideje in abstraktne koncepte in spomine, svoje in drugih, spremljali resničnost izjav obtoženih in prič. Vmesniki ne bodo povezovali zgolj možganov posameznika in računalnika, ampak bo mogoča sinhronizirana aktivnost med možgani enega in možgani drugega človeka; še več, mogoče bo vzpostaviti hierarhično urejene paralelne povezave in sinhrono sodelovanje med številnimi možgani in številnimi računalniki. Rezultat bo izjemno izboljšana učinkovitost pri zdravljenju nevroloških bolezni, nesluteno povečanje inteligence in sposobnosti človeškega uma. Znali bomo vzpodbujati usmerjeno nevrogenezo in krepiti ali pa na novo vzpostavljali določene možganske poti ali centre. Analize funkcijskih slik bodo individualizirane in tako izpopolnjene, da bomo lahko napovedovali bodoč razvoj bolezni in sposobnosti, kot so inteligenca, pozornost, akademske zmožnosti. Ameriška raziskovalna agencija (DARPA) dela na projektu »telepatske čelade«, ki bi omogočila tehnološko osnovo za telepatsko komuniciranje med vojaki. Metode stimulacije možganov, kot je TMS, bodo zmožne draženja posameznih možganskih celic. Vmesniki med možgani in računalniki bodo postali kompleksni in dizajnerski, prilagojeni individualnim željam in usklajeni z drugimi tehnologijami, npr. tehnologijo novih virtualnih resničnosti. Široko bodo razširjene nove metode, kot je npr. optogenetika, ki uporablja na svetlobo občutljive beljakovine. Zgolj s prižigom svetlobe bo moč aktivirati ali zmanjšati določena vedenja, npr. agresivnost. Nevralni prah, ki so ga pred kratkim opisali raziskovalci z

Berkeleyja (Seo in sodelavci, 2016), bo izjemno zmanjšan, v obliki mikro ali nano prahu ga bo moč »zasejati« po širnih predelih možganov, nato pa z ustrezno metodo (morda optogenetiko) brezžično aktivirati, ugašati, povezovati znotraj lastnih možganov, z nevrlnim prahom v možganih drugih ljudi, računalnikov in svetovnega spleta.

Med boleznijo, opazovanjem misli in »novim normalnim«: etika, pravo in družba

Razvoj novih nevrotehnologij je nezadržen in izjemno hiter. So vse bolj razširjene in dosegljive, kmalu pa bodo nekatere tudi poceni in umeščene v naš pametni telefon ali, ustrezno pomanjšane, v naše telo. Mnoge bolezni bomo odkrili leta pred prvimi znaki in mnoge bomo zelo učinkovito zdravili. Imeli bomo dostop do širokega znanja, zmožnost globoke in globalne inteligence, mislili bomo hitreje, spretno bomo uporabljali razne vrste pozornosti in spominov, ko in če jih bomo potrebovali. Nadzor naših možganskih in telesnih aktivnosti ter takojšnje povratne informacije nam bodo omogočili optimalne vedenjske vzorce.

A vprašajmo se: je taka tehnologija res v službi človeka? To je možno, a ne samo po sebi razumljivo. Druga možnost je, da je nevrotehnologija reinkarnacija mogočnega demiurga, ki nas usmerja v navidezno negotovo, neznano nam bodočnost, z nepredvidljivimi tveganji, morda proti razvojnim stopnjam, ki presegajo antropocentričnost in se razvijajo v smeri nove evolucijske stopnje, ki presega biološko in kulturno in posega proti transhumani viziji povezanega organskega in neorganskega.

Zato moramo pazljivo prisluhniti vprašanju, ki se porajajo z novo nevrotehnologijo. Bo le-ta vodila v neenakost in socialno-ekonomski prepad med »nevrokognitivno okrepljenimi« in »zgolj zdravimi«? Bo kognitivna krepitev postala sčasoma dolžnost, pričakovanje (kot sta danes cepljenje in fluoriranje vode) in ne svobodna izbira? Bodo lahko država ali delodajalci nadzirali in modificirali naše misli in občutja? Je z elektronskimi čipi dopolnjen jaz še vedno moj izvorni jaz? Lahko pri vojaku uporabimo možgansko stimulacijo, da ne bo o vojni razmišljal v kontekstu morale? Lahko uporabimo pršilo oksitocina, da bo zločinec kriminalistu bolj zaupal in mu razkril zločin? Se lahko od kirurga pričakuje (ali celo zakonsko zahteva?), da se bo pred operacijo javil na električno stimulacijo možganov ali da bo zaužil modafinil, kar mu bo izostrilo pozornost, izboljšalo motoriko in ugodno vplivalo na uspeh operacije (že danes v primeru utrujenosti oblasti v Queenslandu v priporočilih zdravnikom omenjajo 5 – 6 skodelic kave na noč)? Se z jemanjem zdravila, ki otopi afekt in jemlje spominu čustveno ostrino (kot npr. propranolol) v temeljih spreminja naša identiteta in občutek jaza? Je bolj odgovorno do našega telesa, uma, vloge v družbi, da »igramo z dodeljenimi nam kartami«, taki, kakršne nas je ustvarila narava, ali da skušamo telo in um čim bolj izboljšati in okrepiti? Ali naj bomo do jemanja poživila pri študentu ali propranololu

la pri pianistu enako kritični in netolerantni kot do dopinga pri športu? Je meja (ne)sprejemljivega hitrost in intenzivnost sprememb – naj spremenjeno, okrepljeno, žlahtnjeno sposobnost dovolimo, če tempo in intenzivnost sprememb dovolj jeta adaptacijo in integracijo mojega prejšnjega jaza, sicer pa ne?

Trenutno odnos do krepitev kognitivnih sposobnosti sega od fanatičnega entuziazma, preko previdnega optimizma, do skepse in prestrašenega odpora. Neracionalni strah in odpor bo počasi kopnel, zlasti, če bo znanost dokazala, katere metode so varne in učinkovite, politika pa poskrbela, da bodo dostopne vsem. V biologiji človeka je namreč nezadržno stremljenje k več sreče in k večji ustvarjalnosti. A dokler tudi etika in pravo ne bosta rešila žgočih vprašanj, povsem izginil ta strah ne bo nikoli. Skrajni čas je, da se raziskovanje novih tehnologij iz laboratorijev preseli tudi v akademske prostore etikov in pravnikov. Etika v dobi že prisotnih in porajajočih se nevrotehnologij namreč ne sme biti *post hoc* refleksija ali analiza. (Pro)aktivno mora sooblikovati dostikrat nejasno mejo med znanjem možnega in modro uporabo le-tega, morda v duhu *recta ratio speculabilium*, *recta ratio agibilium*.

Brain, Neurotechnology and Ethics

In the article we describe the genetic-evolutionary and civilizational mechanisms that constantly, albeit on different time scales, shape the structure and functions of the brain. In the last millennia neuroplasticity has become an ever more important mechanism for faster adaptations, particularly with the emergence of great technological and cultural breakthroughs (invention of writing, the Internet) leading to the reorganization of brain structures and to a new mind. The introduction of new neurotechnologies which enable us to dramatically and rapidly change our brain functions in health and disease and to create an interface between the brain and a computer challenges humans with big questions on the essence of being human and on further development and destiny. This is precisely why there is an urgent need for a new ethics, which must (pro)actively help in delineating the often-blurred and fuzzy border between the knowledge of what is possible and a wise use of such knowledge, perhaps in the spirit of *recta ratio speculabilium*, *recta ratio agibilium*.

Keywords: brain, neuroplasticity, neurotechnology, the Internet, ethics

Literatura

- Crockett, M. J., Siegel, J. Z., Kurth-Nelson, Z., Ousdal, O. T., Story, G., Frieband, C., Grosse-Rueskamp, J. M., Dayan, P. in Dolan, R. J. (2015). »Dissociable Effects of Serotonin and Dopamine on the Valuation of Harm in Moral Decision Making«. *Current Biology*, 25(14), str. 1852–1859.
- Gilleen, J. P., Michalopoulou, P. G., Reichenberg, A., Drake, R., Wykes, T., Lewis, S. W. in Kapur, S. (2014). »Modafinil combined with cognitive training is associated with improved learning in healthy volunteers--a randomised controlled trial«. *European Neuropsychopharmacology*, 24(4), str. 529–539.
- Horikawa, T., Tamaki, M., Miyawaki, Y. in Kamitani, Y. (2013). »Neural decoding of visual imagery during sleep«. *Science*, 340(6132), str. 639–642.
- Huth, A. G., Lee, T., Nishimoto, S., Bilenko, N.Y., Vu, A.T. in Gallant, J.L. (2016). »Decoding the Semantic Content of Natural Movies from Human Brain Activity«. *Frontiers in Systems Neuroscience*, 10: 81.eCollection 2016.
- Ijzendoorn, M. H. van, Huffmeijer, R., Alink, L. R., Bakermans-Kranenburg, M. J. in Tops, M. (2011). »The impact of oxytocin administration on charitable donating is moderated by experiences of parental love-withdrawal«. *Frontiers in Psychology*, Oct 13; 2: 258.
- Iuculano, T. in Kadosh, R. C. (2013). »The Mental Cost of Cognitive Enhancement«. *Journal of Neuroscience*, 33(10), str. 4482–4486.
- Jacobson, L, Koslowsky, M. in Lavidor, M. (2012). »tDCS polarity effects in motor and cognitive domains: a meta-analytical review«. *Experimental brain research*, 216, str. 1–10.
- James, W. (1890). *The Principles of Psychology*. New York: Henry Holt and Company.
- Kandel, E. R. (2006). *In search of memory: the emergence of a new science of mind*. New York: W. W. Norton & Company.
- Kindt, M., Soeter, M. in Vervliet, B. (2009). »Beyond extinction: erasing human fear responses and preventing the return of fear«. *Nature Neuroscience*, 12, str. 256–258.
- Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J. in Frith, C. D. (2000). »Navigation-related structural change in the hippocampi of taxi drivers«. *Proceedings of the National Academy of Sciences of the United States of America*, 97(8), str. 4398–4403.

- Maher, B. (2008). »Poll results: Look who's doping«. *Nature*, 452, str. 674–675.
- Meinzer, M., Jähnigen, S., Copland, D. A., Darkow, R., Grittner, U., Avirame, K., Rodriguez, A. D., Lindenberg, R. in Flöel, A. (2014). »Transcranial direct current stimulation over multiple days improves learning and maintenance of a novel vocabulary«. *Cortex*, 50, str. 137–147.
- Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B. in Gallant, J. L. (2011). »Reconstructing visual experiences from brain activity evoked by natural movies«. *Current Biology*, 21(19), str. 1641–1646.
- Ostrosky-Solis, F., García, M. A. in Pérez, M. (2004). »Can learning to read and write change the brain organization? An electrophysiological study«. *International Journal of Psychology*, 39(1), str. 27–35.
- Pascual-Leone A., Nguyet D., Cohen L.G., Brasil-Neto J. P., Cammarota, A. in Hallett, M. (1995). »Modulation of muscle responses evoked by transcranial magnetic stimulation during the acquisition of new fine motor skills«. *Journal of Neurophysiology*, 74(3), str. 1037–1045.
- Paul, R. L., Goodman, H. in Merzenich, M. (1972). »Alterations in mechanoreceptor input to Brodmann's areas 1 and 3 of the postcentral hand area of Macaca mulatta after nerve section and regeneration«. *Brain Research*, 39(1), str. 1–19.
- Paulesu, E., Démonet, J. F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., Cappa, S. F., Cossu, G., Habib, M., Frith, C. D. in Frith, U. (2001). »Dyslexia: cultural diversity and biological unity«. *Science*, 291(5511), str. 2165–2167.
- Seo, D., Neely, R. M., Shen, K., Singhal, U., Alon, E., Rabaey, J. M., Carmena, J. M. in Maharbiz, M. M. (2016). »Wireless Recording in the Peripheral Nervous System with Ultrasonic Neural Dust«. *Neuron*, 91(3), str. 529–539.
- Sio, F. S. de, Robichaud, P. in Vincent, N. A. (2014). »Who Should Enhance?: Conceptual and Normative Dimensions of Cognitive Enhancement«. *Humana.Mente – Journal of Philosophical Studies*, 26, str. 179–197.
- Small, G. W., Moody, T. D., Siddarth, P. in Bookheimer, S. Y. (2009). »Your brain on Google: patterns of cerebral activation during internet searching«. *The American Journal of Geriatric Psychiatry*, 17(2), str. 116–126.
- Speer, N. K., Reynolds, J. R., Swallow, K. M., in Jeffrey, M. in Zacks, J. M. (2009). »Reading Stories Activates Neural Representations of Visual and Motor Experiences«. *Psychological Science*, 20(8), str. 989–999.

Spitzer, M. (2012). *Digitale demenz*. München: Droemer.

Young, L., Camprodon, J., Hauser, M., Pascual-Leone, A. in Saxe, R. (2010). »Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgment«. *Proceedings of the National Academy of Sciences of the United States of America*, 107, str. 6753–6758.

Wager, T. D., Atlas, L. Y., Lindquist, M. A., Roy, M., Woo, C. W. in Kross, E. (2013). »An fMRI-Based Neurologic Signature of Physical Pain«. *New England Journal of Medicine*, 368(15), str. 1388–1397.

Woo, C. W., Roy, M., Buhle, J. T. in Wager, T. D. (2015). »Distinct Brain Systems Mediate the Effects of Nociceptive Input and Self-Regulation on Pain«. *PLoS Biology*, 13(1):e1002036.

Andrej Ule

Filozofska fakulteta Univerze v Ljubljani

Wittgensteinska kritika zamisli o mislečih strojih

V sestavku analiziram nekaj Wittgensteinovih kritik domnev, ki so ključne za velik del kognitivne znanosti. Menim, da teoretiki kognitivne znanosti še niso resno vzeli Wittgensteinovih kritik in začuda pogosto zamenjujejo vprašanje »kako kaj deluje« z vprašanjem »kaj to počne«. Predvsem pa je usodno zamenjavanje notranjih računskih (ali vzporednih) procesov v možganih, ki morda povzročajo mentalna stanja, z zunanjimi, socialno utemeljenimi vsakodnevnimi kriteriji prepoznavanja, klasifikacije in poznavanja vsebine mentalnih stanj.

Ključne besede: računski procesi v možganih, misleči stroj, Wittgenstein, Turing, sledenje pravilu

Znana je Wittgensteinova kritika ideje o strogo zasebnem jeziku, tj. jeziku, ki bi bil razumljiv in »dosegljiv« le posamezniku oz. le njegovi zavesti, možganom ipd. O mišljenju po Wittgensteinu namreč lahko govorimo le v kontekstu intersubjektivnih dejavnosti ljudi, ne zasebno, namreč v okviru sledenja pravil v različnih jezikovnih praksah in v medsebojnem prepletu raznih človeških dejavnosti. Teh dejavnosti ne moremo prevesti nazaj na procese v posameznikih ali na skupen njihovih posamičnih vedenj.

