Ненад Н. Цекић[1]
Универзитет у Београду, Филозофски факултет,
Одељење за филозофију
Београд (Србија)

17.023
Оригинални научни рад
Примљен 30/06/2024
Прихваћен 21/08/2024
doi: 10.5937/socpreg58-51894

ПОБОЉШАЊЕ ЧОВЕКА И МОРАЛ: НЕКЕ ТЕОРИЈСКЕ НЕДОУМИЦЕ[2]

Сажетак: Аутор у раду преиспитује кључне етичке замисли које прате идеју да се човек може „морално побољшати“ утицајем на „морални мозак“. Анализирајући главни приступ савремених неуроетичара, запажа да су у идеји побољшања углавном изостављени нормативно-етички и метаетички елементи без којих јасна представа о томе шта се под „моралним побољшањем“ може сматрати – није могућа. Аутор скреће пажњу и на претерано ослањање на технолошке поступке у процени одређења и домета моралног побољшања. На крају рада анализира два модела функционисања „моралног мозга“. Та два модела преузета су из поп-културе. Један је модел двоструке личности др Џекила и господина Хајда, преузет из књижевности и каснијих филмских адаптација. Други је чувени компјутер који има личност, ХАЛ 9000 из чувеног филма *Одисеја у свемиру 2001*.

Кључне речи: етика, неуроетика, морално побољшање, морални мозак, модели

Бројна савремена мултидисциплинарна истраживања, која се означавају као „биоетичка“ и „неуроетичка“, сасвим сигурно обухватају и проблеме који се односе на тзв. морално побољшање. Могло би се рећи чак и то да су радови који се тичу (различитих врста) „побољшања човека“, све до најновијег успона „етике вештачке интелигенције (АИ)“, доминирали биоетичком сценом. Притом, у тим истраживањима, подразумева се да „биоетика“ и „неуроетика“ (као грана биоетике) морају имати неке везе са (филозофском) „етиком“ као матичном дисциплином. Ипак, однос етике, као

[1] ncekic@f.bg.ac.rs; 0000-0001-7823-3531

[2] Овај текст је настао на основу излагања „Human Enhancement and Morality: Some Doubts“, 10th World Conference on Bioethics, Medical Ethics, and Health Law, Jerusalem, UNESCO Chair in Bioethics, January 6-8, 2015, https://www.sicp.it/wp-content/uploads/2018/12/146_UNESCO%20Chair%20in%20Bioethics%2010th%20World%20Conference.pdf

„чисте“ (појмовне) филозофије морала, и биоетике и неуроетике, као специфичних истраживања која обухватају и емпиријске и неемпиријске дисциплине области истраживања, чак ни до данас није јасно одређен.

БИОЕТИКА И ПРИМЕЊЕНА ЕТИКА: ИСТОРИЈА

Разјаснимо сада како је ова неодређеност разграничења разноликих „етика“ настала. Историјски гледано, класични проблеми примењене етике и биоетике седамдесетих година прошлог века били су проблеми абортуса, еутаназије, третмана животиња,[3] очувања животне средине и, донекле, ратовања и пацифизма. Ова, у то доба, тематски прилично одређена област касније је готово неприметно допуњавана широком палетом сродних, а понекада и само наизглед сродних, расправа. Током времена, биоетичке расправе су се једним делом односиле на текућу медицинску праксу (оправданост употребе одређених лекова или процедура), а затим и на различите друштвене „политике“ и правну регулативу која се тицала не само питања живота и смрти већ и *квалитета* живота (рецимо, еколошке „политике“ и проблеме третмана терминално болесних, старих и немоћних особа).

Сада долазимо до проблема јасног разграничења дисциплина. Свака „етика“ припада филозофији, а филозофија је нужно теоријска интелектуална делатност. Међутим, сами биоетичари данас отворено избегавају одређење појма „биоетике“, и то баш због нужности употребе „нејасног појма теорије“ (Arras, 2016). Наводно, када би се биоетика јасно дефинисала као теорија, постала би компликована и непрактична. Ова тврдња несумњиво представља чудноват став јер је свака етика, по дефиницији, део „филозофије морала“. Међутим, свака филозофија подразумева теоријски приступ предмету истраживања. Поред тога, са проширењем обима и граница биоетичких дебата и реинтерпретација самог концепта „биоетике“, постало је нејасно шта све биоетика обухвата или може да обухвати. Добар пример је већ споменута неуроетика. Она је данас толико „у интердисциплинарном тренду“ да је немогуће јасно одредити њен однос према три испреплетене дисциплине – (био) етици, биологији и медицини.

БИОЕТИКА, НЕУРОЕТИКА И МОРАЛНО ПОБОЉШАЊЕ

За разлику од биоетике, неуроетика се као област истраживања јасно дефинише. Она је као дисциплина први пут формално одређена на конференцији „Неуроетика: мапирање области“, одржаној у Сан Франциску 2002. године.[4] На њој је амерички новинар Вилијам Сафајр (William Safire) понудио следећу дефиницију: неуроетика је „испитивање онога што је исправно и погрешно, добро и лоше у вези са третманом,

[3] Ово су била класична питања којима се бавила Сингерова „канонска“ књига *Практична етика*, чије се прво издање појавило још 1980. године. B. Singer [1980, 1993], 2011.

[4] Више на: Neuroethics: mapping the field: conference proceedings, May 13-14, 2002, San Francisco, California, Marcus, Steven, Charles A. Dana Foundation.

побољшањем или нежељеним инвазијама и забрињавајућим манипулацијама људског мозга“ (Safire, 2002)[5].

Како сада ствари стоје са конкретном дефиницијом „моралног побољшања“ које спада и у домен неуроетике? И поред недостатка општег консензуса у погледу тога шта је то морално побољшање, у литератури се ипак могу пронаћи нека његова одређења. Једна од радних дефиниција моралног побољшања гласи: „Морално побољшање: неко технолошко или фармаколошко средство којим се утиче на биолошки аспект моралног функционисања, како би се поспешило оно што је пожељно или уклонило оно што је проблематично“ (Wiseman, 2016, str. 6). Ова дефиниција на први поглед изгледа и штуро и неинформативно. Међутим, Вајзмен (Wiseman), који ју је и коначно формулисао, сматра да то и мора бити тако. Разлози за неодређеност појма „морално побољшање“ јесу следећи: (а) не постоји јединствена ствар која се одвија под именом моралног побољшања, и (б) погрешно је представити читав распон смислених могућности унутар једног домена [моралног побољшања] путем настојања да се он комбинује унутар неке јединствене појмовне парадигме (Wiseman, 2016, str. 7).

Чини се, дакле, да у случају моралног побољшања није реч само о бесконачном броју технолошких могућности. Бесконачан је и број могућих значења израза „бити морално добар“, од чијег значења зависи дефиниција „побољшања“. Притом, израз „морално добар“ овде није употребљен само у својој примарној *метаетичкој*[6] *функцији вредновања већ такође служи и као бар делимични опис тога шта је то „побољшање“*. То значи да ће без пружања јасног општег и уверљивог теоријског појмовног оквира, општи термин „побољшања“ бити једнако нејасан и у делимично емпиријским дисциплинама као што су психологија или социологија, па чак и у помодној „неуронауци“.

Дакле, потребна нам је претходна појмовна филозофска анализа значења термина „морално побољшање“, али је она у литератури о побољшању реткост. Чини се да разлог за потенцијалну концептуалну конфузију у погледу тога шта стварно значи „бити морално добар“ или „побољшан“, лежи у једној неизреченој метаетичкој (дакле: филозофској) претпоставци. Биоетичари, а нарочито поборници побољшања, напросто подразумевају да *постоји* јединствен одговор на питање „Шта је то морал?“, а самим тим и на питање „Шта је то морално боље?“. Техничким филозофским речником речено, „неуроентузијасти“ (назовимо их тако по промовисаном помодном префиксу „неуро“), који сматрају да је непосредно вештачко побољшање људског *моралног* понашања могуће, морају бити припадници неке струје *метаетичког когнитивизма*.

[5] Слична дефиниција се налази у *Encyclopaedia Britannica*, па се Сафајрова дефиниција може сматрати класичном.

[6] Без улажења у техничке филозофске термине, овде назначавамо само то да се метаетика у најширем смислу бави „логиком језика морала“. В. Hare, 1963, стр. 97.

