

UVOD

V knjigi, ki je pred vami, razmišlja o sodobnih znanstvenih razumevanjih in načinih raziskovanja odločanja v povezavi z vprašanjem osebne avtonomije. Če upoštevamo nekatere pomembne lastnosti odločanja in udejanjanja avtonomije, ki jih znanstvene raziskave pogosto zanemarjajo, ugotovimo, da so odločevalke in odločevalci¹ sposobni soustvarjati to, kar so ali želijo postati. Spoznati jih moramo v njihovih vsakdanjih okoljih, raziskati, kako sami občutijo in razumejo svoje odločanje, ter razmisliti o njihovih možnostih (samo)refleksije. Ker pa avtonomije in njenega udejanjanja ni tako težko spodkopati, moramo v sodobnem svetu, polnem izbir, namernih spreminjanj okolij odločanja in manipulacij vedenja, paziti, da je ne izgubimo. Slednje seveda velja le pod predpostavko, da nam avtonomija predstavlja pomembno vrednoto – ne smemo namreč zanemariti dejstva, da okolja, v katera smo vpeti, bistveno sodoločajo, kako se odločamo, kaj smo kot osebe in kakšne so naše vrednote. Meniva, da sta osebna avtonomija in njeno udejanjanje pomembni tako za posameznika kot za družbo. Prav zato moramo predpostavkam, na katerih slonijo naša razumevanja in načini raziskovanja odločanja, praktičnim uporabam spoznanj o odločanju, ki iz njih izhajajo, ter dilemam, ki se ob tem porajajo glede osebne avtonomije, nameniti kritičen razmislek.

Pojem avtonomije je mnogo širši, kot ga predstavlja v tej knjigi. V njej se osredotočava predvsem na osebno avtonomijo in se večinoma omejujeva na nekatere temeljne vidike, ki se pojavljajo v sodobnih razpravah o uporabi različnih namernih spreminjanj odločitvenih okolij. Vseeno naj nekaj uvodnih besed nameniva tudi bolj klasičnim pogledom na osebno avtonomijo.

V razpravah o osebni avtonomiji se kot izhodišče pogosto navaja znani odlomek iz spisa Immanuela Kanta »Odgovor na vprašanje: Kaj je razsvetljenstvo?«:

Razsvetljenstvo je človekov izhod iz nedoletnosti, ki je je kriv sam.

1 V nadaljevanju knjige kot nevtralna oblika uporablja slovnični moški spol.

Nedoletnost je nezmožnost uporabljati svoj razum brez vodstva koga drugega. Krivda za to nedoletnost pa je lastna, če vzrok zanjo ne zadeva pomanjkanja razuma, temveč odločnosti in poguma, da bi ga človek uporabljal brez vodstva koga drugega. Sapere aude! Bodi pogumen, uporabljal svoj lastni razum, je torej geslo razsvetljenstva. (Kant, 1784/2006: 23)

V odlomku Kant slavi intelektualno samodoločitev in kot nedoletnost označi zavestno sprejemanje vodstva nečesa, kar je zunaj njega. Pri tem izpostavlja predvsem razum. Podobno samozadostnost v eseju o samozanašanju² zagovarja ameriški filozof Ralph Waldo Emerson (1950), ki mnenja drugih in njihov potencialni vpliv jemlje kot grožnjo, ki spodkopava (samo)vrednost, na kateri je utemeljena individualnost. Tudi David Hume (1748/1974) poudarja, da se je treba zanašati nase, bolj kot razum pa izpostavlja vlogo lastnih čutov in spomina, saj naj bi le njim lahko zares zapuili. Adam Carter (2020) pravi, da bi si z upoštevanjem zahtev teh treh mislecev lahko zamislili popolnoma avtonomno osebo. Ta bi morala izpolnjevati naslednje zahteve: (i) nikoli ne bi nekritično verjela drugim; (ii) nikoli ne bi svojemu razumu dovolila, da ga vodi drugi; (iii) aktivno bi si prizadevala, da je ne oblikujejo mnenja drugih (prav tam: 2939). Taka oseba bi sicer imela popolno spoznavno avtonomijo, a vprašanje je, ali bi bilo to zanjo dobro. Oseba bi bila v popolni kognitivni izolaciji od drugih, kar bi jo posledično odrezalo od kognitivne povezanosti, ki je za človeka ravno tako ena od dobroti oziroma vrlin. Carter to ugotovitev povzame kot parafrazo slavne Kantove misli (prav tam: 2940):

Veličastna intelektualna samodoločitev in neodvisnost misli brez ustreznih temeljev znanja – od drugih, tehnologije, interneta, idr. – je dejansko prazna, medtem ko je znanje v odsotnosti ustrezne sposobnosti avtonomne samodoločitve slepo.

