



ČASOPIS  
PRO  
SOUČASNOU  
FILOSOFII

---

ČÍSLO 2  
ROČNÍK 13  
2021

---

## FILOSOFIE DNES

Časopis pro současnou filozofii  
vydává KFSV FF UHK

**Šéfredaktorka:** Iva Svačinová

**Redakční rada:** Ondřej Beran, Jaroslav Daneš, Petr Glombíček, Tomáš Hirt, Juraj Hvorecký, Vojtěch Kolman, Ladislav Kvasz, Róbert Maco, Tomáš Marvan, Martin Paleček, Jaroslav Peregrin, Marie Hrdá, Ondřej Švec, Eugen Zelenák

**Editoři:** Markéta Panoušková, Hynek Kaplan

**Sazba a grafická úprava:** Hynek Kaplan

**Adresa:** Filozofie dnes, KFSV FF UHK Rokitanského 62  
500 03 Hradec Králové

**E-mail:** [iva.svacinova@uhk.cz](mailto:iva.svacinova@uhk.cz)

vychází elektronicky,  
dvakrát ročně

[www.filosofiednes.ff.uhk.cz](http://www.filosofiednes.ff.uhk.cz)

ISSN 1804-0969

Ročník 13, Číslo 2 (2021)

---

## Obsah

### Studie

Ondřej Beran	„Morální intuice“. Co si můžeme nebo máme odnést z četby Haidta	3
Lukáš Mareš	Výchovná a vzdělávací role sportu u myslitelů Sókrata, Platóna a Aristotela ve vztahu k problematice dobrého sportu a vedení dobrého života	44
Jaroslav Malík	Superintelligence a problém kontroly: Skutečný problém nebo pseudo-problém?	73
Miroslav Vacura	Objektivistická etika Ayn Randové	119

# „Morální intuice“. Co si můžeme nebo máme odnést z četby Haidta

Ondřej Beran

*Centrum pro etiku (Katedra filosofie a religionistiky)  
Fakulta filozofická, Univerzita Pardubice  
Pardubice, Studentská 95, 53210  
ondrej.beran@upce.cz*

Text usiluje o kritický přehled některých prací sociálního psychologa Jonathana Haidta, s ohledem na to, jaké důsledky z Haidtových tvrzení vyplývají pro pojem morálky a možnosti filosofické diskuse o morálních problémech. Budu tvrdit, že jakkoli jsou Haidtova zjištění a jejich interpretace pozoruhodné a přínosné, některé aspekty lidské morálky a morálního života zkreslují. V první části textu se dotknu nejasností spjatých s Haidtovým pojmem morální intuice. Vodítkem bude způsob, jakým Haidt zavádí tento pojem a pracuje s ním ve svém příkladu sourozenců Marka a Julie. Ve druhé části naznačím, jaký dopad mají tyto nejasnosti pro Haidtovy politicko-psychologické analýzy. V poslední části poukážu na význam Haidtem přehlížené rozmanitosti filosofických teorií toho, co je to morálka. Společným jmenovatelem mnoha mých kritických poznámek je jednak vágnost Haidtova pojmu „intuice“, jednak to, že i když Haidt označuje svá zkoumání za čistě deskriptivní a pluralistická, implicitně předpokládá zhruba utilitaristické pojetí morálky, včetně jeho normativních aspektů.

Klíčová slova: intuice, morálka, liberalismus, konzervativismus, Jonathan Haidt

## Úvod

Sociální psycholog Jonathan Haidt soustavně nabízí nový a, jak naznačuje, od iluzí oproštěný a vědeckým zkoumáním podložený pohled na morálku. Jeho ambicí je předložit takový obrázek morálky, který by

umožňoval vysvětlit morální fenomény, jež etablované morální filosofie vysvětlit nedokážou – protože jejich představa o tom, co je morálka, je naivní, nepřesná, nebo zastaralá. Pokusím se ukázat, že zatímco Haidtovy experimenty a jejich zjištění jsou nepochybně cenné a zajímavé, méně jasné je, zda z nich skutečně vyplývají Haidtem naznačované důsledky pro chápání morálky. Důvodů může být několik. Jednak Haidtovo vlastní pojetí toho, co je morálka, které není „objektivním faktem“ a je ve vztahu k otázkám, které si klade, fakticky před-pojetím. Jednak jeho pojem „intuice“, který v tomto pojetí hraje klíčovou roli. A jednak specifický význam, který Haidt přikládá způsobu, jakým se to, co považuje za motivace morální povahy, „překládá“ do postojů a pohybů politických. Zde hraje významnou roli jeho čtení politické mapy, zvláště současných USA; tento kontext vztahů mezi morálkou a politikou je ústředním příkladem diskuse v nejznámějším Haidtově díle, knize *The Righteous Mind*.<sup>1</sup>

V první části se dotknu Haidtova samotného pojmu „morální intuice“ a nejasností spojených s tím, jak mu rozumět. Vodítkem bude způsob, jakým Haidt zavádí tento pojem a pracuje s ním ve svém slavném příkladu sourozenců Marka a Julie. Ve druhé části naznačím, jaký dopad mají tyto nejasnosti pro Haidtovy politicko-psychologické analýzy. V poslední části poukážu na význam Haidtem přehlížené rozmanitosti filosofických teorií toho, co je to morálka, respektive možnosti uvažovat o morálce v pojmech překračujících rámec zkoumání morálních intuic v jím zavedeném smyslu.

## 1. Co jsou intuice

Svůj článek „The Emotional Dog and Its Rational Tail“ začíná Haidt tímto slavným příkladem:

„Julie a Mark jsou sourozenci. Během letních prázdnin jsou spolu ve Francii. Jednoho dne zůstanou sami přes noc v chatě poblíž pláže. Rozhodnou se, že by mohlo být zajímavé a zábavné, kdyby se spolu zkusili pomilovat. Přinejmenším by to pro oba byla nová zkušenost. Julie pravidelně bere

1 Haidt (2012).

antikoncepci a Mark pro jistotu použije i kondom. Oba si milování užijí, ale rozhodnou se to už neopakovat. Ta noc zůstane jejich tajemstvím, díky kterému jsou si nyní ještě blíží. Co si o tom myslíte? Bylo OK, že se spolu pomilovali?<sup>2</sup>

Smyslem tohoto příkladu je, jak Haidt naznačuje, testovat morální reakce posluchačů. Má za to, že svůj příklad obezřetně formuloval tak, aby neobsahoval žádný detail, na němž by bylo možné založit morální odsouzení Marka a Julie, protože tu nikdo nedozná žádné *újmy*. Je-li tedy okamžitou reakcí posluchačů, že to „nebylo OK“, dovozuje, že tato reakce musí vyplývat z jiného zdroje, než z racionální reflexe příkladu, protože na *této* rovině se morální odsouzení nemá oč opřít a respondenti mají problémy uvést dobré racionální zdůvodnění svého odsudku.