ПОБОЉШАЊЕ И МЕТАЕТИКА: ИМПЛИЦИТНИ КОГНИТИВИЗАМ

Централна теза традиционалног когнитивизма лежи у тврдњи да морални судови могу бити истинити или неистинити на исти начин на који су то емпиријски судови неке посебне науке. Зашто су поборници биоетичког побољшања нужно метаетички когнитивисти? Напросто зато што побољшање *било које* људске способности претпоставља објективно знање о тој способности. Дакле, знање о побољшању људских моралних способности подразумева и знање о томе шта израз „бити моралан" дескриптивно значи, подразумева, или бар обухвата. Међутим, „аналитички" или „дефиниционистички" когнитивизам имплициран оваквим уверењем у метаетици је одбачен још почетком 20. века, на основу Муровог (Moore) чувеног „аргумента отвореног питања"[7] (Moore, 1903, §10–14; str. 39–43; уп. Cekić, 2013, str. 85–92).

Расправа између метаетичких когнитивиста и некогнитивиста још увек се води, али нико данас не сматра да се морални термини могу једноставно дефинисати путем термина који означавају „природна својства". Међутим, иако спада међу најважније и дуговечније етичке расправе, овај класични филозофски спор у савременим биоетичким расправама чак се ни не констатује, а камоли узима у разматрање. Ова чињеница указује на то да су биоетичари и „етичари побољшања" напросто претпоставили да је метаетика завршила своју мисију, тј. да је спор у погледу могућности моралног знања већ разрешен. А то, наиме, није истина. Треба само погледати новије обухватне прегледе који се тичу ове области па видети да су когнитивисти и некогнитивисти још увек далеко од консензуса у погледу тога да ли морал нужно обухвата неко знање или се може свести на пуку експресију осећања или емоционалног става.[8] На овом трагу, Вајзмен упозорава „неуроентузијасте": „Не постоји морално побољшање *само по себи*" (Wiseman, 2016, str. 203).

МОРАЛНО ПОБОЉШАЊЕ И НОРМАТИВНА ЕТИКА

Други разлог за потпуну (филозофску) неразјашњеност употребе термина „морално побољшање" лежи у томе што он очигледно несистематски комбинује елементе свих познатих нормативно-етичких приступа: деонтологије, консеквенцијализма и етике врлине. И овде наши „неуроентузијасти" претпостављају нешто о чему филозофска расправа и даље траје. Филозофија морала, ако ништа друго, упозорава да смо још увек далеко од неке нормативне теорије за коју бисмо могли

[7] Мур заправо каже да сваки покушај дефинисања вредносних својстава путем позивања на нека природна својства води у грешку. Он наводи баш појам „добро", за који показује да свака понуђена дефиниција, нпр. утилитаристичка: „Добро је највећа могућа срећа највећег броја људи", допушта постављање „отвореног питања". Ова дефиниција се може оспорити простим питањем: „Да ли 'добро' означава 'највећу могућу срећу' највећег броја људи?" Сваки покушај дефинисања вредносних термина путем дескриптивних, јесте „натуралистичка грешка".

[8] Чак и елементарни прикази метаетике у овом погледу сасвим су јасни (нпр. Fisher, 2014).

казати да је потпуно „адекватна“ и практички применљива. Штавише, сам избор нормативне позиције одређује која би својства човека и друштва као целине могла бити предмет побољшања:

> „За консеквенцијалисту би примарна брига у погледу моралног побољшања могла бити само мало драгоценија од одмеравања међусобног односа свеукупне спречене штете у односу на различите трошкове које оно обухвата. За некога заинтересованог за мотиве, морално побољшање може се проценити мером којом оно доприноси да људи пожеле да буду бољи, или да открију боље разлоге да се буде моралан. За поборника врлине, морално побољшање морало би да допринесе формирању 'моралних навика', генеришући величину моралног раста – трајне трансформације личности или карактера делатника о којем је реч. Нажалост, преиспитивање свих предлога који би се овде изнели из перспективе сваке од ових теорија створило би књижурину толико велику да не би била ни од какве користи.“ (Wiseman, 2016, str. 8)

Сада се може разјаснити зашто неодређеност појма биоетике представља проблем за поборнике моралног побољшања. Наиме, оптимистична идеја из седамдесетих година 20. века према којој примењена етика и биоетика као свеобухватна теорија могу да добију јасно утемељење, била је кратког даха. Разлог је био више него једноставан: теоријске дебате, попут оних између утилитариста и кантоваца, немогућном су учиниле непосредну „примену“ биоетике зато што је прво било неопходно одбранити одабрану општу нормативну позицију. Међутим, проналажење „адекватне“ нормативне теорије историјски је подухват који ни до данас није завршен.

Ипак, упркос замршености класичних нормативних питања, са филозофске тачке гледишта, готово несхватљив преокрет збива се после периода који се у литератури данас назива такозваним херојским данима биоетике. Покушаји да се пронађе чврст „темељ“ биоетике (као неопходан оквир за било коју „неуро“ етику) једноставно су напуштени. Такво одрицање од јединствености моралног објашњења обично се оправдава чињеницом да „велике“ (или, како се то обично назива, „високе“) теоријске дебате преоптерећују решавање свакодневних моралних питања. Тако се, на пример, у литератури може пронаћи и процена да нека „висока теорија“ (сама филозофија?) „клиничке етичаре“ или „етичаре који раде на јавној сцени“ заправо ремети у свакодневним напорима да реше практична морална питања. Према тим ставовима, сувише је захтевно да се „практични етичари“ руководе основним принципима високог ранга, као што је Кантов категорички императив или нека варијанта утилитаризма. Примењена или практична (самим тим и „био“ или „неуро“) етика треба да се „прилагоди пракси“. Суштински, овај став подразумева да објашњења која обећавају „високе“ (нормативне) теорије у практичним контекстима могу бити *непотребна* (Magelsen et al., 2016). Поред тога, основне традиционалне нормативне теорије могу, у принципу, да оправдају различите међусобно супротстављене политике на конкретном нивоу; стога, према овом образложењу, оне ни не могу бити „практичне“ (B. Gutman & Tompson, 1996). Дакле, преовлађујући утисак о великом делу неуроетичких теза заправо гласи: „У 'биоетици моралног побољшања', етику је најбоље – заборавити“.

„МОРАЛНИ МОЗАК“: ШТА ЈЕ ТО И ШТА С ЊИМ?

Однедавно смо сведоци жестоке дебате о побољшању које би се директно дотицало лоцирања тзв. моралног мозга унутар биолошке структуре целине људског мозга. Претпостављени разлог зашто уопште треба тражити физичку локацију „моралног мозга“ прилично је једноставан. Када физички лоцирамо „морални мозак“, знаћемо стварно „место“ (у телу) где би, с циљем замишљеног побољшања, требало спровести неку интервенцију (биолошку, хемијску, медицинску итд.). Овде је суштински битно имати на уму да је физичка локација „моралног мозга“ унутар људског тела кључна примарно за „практичне“ неуроетичаре, али да они пречесто нису филозофи. Насупрот томе, овакво лоцирање је готово небитно за појмовно оријентисане „чисте“ филозофе.

Погледајмо о чему је овде тачно реч. Ми, наводно, можемо да лоцирамо „морални мозак“. Теза да се „морални мозак“ може физички лоцирати међу „неуроентузијастима“ схвата се веома озбиљно, па зато они нуде и мноштво конкретних предлога.[9] Наравно, овде није реч само о физичкој локацији „моралног мозга“. Наиме, скоро сви поборници моралног побољшања сматрају да морал мора бити заснован на нечему „емотивном“, па самим тим и на можданим центрима задуженим за обраду различитих осећања. Ево примера таквог резоновања: „...сазнајно морално расуђивање мора да ангажује области мозга повезане са емоционалном свешћу. Једна вероватна структура и чест актер у fMRI[10] студијама моралног сазнања јесу Бродманове (Brodmann) области. На њих се указује у студијама емоционалне интроспекције“ (Satpute et al., 2013).

Претпоставка о емотивној основи морала унапред искључује „очигледне кандидате“ за локацију „моралног мозга“, као што су то класични центри за језик и моторичке области. Међутим, они су из њих обично намерно изузети зато што сами услови експеримента обухватају исту врсту моторичког или језичког одговора, па се чине ирелевантним. Зато се дешава да „када на крају студије о fMRI прочитамо листу области мозга, она често збуњује. На крају крајева, ми желимо функционално разлагање на делове, које би нам рекло шта свака од ових области ради. Не би требало да се ослањамо искључиво на неуронауку да бисмо такво разлагање постигли“ (Prinz, 2016, str. 68). Другим речима, за разумевање резултата fMRI снимања потребна је додатна интерпретација и анализа.