Zato Carter predlaga opredelitev, ki je bolj uravnotežena, čeprav še vedno povzema bistvo omenjenih zahtev. Avtonomija je ogožena, kadar gre za *slepo* zaupanje in kadar prelaganje kognitivnih nalog na zunanje dejavnike na relevantne in netrivialne načine zmanjšuje intelektualno samodoločitev. Povedano na drug način, osebna avtonomija kot vrlina zahteva, da se oseba opira na druge

² Ang.: *Self-Reliance*.

in prelaga kognitivne naloge na zunanje dejavnike, tako da pridobiva znanje in druge epistemske vrline »do točke, ko bi to početje ogrozilo njene sposobnosti samodoločitve« (prav tam). Na tako pojmovanje osebne avtonomije kot nekakšne vrline samoregulacije, ki ni binarna (vse ali nič), ampak zaseda različna mesta na kontinuumu, se opirava tudi v knjigi, predvsem v zadnjih dveh poglavjih.

Naj zdaj na kratko predstaviva posamezna poglavja knjige.

V poglavju *Odločanje: od klasičnih do sodobnih pogledov* najprej na kratko predstaviva splošno razumevanje odločanja, v treh sledečih razdelkih pa še nedavno zgodovino nekaterih ključnih pogledov na odločanje in odločevalce.

V prvem razdelku oriševa temeljne ideje normativnih teorij, ki odločanje pojmujejo kot izrazito racionalen proces, v katerem naj bi bili odločevalci sposobni priti do vsaj približno optimalnih izbir. Herbert A. Simon (1955; 1956) je podal odmevno kritiko takšnega pogleda na odločanje in odločevalce, v kateri trdi, da moramo »globalno racionalnost ekonomskega človeka« zamenjati s podobo, ki odraža dejanske kognitivne zmožnosti odločevalcev, v zakup vzeti omejeno dostopnost relevantnih informacij in specifikke okolij, v katerih se odločevalci odločajo. Izhajajoč iz Simonove upravičene kritike, so empirične znanosti dokončno izpodbile idealizirano podobo odločanja, kot so jo predstavljale nekatere normativne teorije.

V drugem razdelku predstaviva odmeven program hevristik in pristranosti, ki sta ga zasnovala Amos Tversky in Daniel Kahneman (1971; 1974) in z njim vplivala na velik del sodobnih empiričnih pristopov k raziskovanju odločanja kot tudi na mnogotera sodobna razumevanja le-tega. Raziskovalni program poskuša pokazati, da se odločanje povečini odvija kot nezaveden oz. intuitiven proces, v katerem ključno vlogo igrajo hevristike – nekakšne nezavedne mentalne bližnjice³ –, ki v mnogih situacijah vodijo do sistematičnih napak presojanja (pristranosti) in za odločevalca neugodnih odločitev. Ob tem je treba omeniti, da program hevristik in pristranosti – vsaj v svoji zasnovi – ne zanika trditve, da sta intuitivno mišljenje in odločanje v splošnem uspešna (Kahneman, 2003).

V zadnjem razdelku oriševa teorije dvojnega procesiranja. Te v osnovi trdijo, da je potrebno razlikovati med dvema različnima načinoma oz. tipoma procesiranja: hitrimi, avtomatičnimi in neza-

3 Hevristike so sicer lahko tudi zavestne, a se program hevristik in pristranosti osredotoča na njihove nezavedne vidike.

vednimi procesi ter počasnimi, nadzorovanimi in zavestnimi oz. preudarnimi procesi. Znanosti o presojanju in odločanju, izhajajoč iz raziskovalnega programa hevristik in pristranosti, trdijo, da večino dela opravljajo procesi prvega tipa. Takšen pogled stoji v ozadju mnogih sodobnih raziskovalnih praks in razumevanj odločanja. Predvsem sodobnejše izpeljanke programa hevristik in pristranosti poskušajo takšno podobo odločanja in odločevalcev razumeti (in/ali raziskovati) kot trditev, da zavestni in samorefleksivni procesi v odločanju ne igrajo prav velike vloge.