Zdrojem, který Haidt na základě takto nadesignovaného experimentu identifikuje, je „intuice“: „okamžitý záblesk odporu“, který posluchači pocítí a který se mohou ex post pokoušet racionalizovat; pak ovšem jde o zpětnou reinterpetaci. Čím „intuice“ je, Haidt o mnoho více nspecifikuje: je *okamžitá* a je formou *přímého vnímání* relevantní kvality posuzované situace. Intuicí je, zdá se, v zásadě *jakákoli* forma takovéto reakce; přinejmenším Haidt tímto termínem označuje dosti heterogenní soubor různých reakcí sdílejících primárně ten rys, že *nemají* povahu „pomalého“ racionálního promýšlení. Nejsou sice výsledkem racionální reflexe, a jsou tedy a-racionální, nejsou však pouze iracionální. Mimo jiné proto, že Haidtův model je *sociálně-intuicionistický*: pocíťované intuice jsou, vedle svých evolučních kořenů, ovlivňovány rovněž heterogenním souborem faktorů jako *peer-pressure*, akulturace, zvyky v dané společnosti, atd.<sup>3</sup> Zdá se nicméně, že Haidt má tendenci spatřovat i v existenci a v kulturně rozmanité podobě těchto sociálních faktorů fenomén, jehož vysvětlení je přinejmenším zčásti také naturalistické, spíše než kulturní. Zmíním to ještě dále.

Jistě ne všechny intuice jsou nutně *morální* intuice. Haidtovo kritérium či pojem morálky je v této fázi jeho argumentu příznačně formální: morální jsou ty z našich intuitivních reakcí, které spočívají

---

2 Haidt (2008 [2001], s. 1024).

3 Tamtéž, s. 1041.

v „hodnocení (dobré vs. špatné) jednání či charakteru lidí, s ohledem na soubor ctností považovaných v dané kultuře za závazné“.<sup>4</sup> Toto svou povahou formální pojetí je nutné pro postulování Haidtova morálního pluralismu, klíčového pro jeho projekt výzkumu psychologie v politice; více o tom v oddíle 2. Jak však uvidíme, na jiných místech táhne Haidt k předpokladu, že mechanismus ustavení těchto standardů a motivace jejich vymáhání společností má víceméně utilitaristický charakter: různé morálky jsou zacíleny k účelu vzájemně prospěšné spolupráce a prospívání ve společnosti.<sup>5</sup> Více o tom, stejně jako o možných jiných způsobech, jak chápat, co je „morálka“, v oddíle 3. Zde jen zmíním, že Haidt se nehlásí k utilitarismu jako normativní etické teorii, nýbrž spíše za pomoci těchto pojmů analyzuje nevyslovenou, „intuitivní“ rovinu ustavení, motivace a uplatňování našich morálních postojů. Platí ovšem, jak jsem naznačil už výše, že na této rovině jsou pro Haidta i v případě morálky lokalizovány značně rozmanité věci; nemalá část jeho diskuse morálních intuic je vedena ve vztahu k vynášení morálních (hodnotících) *soudů*. Na druhou stranu, jak poněkud mlhavě Haidt upřesní, morální intuice „zahrnují“ také morální *emoce*.<sup>6</sup>

Neurčitost Haidtova pojetí intuice snad poněkud zarazí v porovnání s vlivnými filosofickými autory (Michael Huemer, Robert Audi, aj.), kteří věnovali mnoho úsilí precizní formulaci a obhajobě různých verzí „etického intuicionismu“. Zmíním v krátkosti jen Huemerovo pojetí, které je Haidtovi blíže než to Audiho:<sup>7</sup> podle něho jsou „intuice“ kognitivními stavy, které zachycují realitu, ovšem jiným způsobem než přímé smyslové vnímání nebo soudy či přesvědčení odvozené z důvodů. Lze je ovšem prostřednictvím reflexe „přeložit“ do formátu majícího propoziční obsah. Jakkoli některá naše morální přesvědčení mohou být ovlivněna zkreslením (*bias*) či odvozena z jiných, nemohou být taková všechna – některá vycházejí z intuitivního nahlédnutí výchozích morálních pravd či faktů, které se nám přímo jeví (*seem*).<sup>8</sup>

4 Tamtéž, s. 1028.

5 Např. Haidt (2012, s. 341).

6 Haidt (2008 [2001], s. 1041).

7 Zdařilý *kritický* přehled Huemerova intuicionismu, včetně načrtnutí jeho vazeb k jiným filosofickým formám intuicionismu i empirickému zkoumání intuic (např. Haidtovu), nabízí Chudárková (2020).

8 Např. Huemer (2008, s. 99nn).

Nebudu zde polemizovat s filosofickým intuicionismem, který např. v Huemerově případě sleduje další komplexní agendy, jako obhajobu jistých metafyzických tezí. Bez ohledu na to, zda s Huemerem a dalšími souhlasíme, je třeba ocenit pečlivost, s níž se snaží odlišit morální intuice od přesvědčení a morálních emocí. Mimo jiné proto, že pocítění morální emoce (odpor k X), spontánnímu vynesení morálně hodnotícího soudu („... že jednání X je odsouzeníhodné“) a okamžitému vnímání morálně relevantní kvality nějaké situace (podlost jednání X) předchází rozdílné předpoklady. Haidta tyto rozdíly podle všeho příliš nezajímají. Naopak, podstatným rozdílem, k němuž se opakovaně vrací, je rozlišení inherentní duálnímu modelu lidské kognice, k němuž se Haidt, a ostatně nejen on hlásí. V rámci tohoto modelu jsou tyto reakce „rychlými“ vyjádřeními naší evolučně starší výbavy, víceméně spontánními a nemožerovanými „pomalou“ zpětnou racionalizací, a jako takové jsou také vyjádřeními toho, co má v naší psychologii a jednání skutečnou motivující sílu. Ještě se k tomu níže vrátím.