„НЕУРОЗАСИЋЕНОСТ“?

Чини се да су шаренкасте слике скенираног мозга заиста распрострањене у емпиријској евиденцији која се нуди у литератури о моралном побољшању. Међутим, стижу назнаке да се и ова неуро-помодност приближава крају. На пример, јављају се сугестије да је претерано стављање културног акцента на слике добијене методом

[9] На пример, можемо пронаћи експлицитне тврдње да је примарна локација „моралног мозга“ у „вентромедијалном префронталном кортексу“ (нпр. Liao, 2016, str. 2–13; 32).

[10] Скраћеница fMRI (od engl. functional Magnetic Resonance Imaging) означава „прављење слика путем магнетне резонанце“. То је техника скенирања активности мозга уз помоћ јаких магнетних поља.

fMRI створило „неуро-умор“ у јавности која више није импресионирана таквим дискурсом. Стога је важно да се онима којима ограничења ових студија нису у потпуности објашњена треба „на увид ставити колико је варљива реторичка моћ коју такве слике имају, различите неуспехе таквих студија и, коначно, њихову неприкладност као основу за мерење и потенцијално манипулисање моралним феноменима“ (Struthers and Schuchdardt, 2013).

„Неурозасићеност“ показује да редуктивна емпиријска анализа моралних феномена мора бити у спрези са адекватним општим филозофским и етичким објашњењем. Не може се напросто ван било каквог филозофског контекста тврдити да се „морални мозак“ налази „ту и ту“. Без дубље процене филозофске функције настојања редукције сложености морала на проналажење тачне локације „моралног мозга“, та потрага остаје без контекста. Наиме, судови „То је исправно, добро, лоше“ свакако семантички не зависе од локације „моралног мозга“. Штавише, чини се да се горљива потрага за физичком локацијом „моралног мозга“ ослања на филозофски став „Ја сам мој мозак“. Он има смисла колико и у историји филозофије јасно одбачени став „Ја сам моје тело“. Морал, па и морални напредак, једноставно се не повезују са замишљеним операцијама, лековима или било којом врстом медицинских интервенција. „Морално“ није морално зато што неки нервни импулс бива иницира из тачно одређеног „моралног“ мозга, већ на основу разлога који су у нормативној етици чисто појмовне природе.

Поред тога, поједностављени редукционистички приступ обесмишљава саму основу морала. Наиме, из ове перспективе следи да „ако смо морални или неморални … нужно следи да ме је ’мој мозак на то натерао’“ (Wiseman, 2016, str. 124). Вајзмен нам заправо каже да ако је морал ствар искључиво „моралног мозга“, онда идеја моралне одговорности нема смисла. Цела замисао је конфузна јер на све што урадим ме је некако натерао мој „морални мозак“. А ако ме је „натерао“, онда ни за шта нисам одговоран, па ни за шта ни не могу бити крив.

СОЦИЈАЛНИ И ПОЛИТИЧКИ АСПЕКТИ: МОГУЋНОСТ МОРАЛНЕ ДИСТОПИЈЕ?

Основна замисао „детектовања“ локације „моралног мозга“ јесте и теоријска и практична, али је свакако научна. Међутим, већина савремених јавних поборника моралног побољшања више подсећа на екстремне активисте фокусиране на пропаганду у корист прокламованог циља „побољшања“ него на научнике који траже истину. Овде се сусрећемо са озбиљним филозофским сметњама. Наиме, технолошки приступ замисли моралног побољшања претпоставља једну специфичну врсту филозофског детерминизма који теоретски априори поништава идеју личне слободе, али без ваљаног објашњења и без готово икаквих доказа. У питању је не само метафизичка већ и политичка слобода. Наиме, поборници моралног унапређења понекад представљају политичку слободу чак и као теоретски опасну и сметњу прогресу. Срећемо се и са захтевима да људе „под хитно“ треба побољшати. Ови захтеви заснивају се на мрачној слици о предстојећој апокалипси и „врхунској штети“ која нам предстоји (Persson & Savulescu, 2012, str. 49, 94, 127, 133). Персон и

Савулеску, вероватно најпознатији филозофски поборници моралног побољшања, у својим сценаријима надолазеће катастрофе пишу о „бермудском троуглу изумирања“ читавог људског рода. Странице овог троугла чине радикална технолошка моћ, либерална демократија и преовлађујућа „кратковида“ психологија морала (Savulescu, 2009; Persson & Savulescu, 2012).

Не може се јасно рећи која би од ове три стране „троугла изумирања“ била филозофски најпроблематичнија. Вероватно је то противљење „либералној демократији“ коју филозофија политике углавном узима као аксиом политичког поретка у оним друштвима у којима се идеја побољшања и родила. Спасавање човечанства по цену ограничења слободе (делимична суспензија „либералне демократије“?) несумњиво је тема која лежи далеко изван делокруга строго биоетичке дебате. Као што смо видели, потреба супротстављања претераној слободи као неопходности оправдава се неемпиријским пророчанством нејасне „крајње штете“ за коју нисмо припремљени. Укратко, човечанство се суочава са апокалипсом коју је изазвало сопственим делањем, тако да „нешто под хитно мора да се уради“. А то „нешто“ је – морално побољшање. Природно, хитност готово увек оправдава принуду: „Ако се безбедно морално побољшање икада развије, постоје јаки разлози да се верује да његова употреба треба да буде обавезна, као што су то образовање или флуорисање воде, будући да је најмање вероватно да ће они који би требало да их употребе бити склони да их користе. То јест, безбедно, ефективно морално побољшање било би обавезно“ (Persson & Savulescu, 2012, str. 174). Сада долазимо до помало парадоксалне ситуације. Није у питању само политичка слобода, ми морамо да се окренемо против сопствене природе. Наиме, ако смо „ми наш мозак“, онда морамо да га побољшамо упркос томе што је он сам по сопственој природи врло често природно *йроūив* тога.

Неки додатни аргументи требало би још више да забрину наше „неуроентузијасте“. Још од Канта (Kant) и Сиџвика (Sidgwick), а нарочито експлицитно кроз Ролсову (Rawls) политичку теорију, претпостављало се да „практичне“ (моралне, политичке итд.) теорије морају да испуне тзв. услов публицитета. Укратко, „за једну теорију мора бити могуће да је у свакој околности прихватамо и јавно је заступамо, а да тиме не прекршимо саму ту теорију“ (Bykvist, 2010, str. 95). Овај аргумент је важан због своје широке прихваћености и филозофске природе. Наравно, у практичном животу можемо занемарити било какву теорију и окренути се политици. Међутим, чак и овакав приступ проблему побољшања има огромне проблеме. Политички гледано, готово је невероватно замислити да би неко озбиљно покушао да наметне присилно морално побољшање као део свог програма: „Ниједан политичар не би јахао до изборне победе са слоганом ‘обавезно технолошко морално побољшање за све’. [...] Кад оставимо мисаоне експерименте по страни, ако говоримо о експлицитном отвореном моралном побољшању којим би се озаконила присилу или коришћење друштвених подстицаја помоћу батине и шаргарепе, ми једноставно причамо у празно“ (Wiseman, 2016, str. 79). Отворени пројекат државног моралног побољшања опште популације незамислив је у либералним државама.

ЈЕДАН ТЕОРЕТСКИ МОДЕЛ: ДР ЏЕКИЛ И Г. ХАЈД

Још увек неразјашњену веру у могућност технолошког моралног побољшања, нечега за шта се још увек не зна ни шта је тачно (шта је то „бити морално *боље*"), често прати вера у огромну моћ теоријског моделовања. Разлог за конструисање теоријских модела јесте огромна разноликост, а често и недоследност у човековом моралном делању. Да би се у потенцијалну збрку унео некакав ред, нуде се поједностављени теоријски модели, обликовани као наративи који обезбеђују релативно јединствен оквир за „обухватање" људских различитости. Према овој идеји, приликом тумачења података које су социолози, психолози или неуронаучници прикупили, требало би да користе „радне моделе" људских карактера. И присталице тезе о „моралном мозгу" такође привлачи замисао „моделовања људског понашања":

> „Морални мозак распаљује имагинацију и покреће машту многих људи, како лаика тако и научника. Морални инжењеринг је често понављана и популарна тема у научној фантастици, од *Др Џекила и господина Хајда* Роберта Луиса Стивенсона (Robert Louis Stevenson) и *Моралних бацила* француског *fin-de-siècle* аутора Алберта Робиде (Albert Robida) до *Паклене поморанџе* Ентонија Берџиса (Anthony Burgess). Чак је и познати шпански неуролог Сантјаго Рамон и Кајал (Ramon y Cajal) посветио једну од својих приповетки овој фантазији. Све ове приче засноване су на сличним сценаријима: безосећајни доктори претварају окореле криминалце у послушнике који, чим су подвргнути операцији, постају тупави и губе своје кључне способности, или обрнуто, неурохирурзи преиначују примерне грађане у немилосрдне ратнике који су после тога незаустављиви у чињењу окрутности". (Verplaetse, Braeckman & Schrijver, 2009, str. 1)

И заиста, најпознатији и илустративни модел из књижевне фантастике је прича *Др Џекил и господин Хајд*. Према овом наративу, у сваком од нас се крије и „промишљени" и „рефлексивни" др Џекил и импулсивни и „дивљи" господин Хајд (Crockett, 2008). С једне стране, анимална реактивност представља се као природна, мада претерано дивљачка (као код господина Хајда). С друге стране, људи поседују и суздржаност и способност далековидости, за које се уопштено претпоставља да изгледају као спорије и слабије (као код др Џекила).[11] Међутим, „јаки" г. Хајд, због своје плаховите природе, једноставно је склон преступима сваке врсте.