Wittgenstein je bil zato zelo skeptičen do zamisli o mišljenju (in podobno o zavesti) kot »notranjem procesu«, ki poteka v možganih posameznika. To zamisel je odklanjal prav tako močno kot metafizični psihofizični dualizem. Zgolj notranje-možganski proces mišljenja bi bil nekako sam sebi dovolj. To bi ga v končni fazi naredilo za prav tako »zasebnega«, kot so mentalna stanja in mentalni procesi domnevne nematerialne duševne substance. Dokler nimamo podane intersubjektivno preverljive (javno dostopne) in sistematične zveze med domnevnimi notranjimi stanji ali procesi v možganih in delovanjem posameznika v intersubjektivnih (socialnih) kontekstih, se izpostavljamu Wittgensteinovi kritiki zasebnega jezika, zasebnih stanj in procesov itd. Res lahko domnevamo, da takšna zveza ni načelno izključena in jo bomo morda kdaj lahko eksplicirali, vendar pa ostaja

mučen prelom med javno dostopno prakso delovanja ljudi v različnih situacijah (npr. jezikovnih igrach), ki določa pomen našim doživljanjem, željam, nameram, mislim, in notranjo, duševno ali »možgansko« naravo oz. strukturo teh doživljanj, želj itd.

Opozarjam, da ne gre le za konceptualni razkol med domnevno povsem računsko (ang. *computational*) določenim substratom mentalnih stanj in našim faktičnim doživljanjem teh stanj, temveč za razkol med tem, *kako delujejo* domnevni procesorji, ki tvorijo fiziološko-fizikalno osnovo mentalnih stanj, ter tem, *kaj* in *kako smo mi*, ko doživljamo ta stanja in jih izražamo v svoji družbeni praksi. K temu, kako smo, pa bistveno sodi, da ima naše doživljanje in delovanje svoj *pomen za nas*, da prav skozi svoje doživljanje in delovanje *občutimo svoj obstoj*. To ni le konceptualna razlika, temveč kategorialna razlika, za katero se nam niti ne sanja, kako bi jo lahko premostili.

Že v pojmu delovanja procesorjev tiči dvournost, namreč med opisom vzročnega procesa, tj. zaporedja stanj, procesorjev in pravilom delovanja procesorjev. Prvo je možno doseči s tem, da navedemo fizikalne zakone, ki regulirajo fizikalni potek dogajanja, in začetna stanja procesorjev v določenih časovnih trenutkih, drugo pa s tem, da navedemo npr. operacijsko tabelo prehodov stanj procesorja v odvisnosti od vhodnih podatkov in trenutnega stanja procesorja. Toda ta tabela je povsem idealizirana tvorba, ki kaže, kako *naj bi* procesor deloval v idealnih pogojih, ne pa, kako dejansko deluje. Je en od opisov *pravila* delovanja, ki naj bi ga implementiral procesor, ne opis procesa. Takšna pravila lahko včasih podamo s tabelo operacij, ki jih procesor lahko opravi ob ustreznih vhodnih podatkih in svojih notranjih stanjih. Kakorkoli že opredeljujemo pravila delovanja procesorja, moramo razlikovati med idealiziranim *operacijskim* delovanjem in dejanskim *potekom*, tj. dejanskim zaporedjem fizičnih stanj procesorja. Operacijsko delovanje procesorja je tisto, kar mu mi *pripišemo*, ko ga *tolmačimo* kot implementacijo določenega pravila ali tabelarično opredeljene operacije.

Vsaj v prvi, »računalniški« fazi kognitivne znanosti, so se avtorji pogosto sklicevali na model Turingovega stroja, ki naj bi predstavljal splošno formulacijo za računsko delovanje naravnih ali umetnih procesorjev. Vendar je ta model le idealizirana struktura delovanja sistemov, ki jim *pripisujemo* računsko delovanje, temu modelu pa ne ustreza noben realni sistem. Razlika ni le v tem, da ima Turingov stroj neskončen spominski trak, temveč tudi v tem, da se ta »stroj« lahko po nekaj korakih računanja vrne v natanko ista stanja, medtem ko se realni (končni) avtomati nikoli zares ne vrnejo v isto stanje, pač pa le v istovrstno stanje, ki sodi v enak ekvivalenčni razred stanj glede na določeno ekvivalenčno relacijo med stanji. Zato je vprašanje, ali pri njih sploh gre za računanje v smislu *operiranja s simboli*.

Cris Fields je v svojem prispevku »*Measurement and Computational Description*« (1996) v zborniku, posvečenem Turingu, izrecno poudaril, da moramo

razlikovati med vedenjem fizičnega sistema v času in (dinamičnim) opisom tega sistema s pomočjo algoritma, ki določa spreminjanje sistema v odvisnosti od (diskretnih) sprememb določenega parametra (ali parametrov), ki je opazljiv (merljiv) v času dogajanja. Tu gre za *naš* način opisa, ki omogoča preslikavo sprememb stanj sistema v korake algoritma. Pri tem moramo imeti na voljo določeno interpretacijo, ki preslika stanja sistema v zaporedje abstraktnih računskih stanj, pri čemer vsako stanje a_k sistema pridobimo z uporabo prvih k korakov algoritma ob določenem inputu (Fields, 1996: 170). Fields opozarja, da je takšna preslikava možna le, če vse operacije, ki so navedene v procesu preslikave, algebraično komutirajo (tj. kvantnomehansko gledano, ustrezne meritve ali opazovanja procesa so medsebojno neodvisna). Fields se nanaša na E. Dietricha, ki je opozoril, da lahko takšno interpretacijo sestavimo le tedaj, če sistem že vnaprej in neformalno jemljemo kot sistem, ki umešča (instancializira) funkcijo, ki bo izračunana z zadanim algoritmom, in če k korakov algoritma prepoznamo kot izračune poprej podanih podfunkcij (Dietrich, 1989).

Fields govori o takšni interpretaciji sistema kot o »virtualnem stroju« za dani algoritem. Takšna interpretacija pa je možna le tedaj, če lahko vedenje sistema od začetnega stanja (razumljenega kot vhodni podatek algoritma) vedno interpretiramo kot izvajanje danega algoritma pri danem vhodnem podatku. To je seveda le idealizirana predpostavka, oz. boljše predpostavka idealizacije, ne realni potek dogajanja v sistemu. Sistem torej »računa« (ang. *computes*) le v okviru naše interpretacije kot virtualni stroj, *ne po sebi*. Dinamičen opis sistema predstavlja le določeno relacijo, ki dovolj natančno podaja zaporedje (merljivih, opazljivih) fizičnih stanj sistema v zaporednih časovnih intervalih. To zaporedje pogosto lahko opišemo s kakšno lepo tekočo funkcijo, vendar to ni bistveno, to je že stvar interpretacije. A za računsko delovanje sistema bistveno potrebujemo takšno interpretacijo, ki zaporedna stanja sistema predstavi kot rezultate uporabe določenega algoritma (tj. sistema operacij) na ustreznih vstopnih stanjih sistema, ki predstavlja jo vhodne podatke virtualnemu stroju.

Fields dalje izvaja, da lahko vsak sistem, ki ga interpretiramo kot algoritemsko delujoči stroj, tudi *dejansko je tak* stroj, kajti že *simulirano računanje* je računanje. Torej takšna interpretacija dela tak sistem različnega od neinterpretiranega dinamičnega procesa, kot je npr. tok tekočine (Fields, 1996: 171). Drugi Fieldsov zaključek je »da lahko isti fizični sistem na podlagi različnih množic meritev interpretiramo kot različne virtualne stroje, namreč z interpretacijo vrednosti različnih množic spremenljivk kot pokazateljev stanja sistema« (prav tam).

Fieldsovo tezo, da je vsaka popolna simulacija računanja že tudi računanje, moramo brati tudi v obratni smeri, namreč da je vsako računanje svoja popolna simulacija, namreč takšna *interpretacija delovanja* sistema, ki ga predstavi kot popolno računanje (tj. vsako stanje sistema predstavi kot računski rezultat določenih vhodnih podatkov in notranjih stanj sistema). Potemtakem ni računanja po sebi,

tj. nobeno računanje ni potek dogodkov brez ustrezne (računske) interpretacije. To misel lahko povežemo z zanimivo Wittgensteinovo kritiko izenačevanja operacije in funkcije (predikacije), oz. operacij in relacij, ki jo je podal že v *Tractatusu*. Wittgenstein je namreč trdil, da so operacije sestavina našega simbolizma, oz. načina prikaza odnosov med formami stanj stvari, ne pa stanja stvari, dejstva ali dogodki v svetu. Operacije se lahko nanašajo na lastne rezultate kot bazo za operiranje, omogočajo matematična zaporedja predmetov, lahko izničujejo rezultate drugih operacij ali celo same sebe (Wittgenstein, 1976: 5.251, 5.252, 5.254). Predikacije tega načeloma ne zmorejo. Od tod med drugim izhaja tudi ugotovitev, da operacije niso nobena stanja stvari ali dejstva v svetu, temveč so *naš način prikaza* formalnih odnosov med različnimi opisi stanj stvari. V tem smislu tudi računanje, v kolikor so resnične operacije, podane s shemami Turingovega stroja, ne poteka »v naravi«, temveč le v *simbolnih sistemih* prikaza delovanja stroja. Za računanje je nujen operacijski (simbolni) opis dogajanja. Sledeč Wittgensteinu, mora računanje zajemati zaporedno (včasih tudi zaporedno in vzporedno) izvajanje operacij, ki jih lahko predstavimo le v določenem simbolnem sistemu, tj. ob uporabi ustreznih *formalnih izrazov*, spremenljivk, tabel delovanja ali algoritemskih postopkov. To pa ni isto kot opis »naravnega poteka« dogajanja v »stroju«, kajti ta opis ni nujno operacijski, lahko bi bil npr. le časovni diagram dogajanja.

V svoji kasnejši filozofiji je Wittgenstein opozarjal na podobno razliko med vzročnim delovanjem strojev in operacijskim delovanjem. Slednje po njegovem lahko poteka le v okviru predstave o stroju kot »svojem lastnem simbolu«. Te misli so zlasti pomembne za razpravo o možnosti mislečih strojev oz. za »strogi program« umetne inteligence. Po Wittgensteinu lahko stroj pojmujeemo idealizirano, tj. tako, da abstrahiramo od možnih okvar, nepopolnosti delovanja itd. V tem primeru se zdi, da »[s]troj ali sliko stroja uporabljamo kot simbol za določen način delovanja« (Wittgenstein, 2014: par. 193.).¹ To je povsem različno od napovedi dejanskega vedenja stroja. V slednjem primeru na splošno ne pozabimo na možnost napak, poškodb njegovih delov itd. Realni stroj se vedno lahko giblje povsem drugače kot napoveduje njegov simbolni dvojnik. Za simbolni dvojnik je možnost odpovedi *a priori*, tj. gramatično nemogoča. Wittgenstein zaključuje: »[P]remikanje stroja kot simbola je vnaprej določeno na drug način kakor premikanje danega dejanskega« (prav tam).

Toda tudi govor o idealnem ali idealiziranem stroju, ki naj bi se nahajal v stroju-simbolu, zavaja, npr. v govor, da »stroj ima (poseduje) svoje možnosti premikanja«. Wittgenstein ugotavlja, da ta idealna možnost gibanja ni gibanje, pa tudi ni podajanje fizičnih pogojev gibanja, saj se dejanski stroj lahko giblje povsem drugače, kot mu predpisuje njegova idealna podoba. Wittgenstein pravi, da se nam zdi, kot da je idealna možnost premikanja neke vrste »senca samega premikanja«.

¹ Podobno v Wittgenstein (1964: I., par. 125).

»Rečemo 'Izkustvo bo pokazalo, ali bo to dalo temu tečaju tako možnost premikanja', ne pa tudi 'Izkustvo bo pokazalo, ali je to možnost tega premikanja'« (prav tam: par. 194).²

V takšnem govorjenju napačno tolmačimo svoj način govora, ko zamenjujemo govor, ki opisuje operiranje v stroju-simbolu, z opisom realnega delovanja stroja. Wittgenstein pravi, da »[k]adar filozofiramo, smo kakor divji ljudje, ki slišijo način izražanja civiliziranih ljudi, ga napačno interpretirajo in iz te interpretacije nato izpeljejo najbolj nenavadne zaključke« (prav tam). Skratka v stroju-simbolu gre za poseben *izraz pravil* delovanja, oz. delovanja po pravilu, ne za dejansko delovanje. Podobno bi npr. »stroj za konjunkcijo«, podan z ustrezno tabelo prehodov stanj, predstavljal simbol za delovanje po pravilu, ne pa dejanski potek delovanja kakega »logičnega« stroja, npr. preprostega procesorja, ki implementira konjunkcijo.

Zopet smo torej pred načelno razliko med operacijami v nekem simbolnem sistemu (modelu) in med realnimi procesi v svetu. Še tako velika popolnost v realnem delovanju kakega stroja, npr. stroja, ki bi implementiral kako logično operacijo, ne pomeni, da stroj *izvaja* to operacijo, kajti še tako popolni stroj se lahko polomi, odpove, zastane ipd., medtem ko *izvajanje* te operacije načelno izključuje takšno možnost. Te možnosti so po Wittgensteinu kategorialno, gramatično izključene.

Vendar pa Wittgenstein ni dopuščal, da bi katerikoli izraz pravil oz. sledenja pravilom pomenil tudi dokončno ali izvirno *interpretacijo* ali *razlago* sledenja pravilu. V tem smislu tudi ne bi brezpogojno pristal na Fieldsovo formulacijo, da računska interpretacija delovanja sistema po pravilu računanja predstavlja že dejansko sledenje pravilu računanja in je torej simulacija računanja že računanje. Če namreč računska interpretacija pomeni le simbolni *izraz* preslikave fizičnih stanj sistema v algoritemska stanja računskega sistema, potem še ne gre za računanje. Šele *izvajanje* (ang. *executing*) te preslikave, torej dejansko izvajanje algoritma, je računanje. V tem smislu je računanje šele *dejanska* računske interpretacija, ne pa npr. le zamišljena interpretacija, ki zgolj formalno opiše algoritem računanja. Tabela preslikav ali tabela prehodov še ni operacija, pač pa je njen simbolni izraz. Če nekoliko pomešamo Wittgensteinovo terminologijo iz *Tractatusa* in v *Filozofskih raziskavah*, bi dejali, da se dejansko računanje le *kaže* v uporabi določenega algoritma, npr. v družbeni in javno dostopni *praksi* (Wittgenstein, 2014: par. 202), *tehniki* (par. 199) ali *običaju* (prav tam), v katerih moramo sami smiselno sodelovati, *opišemo* pa jo lahko s kakim formalnim simbolizmom.