Nyní se však podívejme blíže na Haidtovu představu, že experimenty podobné tomu, který předkládá on, jsou nejen dostatečné, ale dokonce lépe způsobilé než úvahy „čistě filosofického“ rázu k tomu, aby odhalily přehlížené skutečnosti o lidské morálce. S tímto rámcem studia morálních fenoménů jsou ale spojeny problémy. V první řadě ten, že se situacemi, k nimž zaujímáme morálně hodnotící postoje, jsme obvykle konfrontováni ve značně odlišné formě, než jakou naznačuje Haidt. Jeho přístup ignoruje význam kontextu:<sup>9</sup> smysl závěrečné otázky příkladu se bude různit v závislosti na mnoha věcech. Např. kdo a komu ji klade: co když ji takto sami sobě kladou Mark a Julie?<sup>10</sup> Radikálnější formou této kritiky může být argument, že hodnocení předloženého příkladu v takto osekané formě, a publikem, které k němu nemá žádný osobní

9 Viz komplexní Strandbergovu (2020) kritiku.

10 Jak upozorňuje Strandberg (2020, s. 202n), v takovém případě by pro respondenty příkladu smyslem závěrečné otázky bylo „Co bych na to řekl já na místě Marka/Julie?“, což je filosoficky obtížnější otázka – jak vůbec „být někým jiným“, popř. jak „být někým jiným“ a přitom být pořád sám sebou natolik, aby otázka znamenala „co bych na to řekl já?“.

Tyto poznámky upozorňují na to, že morálně hodnotit nebo reflektovat jakoukoli situaci znamená dělat to *jako někdo* – pro koho na sebe situace bere určitý význam, neboť konkrétním (rozmanitým) způsobem souvisí s tím, co už se v životě *tohoto* někoho událo, děje, nebo teprve stane.



vztah, nemá nic společného s reálnými situacemi morálního hodnocení. Prokazuje jen tolik, že okamžitá reakce na takto předložený příklad v takovémto kontextu bude takováto. Haidt sám charakterizuje tuto reakci jako „morální ohromení“ či „zmatení“ (*dumbfounding*), a připisuje ji zmatku či zahanbení respondentů *z toho*, že pro svou odmítavou reakci, kterou *sami považují* za podepřenou *důvody*, žádné důvody nedokážou uvést. Pro něho je to důkazem, že morální hodnocení vyplývají z neracionálních intuic. Dalo by se to ale číst také tak, že způsob, jakým svůj příklad podává, je rozporný a chybí mu příliš mnoho detailů pro to, aby představoval „realistický popis možného sexuálního vztahu mezi sestrou a bratrem“.<sup>11</sup>

Uvažme analogický příklad:

Mark a Julie jsou vcelku dobří přátelé, vycházejí spolu bez problémů. Jednoho dne se náhodou potkají na aukci bezprizorných dětí,<sup>12</sup> a rozhodnou se, že by mohlo být zajímavé a zábavné být na jeden den rodičem. (Děti lze do 24 hodin bezplatně vrátit.) Přinejmenším by to pro oba byla zajímavá zkušenost. Nechtějí ve výchově nic zanedbat a chtějí být dobrými rodiči; když vydražený chlapec Markovi na jakousi otázku hrubě odsekne, Mark ho výchovně plácne přes ruku (samozřejmě ani zdaleka ne tak silně, aby mu způsobil zranění, újmu, nebo přetrvávající fyzickou bolest). Co si o tom myslíte? Byl jeho trest OK?

Respondent podobného cvičení by byl pravděpodobně zmaten v tolika různých smyslech, že by ani nevěděl, odkud začít. Snad by podobně jako

---

11 Kronqvist & Elgabsi (rkp.). Autoři předkládají rozbor řady dalších souvislostí normálně přítomných v morální reflexi příkladů incestu (ale toto pozorování platí nejen pro incest): co to znamená, že někdo je pro někoho sourozencem, jak a v jakých situacích se význam tohoto vztahu utvářel a jakým způsobem do toho, co pro někoho jeho sourozenec znamená *jako sourozenec*, zasáhne to, že se spolu vyspí. Sourozenci spolu obvykle vyrůstají od dětství – figurují např. jako motivující pojmy příkazů a zákazů na úvodních stupních morální výchovy (že druzí se nebijí a že se s nimi dělíme o bonbóny, se učíme chápat skrze to, že nemáme bít své sourozence a máme se s nimi dělit o bonbóny). Atd.

12 Haidt svůj obdobně bizarní příklad ponechává bez jakéhokoli dalšího vysvětlení; já mám za to, že trochu větší soucít se čtenáři je na místě, a navrhuji, že si tu mohou představit něco jako výchozí situaci zápletky *Starosty casterbridgeského*.

Haidtovi respondenti odpověděl „Ne“ a podobně jako oni by měl problém vysvětlit proč. Avšak těžko tvrdit, že hlavním důvodem jsou hluboce zakořeněné, nicméně explicitně obtížně obhajitelné intuice o morálně třaskavém tématu, jako jsou fyzické tresty dětí. O pohnutí otázkou, zda je fyzické trestání dětí morálně přípustné, a o to, jakého kognitivního druhu je zdroj tohoto přesvědčení, tu na prvním místě nejde. Není totiž jasné, zda tu vůbec máme co dělat s morálním příkladem, nebo spíše s nepochopením toho, jak vypadá morálně zatížený příklad jinde než v učebně či učebnici.<sup>13</sup>

Jakkoli je tento příklad záměrně bizarní, zdá se mi, že se od Haidtova příkladu neliší tak, jak by se na první pohled zdálo. Můj příklad se zdá matoucí, protože nakládá *zjevně* nepochopitelným způsobem se zkušenostně komplexními a morálně zatíženými pojmy jako rodičovství nebo trest. Trestem je například nějaký typ jednání či interakce, k níž dojde mezi dvěma lidmi, pokud jsou vůči sobě navzájem v adekvátně – i institucionálně – etablovaných vztazích. Jejich vzájemný poměr musí být např. vyjádřením toho, že je tu jistým způsobem oboustranně uznáván vztah legitimní autority. Nezachází ovšem Haidtův příklad podobně problematicky jako já s pojmem trestu, s pojmy jako „rozhodnout se“, že by mohlo být „zábavné“ mít sex, v přesvědčení, že mohou dostatečně informativním způsobem fungovat i ve vzduchoprázdnu, jímž je obklopen? Haidtova volba příkladu je vedena snahou využít incest jako intuitivně fungující příklad, nezávislý na pozdější „pomalé“ racionalizaci. Existují však také výzkumy naznačující, že incestní tabu je spíše myšlenkovou konstrukcí, jejíž pomocí je ve společnostech stabilizován a materializován vnímaný význam příbuzenských vztahů;<sup>14</sup> *někdy* je např. toto tabu vázáno na ukazatel soužití v jedné domácnosti, bez ohledu na to, zda se jedná o pokrevní příbuzné.<sup>15</sup> Z těchto a podobných

13 Viz Leviho (2000, s. 25, 127, 137, 144 aj.) systematickou kritiku nefunkčních školských příkladů, v nichž nefigurují „postavy“, v jejichž ústech by cokoli, co (podle příkladu) mají říkat, mohlo znít přirozeně, nebo naopak takřkajíc *out of tune* (srov. Ebersole, 2002, s. 327). Markovo a Julino „rozhodnutí, že by mohlo být zajímavé a zábavné...“ nezní ani přirozeně, ani *out of tune*, protože Mark a Julie nejsou postavy ve stejném smyslu, v jakém jimi jsou Raskolnikov, Anakin Skywalker nebo skutečné osoby (které také mohou být postavami příkladu). Viz k tomu také Beran (2021, kap. 2.3 aj.).