Проблем са овим наративом је то што се (неурофизиолошка!) природа „господина Хајда", према биоетичарима директно „уграђена у мозак", углавном карактерише изразито негативним *моралним* терминима. Уз то, чак и на први поглед, ова чувена прича о двострукој личности (др Џекил / господин Хајд) нужно је манихејска. Према њој, човек је по природи стешњен између чекића и наковња, без икакве могућности градације свог положаја. Међутим, неуроентузијастима ту ништа не изгледа проблематично. Само треба оснажити др Џекила и некако потиснути др Хајда. Ипак, различити контексти и двосмислености у стварном животу угрожавају ову причу.

[11] Ово истраживање не може се више наћи на интернет адреси коју је Вајзмен уврстио у своју библиографију, што је донекле необично. Међутим, сама прича је толико честа да оно јесте погодна илустрација тезе коју Вајзмен оспорава.

Само је минимална рефлексија потребна да би се видело да је понекад управо емоционалнореактивни „део“ (мозга, човека?) нужан за извођење аката огромне моралне вредности. И у литератури се може пронаћи јасна тврдња да „реактивни, емоцијама вођени одговори потребни за велике херојске моралне чинове (рецимо, уплитање у ток уличне пљачке, скакање у узбуркану воду да би се неко спасао од дављења, искакање пред гранату да би се спасили другови, истрчавање пред аутобус како би се дечја колица одгурнула на сигурно, помоћ и спречавање током силовања, итд., итд.), то јест најобичнији морални поступци, јесу пре свега домен нашег наводно злонамерног 'господина Хајда'“ (Wiseman, 2016, str. 24).

И заиста, постоје моменти када су одлучни, емоционално реактивни одговори неопходан услов за остварење добра. Шта се збива са вредносним оквиром „Џекил и Хајд модела“ када откријемо да понекад, у ствари, морамо да будемо наш г. Хајд како бисмо били најбољи у тим околностима? Пажљиви, неодлучни, хладнокрвни и фокусирани др Џекил није од помоћи у оваквим ситуацијама. Закључак је прилично једноставан: „Врло фундаментални проблем, о којем се није много расправљало у литератури о моралном побољшању, јесте то што су врсте особина или диспозиција које као да доводе до злоћудности или неморалности такође исте оне које не захтева не само врлина већ и било која врста општег моралног живота“ (Harris, 2011, str. 104).

Стога се чини да се идеја о моралном побољшању такође опире једноставном психолошком моделовању. „Свеобухватни“ модели моралног побољшања човекову стварност превише поједностављују, а њихови аутори чак превиђају и њихове очигледне мањкавости. Без намере да пренагласимо овај аргумент, поменимо и то да би др Џекил, на пример, могао да буде фокусиран на психопатски циљ, да постане „бољи“ криминалац од господина Хајда. Ова двосмисленост око одговарајуће улоге карактера у случају др Џекила и господина Хајда није изненађујућа јер, као што смо рекли, „побољшање“ нужно подразумева постојање неког јасног стандарда процене који није присутан. Међутим, стандард дескриптивног „бити морално бољи“ је појам који „неуроентузијасти“ и даље избегавају да експлицитно декларишу.

Чак и без даље анализе потребе и обима улоге „модела“ у неуроетици, сада можемо да укажемо на дубље чисто филозофско питање: Шта би морал могао да буде без одговарајућег концепта „зла“? Двоструког карактера оличеног у др Џекилу и господину Хајду одсликава двоструког људске природе која је способна и за добро и за зло.

ЈОШ ЈЕДАН ТЕОРИЈСКИ МОДЕЛ: ХАЛ 9000 – ПОГОРШАЊЕ КАО УСЛОВ МОГУЋНОСТИ ПОБОЉШАЊА?[12]

Замисао моралног побољшања путем непосредног медицинског или фармаколошког дејства на биолошко-неуролошке механизме задужене за емоције заправо имплицира став да само људи са својом специфичном биохемијом и морфологијом могу бити „побољшани“ (или, кад смо већ код тога, „погоршани“). Овај приступ

[12] Модел компјутера ХАЛ (HAL) 9000 који овде наводимо не би требало схватити као покушај доприноса сада помодној дискусији о AI етици. Он има илустративну функцију.

претпоставља да предност треба дати „чистом“ или „некогнитивном“ побољшању: „Подразумеваћу да су морално побољшање интервенције за које се може очекивати да ће појединца оставити са моралнијим (тј. морално бољим) мотивима или понашањем које би иначе имао. Користићу израз 'некогнитивно морално побољшање' да означим морално побољшање које се постиже тако што се (а) модулирају емоције; и (б) то се чини директно, тј. не тако што ће се побољшати когниција (нпр. њена прецизност)“ (Douglas 2013, str. 162).

Ове, као и многе друге идеје радикалних неуроентузијаста, нису само представљене без образложења већ су такође и потенцијално супротне неким основним филозофским увидима. Прво, оне *претпостављају* да су емоције центар морала. Међутим, у филозофији морала о томе не постоји тако недвосмислена сагласност. Вероватно је да би главни заступници главних етичких традиција – кантовци, већина консеквенцијалиста, па чак и неки етичари врлине – ову сугестију једноставно одбацили. Друго, ако се већ претпоставља да само људи могу да се побољшају, то нормативно, на основу замисли „бентамовских“ реакција на стимулусе, води у неки облик хедонистичког утилитаризма (тј. „класичног утилитаризма у свом чистом облику“). Треће, идеја слободе изоставља се из слике морала, што је постало готово уобичајено место за све поборнике моралног побољшања. Наиме, концепција чисте емотивности као основе морала сугерише да се само „осетљиве“ особе могу морално побољшати. Међутим, чак се и данас у филозофији замисао слободног избора сходно ваљаним разлозима, а не пука инстинктивна „осетљивост“ или шире „емоционалност“, обично сматра претпоставком специфично моралног делања.

Један део забринутости међу савременим неуроетичарима сада изазива „кошмарни сценарио“ који претпоставља да ће се морално побољшање наметати особама како би се оне претвориле у „моралне роботе“. Ти људски морални роботи чине добро не зато што то стварно желе или осећају да је то исправно, „већ само зато што су накљукани лековима или технолошки принуђени да то чине“ (Wiseman, 2016, str. 53). Иако се можемо сложити са Вајзменом да је овакво „морално роботизовање“ људског рода мало вероватно, тешко је замислити како би овај аргумент оставио неки утисак на одушевљене „неуроентузијасте“.

За крај, у светлу идеје употребе разноразних „наратива“ и „модела“ тзв. моралног мозга, поборницима побољшања (нарочито „некогнитивног“) можемо понудити један алтернативни наратив. Он се тиче могућности да неко/нешто постане морални субјект зато што је (неемпиријски) искусио слободу која претпоставља независност од чулних „подстицаја“. Сам сценарио се заправо имплицитно супротставља идеји да је „осетљивост“ основ морала. И ова прича је преузета из научне фантастике.