Drugače kot opisani pogledi zagovarjajo stališče, da teza omejene racionalnosti ni v nasprotju s pogledom, da zavestni, metakognitivni in (samo)refleksivni procesi igrajo pomembno vlogo v odločanju ter so bistveni za ohranjanje in udejanjanje osebne avtonomije. Poudariti bi želela, da se strinja s kritikami idealizirane podobe odločanja, ki so jo podajali normativni modeli in teorije, in nikakor nočeva trditi, da vedenjske znanosti o odločanju ne prinašajo mnogih zelo zanimivih kot tudi uporabnih spoznanj o človeškem vedenju in odločanju (glej npr. poglavje *Avtonomija v svetu spodbud*). Po drugi strani želiva izpostaviti, da so sodobni pogledi in načini raziskovanja odločanja, ki izhajajo iz predstav programa hevristik in pristranosti ter njegovih sodobnih izpeljank, pomanjkljivi: zanemarjajo namreč nekatere ključne lastnosti odločanja, ki se jasneje izrazijo, če opazujemo odločanje v okoljih vsakdanjega življenja in resno upoštevamo vprašanje, kako odločevalci sploh razumevajo okolja, v katerih se odločajo, ter sebe kot odločevalce. Sicer je res, da vedenjske znanosti o odločanju (in tudi ekonomija) dandanes vedno bolj uporabljajo tako imenovane eksperimente v naravnih okoljih⁴ (npr. Al-Ubaydli in List, 2019). Na ta način se dotikajo nekaterih vidikov odločanja v vsakdanjih okoljih in s tem spoznavajo zelo zanimive povezave med določenimi okoliščinami in vedenjem oz. odzivi odločevalcev. Dejstvo pa je, da s tem ne spoznavajo kognitivnih procesov, ki stojijo v ozadju takšnih ali drugačnih povezav. S preučevanjem vedenja množic, ki predstavlja temeljni način raziskovanja vedenjskih znanosti, iz podobe odločanja izpuščajo posameznega odločevalca. Ne nazadnje se, z osredotočenostjo na vedenje, izognejo tistemu, kar jih je pravzaprav v resnici zanimalo: kako odločevalci s svoje perspektive razumevajo in doživljajo odločanje. Na to nakazuje tudi opis »Ciljev in obsega« znan-

4 Ang.: *natural field experiments*.

stvene revije *Nature Human Behaviour*, ki spada pod krovno revijo *Nature*, eno od najvplivnejše revij v svetu znanosti: »[...] *Nature Human Behaviour* izdaja raziskave izrednega pomena, ki se posvečajo kakršnim koli vidikom individualnega ali kolektivnega človeškega vedenja. Kako ljudje zaznavajo, mislijo, čutijo, se odločajo in delujejo? [...]« (*Nature Human Behaviour, Aims & Scope*⁵). Kot vidimo iz navedka, revije ne zanima le vedenje, kot kaže prvi stavek, ampak tudi izkustvo; mišljenje, zaznavanje, občutenje in odločanje so namreč duševni⁶ pojavi, ki so bistveno in neodtujljivo zavestna, subjektivna stanja, ki jih zgolj s preučevanjem vedenja ne moremo razumeti. Nanje brez zavestnosti oz. izkustva sploh ne bi mogli pomisliti. Pomenljivo je tudi, da opis v eni sapi izenači vedenje in izkustvo – kar je kategorialna napaka, kjer stvar, ki pripada eni kategoriji, zmotno predstavimo kot stvar, ki v resnici pripada drugi.

V poglavju *Odločanje med laboratorijem in vsakdanjim življenjem* najprej podrobneje analizirava splošno sprejeto pojmovanje odločanja ter nekatere prevladujoče sodobne načine njegovega empiričnega raziskovanja, predvsem znotraj vedenjske ekonomije, psihologije in kognitivne nevroznanosti. Analiza pokaže, da je večina sodobnih pristopov k razumevanju in raziskovanju odločanja pomanjkljivih. Poudarjajo namreč le določene vidike odločanja, druge, prav tako ključne, pa izpuščajo – močno na primer zanemarjajo razne subjektivne vidike, kot so občutenje ali preudarjanje o lastnih ciljih, vrednotah in željah ali občutenje in razmislek o subjektivnih posledicah odločitev. Poleg tega se večinoma osredotočajo na preučevanje odločitvenih problemov v specifičnih laboratorijskih pogojih, ki nadalje sodoločajo omejeno razumevanje odločanja. V drugem razdelku zato analizirava nekatere odločitvene naloge, s katerimi so udeleženci soočeni v laboratorijskih raziskavah, ter jih primerjava z nekaterimi namišljenimi kot tudi raziskanimi odločitvenimi situacijami iz vsakdanjega življenja. Izkaže se, da prve zgrešijo ključno lastnost vsaj nekaterih odločanj, odločitev in njihovih udejanjanj v vsakdanjem življenju: odločitvene situacije (in odločitve) so za odločevalce bistveno in neločljivo vpete v to, kaj zanje pomenijo in subjektivno prinašajo. Sodobne znanosti o odločanju tako podajajo pomanjkljiv in v nekaterih vidikih napačen po-