14 Viz např. Read (2014, s. 172).

15 Antfolk et al. (2012).

důvodů se zdá namíste opatrnost, pokud jde o výklad přesné *povahy* „intuitivní“ nevole, kterou v respondentech Haidtův příklad budí.

Přesvědčení, že morálka je věcí intuic, které jsou okamžitými reakcemi souhlasu či nesouhlasu, navíc pomíjí fakt, že morální pojmy jsou materiálem také takových forem reflexe, které *nejsou* souhlasem ani nesouhlasem. Jak budu např. uvažovat o životním příběhu Kurta Gersteina a o jeho konci?<sup>16</sup> Mnohé z pojmů, které se v mé reflexi tohoto příkladu, pokusím-li se ji odít do slov, objeví – nenávisť, nacionalismus, výčitky svědomí, přetvářka, odvaha, zoufalství, kolaborace –, jsou pojmy „morálně hutné“ (*thick*),<sup>17</sup> a mé čtení tohoto příkladu je nevyhnutelně morální. Příklad Kurta Gersteina ani není vlastně možné informativně a relevantně popsat bez použití neredukovatelně morálních pojmů; a přece výsledná reakce nemusí mít podobu schválení či odsouzení, a to ne proto, že Gersteinovo „morální skóre“ vychází na „remízu“. Můžu pocítovat lítost či soucit (nejen ve smyslu jednostranné lítosti nad někým, komu pouze bylo ublíženo), či úžas z jakési neproniknutelné morální hlubiny – být „morally dumbfounded“ ve zcela odlišném smyslu, než o jakém mluví Haidt. Haidt, jako psycholog, jistě nemusí znát nejnovější vývoj morální filosofie. Ale vzhledem k tomu, že jeho přiznanou ambicí je právě přerámování diskuse o morálce *vůbec*, resp. neuspokojivého stavu, v němž ji zanechala filosofie, zanechává v člověku rozpaky otázka, co si má s Haidtovými pojmovými nástroji a výsledky počít, chce-li se postavit např. k problému morální náhody (*moral luck*).<sup>18</sup> Jedním z výsledků, k nimž diskuse o morální náhodě vedla, je úvaha, že skutečné jemnosti a ostrosti nabývá naše „morální oko“ tam, kde tváří v tvář reflektované situaci nemůžeme než *zdržet se* soudu či odsouzení.<sup>19</sup> Toto zdržení se přitom nemusí být *ex post* racionalizací; soucítící oko může, přesně jak

16 Gerstein, synek z dobré („čistokrevné“) pruské rodiny s vojenskou tradicí vstoupil na počátku 30. let do NSDAP a SA, aby se s nimi později pravděpodobně z nábožensky motivovaných důvodů rozešel a byl dokonce v druhé polovině 30. let krátce ve vězení a koncentračním táboře. Na počátku 2. světové války však vyvinul značné úsilí, aby byl do strany zpět přijat, a vstoupil také do Waffen-SS, dost možná však již s úmyslem pracovat proti systému „zvnitř“. Za cenu osobního rizika shromáždil informace o nacistickém vyhlazování Židů a předal je druhé straně. Na konci války se vzdal spojeneckým silám, byl uvězněn jako podezřelý z válečných zločinů a ve vazbě spáchal (patrně) sebevraždu.

17 Williams (1985, s. 129n, 140n).

18 Williams (1981), Nagel (1979).

19 Viz např. Browne (1992), Pacovská (2018).

to požaduje Haidt, jaksi přímo vidět, že „kdo jsem já, abych tě soudil“.

Ať už jsou tedy morální intuice čímkoli, zdá se méně jasné, než Haidt předpokládá, že nejlepším zdrojem k jejich porozumění bude jejich *experimentální testování*. Bez ohledu na to však stojí za to vrátit se i k onomu „ať už jsou morální intuice čímkoli“. Už jsem zmínil, že Haidtovo vymezení „intuice“ je *negativní*: to, co nemá povahu racionálního, „pomalého“ rozvažování. To, že si neláme hlavu s koherentní charakteristikou vztahu mezi emocemi a intuicemi, je ovšem nešťastné. Někdy mluví o tom, že intuice emoce „zahrnují“. Jindy prezentuje svou věc jako otázku primátu rozumového uvažování, nebo emocí v morálce; tehdy se zdá, jako by se Haidtova zjištění o intuicích rovnala zjištění o primátu emocí.<sup>20</sup> Jindy mluví o „intuicích a emocích“ jako o dvou různých věcech stejného řádu, které lze klást vedle sebe.<sup>21</sup> *Filosofický* rámec Haidtových sympatií k pojmu emocí tvoří v první řadě Hume a jeho pojetí vášní (*passions*), které Hume popisuje podobným způsobem jako Haidt intuice. Jak už jsem ale výše zmínil, Haidtův projekt negativně poznamenává jeho pojmová nejasnost. *Pouze* pokud je intuitivně zaujatý hodnotící soud (přesvědčení), intuitivní percepce jisté morální kvality situace ve zhruba huemerovském smyslu a intuitivně pocítená morální emoce *tímtéž*, nebo na rozdílech mezi nimi nezáleží, může se toto pozorování dovolávat na svou podporu humovského argumentu o primátu vášní – nebo naopak být uváděno na jeho důkaz.