Seveda se lahko vprašamo, kaj pa mi sami, ljudje, ali vsaj mi *dejansko operiramo* in podobno, vsaj mi *dejansko sledimo pravilu* in ne le zgolj interpretiramo svoje vedenje kot operiranje ali sledenje pravilom? Po Wittgensteinu oboje dejansko

² Enako v Wittgenstein (1964: I par. 125).

počnemo. To pa zato, ker je naše operiranje in sledenje pravilom *skupno uvežbana praksa*, ki jo pridobimo v delovanju, ki je v skladu z *življenjsko formo ljudi*. Kajti sledenje pravilom (in izvajanje operacij, ki je le posebna vrsta sledenja pravilom) je delovanje, ki poteka kot nekaj nedvornega in drugim ljudem razumljivega. To nedvornost moramo enostavno sprejeti kot neko osnovno dejstvo: »To, kar je treba sprejeti, kar je dano – bi lahko rekli –, so *življenjske forme*.« (prav tam II: par. 345). Človeška življenjska forma ni nekaj fiksnega, temveč je naš skupni način delovanja. Je preplet dejavnosti, ki opredeljuje naše pojmovanje in delovanje. O tem piše Wittgenstein v *Zvezkih (Zettel)* (1970):

Kako lahko opišemo človeški način delovanja? Pač, le tako, da prikažemo, kako vseprek mrgoli (ang. *swarms*) delovanje različnih ljudi. Ne, kar *sedaj nekdo misli*, posamezno dejanje (ang. *action*), temveč vse mrgolenje človeškega delovanja, je ozadje, na katerem opazujemo vsako dejanje, ki določa našo sodbo, naše pojme in reakcije. (Wittgenstein, 1970: par. 567)

Pravila oz. sledenja pravilom, prav tako operacije in izvajanja operacij so za Wittgensteina neločljiva sestavina človeške življenjske oblike oz. »mrgolenja« človeškega delovanja. Zato sledenja pravilom ne moremo abstrahirati in ga opazovati ločeno od tega ozadja, npr. tako, da bi se spraševali, ali lahko kak stroj ali organizem sam po sebi sledi pravilom.

Ker mišljenje predpostavlja sposobnost sledenja pravilom, se po Wittgensteinu tudi ni smiselno spraševati o tem, ali npr. stroj misli. Oz. točneje, takšno spraševanje bi predpostavljalo določeno podobnost človeku. O tem je tudi izrecno pisal v *Filozofskih raziskavah*:

Bi stroj lahko mislil? – Bi lahko imel bolečine? – No, ali bi o človeškem telesu lahko dejali, da je tak stroj? Temu, da bi bil tak stroj, je vendar kar sedita blizu. (Wittgenstein, 2014: par. 360).

Toda stroj vendar ne more misliti! – Je to izkustveni stavek? Ne. Le o človeku in tistem, kar mu je podobno, rečemo, da misli. To rečemo tudi o lutkah in celo o duhovih. Oglej si besedo 'misliti' kot orodje! (Prav tam: par. 361)

Wittgenstein torej vendarle ne zavrne povsem možnosti, da stroj misli, a ob predpostavki določene, četudi virtualne podobnosti s človekom. V čem bi bila takšna podobnost? Seveda, v tem, da si lahko zamišljamo delovanje stroja kot sestavino človeške življenjske oblike, tj. kot smiselni del mreže človeške dejavnosti, npr. mreže komuniciranja med ljudmi, sodelovanja ljudi v skupnih projektih, igranja itd. Lahko rečemo le to, če si lahko zamišljamo, da bi kak robot dobro posnemal delovanje ljudi v paradigmatičnih življenjskih situacijah ljudi in se enakopravno vključeval v komunikacijo med njimi, potem si ga lahko zamišljamo kot mislečega. Prav podobno si npr. zamišljamo vedenje »živih« lutk (npr. Ostržka) ali duhov. Toda od zamišljanja do uresničenja je dolga pot, morda nemogoča.

Wittgenstein pa z zadnjim stavkom zgoraj namigne, da je to podobno, kot da bi bilo mišljenje neko orodje. Ali je mišljenje lahko orodje? Morda nekako tako, kot so naše roke »orodje« za delo, ki se je razvilo v evoluciji človeka. Seveda lahko tudi tako gledamo na mišljenje, a to je le en vidik, ki pa ne zajame celote. Lahko gledamo na mišljenje *kot* orodje (evolucije, narave, boga itd.), toda prav to, da gledamo nanj kot na orodje, kaže, da mišljenje samo po sebi ni orodje. Na človeka npr. ne morem gledati, kot da je živ, ali kot da misli, temveč ga imam preprosto za živega, mislečega itd. Lahko pa gledam nanj, kot da je lutka, a to ni njegova narava. Podobno lahko na lutko ne gledam, kot da je simulacija človeka, temveč (za nas) *je* simulacija človeka. Lahko pa gledam nanjo kot živo, mislečo bitje, vendar pa to ni njena narava.

Wittgenstein se je med drugim kritično obračal tudi na Turingovo zamisel računalniških strojev, ki mislijo. V *Beležkah o filozofiji psihologije* je npr. l. 1947 zapisal:

Turingovi »stroji«. Ti stroji so vendar *ljudje*, ki računajo. To, kar pravi, bi lahko izrazili tudi v obliki *iger*. In sicer bi bile zanimive takšne igre, pri katerih bi po določenih pravilih prišli do nesmiselnih napotkov. Mislim na igre, podobne »tekmi«. Recimo, da dobite npr. ukaz »Nadaljuj enako dalje«, in to tedaj, ko je to povsem nesmiselno, npr. ko zaidete v krog; kajti ta ukaz ima smisel, vendar le na določenih mestih (Watson.) (Wittgenstein, 1984: par. 1096).

Wittgenstein sicer ni zapisal, kaj bi se zgodilo z ljudmi, ki bi se znašli v tovrstni igri in pred podobno nesmiselnimi napotki, verjetno pa je domneval, da bi ljudje prepoznali nesmisel takšnih napotkov in nekako »izstopili« iz igre.³ Wittgensteinova tiha domneva je, da v kolikor bi bili Turingovi stroji res zgolj računaljoči stroji, ne pa ljudje, ki računajo, potem ne bi prepoznali nesmisla napotkov in bi še dalje skušali »delovati« v skladu z njimi. Če pa bi »Turingovi stroji« ravnali podobno kot ljudje, tj. bi zaznali nesmisel napotka, potem bi bili dejansko »ljudje, ki računajo«, ne pa (kot je morda domneval Turing in za njim domnevajo zagovorniki stroge umetne inteligence), »stroji, ki se vedejo kot ljudje«. Zato pravi, da so »Turingovi stroji« pravzaprav ljudje, ki računajo.

Wittgenstein je nato dal še primer matematičnega postopka, ki po njegovem privede do podobnega brezpotja, namreč varianto Cantorjevega diagonalnega dokaza, da je množica naravnih (in podobno racionalnih) števil »šibkejša« kot množica realnih števil. Vsak poskus števne »urejanja« vseh števil, zapisanih kot neskončna decimalna števila (npr. vsa realna števila med 0 in 1) privede do paradoksa, ko pridemo do napotka, da zapišemo *n*-to decimalno *n*-tega števila, kajti tedaj bi mora biti *n*-ta decimalna obena enaka in različna od same sebe. To je po Wittgensteinu podobno kot napotek »Nadaljuj enako dalje«, torej do nesmiselnosti.

³ Wittgenstein ni navedel, na kateri Turingov tekst se je tu nanašal, domnevam, da gre za tedaj znan Turingov tekst o izračunljivih številih (Turing, 1937).

ga napotka in s tem do odkritja nesmiselnega pravila. Tudi tu ljudje lahko prepoznamo nesmiselnost, medtem ko stroj tega ne bi zmožel.

Posredno je s tem namignjeno, da Turingovi stroji ne bi prestali »Turingovega testa« (ta test je sicer Turing formuliral šele l. 1950 in ni znano, da bi Wittgenstein poznal njegov sestavek, ampak ideja za ta test je bila pri Turingu prisotna že precej pred tem), ker bi slej ko prej zašli v nesmiselne, neizvršljive korake pravil in se skušali še dalje »ravnati« po njih, medtem ko naj bi ljudje hitro prepoznali njihov nesmisel. Domnevam, da gre za tezo, da se ljudje sicer naučijo pravil s »slepim« (nedvomnim) sledenjem pravil, vendar pa to ne pomeni, da vseskozi slepo sledijo pravilu. Če se izkaže, da jih sledenje pravilu vodi v slepo pot, to opazijo in se ustavijo ali popravijo pravilo, medtem ko naj bi stroji tega ne zmoogli, ker pač zgolj »mehanično« sledijo zadanemu pravilu. To pa zato, ker poteka sledenje pravilom pri ljudeh v skupnostih, kjer se tudi naučijo smiselnega sledenja pravilom, smiselnega seveda glede na življenjsko formo ljudi.

Kot ugotavlja tudi Stuart Shanker v svojem sestavku o Wittgensteinu, Turingu in Churchu, je bila temeljna Turingova napaka (v njegovem spisu o izračunljivih številih iz 1930, na katerega se je Wittgenstein nanašal v svojih kritičnih opazkah o »Turingovem stroju«) v tem, da je nevede prešel od enostavnih podpravil, ki sestavljajo kompleksni aritmetični algoritem, k sklepu, da se takšna pravila lahko uresničujejo povsem mehanično. Vendar pa dejstvo, da ljudje lahko takšnim pravilom sledimo povsem slepo (brez razmišljanja, brez zavesti), nikakor ne dopušča sklepa, da so sama po sebi nekognitivna in da jim lahko *sledi* tudi kak stroj (Shanker, 1987: 641).

Wittgenstein bi seveda zavrnil sleherno sklicevanje na bojda povsem »notranja« (mentalna, možganska) stanja »razumevanja pravil«, ki naj bi omogočali in nadzorovali sledenje pravilom. Seveda bi Turingovi zagovorniki (ali zagovorniki sodobne UI) odgovorili, zakaj pa tega ne bi zmogli tudi kaki sodobni računalniki, še zlasti, ker je danes mogoče tudi računalnike učiti pravil v kaki »računalniški skupnosti« z »računalniško formo življenja«. Ne vem, kako bi Wittgenstein odgovoril na ta izziv, morda bi dejal: Če je res tako in bi res zgradili take računalniške sisteme, ja, potem bi pač imeli opravka z dejanskim sledenjem pravilom in bi morali takšne računalnike *obravnavati kot ljudi*, ne pa kot *stroje*. To bi verjetno vso razpravo postavilo na glavo, saj bi od tu dalje »normalni« zahodnjaki skušali dokazovati, kako stroji pač ne morejo biti ljudje, namreč zavestno misleča bitja. V to razpravo se na tem mestu ne bom spuščal, ker je podobno brezplodna kot razprava o tem, ali lahko stroji mislijo. Poudarjam pa, da Wittgenstein zavrača vsako argumentiranje v korist ljudi zato, ker imajo svoja »notranja mentalna stanja«, stroji pač ne, ker to vnaša v razpravo domnevo o »zasebnem jeziku«, ki jo je Wittgenstein na široko zavračal v *Filozofskih raziskavah* in drugje.

Ali Wittgensteinovo zavračanje mentalnih fenomenov kot notranjih stanj v posameznikovi duševni ali fiziološki strukturi pomeni, da se je odpovedal znanstveni psihologiji? Ali moramo psihološke pojave razlagati kot v celoti družbene ali vsaj medosebne, ne pa kot osebne? Res je, da je predstava o Wittgensteinu kot nasprotniku znanstvene psihologije dokaj običajna (gl. npr. Williams, 1999; Rey, 2003), vendar bi na podlagi Wittgensteinovih ostrih opazk o filozofiji enako ali še bolj upravičeno lahko govorili o njegovem nasprotovanju filozofiji itd. A takšni sklepi so po mojem mnenju preuranjeni. Prej bi lahko dejali, da se je Wittgenstein trudil otresti vseh psiholoških, filozofskih in tudi drugih teorij o človeški duševnosti, človekovem govoru, komunikaciji ipd., ker je skušal izpostaviti tisto, kar po njegovem ni nobena teorija več, temveč izraz »trdnih tal«, tj. nedvomnega obvladanja pravil in jezika. Ker so nekatere psihološke in filozofske teorije skušale zajeti tudi ta »tla« in jih razlagale z naravoslovnimi, psihološkimi in filozofskimi domnevami in teorijami, so s tem le dodatno oteževale »pregleden prikaz« (Wittgenstein, 2014: par. 122) in temu se je Wittgenstein silovito upiral. V *Beležkah o filozofiji psihologije* je Wittgenstein celo skiciral neke vrste načrt za analizo psiholoških pojmov, vendar kot vsakdanjih pojmov, ne pa pojmov, ki bi jih za svoj namen ustvarila znanost (1984: II, par. 63, 148). To pomeni, da pri gradnji psiholoških pojmov ne smemo pretrgati njihove zveze z vsakdanjim razumevanjem. To razlikuje psihologijo od nekaterih naravoslovnih ved, ki težijo k čim večji neodvisnosti od vsakdanjega razumevanja svojega predmetnega območja.

Ohranjanje sistematske zveze z vsakdanjim razumevanjem psiholoških pojmov in pojavov ne pomeni eliminacije znanstvene psihologije ali njenega zvajanja na behaviorizem. Kot meni Mary McGinn, Wittgensteinova kritika psiholoških pojmov ne pomeni, da psihološki pojmi terjajo vedenjske kriterije svoje uporabe, temveč »poskus pokazati, da zgolj z introspekcijo ne moremo izpeljati predstave o tem, kaj je dano psihično stanje« (McGinn, 1997: 130). Po McGinnovi argument zoper zasebni jezik ni v tem, da »morajo naši psihološki pojmi posedovati javne kriterije, temveč da le s spominjanjem na gramatiko naših običajnih psiholoških pojmov lahko dojamemo bistvo ali naravo psihološkega stanja dane vrste« (prav tam). Strinjam se s to interpretacijo.

Seveda je takšno pojmovanje psihologije in psiholoških pojmov v precejšnjem nasprotju z intencami dobršnega dela kognitivne znanosti, zlasti v nasprotju z radikalnimi zastopniki računskega pristopa.