⁵ <https://www.nature.com/nathumbehav/aims>

⁶ »Duševnost« uporabljava v smislu angleškega *mind*, ki se nanaša tako na bolj razumske kot na hotenjske in afektivne procese.

gled na odločanje. Če po drugi strani v razumevanje odločanja pristavimo to, kar nas v resnici zanima – kako se odločamo v vsakdanjem življenju –, razpremo več prostora za osebno avtonomijo, kot ga po najinem mnenju dopuščajo sodobni raziskovalni pristopi in pogledi na odločanje.

V poglavju *Vloga metakognicije v odločanju* na podlagi nekaterih raziskav vloge metakognicije v odločanju in nadaljnega teoretskega razmisleka pokaževa, da človeška sposobnost nanašanja na lastne duševne procese in stanja igra bistveno vlogo pri tem, kako se odločamo in delujemo. Omogoča nam na primer razmislek o preteklih in prihodnjih odločitvah, refleksijo želja, ciljev in vrednot ali občutenje raznih subjektivnih vidikov lastnih izbir. Človeška zmožnost nanašanja na lastno duševnost ter vpetost naših odločitev v širše cilje in vrednote, katerih refleksija je imanentna vsaj nekaterim vsakdanjim (morda pomembnim) odločitvam, sta vidika, ki sta bistvena tudi za udejanjanje avtonomije (kar prikaževa tudi v poglavjih *Avtonomija v svetu spodbud* in *Umetna inteligenca, algokracija in avtonomija*).

V prvem razdelku poglavja predstaviva različna pojmovanja metakognicije – v knjigi jo razumeva kot vse vrste nanašanja na lastne pretekle, sedanje ali prihodnje duševne procese in stanja in tako k njej prištevava metakognitivne presoje, prepričanja, znanje o lastni duševnosti kot tudi občutke. Obravnava metakognitivnih občutkov je v kontekstu preučevanja odločanja po najini presoji zastavljena, čeprav predstavlja zelo pomemben vir informacij o lastni duševnosti in s tem o odločanju. Z obravnavo metakognitivnih občutkov želiva poudariti, da tudi bolj eksplicitnih, preudarnih in racionalnih (meta)kognitivnih procesov ne razumevava ločeno od občutkov. Le-ti so po najinem mnenju značilni za mnoge načine nanašanja na lastno duševnost.

V naslednjem razdelku poglavja se na kratko ustaviva pri vprašanju vloge zavestnih procesov v odločanju. Če bi veljal pesimističen pogled na vlogo zavestnih (meta)kognitivnih procesov v odločanju, kot ga predstaviva na koncu zadnjega razdelka poglavja *Odločanje: od klasičnih do sodobnih pogledov*, bi morali sprejeti, da nanašanje na lastno duševnost v naših odločitvah ne igra pomembne vloge. Morali bi na primer privzeti, da so naša predvidevanja in prepričanja o subjektivnih posledicah odločitev zgolj nekakšni izmisleki, ki z odločitvami nimajo resnične povezave; da so naši občutki o lastnih duševnih stanjih in procesih, ki se nam porajajo med odločanjem, zgolj naključni utrinki, ki ničesar ne prispevajo k izbiram; da ima refleksija o preteklih ali prihodnjih odločitvah le vlogo ra-

cionalizacije že storjenih ali pričakovanih odločitev. Morali bi torej sprejeti, da samozavedanje, razumevanje in (s)poznavanje samih sebe pravzaprav ničesar ne prispevajo k odločanju – stališče, ki se zdi nepojmljivo. Seveda ne moremo zanikati tega, da smo močno določeni z lastnim fizičnim ustrojem, okoljem in z raznovrstnimi nezavednimi mehanizmi, na katere včasih nimamo ne vpliva niti zavestnega dostopa. Ne glede na to pa nimamo ne empiričnih dokazov ne dobrih argumentov za trditev, da zavestno delovanje in izkustvo ničesar ne prispevata k temu, kaj pomeni biti človek, odločevalec in kako ter kakšne odločitve sklepamo. To postane še toliko bolj jasno, če v zakup vzamemo bolj posredne in v času zamaknjene vplive zavestnega nanašanja na lastno duševnost.