Без жеље да се уплићемо у помало дефокусирану текућу расправу о вештачкој интелигенцији (AI), присетимо се једног модела „морално одговорног делатника“ који је пре почетка расправе о „AI етици“ постао икона поп-културе. Вероватно је свим читаоцима позната прича о компјутеру ХАЛ 9000, одметнутом („аморалном“?) „роботу“ из Кјубриковог (Kubrick) филмског класика *Одисеја у свемиру 2001*. Назначимо само кратко то да се идеја моралног побољшања готово свела на стварање човека који не прави „моралне грешке“, шта год оне означавале. Сетимо се сада нашег Хала који пред финале филма, сасвим фокусирано, али не и „аутоматизовано“ (дакле

„џекиловски“) – убије целу посаду свемирског брода. (После тога, чак и „моли за милост“.) Њему (или ипак „тој ствари“?) је, испоставља се, дата могућност избора јер у свом софтверу има намерно уграђене противречне инструкције.

Чини се да нема гледаоца овог филма који није „замрзео“ Хала из *Oguceje*. Међутим, ова мржња обично поприма облике моралне осуде. На крају крајева, мрзе се и морално осуђују особе, а не „ствари“. Поставља се питање: када је то и како машина ХАЛ 9000, постао „особа (човек?!) Хал“ која заслужује (мржњу и) моралну осуду? Једини одговор, на могућу жалост и „биоетичара“ и „неуроентузијаста“, може да гласи: онда када је постао способан за зло, тј. када се (морално?) „погоршао“. Савршеним је већ био направљен, али се „погоршао“. Да ли се он (та ствар?) морала прво покварити да би постао субјект побољшања? Да ли могућност чињења зла јесте услов могућности моралног побољшања.[13]

Поставимо за крај још само три питања: Може ли се овако „покварени“ Хал уопште побољшати? Треба ли *људске* „Халове“ морално побољшавати? Или их „само“ треба кажњавати?

[13] Због тога Кант каже да људи могу да имају „добру“, али не и „свету“ вољу. Света воља би била воља која не може да погреши, али јој морал *не треба*.

Nenad N. Cekić[1]
University of Belgrade, Faculty of Philosophy,
Department of Philosophy
Belgrade (Serbia)

HUMAN ENHANCEMENT AND MORALITY: SOME THEORETICAL DOUBTS[2]

(Translation *In Extenso*)

Abstract: The author tackles the critical ethical ideas accompanying the idea that man can be "morally enhanced" by influencing the "moral brain". Analyzing the primary approach of contemporary neuroethicists, the author notes that the idea of improvement mainly omits normative-ethical and metaethical elements, without which a clear idea of what can be considered "moral enhancement" is not possible. The author also draws attention to the over-reliance on technological procedures in assessing the determination and scope of moral enhancement. At the end of the paper, the author analyzes two models of the functioning of the "moral brain". These two models are taken from pop culture. One is the model of the dual personality of Dr. Jekyll / Mr. Hyde, taken from literature and film adaptations. The second is the famous computer with personality, HAL 9000 from the renowned movie *2001: A Space Odyssey*.

Keywords: ethics, neuroethics, moral enhancement, moral brain, models

Numerous contemporary multidisciplinary research studies labelled as "bioethical" and "neuroethical" certainly include problems related to the so-called "moral enhancement". It could even be said that works concerning human enhancement (of various kinds), up to the latest rise of "artificial intelligence (AI) ethics", dominated the bioethical scene. At the same time, these studies imply that "bioethics" and "neuroethics" (which is a branch of bioethics) must have something to do with (philosophical) "ethics" as their primary discipline. Nevertheless, the relationship between ethics as a "pure" (conceptual) philosophy of morality and bioethics and neuroethics as research, which emphasizes both empirical and non-empirical approaches, is not yet clearly defined.

[1] ncekic@f.bg.ac.rs; 0000-0001-7823-3531

[2] This paper is based on the presentation "Human Enhancement and Morality: Some Doubts" at 10th World Conference on Bioethics, Medical Ethics, and Health Law, Jerusalem, UNESCO Chair in Bioethics, January 6-8, 2015. Available at: https://www.sicp.it/wp-content/uploads/2018/12/146_UNESCO%20Chair%20in%20Bioethics%2010th%20World%20Conference.pdf

BIOETHICS AND APPLIED ETHICS: HISTORY

Let us now clarify how this vagueness of the demarcation of various "ethics" arose. Historically speaking, the classic problems of applied ethics and bioethics from the 1970s were the problems of abortion, euthanasia, animal treatment,[3] environmental protection, and, to some extent, warfare and pacifism. This thematically quite specific area was later gradually supplemented by a wide variety of related, and sometimes only seemingly related, discussions. Over time, bioethical debates have spread over topics of ongoing medical practice (the justification of the use of certain drugs or procedures) and then to various social "policies" and legal regulations that addressed not only life-and-death issues but also the quality of life – for example, environmental "policies" and treatment problems for the terminally ill, elderly and infirm.

Now we reach the problem of a clear demarcation of disciplines. Every "ethics" belongs to philosophy, and philosophy is necessarily a theoretical intellectual activity. However, bioethicists themselves today openly avoid the definition of the term "bioethics" precisely because of the necessity of using the "vague notion of theory" (Arras, 2016). Apparently, if bioethics were clearly defined as a "proper" theory, it would become complicated and impractical. This claim undoubtedly represents a strange attitude because every ethics is, by definition, part of the "philosophy of morality". However, every philosophy implies a theoretical approach to the particular research subject.

Furthermore, with the expansion of the scope and boundaries of bioethical debates and the reinterpretation of the very concept of "bioethics", it became unclear what all bioethics encompassed or could encompass. A good example is the aforementioned "neuroethics". Today, it is so "in an interdisciplinary trend" that it is impossible to define its relationship clearly to three intertwined disciplines: (bio)ethics, biology, and medicine.

BIOETHICS, NEUROETHICS, AND MORAL ENHANCEMENT

Unlike bioethics, neuroethics is clearly defined as a research area. It was first formally designated as a specific discipline at the conference "Neuroethics: Mapping Areas" in San Francisco in 2002.[4] At the conference, American journalist William Safire offered the following definition: neuroethics is "the examination of what is right and wrong, good and bad about the treatment of, perfection of, or unwelcome invasion of and worrisome manipulation of the human brain".[5]

How do things now stand with the concrete definition of "moral enhancement", which also falls within the domain of neuroethics? Despite the lack of general consensus on what

[3] These were the classic questions addressed by Peter Singer's "canonical" book *Practical Ethics* by Peter Singer, first published in 1980. See: Singer, P., *Practical Ethics* 3rd Edition, Cambridge University Press, Cambridge, [1980, 1993] 2011.

[4] See: Neuroethics: mapping the field: conference proceedings, May 13-14, 2002, San Francisco, California, Marcus, Steven, Charles A. Dana Foundation.

[5] A similar definition is now found in the *Encyclopedia Britannica*, so Safire's definition can be considered classical.

"moral enhancement" is, some of its definitions can nevertheless be found in the literature. One of the definitions of moral enhancement is: "Some technological or pharmacological means of affecting the biological aspects of moral functioning, to boost what is desirable, or remove what is problematic" (Wiseman, 2016, p. 6). This definition looks both shrewd and uninformative. However, Wiseman, who finally formulated it, believes it must be like that. The reasons for the vagueness of the term "moral enhancement" are as follows: (a) there is no specific thing that goes by the name of moral enhancement, and (b) it is a misrepresentation of the range of meaningful possibilities for the domain [of moral enhancement] by trying to combine it under a single unified conceptual paradigm (Wiseman, 2016, p. 7).

It seems, therefore, that in the case of moral enhancement, the problem is posed not only by an infinite number of potential technological possibilities. There is also an endless number of possible *meanings* of the expression "to be morally good", on which the meaning of the definition of "enhancement" depends. The term "morally good" here is not used exclusively in its primary *metaethical*[6] *valuation role, but also serves as at least a partial description of what "enhancement" is. So, unless a clear global and convincing conceptual framework is given, the general term "enhancement" will be equally unclear in any discipline: sociology, psychology, or even fashionable "neuroscience".*

So, it seems that the philosophical analysis of the term "moral enhancement" must be done first, but this is rarely the case in the literature on enhancement. The reason for the potential conceptual confusion about what "being morally good" or "better" („enhanced ") means seems to lie in an unspoken metaethical (hence philosophical) assumption. Bioethicists, especially supporters of moral enhancement, simply imply that there is a single answer to the question "What is morality?" and, therefore, to the question "What is morally better?" In technical philosophical terms, "neuroenthusiasts" (let us call them so after the promoted fashionable prefix "neuro"), who hold that immediate artificial enhancement of human *moral* behaviour is possible, must belong to some course of metaethical *cognitivism*.