Wittgensteinova kritika strojne metafore za razlago mišljenja jasno navaja načelno razliko med operacijami v simbolnem modelu stroja in njegovim možnim realnim delovanjem oz. nedelovanjem. Ne gre le za razliko med strojem, razumljnim na idealni način, tj. v abstrakciji od možnih odpovedi, in realnim delovanjem stroja, temveč za kategorialno razliko med sledenjem pravilom, ki jih vsebuje simbolni model, in časovnim zaporedjem njegovih realnih stanj. Če je torej računalniški algoritem res idealizirana razlaga dejanskega delovanja možganov, kot

menijo zagovorniki stroge računalniške koncepcije duha, potem dejansko operiranje ali dejansko sledenje pravilom ne poteka v možganih, temveč to počnemo *mi* kot ljudje, ki uporabljamo navedeni algoritem kot simbolni prikaz mišljenja. Zato je zelo nejasno, če govorimo o »mehanski realizaciji« algoritma v možganih, kajti takšna »realizacija« ni algoritem. Možgani »delujejo« algoritemsko le na podlagi naše (dejanske) interpretacije možganskega procesa kot računskega procesa.⁴ Zato sem kritičen do poskusov zblževanja Wittgensteinovega pojmovanja mišljenja in »računalniškim« pojmovanjem mišljenja (npr. pri Justinu Leiberju (1991) ali Georgu Rayu (2003)).⁵

Čeprav drži, da je Wittgenstein ostro zavračal dualistično pojmovanje odnosa med duševnostjo in telesom, domnevno predjezikovno, notranjo naravo duševnih stanj, introspekcijo kot psihološko metodo in da najdemo podobna stališča tudi pri zagovornikih računalniškega modela duha, pa to ne zadošča za trditev, da je bil Wittgenstein neke vrste »kognitivni naturalist« (kot meni Leiber) ali »funkcionalni mentalist« (kot meni Rey). Seveda se lahko računalnik »vede v skladu s pravilom«, pa vendar še ne *sledi pravilu*, kajti manjka mu ustrezna kompetenca. Ta se po Wittgensteinu razvije v skupnosti, kjer obstaja *utrjena praksa sledenja pravilu*, ki pozna določene običaje, vežbanje v uporabi pravila, razne tehnike sledenja pravilu ipd. Če bi hoteli npr. res dokazati obstoj sistematične podobnosti med (algoritemskim ali vzporednim) delovanjem računalnika in početjem ljudi, ko sledijo pravilom, bi morali razviti nekakšen nov »Turingov test«, ki bi pokazal, kako lahko takšno početje kar se da realistično predstavimo ali »realiziramo« kot računske procese, ki jih morda lahko simuliramo na računalniku. Podobno velja za »modernejš« poskuse modeliranja mišljenja ob pomoči nealgoritemskih, vzporednih informacijskih procesov v umetnih nevronskih mrežah. Tudi v tem primeru ne moremo trditi, da obstoj določenih funkcionalnih podobnosti med delovanjem možganov in delovanjem umetnih nevronskih mrež zadošča za utemeljitev teze, da je mišljenje (eden od) vzporedni(h) proces(ov) v biološki nevronske mreži, v človekovih možganih. Tudi tu bi morali poprej izvesti serijo testov, ki bi npr. pokazali, da obstaja sistematična podobnost med človeško prakso uvežbavanja pravil v skupnosti, kjer obstaja utrjena praksa sledenja pravilom, in procesom »učenja pravil« določenih nevronskih mrež. Vendar vsaj meni niso znani nobeni sistematični poskusi te vrste. Morda bi kdo dejal, da so dosedanji uspehi umetne inteligence, npr. v robotiki, kjer aparati umetne inteligence na vse več področjih uspešno zamenjujejo ljudi, vsaj posreden dokaz te vrste, namreč dokaz teze, da je človeška inteligenca »konec koncev« le nekoliko bolj kompleksna in bolj zmogljiva »računalniška« inteligenca.

⁴ Problematiko sledenja pravilom in Wittgensteinovo evolucijo pojma operacije obravnavam bolj obsežno v svoji knjigi *Operationen und Regeln bei Wittgenstein* (1997).

⁵ Obsežnejšo kritiko Leiberja in Reya sem podal v Ule (2005), kjer se opiram tudi na delo Diane Proudfoot in Jacka Copelandja »Turing, Wittgenstein and the Science of the Mind« (1994).

Teoretiki kognitivne znanosti in umetne inteligence po mojem mnenju še niso resno vzeli Wittgensteinovih kritik mišljenja kot »notranjega procesa« in začuda pogosto zamenjujejo vprašanja »kako kaj deluje« (ang. *how it works*) z vprašanjem »kaj počne« (ang. *what it does*), podobno zamenjujejo vzročno razlago mišljenja z operacijami v simbolnem modelu oz., splošneje, s sledenjem pravilom ter podobno zamenjujejo ravnanje, ki je v skladu s pravilom z dejanskim sledenjem pravilu. Predvsem pa je po moje usodno zamenjavanje notranjih računskih (ali vzporednih) procesov v možganih, ki morda povzročajo mentalna stanja, z zunanjimi, socialno utemeljenimi vsakodnevnimi kriteriji prepoznavanja, klasifikacije in poznavanja vsebine mentalnih stanj.

Wittgensteinian Criticism of Ideas on Thinking Machines

In this paper I analyze Wittgenstein's criticism of several assumptions that are crucial for a large part of cognitive science. In my opinion, theorists of cognitive science have not yet seriously considered Wittgenstein's criticisms and, quite surprisingly, frequently confuse the question "how does it work?" with "what does it do?" Their most "deleterious" mistake, however, is their confusion of the internal computational (or parallel) processes taking place in the brain (possibly causing mental states) with the external, socially-based, everyday criteria of recognition and classification and knowledge of the content of mental states.

Keywords: computational processes in the brain, thinking machine, Wittgenstein, Turing, rule-following

Literatura

Dietrich, E. (1989). »Semantics and the Computational Paradigm in Cognitive Psychology«. *Synthese*, 79, str. 119–141.

Fields, C. (1996). »Measurement and Computational Description«. V Millican, P. J. R. in Clark, A. (ur.), *Machines and Thought*, vol. 1, Oxford: Clarendon Press, str.165–177.

McGinn, M. (1997). *Wittgenstein and the Philosophical Investigations*. London in New York: Routledge.

Leiber, J. (1991). *An Invitation to Cognitive Science*. Oxford: Blackwell.

Proudfoot, D. in Copeland, B. J. (1994). »Turing, Wittgenstein and the Sciences of the Mind«. *Australasian Journal of Philosophy*, 72(4), str. 497–519.

Rey, G. (2003). »Why Wittgenstein Ought to Have Been a Computationalist?«. *Croatian Journal of Philosophy*, 3(9), str. 231–264.

- Shanker, S. G. (1987). »Wittgenstein versus Turing on the Nature of Church's Thesis«. *Notre Dame Journal of Formal Logic*, 28(4), str. 615–649.
- Turing, A. M. (1937). »On computable numbers, with an application to the Entscheidungsproblem«. *Proceedings of the London Mathematical Society*, 2(42), str. 230–265.
- Ule, A. (1977). *Operationen und Regeln bei Wittgenstein*. Frankfurt/M: Peter Lang.
- Ule, A. (2005). »How can one catch a thought-bird?: Some Wittgensteinian comments to computational modeling of mind«. *Synthesis Philosophica*, 20(2), str. 373–388.
- Williams, M. (1999). *Wittgenstein, Mind and Meaning: Towards a Social Conception on Mind*. London in New York: Routledge.
- Wittgenstein, L. (1964). *Remarks on the Foundations of Mathematics*. Cambridge, Mass.: The MIT Press.
- Wittgenstein, L. (1970). *Zettel*. V Wittgenstein, L., *Werke 5*, Frankfurt/M.: Suhrkamp.
- Wittgenstein, L. (1976). *Logično filozofski traktat*. Ljubljana: Mladinska knjiga.
- Wittgenstein, L. (1984). *Bemerkungen über die Philosophie der Psychologie, Bd. I. V Wittgenstein, L. Werkausgabe, Bd. 7*, Frankfurt/M: Suhrkamp.
- Wittgenstein, L. (2014). *Filozofske raziskave*. Ljubljana: Krtina.

Urh Vele

Filozofska fakulteta Univerze v Ljubljani

Forma v službi metode: kako brati Wittgensteinove *Filozofske raziskave*

V tem članku bom dokazoval, da so diametralno nasprotna razumevanja Wittgensteinove filozofske dejavnosti v *Filozofskih raziskavah* posledica specifične dialoške forme samega teksta. Iz te je namreč nemogoče povleči eksplicitna in nedvoumna stališča, ki bi jih lahko pripisali samemu avtorju in bi bila konsistentna s tekstom kot celoto. Toda to ni slabost teksta, ki bi jo veljalo pripisati Wittgensteinovim psihološkim lastnostim, temveč sestavni del njegove avtorske strategije. Če enkrat izhajamo iz te perspektive, potem se *Raziskave* pokažejo kot polifonično besedilo, ki prav skozi igro različnih glasov, ki govorijo z različnih perspektiv, obravnava različne filozofske probleme, pri čemer pa ne daje dokončnih odgovorov nanje. Polifonična dialoškost je tako lastna Wittgensteinovi metodi. Še več: neločljivo je povezana z Wittgensteinovim pojmovanjem – oziroma spreobračanjem – same filozofije.

Ključne besede: forma *Filozofskih raziskav*, enigmatičnost Wittgensteinovega stališča, Wittgensteinova metoda, paradoks metafizofije, polifoničnost

Uvod

Čeprav je bilo spisanih na tisoče in tisoče akademskih člankov in knjig o Wittgensteinovih *Filozofskih raziskavah*, interpreti niti približno ne soglašajo o tem, kaj je avtorjevo stališče v njih. Še več: nesoglasja med interpreti se ne nanašajo zgolj na razumevanje tega ali onega paragrafa, tega ali onega argumenta, temveč na opredelitev Wittgensteinove filozofske dejavnosti kot take.

Medtem ko nekateri razlagalci dokazujejo, da moramo na Wittgensteinovo pozno filozofijo, kakršno najdemo v *Raziskavah*, gledati kot na povsem konvencionalno filozofijo jezika, posredovano v nenavadni obliki (Wiggins, 1997), da Wittgenstein torej, čeprav iz takšnih ali drugačnih razlogov tega ne priznava, argumentira za določeno teorijo jezika, ki jo moramo izluščiti iz njegovih tekstov, na primer teorijo pomena kot besedne rabe (Kenny, 2006); teorijo jezikovnih iger in tezo o transcendentalnosti semantike (Hintikka, 1986); tezo o jeziku kot holističnem sistemu prakse v življenjski obliki (Stern, 1995); antirealizem oz. »čistokrvni je-

zikovni konvencionalizem« (Dummett, 1959); intersubjektivno teorijo pomena (Kripke, 1982); tezo o avtonomnosti in arbitrarnosti slovnice (Baker in Hacker, 2009) itd., ga drugi interpretirajo kot neteoretskega, postmetafizičnega filozofa, ki vendar zasleduje pozitivne cilje, s tem ko poskuša zakoličiti meje jezikovnega smisla, in sicer s topografijo slovnice našega jezika (Baker in Hacker, 2009). Spet tretji govorijo o Wittgensteinovem »kvietizmu«, po njihovem mnenju naj bi Wittgenstein sicer na nadvse izviren način zastavljal prava filozofska vprašanja, a nanje ni ponudil, vsaj ne eksplicitno, nobenih odgovorov, zato je na nas, da sestavimo njegovo filozofijo jezika (Wright, 2001; McDowell, 2000; Putnam, 1975); četrtri v Wittgensteinovi pozni filozofiji zaznavajo novo obliko filozofskega skepticizma (Fogelin, 1987; Kripke, 1982); peti poudarjajo Wittgensteinovo terapevtsko antifilozofsko usmeritev, zanje je pozni Wittgenstein predvsem potrpežljivi antifilozof (Stern, 2004; Pichler, 2004), ki nas želi s pomočjo stroge filozofske argumentacije, s pomočjo pojmovne terapije (Baker, 2006) odvrniti od tradicionalne filozofije; šesti, še radikalnejši, so mnenja, da je cilj njegove filozofije povsem in docela terapevtski: njegova domnevna filozofska prizadevanja so en sam velik *nonsens*, nekakšna satirična imitacija filozofske razprave, dobro zveneča lupina, ki je zavestno konstruirana tako, da deluje na prvi pogled koherentno, a se na koncu, ko na točki razkritja sprevidimo avtorjev postopek, zvijačno sesuje sama vase. Po mnenju radikalnih terapevtskih bralcev Wittgenstein tako pokaže – po principu tistega slavnega predzadnjega zenovskega gnom *Logično filozofskega traktata*: »Moji stavki ponujajo pojasnitev s tem, da jih bo tisti, ki me je razumel, na koncu spoznal za nesmiselne« (Wittgenstein, 2006: LFT 6.54)¹ – na iluzornost, absurdnost in nekoherentnost teoretske filozofije kot take (Diamond, 1991; Conant, 2007).

Nobenega dvoma ni, da so diametralno nasprotna stališča Wittgensteinovih interpretov posledica specifične forme *Raziskav*, v katerih ne najdemo klasične linearne argumentacije, ki bi iz skrbno izbranih premis izvajala enoznačne sklepe, temveč mukotrpne dialoge, ki ne privedejo do jasnih stališč, ki bi jih lahko pripisali samemu avtorju in bi bila konsistentna s tekstom kot celoto. Wittgenstein namreč v *Raziskavah* postopa tako, da postavi neko filozofsko tezo oz. trditev – neposredno, v oklepajih kot vprašanje, jo položi v usta sogovorniku itd. –, a se že takoj v naslednjem koraku distancira od nje bodisi tako, da jo začne reflektirati in nizati

¹ Wittgensteinova dela navajam po nemški izdaji njegovih zbranih del v osmih knjigah, ki je izšla pri založbi Suhrkamp, pri čemer pri *Logično filozofskem traktatu* (LFT) in *Filozofskih raziskavah* (FR) uporabljam Wittgensteinovo oštevilčenje in ne strani, na kateri se citat nahaja, razen ko navajam iz predgovora. Pri Wittgensteinovi *Zapuščini* (»Nachlass«) uporabljam standardno von Wrightovo oštevilčenje. Pri vseh ostalih delih navajam številko strani, na kateri se citat nahaja.

Kjer je le mogoče, navajam že obstoječe slovenske prevode, razen pri *Filozofskih raziskavah*, kjer so prevodi delo avtorja tega zapisa in služijo samo namenom tega članka. To ne pomeni, da nisem upošteval že obstoječih slovenskih prevodov Wittgensteinovih del – na tem mestu bi se še posebej zahvalil Borutu Cerkovniku, ki mi je posodil skripto svojega prevoda *Filozofskih raziskav* –, a sem te pogosto modificiral in preoblikoval.

možne ugovore in protiprimere zoper njo, bodisi z miselnimi eksperimenti, zamišljenimi jezikovnimi igrami, analogijami, izmišljenimi etnologijami, ki pokažejo njene meje, bodisi jo odpravi z ironičnimi aforizmi, retoričnimi vprašanji itd. Poleg tega proizvaja dialoškost na tak način, da različne dramske osebe, ki se zapletajo v razpravo in zastopajo različne, pogosto protislovne filozofske poglede, niso jasno opredeljene z interpunkcijo, njihovi glasovi so brezimni, zato je iz samega teksta pravzaprav nemogoče razbrati, koliko različnih glasov nastopa v *Raziskavah*, niti tega ne, kateri izmed teh glasov je tisti, ki podaja stališča njihovega avtorja, katere so sogovorniku položena v usta znotraj argumentacije *reductio ad absurdum* zato, da bi bila kasneje ovržena, katera misel je navržena kar tako, brez pravega razloga, katera aforistično, katera kot komentar, katera kot osnova za nadaljnji razmislek, katera je mišljena ironično itd.

V tem članku bom dokazoval, da specifična forma *Raziskav* ni slučajen odklon od konvencionalne forme filozofskih razprav, ki bi bil posledica psiholoških značilnosti njihovega avtorja, njegovih literarnih aspiracij oz. ki bi izviral v njegovi želji po ekscentričnosti ali enigmatičnosti, ampak izhaja iz neke notranje nuje filozofiranja samega, iz radikalnega povzetja vzgiba, ki je Wittgensteina pripeljal do tega, da je filozofijo utemeljil in prakticiral kot opisno dejavnost pojasnjevanja jezika, ki ni ne teoretska, ne sistemska in ne empirična. Trdim torej, da je forma *Raziskav* inherentna Wittgensteinovi metodi in neločljivo povezana z Wittgensteinovim pojmovanjem – oziroma spreobračanjem – filozofije.