V naslednjih dveh razdelkih podrobneje predstaviva in analizirava raziskave, ki kažejo na pomembno posredno in v času zamaknjeno vlogo metakognicije v odločanju. Na primeru raziskav o izvedbenih namerah⁷ pokaževa na pomembnost zmožnosti nanašanja na prihodnje odločitve in dejanja, na primeru raziskav o pričakovanih čustvih pa na pomembnost nanašanja na prihodnja čustvena stanja, povezana s potencialnimi izbirami.

V zadnjem razdelku poglavja obravnavava vlogo širših ciljev in vrednot ter človeške sposobnosti, da se med odločanjem nanašamo nanje. Trdiva, da cilji in vrednote odločevalcev predstavljajo okvir, širšo strukturo, v katero so posamične odločitve vpete, nanašanje nanje pa je za odločanje oz. vsaj nekatere odločitve v vsakdanjem življenju bistvenega pomena. Vsaj včasih je ključno, da jih poznamo, smo jih sposobni občutiti ali o njih tvoriti prepričanja, da sploh lahko pridemo do odločitev, ki so vsaj približno v skladu z nami kot osebami. Le tako se lahko vsaj do določene mere odločamo in delujemo v skladu s samimi sabo – kot osebe, ki posedujejo določeno mero osebne avtonomije. Čeprav so naše širše želje, cilji in vrednote odvisne od mnogih dejavnikov, kot so kulturno okolje, vzgoja ali pretekle izkušnje, pa je njihovo (s)poznavanje ključen del odločanja.

Znanosti o odločanju nasprotno še vedno privzemajo pogled, da ljudje nimamo prav veliko vpliva na lastne odločitve, ki so – pa naj si bodo sprejete preudarno ali intuitivno – podvržene najrazličnejšim pristranostim in vodijo do neugodnih izidov ali subjektivnih posledic. Na podlagi takšne podobe odločanja se mnogi raziskovalci

⁷ Ang.: *implementation intentions*.

zavzemajo za vpeljavo specifičnih ukrepov, ki bi posameznikom in/ali družbi s pomočjo spreminjanja odločitvenih okolij omogočile boljše izbire.

V poglavju *Avtonomija v svetu spodbud* razmisliiva o tem, ali odločitvene spodbude, vse bolj prisotne dobronamerne spremembe odločitvenih okolij, spodkopavajo avtonomijo.

V prvem razdelku najprej predstaviva program odločitvenih spodbud, ki poskuša na podlagi izsledkov vedenjskih znanosti o odločanju ljudem pomagati do boljših izbir, ne da bi jim pri tem krnil svobodo izbire. Zagovorniki programa na primer trdijo, da se ljudje neradi aktivno odločajo, ko pa se, se mnogokrat ne odločajo sebi v prid. Kot enega izmed bolj učinkovitih načinov uravnavanja odločitev in vedenja tako predlagajo spremembo privzetih pravil oz. izbir (npr. namesto, da se moramo aktivno odločiti za varčevanje, to postane privzeta oz. začetna izbira, od katere lahko kadarkoli odstopimo). Program odločitvenih spodbud tako predlaga, da v odločitvena okolja vgradimo takšne privzete oz. začetne izbire, ki ne zahtevajo preudarka ali aktivnega odločanja, a hkrati prinašajo ugodnejše izide kot izbire v okoljih, kjer bi se morali aktivno odločati.