Nedávná velmi diverzifikovaná debata, již do filosofie vnesl „obrat k emocím“, zůstává v Haidtově analýze opomenuta. Haidt tak ignoruje jemné rozbory emocí jako specifických *kognitivních* postojů, které navzdory svým instinktivním kořenům vyžadují dlouhé zrání a kultivaci, tu kulturně „osmotickou“, tu vědomou, aby skutečně byly tím, čím jsou.<sup>22</sup> To platí specificky právě pro *morální* emoce jako slitování nebo výčitky svědomí. Tyto emoce jsou tím, čím jsou, *bez ohledu na to*, zda „naskočí okamžitě“, nebo zda se člověk potřebuje „na okamžik zastavit“ a cosi podstatného si připomenout. Ke schopnosti slitování s provinilcem či schopnosti pociťovat výčitky svědomí (třeba proto,

---

20 Haidt (2008 [2001], s. 1036nn).

21 Tamtéž, s. 1029, 1044.

22 K tomu viz např. Solomon (1973), Nussbaum (2004).

že jsem s druhým neměl dost slitování), vede značně dlouhá cesta, kterou popisujeme pojmy jako dospívání, osobní vývoj, „zmoudření“. Je těžko představitelné, že by takové emoce vůbec mohly *nestát* až na konci dlouhé cesty; a schopnost pociťovat je rozhodně není triviální charakteristikou každého „normálního“ dospělého v té které kultuře. Autentičnost jejich statusu jako morálních emocí nebo míra jejich významu v našem morálním životě se ale nepoměřuje délkou cesty, kterou je k emoci nutno urazit, přinejmenším pokud nám jde o něco víc, než jen konstatovat primát emocí, k nimž vede cesta kratší, jako je odpor vyvolávaný, za určitých okolností, představou incestu. Jak jsem řekl, není zjevné, jaký přesně je podle Haidta vztah mezi emocemi a intuicemi, ale vzhledem k tomu, že *poučená* filosofická debata o emocích pracuje se sofistikovaným a přesným pojmem emoce, je Haidtova mlhavost pro jeho vlastní projekt kontraproduktivní.<sup>23</sup>

Představuje totiž problém pro strukturu jeho argumentu. Pro Haidta – jak uvidíme dále – je podstatné, že vše to, o čem mluví jako o intuici, resp. intuitivním zdroji morálního postoje, je fenoménem jednoho druhu, jehož původ či takřkajíc racionále a mechanismus ustavení je také vždy téže povahy, navzdory obsahové rozdílnosti těchto postojů. Tento hypotetizovaný jednotný druh původu je klíčový pro Haidtovu interpretaci jednak samotného charakteru našich intuitivních morálně hodnotících reakcí, a jednak – to zvláště – případů, kdy jsou tyto reakce ve vzájemném rozporu. Haidt případy tohoto typu diskutuje v klíčovém kontextu politických střetů, jemuž budu věnovat následující oddíl. Pokud by však jeho předpoklad jednotné povahy morálních intuic, a tedy jednotného zdroje všeho, co klasifikuje jako intuici nebo intuitivní, byl zproblematizován, kompromitovalo by to podstatně i základní soudržnost Haidtových morálně-psychologicko-politických analýz.

Nechci podezírat Haidta, že design jeho výzkumů je podřízený úmyslu prokázat, co si umínil prokázat. Okamžité reakce na podobně formulovaná cvičení jsou nepochybně zajímavým fenoménem samy o sobě a studium si zaslouží. V nějakém smyslu však čtenáře za nos

---

23 Pro obdobné (příkřejší) hodnocení vágnosti Haidtovy teorie, včetně užitečného shrnutí řady praci podávajících podobnou kritiku, viz Narvaez (2010, s. 166n).

vodí, neboť jim sugeruje vágní pojem intuice, prezentovaný ovšem jako jednoznačný, i díky tomu, že na rozdíl od filosofů hloubajících ve svých kreslech on „dělá vědu“. Haidtův předpoklad primární důležitosti toho, odkud se intuice berou, se přitom promítá do důsledků jeho přesvědčení, že jejich experimentálním zkoumáním se lze dovědět něco zásadního o lidské morálce a morálním životě vůbec. K tomu se blíž dostanu v oddíle 3; teď ale: odkud se tedy naše intuitivní hodnotící reakce berou? Jsou-li svou povahou automatické, do značné míry vrozené a zároveň vzhledem ke kulturně rozmanitým morálním fenoménům *explanačně primární* na této své vrozené, před-kulturní úrovni, mohou být jediné výsledkem sdíleného evolučního dědictví lidského rodu. Jsou jakýmsi „prodloužením“ a reformulací úspěšných adaptačních strategií, s jejichž pomocí se lidstvo muselo vypořádat s překážkami, které potkalo na cestě svého vývoje. Tento důraz na dědictví našich evolučních předků umožňuje Haidtovi jednak implicitně obhájit podstatný rys našich morálních reakcí – jejich okamžitou, před-rationální automaticnost – a jednak na něm dále stavět. Z toho vyplývají určité důsledky pro Haidtovy pozoruhodné analýzy našeho politického života, resp. jeho morální, či moralistické dimenze, jimiž se budu zabývat v následujícím oddíle.

## 2. Intuice v politickém životě

V tomto oddíle se budu zabývat Haidtovými analýzami „politického života“ (zvláště v USA) a jeho komentáři k němu. Chci nicméně předeslat, že mi nejde o „politickou polemiku“ s Haidtem v běžném smyslu, v němž spolu političtí analytici či komentátoři nebo lidé vůbec „nesouhlasí o politice“. K některým Haidtovým poznámkám a úvahám mám i výhrady tohoto typu, primárně mi však jde o něco jiného: o diskusi jeho systematického pojetí toho, odkud se berou motivace našich politických postojů, a důsledků, které z tohoto pojetí vyplývají pro analýzu případů politického nesouhlasu a pro možnosti politické kritiky či obecněji proměn politických postojů.

Haidt své výzkumy morálních intuic využil k pozoruhodné aplikaci na otázky politické kultury a neshody ve své knize *The Righteous Mind*.

Probírá v ní rozmanité aspekty klinče, v němž se ocitli američtí republikáni a demokrati v průběhu posledních dekád. Haidtovým přesvědčením je, že obě strany se navzájem nechápou a neprávem se navzájem podezírají z nemorálních motivací svého politického stanoviska. Adekvátní analýza aplikuje důsledně deskriptivní metody studia morálky; v tomto je Haidt poučen pracemi antropologa Richarda Shwedera,<sup>24</sup> který byl ve své generaci sociálních vědců jedním z nejvýznamnějších zastánců myšlenky kulturního pluralismu.<sup>25</sup>