Forma *Filozofskih raziskav*

Če je za Wittgensteinovo zgodnje delo, *Logično filozofski traktat*, značilen avtoritativen in formulaičen ton, je glavna poteza *Raziskav* brez dvoma njihova neprenehna interogativnost:

Raziskave so napisane kot nenehna dialektična razprava, kjer pomisleki, ki jih spet in spet izreka imaginarni sogovornik, nikoli dokončno ne potihnejo. Ker niso zgrajene kot deduktivna argumentacija z enoznačnimi tezami, ki stojijo na mestu sklepov, se vedno znova vračamo – z vidika različnih posebnih primerov in mnogih zornih kotov – k isti snovi, v upanju, da se nam bodo s pomočjo celotnega postopka problemi pokazali v pravi luči. (Kripke, 1982: 76)

Ker v njih ni vsevednega glasu avtorja/pripovedovalca, ki bi eksplicitno in nedvoumno argumentiral za določene filozofske teorije, kakor je ustaljena praksa akademskih filozofskih razprav, so *Raziskave* filozofsko besedilo, ki kar kliče po osmišljenju. Dokler namreč ne razjasnimo Wittgensteinovega početja, »neposvečenim« nikakor ni jasno, kaj je sploh smisel in namen teh dialogov, *Raziskave* se kažejo – govorim iz lastne izkušnje – kot nekakšen kaotični karneval filozofije, v katerem sicer v neopredeljenih glasovih, ki v *Raziskavah* izmenjujejo svoja stali-

šča, slutimo odmeve slavnih filozofov preteklosti – vsi so tukaj: od Platona do Nietzscheja, od Fregeja do Russella do »avtorja Logično filozofske razprave«, od Descartesa do Kierkegaarda, od Huma do Kanta, od Avgušтина do Schopenhauera itd. –; slutimo jih, čutimo jih, čeprav skoraj nikoli niso omenjena njihova imena in jih torej ne moremo dokončno identificirati.² Prav tako slutimo in čutimo, da se nekje med njimi – tako rekoč med vrsticami – mora skrivati tudi sam avtor *Raziskav*, vendar ga v kakofoničnem hrupu in zmedi *Filozofskih raziskav* ne zmoremo jasno in dokončno identificirati: je to glas, ki zatrjuje, da je filozofija neteoretska dejavnost, ki ne daje nobenih rezultatov v smislu tez?; ali je morda avtorjev glas tisti, ki jezik primerja z igrami in pravi, da je pomen besede njena uporaba v jeziku, da je jezik bolj praktična veščina kot diskurzivno znanje?; ali morda oni, ki, kakor nekakšen nadležni histerični sokratski obad, tako vztrajno, tako nepopustljivo, tako rigorozno niza ugovore in protiprimere in vprašanja zoper filozofske intuicije in s tem ohranja pogovor v teku, da včasih spominja na blazneža, ko ruši vse »zanimivo, veliko in pomembno« (Wittgenstein, 2006: FR 118); in kakšno vlogo ima tisti antropolog oz. etnolog, ki si zamišlja alternativna plemena in nas s tem opozarja, kako nenavadne, neutemeljene in arbitrarne so naše kulturne prakse s stališča tujih, drugačnih kultur, a da je vendar naša »življenjska oblika« tisto, kar nam je dano, kar moramo sprejeti in o čemer, vsaj načeloma, ni smiselno dvomiti; so morda vsi glasovi Wittgensteinovi, imamo morda opravka z nekakšnim avtorjevim notranjim filozofskim dialogom, z njegovo izpovedjo?; morebiti avtorskega glasu v tekstu sploh ni, morda je Wittgenstein samo koreograf te povorke, ki izza scene vleče niti te filozofske drame?

In prav zaradi tega, ker lahko Wittgensteinu na podlagi nejasnosti samega teksta utemeljeno pripišemo različna filozofska stališča, pravzaprav ne čudi, da je – kot se je izrazil nek interpret – »toliko zasebnih različic Wittgensteina, kot je njegovih bralcev in raziskovalcev«:

Privlačnost *Filozofskih raziskav*, dejstvo, da so filozofi in teoretiki raznoraznih usmeritev v njih našli podporo lastnim stališčem, lahko razumemo samo, če uvidimo, da je knjiga napisana tako, da kar kliče po tem, da vsak v njej najde svojega Wittgensteina. Ker je knjiga napisana v obliki dialoga, v katerem je nemogoče jasno opredeliti glasove in potegniti meje med njimi, mora vsak razlagalec sam iz njih izvleči, katera stališča Wittgenstein napada in katera zagovarja. (Stern, 2004: 7)

Na podlagi tega bi lahko sklepali, da so *Raziskave* »zgolj opazke«, se pravi le skupek aforizmov, gmota razcefranih filozofskih zapiskov, ki so sicer res – morda

² Samo 21 paragrafov *Filozofskih raziskav* (od 694-ih) vsebuje reference na druge filozofe, od tega jih le 11 vsebuje referenco na besedilo: trije vsebujejo kratke navedke, ne da bi Wittgenstein omenil vir (342, 413, 436), pet (23, 46, 48, 89, 518) se jih nanaša na filozofska dela v celoti (*Traktat*, Platonov *Teajtet* in Avguštinove *Izpovedi*), samo trije paragrafi pa neposredno navajajo točno določen odlomek iz nekega dela skupaj z referenco (1, 97, 114; *Izpovedi* in *Traktat*).

prav zaradi njihove kriptičnosti – neusahljiv vir filozofskega navdiha, a nič manj in nič več kot to. Ta perspektiva se mi zdi docela napačna. *Filozofske raziskave* so po mojem mnenju premišljeno, celovito, sistematično in metodično delo, avtorska stvaritev, ki jo je potrebno racionalno rekonstruirati v skladu z intencami avtorja. To nadalje preprosto pomeni, da je dialog točno tista oblika, v kateri je želel Wittgenstein posredovati svoja filozofska raziskovanja, še več, ne samo, da je specifična forma *Raziskav* zavestna odločitev avtorja, ta forma je tudi najboljši pokazatelj avtorjevega pojmovanja filozofije in njegove metode:

Wittgensteinovo mišljenje ni aforistično, ni v kratkih, nepovezanih pasajah ali delih teksta. Wittgenstein namreč razvija neverjetno stroga premišljevanja o posameznih primerih, nekakšno fenomenologijo posamičnih primerov, ki ne bi bila možna, če bi vpraševal splošneje. Zaradi tega ne smemo oblike, v kateri predstavlja svoje misli, samo sprejeti, temveč jo moramo motriti kot tisto tekstovno obliko, ki ustreza njegovemu mišljenju. Wittgenstein ni nikakršen literarni aforistik. Zunanja, aforistična oblika njegovih tekstov ne pomeni, da posamezni deli teksta niso med seboj globlje vsebinsko povezani. (Vossenkuhl, 2003: 58)

Temu v prid govori že samo dejstvo, da imajo *Filozofske raziskave* odlikovano mesto v opusu Wittgensteinove pozne filozofije. Medtem ko so ostala dela poznega Wittgensteina, ki so po njegovi smrti izšla pod njegovim imenom, bodisi spodleteli poskusi, da bi napisal klasično filozofsko knjigo, bodisi zgolj uredniški izbor opazk, ki jih je Wittgenstein o določeni temi zapisoval v svoje dnevnike (npr. o matematiki ali filozofiji duha), so *Raziskave* izšle, čeprav jih ni izdal Wittgenstein sam za časa svojega življenja, praktično brez uredniških intervencij – Henrik von Wright, Wittgensteinov učenec in eden izmed skrbnikov Wittgensteinove filozofske zapuščine, trdi, da so *Raziskave* edino delo poznega Wittgensteina, ki je celovito zaključeno in ki ga je Wittgenstein avtoriziral za izdajo: »Sam sem mnenja, da je prvi del *Raziskav* zaključeno filozofsko delo.« Ne smemo pozabiti, da se je Wittgenstein večino izmed dvaindvajsetih let – med letom 1929, ko se je začel ponovno ukvarjati s filozofijo, in letom 1951, ko je umrl – ukvarjal s tem, da bi napisal knjigo, ki jo danes poznamo kot *Filozofske raziskave*, zato lahko že iz tega sklepamo, da so prav *Raziskave* tisto delo, v katerem najdemo srž Wittgensteinove pozne filozofije. To potrjuje tudi sam, ko takoj na začetku predgovora zatrjuje, da so *Raziskave* izbor šestnajstletnega filozofskega raziskovanja, usedlina vsega, kar je v tem času počel, album, v katerem je ovekovečil svoje filozofsko raziskovanje in, kot pravi kasneje, »na konkretnih primerih demonstriral svojo metodo« (Wittgenstein, 2006: FR 133).

Prav tako je že na prvi pogled očitno, da so *Raziskave* edino Wittgensteinovo delo, ki je celostno dialoško. Wittgenstein je sicer praktično vse življenje v svoje dnevnike obsesivno zapisoval filozofske opazke – lahko rečemo, da je forma opazk značilna forma Wittgensteinovega filozofiranja –, ki imajo že same po sebi

običajno – s pomočjo interpunkcije – dialoško strukturo – Wittgenstein običajno postopa tako, da zapiše neko trditev (filozofsko tezo, filozofsko vprašanje, filozofski izziv), nato pa zoper njo niza ugovore, protiprimer, pomisleke, ironične komentarje, aforizme, miselne eksperimente, analogije ...; nato sledi nova trditev, in spet ugovori itd. – vendar *Raziskave* niso samo amalgam opazk z notranjo dialoško strukturo, temveč iz takšnih opazk skrbno sestavljena in strukturirana celota, vsak fragment, vsak glas v njih ima točno odmerjeno mesto znotraj celostne naracije knjige.

Morda te trditve najboljše osvetljuje primerjava z *Rjavo knjigo* (1934). Mnogi razlagalci trdijo, da so *Raziskave* pravzaprav le nadaljevanje *Rjave knjige* v drugačni obliki, vsebina obeh del pa da se ne razlikuje bistveno. Res je sicer, da lahko opazujemo precejšnjo podobnost med obema deloma na ravni same vsebine, a razlika v formi in metodi med obema deloma vpliva tudi na dojetje vsebine: *Rjava knjiga* napreduje sistematično, linearno, korak za korakom, po uniformni metodi, od ene jezikovne igre do druge, od enostavnih jezikovnih iger do kompleksnejših – morda lahko v *Rjavi knjigi* najjasneje vidimo tako imenovano metodo jezikovnih iger –, in ponuja s tem univerzalni metodološki postopek za razreševanje filozofskih problemov, medtem ko raziskovanje v *Raziskavah* poteka vseprek, z različnimi sredstvi, po drobcih, od ene teme do druge, nato do tretje, in spet nazaj do prve, od enega predmeta do drugega itd. Wittgenstein torej v *Raziskavah* ne cilja na zaključeno, hierarhično urejeno celoto kot v *Rjavi knjigi* – ali *Traktatu* –, na absolutno perspektivo, ampak svoje raziskovanje delinealizira in fragmentira z ozirom na konkretne filozofske probleme, in sicer na način, da vendarle nastane sistematično in celotno delo z dobro premišljeno naracijo, ki ponuja več različnih perspektiv na določene probleme. S tem po mojem mnenju nastane povsem novo in drugačno filozofsko delo od vseh prejšnjih Wittgensteinovih filozofskih poskusov, delo, v katerem je najboljše razvidna metoda Wittgensteinove filozofije.

Paradoks Wittgensteinove (meta)filozofije

Po ustaljenem razumevanju, ki izhaja prvenstveno iz Cavellove interpretacije, dialoškost *Raziskav* razumemo kot pogovor med »traktarnim«¹ sogovornikom oziroma učencem/pacientom, ki predlaga različne filozofske teze (»glas skušnjave«, kot ga je poimenoval Cavell (1962)), še zlasti reprezentacijske teorije jezika, kot je na primer avguštinska teorija jezika, kartezijske poglede na jezik občutkov itd., in »posttraktarnim«² filozofskim terapevtom, ki vodi celotno igro tako, da spodbija trditve sogovornika in s tem – posredno – argumentira za svoje alternativno postmetafizično, neteoretsko, nefundamentalistično, pragmatično-praktično pojmovanje jezika (Cavell: »glas korekcije«³). Glas terapevta identificiramo kot Wittgensteinov avtorski glas, medtem ko naj bi učenčev glas utelešal metafizično filozofsko izročilo, katerega glavni predstavnik je prav zgodnji Wittgenstein, avtor *Logično filozofskega traktata*. Po tem pojmovanju je dialog predvsem didak-

tično sredstvo, ki nas po posredni poti pripelje do pravilnega postmetafizičnega filozofskega razumevanja nekega pojava.

Toda takšno razumevanje *Filozofskih raziskav* je nadvse problematično, kajti če natančneje pogledamo vanje, vidimo, da je domnevni Wittgensteinov glas v protislovju s samim sabo. Po eni strani vehementno zatrjuje:

»Filozofski problemi nastanejo, ko jezik praznuje« (Wittgenstein, 2006: FR 38).