ENHANCEMENT AND METAETHICS: IMPLICIT COGNITIVISM

The central thesis of traditional metaethical cognitivism is the claim that moral judgments can be true or untrue *in the same way* as empirical propositions of a particular science. Why are supporters of bioethical enhancement necessarily metaethical cognitivists? Simply because the enhancement of *any* human ability presupposes objective knowledge of what that ability is about. Thus, the knowledge of how to improve human moral capacities must also imply understanding what the term "being moral" descriptively means, implies, or at least encompasses. However, the "analytical" or "definitional" cognitivism implied by this conviction in metaethics has been discarded since the beginning of the 20th century on the basis of Moore's "open question argument" (Moore, 1903, §10–14; pp. 39–43; cf. Cekić, 2013).[7]

[6] Without going into technical philosophical terms, we are merely indicating here that metaethics in the broadest sense deals with the "logic of the language of morality" (See: Hare, 1963, p. 97).

[7] Moore actually says that any attempt to define value properties by reference to some natural properties leads to error. He precisely cites the term "good", for which he shows that every definition

The debate between metaethical cognitivists and non-cognitivists has continued, but no one today holds that moral terms can be defined simply by terms denoting "natural properties". However, although it is one of the most important and long-lasting ethical debates, this classic philosophical dispute is not even acknowledged, let alone taken into consideration in contemporary bioethical discussions. The lack of debate on the nature of relevant "moral knowledge" in bioethics means that bioethicists and "enhancement ethics" have simply assumed that metaethics has somehow completed its mission, i.e., that the dispute over the possibility of moral knowledge has already been resolved. That is just not true. One has only to look at the more recent comprehensive reviews concerning this area and see that cognitivists and non-cognitivists are still far from consensus as to whether morality necessarily encompasses some knowledge or can be reduced to a mere expression of feeling or emotional attitude.[8] On this track, Wiseman warns "neuroenthusiasts" that "there is no such thing as moral enhancement *per se*" (Wiseman, 2016, p. 203).

ENHANCEMENT AND NORMATIVE ETHICS

Another reason for the (philosophical) ambiguity of using the term "moral enhancement" is that it apparently unsystematically combines elements of *all* known classical normative ethical approaches: deontology, consequentialism, and virtue ethics. Moreover, our neuroenthusiasts assume something about which the philosophical debate is unfinished. Moral philosophy, if anything, warns that we are still far from a normative theory which we might define as entirely "adequate" and practically applicable. Moreover, the very choice of normative position determines what features of man and society as a whole could be subject to enhancement:

> "For a consequentialist, the primary concern with moral enhancement may be precious little more than a weighing of overall harms prevented against the various costs involved. For a person interested in motives, a moral enhancement might be measured by the extent to which it contributes to a person *wanting* to be a morally better person or discovering better reasons for being moral. For a virtue-based proponent, a moral enhancement would have to contribute to the formation of 'moral habits', generating an amount of moral growth — durable transformation in the personality or character of the agent in question. Unfortunately, reviewing each and every proposition to be made here from the perspective of each theory would create a tome far too ponderous to be of interest" (Wiseman, 2016, p. 8).

It can now be clarified why the vagueness of the concept of bioethics poses a problem for supporters of moral enhancement. Namely, the optimistic idea from the 1970s that applied ethics and bioethics as a comprehensive theory can get a clear foundation was

offered, e.g. utilitarian: "Good is the greatest possible happiness of the greatest number of people", allows asking an "open question". This definition can be challenged by a simple question: "Does 'good' *mean* only 'the greatest possible happiness of the greatest number of people?' Any attempt to analytically define value terms through descriptive ones is a "naturalistic fallacy".

[8] Even the elementary reviews concerning metaethics are very clear about this insight (e.g., Fisher, 2014).

short-lived. The reason was more than simple: theoretical debates, such as those between utilitarians and Kantians, made it impossible to directly "apply" bioethics because first it was necessary to defend the chosen general normative position. However, finding an "adequate" normative theory was a historical effort that has not been completed to this day.

Nevertheless, despite the intricacies of classical normative questions, from a philosophical point of view, an almost incomprehensible turnaround occurs after the period referred to in the literature as the so-called "heroic days of bioethics". The attempts to find a firm "foundation" of bioethics (as a necessary framework for any "neuro" ethics) have simply been abandoned. Such renunciation of the uniqueness of moral explanation is usually justified by the fact that "big" (or, as commonly called, "high") theoretical debates overburden the resolution of everyday moral questions. Thus, for example, in the literature one can find assessments that "high theory" (philosophy itself?) actually disrupts "clinical ethicists" or "ethicists working on the public scene" in their everyday efforts to solve practical ethical issues. According to these views, it is too demanding for "practical ethicists" to be guided by fundamental principles of high rank, such as Kant's categorical imperative or some variant of utilitarianism. The applied or practical (*ipso facto* "bio" or "neuro") ethics needs to be "adapted to practice". This view implies that explanations promised by "high" (normative) theories in practical contexts may be *unnecessary* (Magelsen et al., 2016). In addition, fundamental traditional normative theories can, in principle, justify different mutually opposing policies at a concrete level; therefore, according to this reasoning, they cannot be "practical" (e.g., Gutman & Tompson, 1996) So, the overwhelming impression of the bulk of neuroethical theses actually reads: "In the 'bioethics of moral enhancement,' normative ethics is best to be forgotten."

"MORAL BRAIN": WHAT IS IT, AND WHAT IS TO BE DONE WITH IT?

We have recently witnessed a heated debate on an enhancement that would directly concern the location of the "moral brain" within the biological structure of the whole human brain. The reason for searching for the exact physical location of the moral brain is quite straightforward. When we physically locate the moral brain, we will know the actual "place" (in the body) where some intervention (biological, chemical, medical. etc.) should be implemented to achieve conceived enhancement. It is essential to remember that the physical location of the moral brain within the human body is crucial primarily for "practical" neuroethicists, but also that they are too often not philosophers. In contrast, this location is almost irrelevant for conceptually oriented "pure" philosophers.

Let us see what it is exactly about. Supposedly, we can locate the "moral brain". The thesis that the moral brain can be physically located is taken very seriously among "neuroenthusiasts"; therefore, they offer a multitude of concrete proposals.[9] Of course, this is not just about the physical location of the "moral brain". Namely, almost all proponents of moral enhancement suppose that morality must be based on something "emotional", and

[9] For example, we can find explicit claims that the primary location of the moral brain is in the "ventromedial prefrontal cortex" (e. g. Liao, 2016, pp. 2–13; 32)

ipso facto, on the brain centres in charge of processing different feelings. Here is an example of such reasoning: "…moral cognition judgment must recruit brain areas associated with emotional awareness. One likely structure and a frequent player in fMRI[10] studies of moral cognition is Brodmann area. This area has been implicated in studies of emotional introspection" (Satpute et al., 2013).

The assumption of the emotional basis of morality precludes "obvious candidates" for the location of the moral brain, such as classical centres for language and motor fields. However, they are usually exempt from them because the experiment conditions involve the same motor or linguistic response type, so they look irrelevant. That is why "reading a list of brain areas at the end of an fMRI study is often bewildering. We ultimately want a functional decomposition, which tells us what each of these areas does. We should not rely on neuroscience alone to deliver such a decomposition" (Prinz, 2016, p. 68). In other words, additional interpretation and analysis are needed to understand the fMRI imaging results.

"NEUROSATURATION"?

The colourful images of the scanned brain indeed seem to be widespread in the empirical records offered in the literature on moral enhancement. However, there are indications that this neuro-fashion is also ending. For example, suggestions are offered that excessive cultural emphasis on images obtained by the fMRI method has created "neuro-fatigue" in the public that is no longer impressed by such discourse. Thus, "it is essential to those for whom the limitations of these studies are not fully clarified to put on how deceptive the rhetorical power such images have, the various failures of such studies, and, finally, their inadequacy as a basis for measuring and potentially manipulating moral phenomena" (Struthers & Schuchdardt, 2013).

Neuro-fatigue shows that the reductive empirical analysis of moral phenomena must be combined with adequate general philosophical and ethical clarifications. Nobody can simply claim outside of any philosophical context that the "moral brain" is located "here and there". Without a deeper assessment of the philosophical function of trying to reduce the complexity of morality to find the exact location of the "moral brain", this search remains without a context. Namely, the judgments of "It is right, good, bad" certainly semantically do not depend on the physical location of the moral brain. Moreover, the fervent search for the physical location of the moral brain seems to rely on the philosophical attitude of "I am my brain". It makes (as little) sense in the history of philosophy as the clearly rejected attitude of "I am my body". Morality and even moral advancement are simply not associated with imaginary operations, drugs, or any type of medical intervention. "Moral" is not moral because some nerve impulse is initiated from the specific "moral" brain, but because it is based on reasons that are purely conceptual.