Haidt líčí tento náhled v rámci autobiografického narativu o tom, jak byl zprvu přesvědčen o univerzální platnosti liberálních morálně-politických hodnot a reagoval odmítavě při své návštěvě Indie na podřízené postavení ženy v tamní společnosti. Časem mu však došlo, že se nedostal do společnosti nemorální, nýbrž společnosti řídící se morálkou odlišnou, v jejímž rámci není ústřední důraz kladen na autonomii jednotlivce, nýbrž na komunitu a její vnitřní vazby.<sup>26</sup> Kulturní rozmanitost morálních kódů pak vyplývá z toho, že morálka čerpá sice ze zdrojů jednoho druhu, nejedná se však o zdroj jediný, nýbrž několik, a v různých kulturách jsou akcentovány různé dílčí zdroje nebo kombinace zdrojů. Haidt identifikuje zprvu 5 zdrojů morálky, spočívajících v intuitivních hodnotících reakcích aplikujících morálně silně zatížené pojmové protiklady: i) péče vs. ublížení, ii) férovost či spravedlnost vs. podvádění, iii) loajalita vs. zrada, iv) autorita (respekt k autoritě) vs. její podvracení, a v) posvátnost vs. její degradace či profanace.<sup>27</sup> Liberální Haidt byl pohoršen společenským postavením indických žen z důvodů akcentujících i) a ii) – tj. vnímal je jako příklad ubližující nespravedlnosti –, „zmoudřelý“ Haidt tento postoj korigoval uvědoměním si důležitosti iii) a iv). Tj., uvědomil si, že jde o příklad společenského uspořádání vyjadřujícího široce sdílené hodnoty ne-individualistické zodpovědnosti vůči většímu celku a uznávání autorit, bez jejichž přítomnosti a ručení není dobré společenské uspořádání možné.

24 Haidt (2012, s. 21nn).

25 V 70. letech byl Shweder jedním z prvních sociálních vědců, kteří opět po čase psali se sympatiemi o Whorfovi.

26 Haidt (2012, s. 118nn).

27 Tamtéž (kap. 7; přehledné shrnutí na s. 178n).

Tyto hodnotící reakce či perspektivy jednotliví členové společnosti uplatňují intuitivně, nereflektovaně; nejsou primárně výsledkem deliberace, nýbrž vyjádřením evolučně etablovaných dispozic vnímat a ve svém vnímání a reakci zdůrazňovat salientní rysy situací, v nichž se člověk nachází. V podstatě je přimknutí se k hodnotám péče, férovosti, loajality, respektu k autoritě a posvátnosti podle Haidta výsledkem strategií adaptace na formy stresu, které kdysi ohrožovaly komunitu a její *fitness*. Dnes jsou sice typickými spouštěči těchto reakcí jiné fenomény, ale původně šlo o to, jakým způsobem správně reagovat na podněty jako nářek dítěte, ohrožení pro skupinu, nebo přítomnost nebezpečné nemoci či znečištění. Různé formy našich morálně akcentovaných reakcí na podněty toho typu, o nichž bývá vedena politická debata, jsou transformacemi stejných zdrojů, z nichž vycházejí všechny naše morální intuice. Oproti případu incestu, kde Haidt shledává jednotnou intuitivní reakci, jsou politické reakce markantně diferencované. Ne však proto, že by se opíraly o různorodé zdroje toho druhu, že někteří lidé mají k dispozici pouze některý z nich a jiný ne, a jiní lidé zas naopak. Pro Haidta je podstatné, že zdroje těchto reakcí jsou stejného druhu; tento předpoklad motivuje jeho akceptaci statusu jich všech jako postojů *morálních*, resp. jeho odmítnutí rozlišovat mezi různými protichůdnými postoji. Dále uvidíme, že Haidtovo pojetí se jen za cenu jistých potíží vypořádává s protipříklady: intuitivně působícími postoji, které ale můžeme být neochotni klasifikovat jako morální, nebo okamžitými morálními postoji, které ale nemusí být v Haidtově smyslu intuitivní.

Jelikož různost politických postojů nevyplývá podle Haidta ze zakotvení ve zdrojích různého druhu, jeho výklad v nich spatřuje výsledek nestejně *míry* implementace toho či onoho intuitivního zdroje v praxi. Důvody této nestejnosti vyplývají víceméně z různorodosti kulturního zázemí a původu různých lidí. Nikoli tedy primárně z rozdílů ve vzdělání, znalosti různých cizích jazyků, specifických typů životní zkušenosti (profesní, cestovatelské, válečné, spirituální aj.), rozdílů socioekonomických a třídních, atd.

V tomto rámci poskytuje Haidtovo autobiografické vyprávění o Indii vstupní bránu ke komplexní analýze střetu na americké politické scéně. Liberální moralismus, či moralismus demokratů se opírá o význam



klíčových protikladů i) a ii), a iii)–v) do značné míry ignoruje, zatímco konzervativní, či republikánský moralismus akcentuje výrazněji právě iii)–v), byť ne zcela jednostranně; podle Haidta je konzervativně-republikánský moralismus vcelku úspěšnou snahou o alespoň nějaké zapojení všech morálních os.<sup>28</sup> Toto pozorování je podle Haidta samého deskriptivní: popisuje, jaké jsou *reálně* základy našich morálních intuic. Jeho konstatování o větší vyváženosti politicko-moralistní krmě servírované republikány má tedy ambici být primárně observačním konstatováním, které zahrnuje hodnotící složky nanejvýš ve smyslu „co je psychologicky účinná politická rétorika“, spíše než „co je morálně správné“.

Určité problémy – Haidtem neřešené – pro toto pojetí vyplývají z případů, kdy jdou různé osy morálky proti sobě. Haidtova diagnóza je prostá: v pozadí stojí jednostranný důraz na pouze některou morální osu. Představme si zde typické příklady, které jsou palivem aktuálně se zostřující kulturně-politické rozepře: Právo na interrupce (svobodné sebeurčení ženy vs. posvátnost nenarozeného života)? Právo na kritiku náboženského přesvědčení (svoboda slova vs. degradace posvátného)? Protesty Black Lives Matter (rovnost a spravedlnost vs. autorita a hierarchie)? Komprehensivní veřejný systém zdravotního a sociálního pojištění (péče o potřebné vs. svoboda a útlak, resp. spravedlnost a proporcionalita ve svém libertariánském čtení)? Z Haidtovy diagnózy vyplývá, že smíru lze docílit pouze tak, že se podaří zohlednit odpovídající aspekt všech pěti, či šesti základních morálních motivů. Právě u výše naznačených případů – které jsou spíše charakteristické, než výjimečné – ale není jasné, jak to udělat. Právo na svobodné sebeurčení ženy v nakládání s vlastním tělem jednoduše dost dobře se zohledněním posvátnosti nenarozeného života v běžném nábožensky inspirovaném pojetí smířit nelze. Zdánlivě „kompromisní“ řešení běžná v řadě evropských zemí, umožňující legální interrupci pouze do určitého týdne těhotenství, jsou všechna nábožensky nepřijatelná úplně

---

28 Tamtéž (kap. 8). Po určitou část knihy pracuje Haidt s tímto pětiosým modelem; postupně jej však aktualizuje i na popud svých pozorování americké politiky a přidává osu šestou: svoboda vs. útlak, kterou (ovšem v různých interpretacích) sdílejí obě politické strany (s. 197nn), což si zároveň vyžaduje jistou reinterpretaci osy ii), díky níž ji lze (opět v různých čteních) identifikovat jako silně akcentovanou oběma stranami (s. 205nn).

stejně, bez ohledu na to, kterým týdnem je vedena hranice. Jako jediná možnost by se zdál svět bez interrupcí, ve kterém je zároveň každé těhotenství dobrovolné, chtěné a aktivně přijímané; což ale, příznačně, je „kompromis“ *de facto* odpovídající preferovanému obrazu světa pouze jedné ze znesvářených stran.