»In nikakršne teorije ne smemo postaviti. /.../ Izginiti morajo vse razlage, nadomestiti jih mora le opis.« (FR 109)

»Rezultati filozofije so odkritje nekega čistega nesmisla in bušk, ki jih je razum staknil pri zaletavanju v meje našega jezika.« (FR 119)

»Filozofija pač vse samo postavi pred nas, ničesar ne razlaga in ne izpeljuje. – Ker imamo vse kot na dlani, tudi ni kaj razlagati. Kajti tisto, kar je nemara skrito, nas ne zanima.« (FR 126)

»Zmešnjave, s katerimi se ukvarjamo, nastanejo takrat, ko je jezik tako rekoč v prostem teku, in ne, ko deluje.« (FR 132)

»Ne obstaja ena metoda filozofije, pač pa obstajajo metode, tako rekoč različne terapije.« (FR 133)

»Filozof obravnava vprašanje; kakor bolezen.« (FR 255)

»V filozofiji ne delamo nobenih zaključkov. /.../ Ugotavljamo le tisto, s čimer se vsi strinjajo.« (FR 599)

Po drugi strani pa izreka trditve – z glasom, ki zagotovo ni sogovornikov, saj so te trditve znotraj teksta postavljene tako, da mu oponirajo –, ki zvenijo kot filozofske teze oziroma sklepi:

»Pomen besede je njena raba v jeziku.« (Prav tam: FR 43)

»Razumeti stavek pomeni razumeti jezik. Razumeti jezik pomeni obvladati neko veččino.« (FR 199)

»Zato je 'sledenje pravilu' praksa.« (FR 202)

»Pravilno in napačno je, kar rečejo ljudje; in v jeziku so ljudje soglasni. To ni soglasje mnenj, temveč življenjske oblike.« (FR 242)

»Bistvo je izrečeno v slovnici.« (FR 371)

Več kot očitno je, da sta ta glasova, ki naj bi bila oba Wittgensteinova, eklatantno protislovna in da drug drugega izključujeta – če ima prav eden, ne more imeti prav drugi, in seveda obratno. Na eni strani imamo torej Wittgensteinovo »metafilozofijo«, ki prepoveduje vsakršno postavljanje filozofskih tez; na drugi strani pa

njegovo dejansko filozofsko prakso, za katero se zdi, da je v nasprotju z metafilozofijo, saj v njej Wittgenstein prav gotovo postavlja določene filozofske teze. Če vzamemo zares prvi izmed obeh glasov in sledimo Wittgensteinovi metafilozofiji, potem moramo zaključiti, da *Raziskave* ne ponujajo nikakršnih tez ali teorij, temveč so v prvi vrsti terapevtske, z metodo gramatičnega opisovanja delovanja jezika – »naša raziskava je zato gramatična« (Wittgenstein, 2006: FR 90) – nas želijo odvrniti od filozofije kot iskanja teorij. Neizbežen je sklep, da je »le terapevtsko branje pravično do samega teksta *Raziskav*, kakor tudi do Wittgensteina kot filozofa« (Hutchinson, 2007: 710). Toda v tem primeru moramo pojasniti, zakaj nastaja vtis, da Wittgenstein vendarle ne počne samo tega, kar razglša, pojasniti moramo, zakaj se zdi, da vendarle postavlja določene filozofske teze. Če pa se po drugi strani odločimo za Wittgensteinovo prakso in jo razumemo tako, da Wittgenstein z njo postavlja določene teze, potem se ne moremo izogniti sklepu, da je »Wittgensteinova filozofija teoretska, ne pa terapevtska« (Hintikka, 1986: 11). In potem smo primorani pojasniti, zakaj se »v njegovi filozofiji dogaja več kot dopušča njegova metafilozofija« (Kenny, 2006: xx), zakaj Wittgenstein trdi, da počne nekaj, česar v bistvu ne počne (se spreneveda? nas zavaja? zavaja sebe?). Znašli smo se torej pred naslednjim paradoksom: če je Wittgenstein filozof, ki argumentira filozofske teze, to nasprotuje njegovi metafilozofiji; če je po drugi strani terapevt, ki želi preko drame, ki jo spremljamo v *Raziskavah*, razrešiti filozofske probleme s tem, da pokaže, da je vsako iskanje filozofskih teorij obsojeno na propad, je to v nasprotju s tem, kar se zdi, da dejansko počne.

Mislim, da smo se s tem dotaknili osrednjega paradoksa Wittgensteinove (pozne) filozofije, in sicer nasprotja med tem, kar Wittgenstein pravi, da počne, in tem, kar se zdi, da počne. Obe alternativni sta tekstovno podprti, zato se zdi, da je ta paradoks nerazrešljiv. Anthony Kenny, eden najeminentnejših razlagalcev Wittgensteinove filozofije, je zato resignirano sklenil:

Čeprav sem se po svojih najboljših močeh trudil, da bi mi uspelo, mislim, da na koncu ni mogoče pomiriti Wittgensteinovega pojmovanja filozofije s celoto njegovih filozofskih dejavnosti v *Raziskavah*. Drugi bodo morda v tem uspešnejši. Tisti pa, ki mislimo, da do pomiritve ne more priti, smo na koncu prisiljeni, da se odločimo, ali bomo sprejeli njegovo teorijo, ali pa bomo sledili njegovi praksi. /.../ Sam se težko odločim med tema alternativama; včasih me privlači ena, včasih druga možnost. (Kenny, 2006: xix)

Kenny še dodaja, da je ta tenzija najverjetneje nerazrešljiva, ker je ni razrešil niti Wittgenstein sam.

Polifoničnost *Filozofskih raziskav*

S to opazko se je Kenny dotaknil bistva problema interpretacije Wittgensteinove filozofije, namignil je namreč, da je bil sam Wittgenstein razpet med filozofskimi in antifilozofskimi impulzi, da je on sam vseskozi nihal med »metafiziko in kritiko jezika«, da je želel po eni strani enkrat za vselej prekiniti s filozofijo, a si vendar ni mogel kaj, da se ne bi znova in znova, iz dneva v dan, vračal k filozofiji. Nasprotujoče si razlage Wittgensteinove filozofije potemtakem niso posledica nezmožnosti interpretov, da bi sprevideli Wittgensteinov postopek, ampak je dvojnost vpisana v to filozofijo samo: že vse od začetka svoje filozofske poti se je Wittgenstein po eni strani zapletal v resno in strogo filozofsko argumentacijo, s katero je želel rešiti filozofske probleme, po drugi strani pa istočasno v diagnostiko izvorov filozofskih problemov, s katero je želel pokazati, da so filozofska vprašanja nesmiselna, ker temelji njihovo zastavljanje »na nerazumevanju logike našega jezika« (Wittgenstein, 2006: LFT 23). Wittgenstein se je torej tako rekoč potapljal v filozofske probleme in argumentiral za določena stališča, vendar je istočasno tudi sejal kal destrukcije in dvoma z nenehnim nizanem pomislekov, protiprimerov, miselnih eksperimentov, nenavadnih analogij, ironičnih komentarjev itd. Tudi *Raziskave* so na ta način dvoumne oz. večumne, tudi v njih spremljamo boj med filozofskimi in antifilozofskimi težnjami, zato jih lahko beremo tako kot resno filozofsko razpravo, ki argumentira določene teze, kakor tudi kot delo, ki »razrešuje filozofijo kot postavljanje tez in teorij z mnogimi terapijami, ki so usmerjene v konkretne probleme« (Pichler, 2004: 123).

Prvi korak k temu, da se nehamo zapletati v nesmiselne interpretacijske spopade in *Raziskave* uzremo v pravi luči, je ta, da se znebimo ustaljene predstave, po kateri so *Raziskave* pogovor med dvema osebama, med filozofskim terapevtom – lahko bi rekli: »subjekt, za katerega se predpostavlja, da ve« –, ki nas poskuša s svojo sokratsko metodo prevpraševanja filozofskih mnenj učenca/pacienta – ki v tej igri igra bolj ali manj samo vlogo z metafiziko zaslepljenega slehernika, ki ga je treba ozdraviti, torej »vlogo marionete oziroma grešnega kozla« (Stern) – pripeljati do tega, da uvidimo svojo nevednost, se znebimo napačnih metafizičnih pojmovanj in sprejmemo njegova nemetafizična pojmovanja, ki nam jih iz taktičnih razlogov ne želi posredovati neposredno, ampak šele skozi tovrstno terapijo. Po mojem mnenju v *Raziskavah* najdemo več glasov, ne samo dva.

Stern temu pritrjuje:

Namesto da argumente, predstavljene v knjigi, vidimo kot izmenjavo med Wittgensteinom in njegovim sogovornikom, predlagam, da jih vidimo kot izmenjavo med mnogimi glasovi, pri čemer nobenega od njih ne moremo neproblematično pripisati avtorju knjige. /.../

Poleg teh nasprotujočih si glasov, ki igrajo različne vloge na različnih mestih v knjigi, namreč najdemo še tretji glas. Ta glas, ki ga ni mogoče vedno

razlikovati od glasu pripovedovalca, ponuja ironične komentarje dialogov, komentarje, ki so deloma sestavljeni iz ugovorov predpostavkam, ki jih sogovornika sprejemata kot samoumevne, deloma pa iz truizmov o jeziku in vsakdanjem življenju, ki sta jih sogovornika spregledala. Večina bralcev obravnava oba glasova kot izraz Wittgensteinovih stališč, zaradi česar ne morejo uskladiti pronicljivih in provokativnih tez, ki jih zagovarja pripovedovalec, in komentatorjevega zavračanja vsakršnih filozofskih tez. /.../

Trdim, da Wittgenstein ni ponudil nikakršnih rešitev filozofskih problemov, ki jih obravnava v *Raziskavah*, temveč poskuša raz-rešiti te problem s pomočjo dialoga med nasprotujočimi si glasovi. (Stern, 2004: 22–23)

Čeprav se načeloma strinjam s Sternom, nas ta razmislek ne sme zavesti, ta delitev je samo način govora, kajti v končni fazi je samo navidezna, provizorična, tako rekoč od zunaj vsiljena tekstu, da ga lažje osmislimo. Tudi Cavell, ki je avtor tega klasičnega razumevanja, je opozoril, da sta oba glasova Wittgensteinova: »Ker sta oba glasova Wittgensteinova, ni noben (izključno) njegov, obstaja namreč onkraj teh glasov, prej in potem, ki ga v Wittgensteinovem delu zasedajo prilike, paradoksi, fantazije, aforizmi itd.« (Cavell, 1996: 374)

Zato nikakor ne želim reči, da je treba strogo ločevati med Wittgensteinom pripovedovalcem in Wittgensteinom komentatorjem oz. avtorjem, to v bistvu niti ni mogoče, tudi Stern ugotavlja, da glasova pogosto »ni mogoče jasno ločiti«. Prav tako ne želim reči, da je bil Wittgenstein razdvojena osebnost, danes resen filozof, ki postavlja teorije, jutri ironični antifilozof, ki želi prekiniti s filozofijo. Moja panta je obratna: v *Raziskavah* ne vlada le dvoglasje, ampak večglasje, v katerem ima vsak glas svojo težo, svoja stališča in točno določeno vlogo znotraj *Raziskav*. *Raziskave* so torej polifonične (Pichler, 2004; Stern, 2004), zato se v primeru, če se odločimo, da je eden izmed glasov v *Raziskavah* tisti, ki predstavlja Wittgensteinova izdelana filozofska stališča in njegove dokončne odgovore na bistvena filozofska vprašanja, slej ko prej najdemo v protislovju z ostalimi glasovi; če vse stavimo samo na en glas, pa najsibo ta filozofski, skeptični, ironični ali antifilozofski, mu lahko vedno oponiramo z navedki iz teksta, zato je potem neizbežen sklep, da so *Raziskave* nekonsistentno in nekoherentno delo.

In zato sem prepričan, da moramo nehati znotraj *Raziskav* – če jih želimo razumeti tako, kakor si je to zamislil njihov avtor – obsesivno iskati eden in edini Wittgensteinov avtorski filozofski glas. Uvideti moramo, da je Wittgenstein avtor tega kontinuiranega in odprtega večglasnega razgovora, ki ga spremljamo v *Raziskavah*, in da je bistvo *Raziskav* – in s tem Wittgensteinove metode – prav v tej medsebojni igri različnih glasov, ki govorijo z različnih perspektiv, nobeden izmed njih pa nima zadnje in dokončne besede.

Dialoška forma tako ni didaktično sredstvo, ki nas poskuša z igro različnih, nasprotujočih si glasov pripeljati do nekega višjega stališča – sinteze –, dialog se ne razreši v monolog, kjer eden izmed glasov prevzame glavno besedo in utiša vse

ostale, ampak je bistveno dialoga to, da nek filozofski problem s pomočjo različnih glasov osvetlimo z vseh strani, nobeden izmed glasov pa na koncu ne zmaga – dialog ostaja nerazrešen. Šele ko enkrat prevzamemo to perspektivo, se nam po mojem mnenju pokaže Wittgensteinov filozofski zastavek v vsej svoji izvirnosti in radikalnosti.

Sklep

Morda nam lahko pri preseganju te dileme pomaga primerjava s Kierkegaardom. Kierkegaard je, kot je znano, večino svoji del podpisoval s psevdonimi (včasih celo v isti knjigi najdemo več različnih psevdonimnih avtorjev, kot na primer v *Ali-ali*), vendar tega po lastnem pričevanju ni počel, ker bi se bal cenzure in bi zato skrival svojo identiteto, ampak je preprosto – iz razlogov, v katere se na tem mestu ne bomo spuščali – ustvarjal različne filozofske avtorje, ki so pisali vsak s svojega stališča in imeli vsak svoje življenje in ki jih je Kierkegaard zapletal v medsebojne dialoge in polemike. A kljub njegovim jasnim opozorilom, da so »avtorji pesniška stvaritev« (Kierkegaard, 2009: 528), so interpreti Kierkegaardove filozofije obsesivno iskali tistega psevdonimnega avtorja, ki naj bi bil najbližje Kierkegaardovim lastnim pogledom, torej nekakšne njegov alterego, skrit za psevdonimom. Sam Kierkegaard je – v postskriptu, ki je podpisan z njegovim imenom – takole razložil svojo rabo psevdonimov:

Moja psevdonimnost oziroma večimenskost ni naključno temeljila na moji osebi, temveč bistveno v produkciji sami. /.../ To pomeni, da sem sam neosebno ali osebno v tretjeosebni vlogi prišepetovalca, ki je pesniško proizvedel avtorje, katerih predgovori so njihov lasten proizvod, kakor tudi njihova imena. V knjigah, ki so podpisane s psevdonimi, torej ni niti ena sama beseda moja. O teh avtorjih nimam nobenega mnenja, razen kot zunanji opazovalec, ničesar ne vem o tem, kaj so želeli povedati, razen kot bralec, z njimi tudi nimam nobene osebne vezi. /.../ Moja vloga je zgolj ta, da sem hkrati tajnik in, precej ironično, dialektično podvojeni avtor avtorja oz. avtorjev. (Kierkegaard, 2009: 527–529)

Čeprav je jasno, da Kierkegaard pretirava, se mi zdi ta razlaga plodna tudi za razumevanje *Filozofskih raziskav*. Po mojem mnenju ne smemo na *Raziskave* gledati kot na delo, v katerem nam želi avtor skozi enega izmed glasov posredovati mnenja in stališča, kaj šele že izdelan in dokončni filozofski sistem, temveč moramo nanje gledati kot na celostno delo, katerega bistvo je prav večglasje in v katerem argumentativne in terapevtske prvine niso strogo ločene, ampak so povezane in komplementarne. Wittgenstein potemtakem ni eden izmed likov v *Raziskavah*, ampak je snovalec besedila, ki tako rekoč pred našimi očmi uprizori filozofsko raziskovanje v obliki večglasnega razgovora, katerega cilj je, da se filozofski problemi ne samo rešijo, ampak, kot neštetočrat poudarja, razrešijo in razpustijo.

Če moje razmišljanje drži, potem je Wittgenstein predvsem filozofski terapevt, ki nas poskuša ozdraviti filozofskih problemov tako, da uprizoni razpravo o določenem problemu in ga s tem, ko neprenehoma kroži okrog njega z različnih perspektiv, osvetli z vseh strani, ne ponudi pa nobenega dokončnega odgovora nanj – kakor hitro mislimo, da smo se v *Raziskavah* končno dokopali do neke filozofske pozicije, se že v naslednjem momentu ta izkaže za nezadostno, zoper njo se začnejo nizati ugovori in pomisleki.