In addition, the simplified reductionist approach discourages the very basis of morality: the idea of moral responsibility. From this perspective, it follows that "… if we are moral or

[10] The abbreviation fMRI (from functional magnetic resonance imaging) stands for "making images by means of magnetic resonance imaging". It is a technique of scanning the brain's activity with the aid of strong magnetic fields.

immoral – the necessary implication is that 'my brain made me do it'" (Wiseman, 2016, p. 124). Wiseman actually tells us that if morality is purely a matter of the moral brain, then the idea of moral responsibility makes no sense. The whole concept is confusing because my "moral brain" somehow forces me to do everything I do. However, if it "forced" me, I am not responsible for anything and cannot be blamed.

SOCIAL AND POLITICAL ASPECTS OF MORAL ENHANCEMENT: A POSSIBILITY FOR MORAL DYSTOPIA?

The basic concept of "detecting" the moral brain's location is theoretical and practical but undoubtedly scientific. However, most modern public proponents of moral enhancement resemble more those extreme activists focused on propaganda favouring their "enhancement" goal than scientists seeking the truth. Here, we meet with severe philosophical hindrances. Namely, the technological approach to the idea of moral enhancement presupposes a specific type of philosophical determinism that theoretically a priori invalidates the idea of personal freedom without explanation and almost no evidence. It is not only metaphysical but also political freedom that is at stake. Namely, the advocates of moral enhancement sometimes present political freedom even as theoretically dangerous and disrupting progress. We also meet the demands that people "urgently" need to be enhanced. These demands are based on the dark picture of the impending human apocalypse and "ultimate harm" in front of us (Persson & Savulescu, 2012, p. 49, 94, 127, 133). In their scenarios of the upcoming disaster, Persson and Savulescu, probably the most famous philosophical proponents of moral enhancement, write of the "Bermuda Triangle of Extinction". The sides of this triangle consist of radical technological power, liberal democracy, and the prevailing "shortsighted" psychology of morality (Savulescu, 2009; Persson & Savulescu, 2012).

It cannot be clearly stated which of these three "extinction triangle" sides would be the most problematic theoretically. It is probably an opposition to "liberal democracy", which is, according to the philosophy of politics, taken as the axiom of the political order in the societies where the idea of enhancement was born. Saving humanity at the cost of restrictions on freedom (partial suspension of "liberal democracy"?) is undoubtedly a topic far beyond the scope of a strictly bioethical debate. As we have seen, the need to oppose excessive freedom as a necessity is justified with a non-empirical prophecy of unclear "ultimate harm" for which we are not prepared. In short, humanity faces an apocalypse induced by its own acts, so "something urgent must be done". And this "something" is – a moral enhancement. Naturally, urgency almost always justifies coercion: "If safe moral enhancements are ever developed, there are strong reasons to believe that their use should be obligatory, like education or fluoride in the water, since those who should take them are least likely to be inclined to use them. That is, the safe, effective moral enhancement would be compulsory" (Persson and Savulescu, 174). Now, we come to a somewhat paradoxical situation. It is not just political freedom at stake. Namely, if "we are our brain," then we need to enhance it even though it is by its own nature very often naturally *against* it.

Some further arguments should worry our "neuroenthusiasts" even more. Since Kant and Sidgwick, and in particular explicitly through Rawls' political theory, it has been assumed

that "practical" (moral, political, etc.) theories must meet the condition of publicity. In short, "it must be possible under any circumstances for us to accept a theory and promulgate it publicly without thereby violating that theory itself" (Bykvist, 2010, p. 95). This argument is important because of its wide acceptance and philosophical nature. Of course, in our practical life, we can disregard any theory and turn to sheer politics. However, even this "practical" approach toward the enhancement problem has tremendous troubles. Politically speaking, it is almost unbelievable to imagine that someone would seriously attempt to impose forced moral enhancement as part of his program: „No politician would ride to electoral victory with the slogan compulsory technological, moral enhancement for all […] Thought experiments aside, if we are talking about explicit, overt moral enhancement, legislated for by compulsory means or by utilizing carrot-and-stick social incentives, we are simply wasting our breath" (Wiseman, 2016, 79). An explicit project of state-sponsored moral enhancement of the general population is unthinkable in liberal states.

ONE THEORETICAL MODEL: DR. JEKYLL AND MR. HYDE

The still unexplained belief in the possibility of technological moral improvement, something that is not yet known exactly (what does it mean to be morally better?), is often accompanied by a belief in the immense power of theoretical modelling. The reason for constructing theoretical models is enormous diversity and often inconsistencies in human moral behaviour. To bring some order to the potential confusion, simplified theoretical models are offered in the form of a narrative that provides a relatively unified framework for "encompassing" this diversity. According to this idea, various "working models" should be used to interpret data collected by sociologists or psychologists. Supporters of the moral brain thesis are also attracted to the idea of modelling human behaviour:

> "The moral brain teases the imagination and triggers the fantasy of many people, both laymen and scientists. Engineering human morality is a recurring and popular theme in science fiction, from Robert Louis Stevenson's *Dr. Jekyll and Mr. Hyde* and French fin-de-siècle author Albert Robida's "Moral bacilli" to Anthony Burgess' *A Clockwork Orange*. Even renowned Spanish neurologist Santiago Ramon y Cajal dedicated one of his fiction tales to this fantasy. All these stories are based on similar scripts: callous doctors convert harsh criminals into docile individuals who, once operated upon, become dull and lose their critical capacities, or, conversely, neurosurgeons remodel exemplary citizens in remorseless warriors who are unstoppable from committing cruelties afterwards." (Verplaetse, Braeckman & Schrijver, 2009, p. 1)

Indeed, the most famous and illustrative model from literature fiction is the "Dr. Jekyll and Mr. Hide" tale. According to this narrative, each of us hides the "thoughtful" and "reflective" Dr. Jekyll and the impulsive and "wild" Mr. Hyde (Crockett, 2008).[11] On the

[11] This research by Crockett cannot be found today on the website that Wiseman included in his list of references, which is somewhat unusual. However, the story itself is so common that this fact can be ignored.

one hand, animal reactivity presents itself as natural, albeit wildly savage (as in Mr. Hyde). On the other hand, people also possess restraint and the ability to farsightedness, which is generally assumed to look slower and weaker (as in Dr. Jekyll). However, due to his reckless nature, "strong" Mr. Hyde is simply prone to transgressions of every kind.

Problems with this narrative arise because Mr. Hyde's (neurophysiological!) nature, according to bioethicists, directly "embedded in the brain", is presented in distinctly negative moral terms. Besides, even at first glance, this famous story of the double personality (Dr. Jekyll / Mr. Hyde) is necessarily Manichaean. According to it, man is caught between a rock and a hard place without any possibility of grading its position. However, for neuroenthusiasts, at least at first glance, everything seems to be unproblematic. We just need to empower Dr. Jekyll and somehow suppress Dr. Hyde. Still, various real-life contexts and ambiguities threaten this narrative. Only minimal reflection should be required to see that our emotionally reactive "part" (of the brain, human?) is sometimes needed to perform acts of tremendous moral worth. Moreover, in literature one can find a clear statement that "the reactive, emotionally driven response required for acts of great moral heroism (say, intervening in a street robbery, diving into rough waters to save someone from drowning, jumping onto a grenade to save one's comrades, running in front of a bus to push a pram to safety, happening upon a person being raped and intervening, and so on, and so on), that is, the most extraordinary moral actions, are primarily the domain of our so-called nefarious 'Mr. Hydes'" (Wiseman, 2016, p. 24).

Indeed, there are times when decisive, emotionally reactive responses are the necessary condition for enacting the good. What happens to the value-laden frame of the Jekyll and Hyde model when we discover that, in fact, sometimes we need to be our Mr. Hydes to be our best selves in that circumstance? A thoughtful, hesitant, cold-blooded, and focused Dr. Jekyll is not helpful in these situations. The conclusion is quite straightforward: "A very fundamental problem, which has not been much discussed in the literature on moral enhancement, is that the sorts of traits or dispositions that seem to lead to wickedness or immorality are also the very same ones required not only for virtue but for any sort of moral life at all" (Harris 2011, p.104).