V naslednjem razdelku razmisliiva, ali dobronamerne odločitvene spodbude na kratki rok krnijo osebno avtonomijo v štirih dimenzijah, ki jih obravnavava: svoboda izbire, hotenjska avtonomija, racionalna dejavnost in odsotnost nadvlade. Ugotoviva, da so skrbi povečini odveč, če predpostavimo vsaj približno dobronamerne arhitekta izbire, ki podpira odkrite, transparentne spodbude in stremi k blaginji posameznika ali družbe. To drži posebej, če si predstavljamo spodbude, ki so odprte za racionalen razmislek, refleksijo in širšo družbeno razpravo. Vseeno meniva, da je potrebno v kontekstu uporabe spodbud najti pravo razmerje med tistimi, ki omogočajo izogibanje raznim pristranostim in neugodnim odločitvam, ter takšnimi, ki spodbujajo in negujejo naše sposobnosti aktivnega odločanja, razmisleka in samorefleksije. Skleneva, da je program odločitvenih spodbud v tem pomanjkljiv, saj večinoma ne uporablja spodbud, ki bi vzpodbujale slednje – razlog tiči v njegovem relativno ozkem razumevanju odločanja in človeške duševnosti nasploh.

V zadnjem razdelku izpostaviva še en manko mnogih sodobnih razprav o vprašanju osebne avtonomije v kontekstu takšnih ali drugačnih posegov spreminjanja odločitvenih okolij. Porast namerne spreminjanja odločitvenih okolij namreč odpira potencialen prostor, v katerem spremembe odločitvenih okolij na dolgi rok

spreminjajo same odločevalce ter s tem do določene mere spodkopavajo avtonomijo. Tukaj razmisliiva o možnosti, da svet, poln spodbud, ki ne podpirajo aktivnega odločanja ali prakticiranja (samo)refleksije, potencialno krni našo sposobnost avtonomnega odločanja. Razmisleke utemeljiva na drugačnem pogledu na človeško duševnost, kot so ga vajene discipline kognitivne znanosti, kot so psihologija, nevroznanost ali vedenjska ekonomija, na katerih slonijo razni posegi spreminjanja odločitvenih okolij. Izhajava predvsem iz ideje Iana Hackinga (npr. 1995a), da se duševni (kot tudi družbeni) pojavi bistveno odzivajo – in zato spreminjajo – na to, kako jih razumevamo, opredeljujem ali spodbujamo. S tem ne želiva kritizirati v vsaj nekaterih primerih jasne uporabnosti dobronamerne spreminjanja odločitvenih okolij. Vseeno meniva, da je potrebno predpostavkam o duševnosti, na katerih temeljijo sodobne znanosti o odločanju in posegi spreminjanja odločitvenih okolij, nameniti kritični razmislek.

Ne nazadnje ta drugačen pogled na človeško duševnost razkriva še eno ključno lastnost odločanja, osebne avtonomije in nasploh tega, kar smo kot ljudje: neodtujljivo vpetost v okolja, s katerimi smo v vsakdanjem življenju v nenehnih interakcijah, bistveno sodeloča to, kako razumemo sami sebe, kako se odločamo in kako sploh razumemo ter udejanjamo svojo avtonomijo.

V zadnjem poglavju *Umetna inteligenca, algokracija in avtonomija* se ukvarjava z vprašanjem, kako orodja umetne inteligence vplivajo na odločevalce in njihovo avtonomijo.

V prvem razdelku na kratko predstaviva temeljne značilnosti in razlike med dvema valoma umetne inteligence. Poudariva, da prvi val, ki temelji na računski metafori (mišljenje je računanje), kot izhodišče jemlje deduktivno logiko in reprezentacijsko teorijo duševnosti. V drugem valu pa gre predvsem za razvijanje orodij, ki temeljijo na induktivni logiki. Modeli ne delujejo po določenih, vnaprej danih pravilih, ampak se na podlagi algoritmov učenja učijo iz množice izbranih podatkov in napovedujejo rešitve na podlagi prepoznavanja vzorcev.

V naslednjih dveh razdelkih izpostaviva dve nezaželeni posledici takega pristopa: netransparentnost in pristranosti. Netransparentnost se nanaša na način, kako je sistem prišel do rešitve naloge. Sistem je za človeka pogosto nekakšna črna škatla, zato je postopek za človeka nerazumljiv in nerazložljiv. Posledica je, da človek rezultat sprejema na zaupanje, ki je sicer lahko utemeljeno na izpričano zanesljivem delovanju sistema, ni pa mogoče izključiti namernih manipulacij, ki jih uporabnik le težko odkrije. Še posebej,