Delo Wittgensteinovega gramatičnega filozofskega raziskovanja, ki smo mu priča preko drame *Filozofskih raziskav*, je tako dvojno, a sopovezano: po eni strani spremljamo rigorozno raziskovanje jezikovnih blodnjakov, s katerim Wittgenstein demonstrira različne postopke in tehnike za razčlenjevanje našega jezika. To siziofsko početje, ki sicer res, kakor vseskozi poudarja Wittgenstein, v prvi vrsti zasleduje terapevtske cilje, vodi do presenetljivih in kritičnih opažanj o delovanju našega jezika, kakor tudi do uvidov o tem, kako se v različnih kontekstih uporabljajo različni filozofski pojmi – na dan pridejo pravila njihove rabe. Lahko bi rekli, da je to »leitmotiv« Wittgensteinove pozne filozofije: ni nečesa takega kot pomena besede na sebi, besede in pojmi so razumljivi samo v njihovi rabi znotraj točno določenega konteksta, ki ni samo strogo jezikoven, ampak je vseživljenjski (kar je konec koncev poanta koncepta jezikovnih iger).

Po drugi strani je sestavni del Wittgensteinovega raziskovanja diagnostika, ki išče izvore in vzroke filozofskih problemov – v *Raziskavah* spremljamo samo genezo določenega filozofskega problema. Cilj večglasnega dialoga v *Raziskavah* torej ni zgolj neposredno terapevtski, ampak tudi diagnostičen: Wittgenstein dobesedno uprizarja razmišljanje, ki privede do nerešljivih filozofskih vprašanj, v katere se po naravi stvari zapleta naš um pri filozofiranju. In v tem smislu so *Raziskave* dobesedno in v prenesenem pomenu dramske; ne samo, da spremljamo filozofsko raziskovanje v značilni dramski obliki dialoga, ampak spremljamo dramatizacijo samega filozofskega mišljenja, ki ga Wittgenstein oživi tako, da predstavi miselne tokove, ki nas zapeljejo v filozofske težave, tako karakteristično, da se bralec prepozna v njih in prizna: »Ja, točno tako sem to misli.« (Wittgenstein, 2000a: 164) Wittgenstein torej v *Raziskavah* dobesedno uprizoni filozofski problem in njegovo genezo, pri čemer je njegov uspeh toliko večji, »kolikor bolj je žebelj zadet na glavico« (Wittgenstein, 2006: LFT 23), kolikor bolj torej v Wittgensteinovih dramskih osebah prepoznamo skušnjave našega lastnega mišljenja.

Ne smemo namreč pozabiti na Kierkegaardovo opozorilo: človeka, ki je žrtev neke iluzije, ne moremo ozdraviti tako, da ga neposredno soočimo z njegovo zmoto, neposredni napad ga bo najverjetneje le še okrepil v njegovi iluziji. Približati se mu moramo neočitno, tako rekoč s hrbta, zvabiti ga moramo v našo igro in mu tako pomoliti ogledalo, v odsevu katerega bo navsezadnje le spoznal samega sebe in uvidel svojo zmoto: »Jaz naj bi bil le zrcalo, v katerem bo moj bralec videl lastno mišljenje z vsemi njegovimi deformacijami in s pomočjo katerega ga bo

lahko spravil v red.« (Wittgenstein, 2005: 78) Šele ko enkrat spoznamo, zakaj smo se zapletli v filozofske probleme in kaj nas je zavedlo k takšnemu mišljenju, lahko končno najdemo »pot iz muholovca« (Wittgenstein, 2006: FR 309), se izkoplujemo iz primeža določenega filozofskega problema in ga za vedno pustimo za sabo. Zato je ena najpomembnejših značilnosti Wittgensteinove specifične filozofske dejavnosti, ki jo demonstrira v *Raziskavah*, diagnostično delo (samo)demistifikacije tako naših učenih kot spontanih teoremov: »Filozofija je pripomoček, ki je uporaben samo proti filozofom in proti filozofu v nas.« (Wittgenstein, 2000b: MS 219, 11)

In zato se mi zdi veliko bolj smiselno govoriti o Wittgensteinovi »avtorski strategiji« (Conant) kot o Wittgensteinovi argumentaciji za določene teze. S tem nika-
kor ne trdim, da v *Raziskavah* ni argumentacije, nasprotno, v *Raziskavah* je mnogo raznoraznih argumentov, predvsem tako imenovanih posrednih, ogromno je miselnih eksperimentov, analogij, različnih opažanj o jeziku ipd. Toda ločevati moramo med tem, kaj ti argumenti predstavljajo na mikro in kaj na makro ravni *Raziskav*. Če ostanemo na mikro ravni določenega miselnega toka v *Raziskavah*, potem popolnoma upravičeno na njih gledamo kot na postopke, s katerimi eden izmed glasov argumentira za določeno filozofsko tezo oziroma proti njej. Če pa gledamo na *Raziskave* kot na celostno avtorsko stvaritev, potem ti postopki niso v prvi vrsti poti do Wittgensteinovih filozofskih rezultatov ali njegovega filozofskega sistema, ampak so poteze znotraj Wittgensteinovega nenehnega rigoroznega prepraševanja in samoizpraševanja, koraki, ki sledijo Wittgensteinovi strategiji razreševanja filozofskih problemov.

Form in the Service of Method: How to Read Wittgenstein's *Philosophical Investigations*

This paper argues that contradictory interpretations of Wittgenstein's philosophical activity in the *Philosophical Investigations* are due to the specific dialogue form of the text. This makes it impossible to discern any explicit and unambiguous position that we can attribute to the author and is consistent with the text itself. However, this is not a shortcoming of the text based on its author's psychological features – on the contrary, it is essential to Wittgenstein's authorial strategy. Once such a perspective is established, the *Investigations* are revealed as a polyphonic text which examines divergent philosophical problems from multiple perspectives through the interplay of different voices, thereby not giving any definite answers to them. Furthermore, the polyphonic dialogical form is inseparably connected to Wittgenstein's conception or, better yet, conversion of philosophy itself.

Keywords: form of *Philosophical Investigations*, enigma of Wittgenstein's position, Wittgenstein's method, the paradox of metaphilosophy, polyphony

Literatura

- Baker, G. P. (2006). *Wittgenstein's Method*. Oxford: Blackwell.
- Baker, G. P. in Hacker P. M. S. (2009). *Wittgenstein: Understanding and Meaning*. Oxford: Willey-Blackwell.
- Cavell, S. (1962). »The Availability of Wittgenstein's Later Philosophy«. *Philosophical Review* 71 (1): 67–93.
- Cavell, S. (1996). »The *Investigations*' Everyday Aesthetics of Itself«. V Mulhall, S. (ur), *The Cavell Reader*, Cambridge: Blackwell, 369–389.
- Conant, J. (2007). »Mild Mono-Wittgensteinianism«. V Crary, A. (ur), *Wittgenstein and the Moral Life: Essays in Honor of Cora Diamond*, Cambridge: The MIT Press, str. 31–142.
- Diamond, C. (1991). *The Realistic Spirit: Wittgenstein, Philosophy and the Mind*. Cambridge: The MIT Press.
- Dummett, M. (1959). »Wittgenstein's Philosophy of Mathematics«. *Philosophical Review*, 68(3), str. 324–348.
- Fogelin, R. (1987). *Wittgenstein*. Druga izdaja. London: Routledge & Kegan Paul.
- Hintikka, M. B. in Hintikka, J. (1986). *Investigating Wittgenstein*. Oxford: Blackwell.
- Hutchinson, P. (2007). »What's the Point of Elucidation?«. *Metaphilosophy*, 38(5), str. 691–713.
- Kenny, A. (2006). *Wittgenstein*. Oxford: Blackwell.
- Kierkegaard, S. (2009). *Concluding Unscientific Postscript to the Philosophical Crumbs*. Cambridge: Cambridge University Press.
- Kripke, S. (1982). *Wittgenstein on Rules and Private Language*. Cambridge, Mass.: Harvard University Press.
- McDowell, J. (2000). »Non-Cognitivism and Rule-Following«. V Crary, A. in Read, R. (ur.), *The New Wittgenstein*, str. 38–53. London: Routledge.
- Pichler, A. (2004). *Wittgensteins 'Philosophische Untersuchungen': Vom Buch zum Album (Studien zur Österreichischen Philosophie 36)*. Amsterdam: Rodopi.
- Putnam, H. (1975). *Mind, Language, and Reality*. Cambridge: Cambridge University Press.
- Stern, D. G. (1995). *Wittgenstein on Mind and Language*. Oxford: Oxford University Press.

- Stern, D. G. (2004). Wittgenstein's *Philosophical Investigations*: An Introduction. Cambridge, Mass.: Cambridge University Press.
- Vossenkuhl, W. (2003). *Ludwig Wittgenstein*. München: Verlag C. H. Beck.
- Wiggins, D. (1997). *Needs, Values, Truth: Essays in the Philosophy of Value*. Oxford: Oxford University Press.
- Wittgenstein, L. (1984). *Über Gewißheit. Bemerkungen über die Farben Zettel. Vermischte Bemerkungen*. Frankfurt am Main: Suhrkamp.
- Wittgenstein, L. (2000a). *Wiener Ausgabe: The Big Typescript*. Frankfurt am Main: Zweitausendeins.
- Wittgenstein, L. (2000b). *Wittgenstein's Nachlass. The Bergen Electronic Edition*. Oxford: Oxford University Press.
- Wittgenstein, L. (2005). *Kultura in vrednota*. Ljubljana: Študentska založba.
- Wittgenstein, L. (2006). *Tractatus logico-philosophicus. Tagebücher 1914-1916. Philosophische Untersuchungen*. V Wittgenstein, L., *Werkausgabe Band 1*, Frankfurt am Main: Suhrkamp.
- Wright, C. (2001). *Rails to Infinity*. Cambridge, Mass.: Harvard University Press.

Borut Ošlaj

Filozofska fakulteta Univerze v Ljubljani

Svoboda svobode govora

Prispevek obravnava vprašanje, v kolikšni meri je svoboda govora danes resnično svoboda. S tem vprašanjem v polju svobode govora odpira problem njene etičnosti, problem njenih umno določenih in univerzalno sprejemljivih meja, ki jih zlasti danes močno določajo koncepti politične korektnosti, sovražnega govora in strpnosti.

Ključne besede: svoboda govora, politična korektnost, sovražni govor, strpnost, etika

Govoriti o svobodi svobode govora, kar se na prvi pogled nemara sliši dokaj sofisticirano, pomeni govoriti o razsežnostih in mejah svobode govora, pomeni govoriti o kriterijih, ki določajo to, kar naj svoboda govora ne le omogoča in dopušča, temveč tudi to, do kod se lahko razteza, da bi na ta način v enaki meri drugega pred drugim ščitila vse govorne subjekte. V ozadju je torej načelno in hkrati tudi temeljno vprašanje, v kolikšni meri je svoboda govora kot ena izmed ključnih civilizacijskih in pravno reguliranih pravic znotraj vsake demokratične družbe resnično svobodna.

Zastaviti si vprašanje o svobodi svobode govora tako nujno pomeni, da tisti, ki takšno vprašanje zastavlja, hkrati tudi predpostavlja, da svoboda govora ni nujno svobodna oz. da svoboda premore različne stopnje, ki nek govor bolj ali manj približujejo ali oddaljujejo od tega, kar naj bi svoboda bila. Če se nek govor, misel, prepričanje in mnenje kitijo s pridevnikom svobode, torej s pojmom, ki se že dobrih 2500 let trudi označevati morda celo temeljno določilo ne le človeškosti, temveč hkrati tudi človečnosti človeka, potem bi se morali hkrati tudi zavedati, da se na ta način podajajo v samo središče potencialno neskončnega bojnega polja možnosti in zmožnosti, želja in interesov, vednosti in vere, predvsem pa vsega tistega, kar sta človek in človeštvo bila, kar sta in kar bi nekoč morda želela postati. Da, zavedati bi se morali, da na ta način vstopajo v odprt prostor duhovne prostosti, ki v svoji načelni eksperimentalnosti in negotovosti ne omogoča niti zanesljive vnaprejšnje napovedi določenega cilja, kaj šele njegove uresničitve.

Zadrega in tveganje tistega, ki želi svobodo govora ne le udejanjati, temveč jo hkrati tudi misliti, pa ni nič manjša, saj se pri kritični presoji svobode govora ne

more naslanjati na prav nobeno formo gotovosti, na noben zakon ali aksiom, razen na svobodo. Mera in kriterij svobode je namreč lahko le svoboda sama. Svoboda, ki naj kot kriterij meri in presoja svobodo govora, pa vendarle ni identična samemu predmetu presoje; kot kriterij je svoboda namreč ideja, medtem ko v govoru nastopa kot dejstvo oz. konkretno stanje. Misliti svobodo svobode govora torej pomeni, uporabiti idejo svobode za kriterij presoje njene konkretnosti v polju govora. Toda na katero idejo svobode s tem namigujemo?

Zaradi formalnih in pragmatičnih razlogov naj zgolj omenim, da se bom posluževal osnovnih kontur Kantove ideje svobode; predvsem zato, ker doslej še nihče ni prepričljivo predstavil boljšega modela (Kant, 1993: 31, 33 in 36; gl. tudi Kant, 2005). Poleg tega je Kantova opredelitev svobode vse do danes ostala ne le najbolj vplivna, temveč je s temeljnimi filozofsko-antropološkimi dejstvi človeka kot *homo symbolicus* hkrati tudi najbolje usklajena. Izhodiščna forma človekove svobode je pri Kantu, tako kot že pri Aristotelu in Hobbesu negativna. Negativna svoboda pomeni svobodo od, pomeni prostost in neodvisnost v najširšem smislu ter s tem načelno neskončno odprtost možnosti. Kdor ostaja na ravni negativne svobode, še ni resnično svoboden in ni dosegel njenega celotnega pojmovnega obsega. To lahko postane šele takrat, ko z umno spoznanim moralnim zakonom svoji volji v vrtincih njene negativne svobode kot samovolje ponudi univerzalno merilo odločanja; s slednjim njegova svoboda postane pozitivna in na ta način uresničena. Pozitivna svoboda je torej pri Kantu oblika etične razrešitve in dovršitve negativne svobode v njeni načelni prostosti in vrednotni indiferentnosti, je meja, ki jo svobodna volja v formi moralnega zakona postavlja sama sebi ter pri tem sledi kriteriju univerzalnosti uma. Pojem meje pa je tu vendarle paradoksen in deloma zavajajoč, saj je prav meja, ki jo umna volja postavlja sama sebi, nujni pogoj brezmejnosti vladavine etične umnosti. Brezmejnost torej zahteva meje, svoboda pa posledično nujnost.