Nemám zde v úmyslu vést diskusi o problému morální přípustnosti interrupce; spíše o tom, jaké světlo tyto příklady vrhají na Haidtův předpoklad *integrovatelnosti* morálky. Zajímavá otázka je, jak je možné, že různé morální motivace, které byly – podle Haidtovy teorie – *původně* motivovány *koherentními* ohledy (přežití a prospívání skupiny), nyní koherentní nejsou. Patrně nejjednodušší hypotéza je, že jejich vývoj – ať už jen některých z nich, nebo všech – byl natolik dramatický, že se od sebe vzdálily a dnes už jednoduše slučitelné nejsou. Pak se nabízí dvě možnosti: Jednou z nich je myšlenku integrovatelnosti morálky opustit, a to i v oslabenější podobě haidtovského pluralismu. Protože dokonce i když přistoupíme na myšlenku různých lokálně specifických morálek, nebude jasné, na základě čeho jim ještě říkat morálky – mnoho případů takové různosti bude toho druhu, že dvě různé sledované „morálky“ nebudou sdílet žádný rys, který by umožnil komukoli z jedné strany myslet cizí morálku jako něco, co je stále ještě variantou téhož, co má on. Pouhá pojmová „setrvačnost“ podložená sdíleným původem v kdysi společném projektu přežití a prospívání skupiny se totiž ve výsledku zdá podobně matoucí, jako kdybychom dnešním savcům, ptákům a plazům říkali stále „dinosauři“.

Druhá možnost je připustit, že potřeba takového pojmu morálky, který svým uživatelům dává koherentní smysl, vyžaduje možnost popření legitimacy některé z Haidtem předpokládaných „os“. Pokud je např. možné aplikovat určitou „metriku“ rozdílu či rozporu mezi originální expresí té které adaptační motivace a její současnou podobou, můžeme připustit, že internalizace morálně významného rozdílu mezi péčí a ublížením má dnes podobnou podobu a smysl jako (pravděpodobně) kdysi. Naopak formy, jež respekt k posvátnosti na sebe bere dnes, mají často velmi málo společného s Haidtem hypotetizovaným komunitním tíhnutím k hygieně. Problémem, byť poněkud odlišným, může trpět i hodnota respektu k autoritě. Respekt k autoritě dnes sice patrně vypadá podobně

jako „kdysi“, méně jisté než kdysi však je, že ve všech svých dnes typických podobách je ku prospěchu společenství. Sama *možnost klást si tyto otázky* – „(kdy) je poslušnost vůči autoritě morálně správná?“ – je pro morální život dnešních lidí charakteristická.<sup>29</sup> Jedním možným vysvětlením toho, proč liberálové odmítají akceptovat politickou hodnotu respektu k autoritě, může být to, že to žádná nepodmíněná hodnota není. Respekt k autoritě není sám o sobě hodnotou, na níž by podle nich bylo možné založit politiku usilující o to, co je správné. Haidtovou okamžitou námitkou by bylo, že nechce stanovit, která politika je morálně správná, nýbrž ukázat, jaké moralistické akcenty mohou určitým způsobem formulované politice chybět, a tím ji oslabovat. V tom má nepochybně do velké míry pravdu, ale pak je jeho analýza spíše analýzou psychologicky úspěšné politické rétoriky, než analýzou morálních postojů; alespoň pokud má „morálka“ samostatný smysl, který není vyčerpán tím, v čem spočívá psychologicky úspěšná politická rétorika.<sup>30</sup>

Úspěšnou politickou strategii lze totiž postavit na velmi různých základech, a – což je významné – i o nich má smysl klást vůbec otázku, zda

---

29 Blum (2013, s. 305) poznamenává, že zatímco odstranění utrpení a spravedlnost jsou takřka jisté nepodmíněné morální hodnoty, klíčové konzervativní hodnoty – s nimiž mladší, levicově-liberálně smýšlející lidé mívají „problém“ – jako loajalita vůči celku a respekt vůči autoritě právě připouštějí další otázky tohoto typu: Jaké autority bych měl respektovat? Ke komu bych měl být loajální? Autorita musí být morální autorita, větší celek musí sledovat morálně správný cíl, aby požadavek loajality a respektu k autoritě mohl být *morálně* oprávněným požadavkem.

Jsem si méně jistý než Blum tím, že pojem morálky neposkytuje prostor pro kladení otázek jako „(kdy) je morálně správné neublížovat druhým?“, zdá se mi ale, že tento prostor přece jen je, v porovnání s otázkou autority, poněkud užší. Je vázán např. na specifické debaty typu „spravedlivé války“, spíše než aby byl morálnímu tázání, použijeme-li tento nadnesený výraz, vlastní „gramaticky“.

Blumův zlehka načrtnutý kritický argument lze expandovat např. takto: Jsou-li jednotlivé díle morální osy podřízeny celkovému účelu morálky, musí být možno tu či onu dočasně s ohledem na tento celkový účel suspendovat. Je ovšem značný rozdíl, zdali je s ohledem na celkový prospěch komunity ublíženo nějakému jednotlivci (srv. Gaitovu [2006, s. 261] diskusi problematiky mučení), nebo zda je s ohledem na celkový prospěch komunity nerespektována některá náboženská zvyklost (důsledky amerických zákonů proti polygamii pro mormonskou církev). Oboje může být nutné, a politická praxe také tak často v reálu vypadá, tato nutnost je ovšem právě politicko-pragmatického charakteru. Zatímco v prvním případě se však zdá, že *morální* problematičnost takového jednání zůstává problematičností i přesto, že jednání akceptujeme jako nezbytné z jiných důvodů, politicko-pragmatických, druhý z těchto případů se ukazuje jako střet politické autority s náboženskou, bez morální dimenze.