ko gre za visoko tvegana področja, kjer so v igri človekove pravice (npr. pri uporabi prepoznavanja obrazov), je treba biti zelo pazljiv, uporabo regulirati in jo morda tudi prepovedati. Pristranosti v sistemu pa so največkrat posledica načina izbora podatkov. Računalniškemu odločanju se pogosto pripisuje objektivnost in nepristranost, izkaže pa se, da temu še zdaleč ni tako. Podatki, na katerih se sistem uči, niso preprosta *data* (surovi podatki), ampak so *capta* (podatki izbrani iz celote vseh možnih podatkov), ki odražajo stališča in prepričanja razvijalcev in naročnikov ter so vpeta v družbeni kontekst. Pristranosti se tako prenesejo v sam sistem in vplivajo na naslednje odločitve, kar posledično le še krepi pristranosti in lahko vodi do nepravilnih in etično spornih rešitev. Cathy O'Neil (2016) je take sisteme poimenovala »orožja matematičnega uničenja«.

V naslednjem razdelku predstaviva pojem algokracije Johna Danaherja (2020), s katerim označuje vladavino algoritmov. Primerov, ko se v vsakdanjem življenju prepustimo avtoriteti algoritmov, je veliko, a se med seboj pomembno razlikujejo po tem, kakšna je delitev dela med algoritmom in človekom ter v kateri fazi odločanja človek prepusti delo algoritmom. Predstaviva logični prostor možnosti, ki kaže na kompleksnost relacij (odnosov), nato pa ugotavlja, kakšen je vpliv algokracije na štiri dimenzije avtonomije, na katere se nanašava tudi v poglavju *Avtonomija v svetu spodbud*. Pokazuje, da zaradi pomanjkanja transparentnosti in razložljivosti, podatkovnega fundamentalizma, mita o objektivnosti pri izboru metod analize in algoritmov, nadzora ter predvsem manipulacije, v algokratskih sistemih lahko prihaja do zmanjševanja avtonomije.

Poglavje skleneva z razpravo o razširjeni kogniciji, kjer pojem kognicije vključuje tudi dele okolja. Opozarjava, da je treba biti pri načrtovanju in uporabi pametnih orodij previden, saj pretirana in dolgotrajna uporaba lahko tako oslabi duševne sposobnosti, da pride do »kognitivne atrofije«. Takrat uporabnik sam ni več zmožen samostojnega delovanja, zapade v neke vrste intelektualno neobogljnost in tako drastično zmanjša svojo avtonomijo. Meniva, da bi ob pazljivi uporabi sistemi umetne inteligence lahko okrepili kognitivne sposobnosti ljudi ter pripomogli k več znanja in boljšemu odločanju, a hkrati se moč, ki jo imajo lastniki in upravljalci, lahko hitro zlorabi v druge namene (npr. čim večji dobiček in nadzor nad uporabniki). Gre za neke vrste tehnodružbeni inženiring, kjer oblikovanje okolja navaja ljudi na to, da se vedejo kot preprost stroj.

Ne glede na to, da knjiga morda odpira več vprašanj kot daje odgovorov, misliva, da je avtonomija, čeprav omejena, pomembna vrednota, ki se ji ne bi smeli kar tako odreči.

*

V knjigi črpava iz svojih preteklih razmislekov in raziskovanj. Poglavlji *Odločanje: od klasičnih do sodobnih pogledov* in *Vloga metakognicije v odločanju* predstavljata posodobljen izsek dela doktorske disertacije Tomaja Strleta *Metakognicija kot povezovalac mišljenja in intuicije v procesu odločanja* (Strle, 2015), poglavje *Odločanje med laboratorijem in vsakdanjim življenjem* pa posodobljen izsek članka Tomaja Strleta »Odločanje: od laboratorija do resničnosti vsakdanjega življenja« (Strle, 2016a). V poglavju *Avtonomija v svetu spodbud* deloma izhajava iz razmislekov, ki sva jih razvila v prispevkih »Looping effects of neurolaw and the precarious marriage between neuroscience and the law« (Strle in Markič, 2018), »Looping minds: How cognitive science exerts influence on its findings« (Strle, 2018) in »Podoba omejene racionalnosti in povratni učinki spreminjanja odločitvenih okolij« (Strle, 2019). Poglavlje *Umetna inteligenca, algokracija in avtonomija* deloma temelji na prispevkih Olge Markič »Prvi in drugi val umetne inteligence« (Markič, 2021) in »Extended Cognition: Feedback Loops and Coupled Systems« (Markič, 2017).