Če zdaj Kantovo dvodelno idejo svobode¹ prenesem nazaj na izhodiščno vprašanje, v kolikšni meri je svoboda govora resnično svobodna, potem na ta način v polju svobode govora odpiram vprašanje njene etičnosti, vprašanje njenih umno določenih in univerzalno sprejemljivih meja. Oglejmo si najprej aktualno stanje meja svobode govora. Z aktualnimi mejami mislim zlasti in predvsem tiste meje, ki so v zadnjem času postale predmet najrazličnejših polemik; to so politična korektnost, strpnost in sovražni govor. Vsi trije koncepti danes veljajo za nekakšne samoumevne moralno-politične kode, s katerimi udeleženci javnega prostora na-

¹ V izogib možnim nesporazumom naj ponovim, da sledim zgolj osnovnim konturam njegovega dvodelnega pojma svobode, ne pa tudi tistim moralnim implikacijam, ki jih Kant zakoliči v specifično razumljenem moralnem zakonu (kategorični imperativ); zgolj ta je namreč pri njem identičen uresničen pozitivni svobodni. Da je pozitivna svoboda v končni fazi tesno povezana z vdorom etičnosti v načelno odprtost negativne svobode, ostaja nedvomno; kaj to etično po svojem bistvu je, pa je že neko drugo vprašanje, ki ga na tem mestu puščam odprtega, in ga zato nikakor ne enačim s Kantovim moralnim zakonom. Več o opredelitvi svobode v Ošljaj (2015).

čeloma uravnavajo odprte prostore prostosti svobode govora, ji postavljajo smiselne meje in na ta način ščitijo javni prostor pred možno samovoljo in ekscesnostjo svobode govora v njeni izhodiščni negativni odprtosti. Formalno gledano vse tri kode pomenijo določeno konkretizacijo pozitivne svobode, seveda ne v smislu Kantovega moralnega zakona, vsekakor pa so to norme, ki naj uravnavajo in omejujejo svobodo govora tako, da v intersubjektivni mreži različnih govorno izraženih mnenj, stališč in interesov v vzajemni meri ščitijo svobodo pred svobodo; prav to pa načeloma ni nič drugega kot način etičnega samoomejevanja svobode v dobro drugih in nas samih, vsaj potencialno pa tudi v dobro uveljavljanja brezmejnosti vladavine uma. Seveda pri tem ne smemo prezreti, da je to le etična ideja, ki zgolj nakazuje smiselnost in nujnost omenjenih konceptov oz. kod. Toda ne le v tem primeru, tudi nasploh velja, da imamo ljudje velike težave z udejanjanjem posameznih idej, pa naj so te še tako smiselne. Politična korektnost, strpnost in sovražni govor so v javni prostor vsaj načeloma vstopili in se v njem uveljavili kot regulatorji svobode govora, kot imena pozitivne svobode, ki naj pomagajo zagotavljati in varovati enakost, pravičnost in svobodo v njeni pozitivni, etični oz. humani naravnosti. Vprašajmo se, ali je temu res tako, ali omenjeni samoomejevalni regulatorji svobode govora dejansko pomenijo uresničevanje svobode kot etosa, jim res uspeva braniti svobodo enih pred svobodo drugih?

Prepričan sem, da med kritično mislečimi ni nikogar, ki bi na omenjeno vprašanje brez slehernih zadržkov odgovoril pritrdilno. Se morda razlog našega nezadovoljstva nad prepričljivostjo in uspešnostjo omenjenih moralno-političnih kod skriva v njihovi nezadostnosti, in bi bilo zato nemara bolje, če bi še naprej vzdrževali in dosledneje uporabljali klasične moralne kode razsvetljenstva, kot so dostojanstvo, enakost, pravičnost, spoštljivost ipd.? Bi z odločnejšo umestitvijo klasičnih norm v jedro javnega prostora res lahko zmanjšali obseg anomalij svobode govora? Bojim se, da ne. Vsekakor je res, da lahko s trenutno modnimi koncepti obidemo niz temeljnih standardov demokratične kulture, kot so univerzalnost presojanja, domnevo nedolžnosti, jasen mandat razsojevalca za presojanje, pravico do pravnega sredstva itd., zaradi česar so ne po naključju predmet številnih pomislekov. Rečeno velja še toliko bolj, ker za moderne regulatorje svobode v aktualni dobi postfaktičnosti upravičeno velja, da so interpretacijsko bolj odprti in pomensko manj jasni, kot to velja za klasične norme, in da so zato posledično tudi bolj odporni na različne ugovore in morebitne obrambe; ja, z njimi je možno lažje manipulirati. Kljub nedvomni razliki med klasičnimi in sodobnimi normativnimi regulatorji svobode, pa lahko temeljne standarde demokratične kulture brez večjih težav vendarle obidemo tudi pri uporabi klasičnih razsvetljenskih moralnih norm. Tudi te namreč dopuščajo diskvalifikacijo posameznika, tudi v tem primeru lahko že najmanjša slutnja nesoglasja zadostuje, da se nekoga ožigosa, pri čemer se bosta tako žrtev kot tudi diskvalifikator lahko enako strastno sklicevala na enakost, dostojanstvo, pravičnost in spoštljivost. Če je nekdo trn v peti javnemu mnenju, če se upira konformnosti, ga bo njegova svoboda govora zelo verjetno

stala izolacije ali še česa hujšega, in to ne glede na to, katere regulativne koncepte zagotavljanja in omejevanja svobode govora družba pri tem uporablja. Prislunhni-mo naslednjim besedam Alexisa de Tocquevilla iz davnega leta 1835, besedam avtorja, ki je bolj kot kdorkoli drug razumel, kako ranljiva je svoboda (govora) tudi znotraj demokratičnega družbenega reda:

/.../ [večina] sklepa /.../ okrog misli strašanski obroč. Znotraj njegovih meja je pisec svoboden; a gorje tistemu, ki si ga upa prestopiti. Ni se mu sicer treba bati grmade, se pa izpostavi vsakovrstnemu zgražanju in vsakodnevemu preganjanju. Politična kariera je zanj zaprta /.../ Odrekajo se mu vse, tudi slavo. /.../ Gospodar nič več ne govori: »Mislili boste tako kot jaz ali pa boste umrli.« Temveč: »Ni vam treba misliti tako kot jaz; vaše življenje, imetje, vse vam ostaja; a od danes ste med nami tujec. Ohranili boste pravice, ki vam pripadajo v [meščanski] družbi, vendar vam ne bodo koristile /.../ Ostali boste med ljudmi, a pravice človeškega rodu boste izgubili. /.../ Pojdite v miru, puščam vam življenje, toda življenje, ki vam ga puščam, je slabše od smrti.« (Tocqueville, 1996: 253–254)

Ne le takrat, pred skoraj 200 leti, tudi danes se nekaterim posameznikom godi podobno, le da bi danes namesto večine prej govorili o javnem mnenju, ki ga ustvarjajo mediji v povezavi z določeno vredno(s)tno naravnano politiko, medtem ko namesto pisateljev vlogo žrtve javnega mnenja danes lahko igra vsak, ki vstopa v javni prostor in v njem sklicujoč se na svobodo govora zastopa določena javnemu mnenju neprijetna stališča. Četudi je bila koda politične korektnosti pred 200 leti šele v povojih, in se je praviloma še vedno uporabljal klasični razsvetljski moralni besednjak, pa so bile oblike sankcioniranja avtorjev neprilagojenih stališč podobne današnjim. Če so tovrstne sankcije takrat in danes posledica svobode govora in njen možni končni rezultat, potem v takšni svobodi govora gotovo ne more biti prav veliko svobode.

Zakaj regulatorji svobode govora ne takrat ne danes, kljub njihovim razlikam, ne delujejo kot bi morali; še več, zakaj jih je mogoče uporabljati in celo zlorabljati za različne namene, in sicer ne glede na to, kakšni ti regulatorji so? Zakaj danes nismo nič bližje ideji pozitivne svobode živetega družbenega etosa kot pred 200 ali 1000 leti? Zaradi sprenevedanja, se glasi odgovor; zato, ker ne vidimo ali ne želimo videti, da stopnja svobode določene družbe, četudi je ta formalno ponotranjila temeljne demokratične standarde, v glavnem ni odvisna od teh ali onih moralno-političnih konceptov, kod ali norm. Osrednji problem svobode govora se praviloma namreč nikoli ne skriva v njenih aktualnih regulatorjih, temveč pretežno ali celo izključno v ljudeh, ki se in kako se teh regulatorjev poslužujejo oz. se jih morebiti sploh ne poslužujejo, s čimer ravnaajo izključno v polju negativne svobode kot samovolje. Toda slednja možnost, ravnanje v polju samovolje negativne svobode brez slehernih regulatorjev in samoomejevalnih norm, tu ni predmet mojega razmisleka, kakorkoli je seveda res, da ne tako majhno število ljudi,

zlasti v našem kulturnem krogu, svobodo govora pojmuje prav na tovrstno bizaren način. Osredotočimo se torej zgolj na vprašanje razumevanja in uporabe aktualnih normativnih regulatorjev svobode govora. Politična korektnost, strpnost in sovražni govor so moralno-politične norme z visoko stopnjo pomenske odprtosti in nejasnosti, in so prav zato tudi, bolj kot to velja za klasične norme, povezane z zahtevo, da smo jih v posameznih kontekstih zmožni ustrezno razumeti in uporabiti. Ustrezno razumevanje in način uporabe teh konceptov zato terjata kritično in etično zrelo osebo. Toda načeloma to velja za vse koncepte, norme in pravila. V vseh obdobjih so ti – to je treba zdaj jasno poudariti – zgolj orodja, ki se jih lahko poslužujemo ali pa tudi ne; če da, pa jih lahko spet uporabljamo tako ali drugače. Zgolj zaradi dobrih konceptualnih orodij namreč še nobena družba ni postala dobra. Zato tudi slaba ne pomenijo nujno, da se bo slabo dogajalo. Če nekdo hodi naokrog z orodji uničevanja (pištola), to še ne pomeni, da bo uničeval. Možno je, da bo življenja celo reševal. Prav tako tisti, ki je oborožen z najplemenitejšimi koncepti človečnosti, človečnost ne bo nujno sejal, temveč bo prav s pomočjo teh konceptov in načinov njihove uporabe ali zlorabe, življenja lahko prefinjeno uničeval – še zlasti, če ima vpliv in moč. Osrednji problem svobode govora torej ni v nezadostnosti in omejitvah trenutno modnih moralno-političnih norm, temveč v nezadostnosti in omejitvah umnosti in človečnosti tistih, ki se teh kriterijev poslužujejo.

Vloga moralno-etičnih regulatorjev naj bi praviloma vselej služila predvsem zaščiti šibkejših članov družbe. Naj bi, praviloma; dejansko pa se vedno znova izkaže, in v primeru politične korektnosti, strpnosti in sovražnega govora ni nič drugega, da ti najbolj učinkovito služijo prav tistim, ki njihovo zaščito najmanj potrebujejo, torej tistim, ki imajo največ vpliva in moči, tistim, ki oblikujejo javno mnenje in preko njega zasledujejo te ali one razmeroma ozko odmerjene interese. Z njihovo pomočjo se trudijo vzdrževati videz lastne etične neoporečnosti, dejansko pa uveljavljene regulativne norme v njihovi pomenski odprtosti zlorablajo; predelujejo jih v sofisticirana kvazi-intelektualna orožja, ki so rezistentna na vsak umni ugovor, z njimi lahko brez posebnih težav onemogočajo drugače misleče. Svoboda govora se tako zdi svobodna samo za nosilce vpliva in moči; oni lahko rečejo, kar želijo, uporabljajo interpretacijsko odprte moralne kriterije, kot jim ustreza, ne da bi se morali bati za morebitne posledice. Tistim, ki niso v celoti usklajeni z javnim mnenjem, svoboda govora sicer dopušča, da izrečejo, kar želijo, čutijo ali vedo, toda s tem tvegajo različne sankcije ali celo popolno osamitev, ki jo širša javnost morda sploh ne bo zaznala. Njihova svoboda govora je zato golo slepilo, ki pomaga vzdrževati videz demokratične družbe v njeni opevani odprtosti in svobodi. Toda tudi navidezna svoboda govora subjektov vpliva in moči kot tudi vseh tistih, ki se jim mnenjsko pridružujejo, ni resnično svobodna. Če bi to naj bila, potem bi morala biti nedvoumni izraz tiste etično kodirane pozitivne svobode, ki lahko prestane test umne univerzalizacije. Ko javno mnenje samo sebe določi za normo politične korektnosti, ko njegovi apologeti lastno svo-

bodo govora in kode njenega samoomejevanja uporabljajo kot orožje, s katerim onemogočajo drugače misleče, ko torej svobode ne uporabljajo kot smoter, kar edino lahko ustvarja prostor etosa, svoboda v prostoru, ki ga obvladujejo, izpareva kot vodne kapljice na razgreti plošči. Če je svoboda svobode govora kritično mislečih ne redko zgolj slepilo, pa je svoboda svobode govora nosilec vpliva in moči v veliki meri farsa. Sovražni govor, politična korektnost in nestrpnost kot samovoljno uporabljana orožja so v tovrstnih primerih simptomi manjka (pozitivne) svobode v osrčju svobode govora. V takšnem govoru svoboda, ker je sama sebi (ideološko-politično) sredstvo in ne cilj, razkraja in niči samo sebe. Za fasado lažne svobode se z navidezno etičnimi sredstvi tako lahko mirno nadaljujejo različni ideološki boji.

Nedvomno dejstvo, ki opredeljuje vrh vseh struktur, ki ustvarjajo »politično korektno«¹ javno mnenje, so igre vplivov in moči v povezavi s svojevrstno mešanico umnosti in neumnosti; vse drugo so pretežno sredstva, s katerimi te igre igrajo tisti, ki vpliv in moč imajo. Ko bo te igre igralo več umnih oz. etično odgovornih ljudi, bodo tudi igre lepše in za večino boljše, dokler pa ne, pa je pravzaprav tudi čudenje nad tem, da je tako, kot je, povsem odveč. Nikakor pa ne more biti odveč prizadevanje, da bi in kako bi lahko bilo ne le drugače, temveč tudi boljše. To pa je možno le tako, da svobodo kot slepilo in farso postopoma nadomeščamo z več resnične in udejanjane pozitivne svobode, enakovredno za vse ljudi dialoške volje. Prvi, vendar hkrati tudi najlažji korak v tej smeri pa je prav gotovo razkrivanje različnih anomalij svobode v njeni praksi.

Naj sklenem: Politična korektnost, strpnost in sovražni govor so primerni mehanizmi reguliranja svobode govora takrat, ko so del prevladujoče umnega in odgovornega javnega prostora in se jih poslužuje zadostno število etičnih oseb. Samo v tem primeru lahko svoboda govora postane resnično svobodna.

Freedom of Free Speech

The paper discusses to what extent free speech is really free today. Thus, it introduces to the field of free speech the problem of freedom's ethical nature, its rationally delineated and universally acceptable boundaries that are especially today strongly determined by the concepts of political correctness, hate speech and tolerance.

Keywords: free speech, political correctness, hate speech, tolerance, ethics

Literatura

Kant, I. (1993). *Kritika praktičnega uma*. Ljubljana: Društvo za teoretsko psihoanalizo.

Kant, I. (2005). *Utemeljitev metafizike nravi*. Ljubljana: Založba ZRC SAZU.

Ošlaj, B. (2015). *Človek, svet in etos: študije o postsekularni filozofiji in svetovnem etosu*. Ljubljana: Znanstvena založba Filozofske fakultete.

Tocqueville, A. de (1996). *Demokracija v Ameriki I*. Ljubljana: Krtina.