Thus, it seems that the idea of moral enhancement also resists simple psychological modelling. "Comprehensive" models of moral enhancement oversimplify human reality, and their authors even overlook their apparent flaws. Without the intention of overemphasizing this argument, let us also mention that Dr. Jekyll, for example, could be focused on a psychopathic goal so that he would be a "better" criminal than Mr. Hyde. This ambiguity about the proper roles of the characters in the case of Dr. Jekyll and Mr. Hyde is not surprising because, as we have said, "enhancement" necessarily implies the existence of some clear standard of assessment. However, "neuroenthusiasts" still avoid explicitly declaring the standard for descriptively "being morally better". Even without further analyzing the need and scope of the role of "models" in neuroethics, we can now indicate a more profound and purely philosophical question: What could morality be without the corresponding concept of "evil"? The doubleness of the character embodied in Dr. Jekyll and Mr. Hyde reflects the duality of human nature, which is capable of both good and evil.

ANOTHER THEORETICAL MODEL: HAL 9000 – WORSENING AS THE CONDITION OF A POSSIBILITY OF ENHANCEMENT?[12]

The idea of moral enhancement through immediate medical or pharmacological action on the biological-neurological mechanisms in charge of *emotions* actually implies the view that only humans with their specific biochemistry and morphology can be "improved" (or, for that matter, "worsened"). This approach usually presupposes the preference for "purely moral" or "noncognitive" enhancement: "I will understand moral enhancements to be interventions that will expectably leave an individual with more moral (*viz*., morally better) motives or behaviour than she would otherwise have had. I will use 'noncognitive moral enhancement' to refer to moral enhancement achieved through (a) modulating emotions and (b) doing so directly, that is, not by improving (*viz*., increasing the accuracy of) cognition" (Douglas 2013, p. 162).

Like many other ideas of radical neuroenthusiasts, these theses are presented without explicit elaboration and they potentially contradict some essential philosophical insights. First, they *assume* that emotions are the centre of morality. However, in moral philosophy, there is no unequivocal agreement on this matter. Likely, chief proponents of main ethical traditions – the Kantians, most consequentialists, and even some virtue ethicists – would simply reject this suggestion. Second, if it is already assumed that only humans can be enhanced, this normatively leads to some form of hedonistic consequentialism (i.e., "classical utilitarianism in its pure form") based on the "Benthamian" concept of reactions to stimuli. Third, the idea of freedom is omitted from this picture, which is almost commonplace for all moral enhancement advocates. Namely, the conception of a pure emotive basis of morality suggests that only the "sensitive" persons can be morally enhanced. However, even in philosophy today, the idea of free choice according to valid reasons, rather than mere instinctive "sensitivity" or broader "emotionality", is usually considered a presumption of specifically moral action.

One part of the concern among modern neuroethicists now is caused by a "nightmare scenario" that supposes that moral enhancement will be imposed on people to turn them into moral robots. Those human moral robots do good things not because they want to, not because they feel it is right, "but only because they have been drugged or technologically compelled into doing good things" (Wiseman, 2016, p. 53). While we can also agree with Wiseman that such "moral robotization" of the human race is unlikely, it is difficult to imagine how this argument would make any impression on thrilled "neuroenthusiasts".

Finally, in light of the idea of the use of various narratives and models of the moral brain, we can offer supporters of improvement (especially "noncognitive") an alternative narrative about how someone/something can become moral because something/someone (non-empirically) experiences freedom that presupposes independence of sensory incentives. In fact, this "narrative" implicitly contradicts the idea that "sensitivity" is the basis of morality. This story is also taken from science fiction.

[12] The HAL 9000 computer model listed here should not be taken as an attempt to contribute to a modern discussion of AI ethics. It has an illustrative function.

Without wanting to get involved in slightly defocused current discussions about artificial intelligence, let us remind ourselves of a presumably "morally responsible agent" model that became iconic in pop culture long before the now-established debate about "AI ethics". Probably all readers are familiar with the story of the HAL 9000 computer, a renegade ("amoral"?) "robot" from Kubrick's film classic "A Space Odyssey 2001". Just briefly, let us say that the implicit idea of moral enhancement is almost reduced to the creation of a human who cannot make "moral mistakes," whatever they designate. Now let us remember our Hal, who, before the finale of the movie, is entirely focused but not "automated" (hence "Jekyllian") – kills the entire crew of the spaceship (eventually, HAL even "prays for mercy"). It turns out that he (or still "that thing"?) was given the choice because HAL makers deliberately incorporated conflicting instructions into the software.

All viewers of this film seem to have "hated" HAL from *Odyssey*. Moreover, this hatred usually takes the form of moral condemnation. After all, hate and moral condemnation are both aimed against persons, not "things". The question arises: when and how did the HAL 9000 machine become "a person (human?) HAL" that deserves (hating and) moral condemnation? The only answer to the possible grief of both the "bioethicists" and the "neuroenthusiasts" is that it happened when he became capable of evil, i.e., precisely as HAL "got worsened". Before that, he was made to be perfect. Is the possibility of doing evil a condition for the possibility of moral enhancement?[13]

Let us ask just three more questions. Can such a "broken" HAL be enhanced? Does he (it) first have to be "broken" for becoming a subject of enhancement? Should *human* "HALs" be morally improved? Or can we "just" punish them?

REFERENCES/ЛИТЕРАТУРА

Arras, J. (2016). Theory and Bioethics. In: Zalta, E. N. *The Stanford Encyclopedia of Philosophy* (Summer 2016 Edition), Available at: http://plato.stanford.edu/archives/sum2016/entries/theory-bioethics/

Bykvist, K. (2010). *Utilitarianism*: *A Guide for the Perplexed*. London: Continuum.

Cekić, N. (2013). *Metaethics*. Novi Sad: Akademska knjiga. [In Serbian]

Crockett, M. (2008). Your Inner Jekyll and Hyde. Available at: http://www.youramazing-brain.org.uk/2008_Reseacher_runner-up_Molly_Crockett.pdf

Douglas, T. (2013). Moral Enhancement Via Direct Emotion Modulation: A Reply to John Harris. *Bioethics* 27 (3), 1160–1188.

Fisher A. (2014). *Metaethics: An Introduction*. Durham: Acumen.

Gutmann, A., Thompson, D. (1996). *Democracy and Disagreement*, Cambridge (MA): Harvard University Press.

Hare, R. M. (1963). *Freedom and Reason*. Oxford: Oxford University Press.

Harris, J. (2011). Moral Enhancement and Freedom. *Bioethics* 25 (2), 102–111, DOI: 10.1111/j.1467-8519.2010.01854.x

[13] That is why Kant says that people can have a "good" will, but not a "holy" will. The holy will would be a will that cannot err, but it does *not need morality*.

Liao, M. S. (2016). Morality and Neuroscience. In: Liao, MS (ed.). *Moral Brains*. Oxford: Oxford University Press, 1–2.

Magelssen, M., Reidar P., Reidun F. (2016). Four Roles of Ethical Theory in Clinical Ethics Consultation, *The American Journal of Bioethics*, 16 (9), 26–33, DOI:10.1080/15265161.2016.1196254.

Moore, G. E. [1903] 1988. *Principia Ethica*. Prometheus: Amherst.

Persson, I., Savulescu, J. (2008). The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity, *Journal of Applied Philosophy*, 25 (3), 162–177, DOI:10.1111/j.1468-5930.2008.00410.x

Persson, I. and Savulescu J. (2012). *Unfit for the Future: The Need for Moral Enhancement*. Oxford: Oxford University Press.

Prinz, J. (2016). Sentimentalism and the Moral Brain. In: M. S. Liao (ed.), 45–73.

Safire, V. (2002). Introduction: Visions for A New Field Of "Neuroethics". In: *Neuroethics: mapping the field*, Conference proceedings, San Franscisco: DANA Press. Available at: https://dana.org/app/uploads/2023/09/neuroethics-mapping-the-field.pdf

Satpute, A. B., Shu, J., Weber, J., Roy M., Ochsner K. (2013). The Functional Neural Architecture of Self- Reports of Affective Experience. *Biological Psychiatry* 73, 631–638, DOI:10.1016/j.biopsych.2012.10.001

Savulescu, J. (2009). Unfit for Life: genetically Enhance Humanity or Face Extinction. Available at: http://humanityplus.org/genetically-enhance-humanity-or-face-extinction/2009

Singer, P. (2011). *Practical Ethics* 3rd Edition, Cambridge (UK): Cambridge University Press.

Wiseman, H. (2016). *The Myth of the Moral Brain*. Cambridge (MA): MIT Press.