30 I pak ovšem není jasné, jak vůbec mohou navzdory tomuto svému dlouhodobému znevýhodnění demokratičtější kandidáti vítězit ve volbách (Blum 2013, s. 307).

jsou morálně správné. Bez ohledu na to, jak v tom kterém případě bude znít odpověď, *sama možnost* tuto otázku klást a zásadní význam, který pro nás tato možnost má, se Haidtovu pojetí podstatně vymyká, resp. není jasné, jaké by v něm měla mít místo. Uvažme tento příklad: Četné politickovědné studie zkoumají vzestup populistických politiků a hnutí v posledních letech, ovšem v první řadě jako otázku použité politické rétoriky, její synergie s historickou, kulturní, sociální a ekonomickou situací daných zemí, atp. Úspěšný politik je bezpochyby – z definice – úspěšný díky tomu, že oslovil víc voličů než politik méně úspěšný či zhola neúspěšný; to je ale víceméně všechno. Pokud budeme předpokládat, že *úspěšného* politika činí úspěšným *nutně* to, že brnká na větší množství *morálních* strunek, ztratíme jakýkoli nástroj rozlišování mezi morálkou a popularitou či konformitou. Haidt ovšem, příznačně, ztotožňuje morální výchovu, či dosažení ctnosti, s „plnou enkulturací“.<sup>31</sup>

Jedním z důsledků bude, že z pojmu morálně motivované politické kritiky se stane *contradictio in adiecto*. To platí přinejmenším v okamžiku *vystoupení* s takovou kritikou. Haidt sám opakovaně zmiňuje inspirativní postavy hnutí za občanská práva 60. let jako Martin Luther King; o tom, zda je motivace či sdělení takových kritických postojů skutečně morální, ovšem jako by rozhodoval „test přesvědčivosti“. Teprve v okamžiku, kdy se takový postoj dostatečně rozšíří alespoň v jedné rozumně velké společnosti, můžeme konstatovat, že prokázal, že se opírá alespoň o některou složku našich vrozených morálních intuic. I Haidt přiznává lidské společnosti potenciál plasticity a vývoje; tj. připouští možnost přesvědčit většinu společnosti o nějakém morálním postoji, který předtím nesdílela. I tato možnost je však podmíněna tím, že využívá potenciál dočasně „spícího“, nevytěžovaného intuitivního důrazu na některý z Haidtem mapovaných morálních protikladů. Naproti tomu představa morálně motivované kritiky dlouhodobě většinově úspěšné politiky, kritiky, která *má smysl*, i když se jí dlouhodobě nedaří ničeho podstatného dosáhnout v podobě širší proměny společenského mínění a postoje, předpokládá jiný pojem

31 Viz k tomu kritiku v Narvaez (2010, s. 166). Haidt v těchto postřezích o významu enkulturatione v morálce deklaruje svou blízkost etice ctností; její přední představitelé jsou si však důležitosti udržení rozdílu mezi konformitou a ctností ostře vědomi – srv. např. Annas (2001, kap. 3), rozpracovávající Aristotelův postřeh, že každý usiluje o dobro, nikoli o opakování toho, co dělali jeho rodiče.

morálky než Haidtův. V intencích Haidtova pojetí takový pojem morálky nedává dobrý smysl. Je otázka, co z toho vyplývá pro chápání *morálního* statusu např. utopických myšlenek lidské rovnosti, majících dlouhou náboženskou i filosofickou tradici, jestliže různé formy rasové, etnické, národnostní nebo sexuální diskriminace se historicky i v současnosti zdají být spíše pravidlem než výjimkou.

S tím souvisí, že Haidtův model neumožňuje či se příliš nesnaží rozlišovat mezi morální motivací politických postojů, a tím, že sami političtí aktéři považují motivace svých politických postojů za morální. Či jinými slovy, v Haidtově pojetí není možné něco za morálně správné – pakliže se jedná o sentiment dostatečně rozšířený a společensky etablovaný, a zároveň spontánně vyjadřovaný – *pouze považovat*. Nelze dost dobře vysvětlit, proč je přesvědčení o správnosti rasové segregace na americkém Jihu rasistický předsudek, a ne prostě lokálně platná morálka<sup>32</sup> – vzpomeňme na „zmoudřelého“ Haidta v Indii. Příkladnějším rasistickým předsudkem dost dobře nemohlo být do té doby, než hnutí za občanská práva dosáhlo v šedesátých letech jisté širší proměny společenské atmosféry.

Příklady, na nichž Haidt dokládá nutnost reformulovat pětiosý systém v šestiosý, jsou v tomto ohledu výmluvné. Jak se ukázalo, jeho původní model totiž nepostihoval morální stanovisko lidí, které trápí např. tohle (a dále cituje z emailů svých čtenářů):

„Volím republikány, protože jsem proti tomu, aby mi někdo (úřady) bral moje těžce vydělané peníze a dával je neproduktivním pobíračům dávek, svobodným matkám rodičím drogově závislé děti, ze kterých nebude nic jiného než demokrati... Vy jste jenom převzdělaný ,filosof s pěstěnýma rukama, kterého platí za kladení stupidních otázek a za hledání ,rozumných‘ odpovědí... Přestaňte brát drogy a přečtěte si Junga.“

Jiní jeho nespokojení čtenáři ventilují nesouhlas s volbou demokratů takto: „Nesnášíte lidi, kteří si své peníze vydělávají tvrdou prací, žijí si

<sup>32</sup> LaFollette & Woodruff (2015, s. 459).

vlastní život a nespolehají na vládu od kolébky do hrobu.“ Nebo: „máte pět dětí s třemi různými chlapy a potřebujete sociální dávky.“ Haidt výslovně říká, že tyto projevy doslova „překypují morálním obsahem“, který se „vesměs týká spravedlnosti (férovosti)“.<sup>33</sup> Proč ale klasifikovat tyto projevy jako projev morálního zaujetí, a ne např. jako výlev třídní či rasové nesnášenlivosti?<sup>34</sup>

Haidtova v této pasáži spíše implicitní pozice je taková, že se o projev morálního zaujetí jedná díky tomu, že je tento postoj motivovaný *intuicí*, a postoj *hodnotící* – reagující chválou či odsouzením. Pokud není možné, aby intuitivní reakce tohoto široce rozšířeného druhu byly motivované jinak nebo vycházely z jiného zdroje než takového, který zásadním způsobem reprezentuje to, „o čem“ je morálka, je ovšem Haidtova pozice nevyhnutelná.

Haidtův relativní nezáměr o význam či samu existenci možnosti, že jde o výlev rasové nesnášenlivosti, zbytečně uzavírá prostor pro důležité otázky. Např.: je možné, aby rasistický výlev byl *zároveň* morálním postojem? Popř. *jak* je to možné? Pokud to možné je, naznačuje to, že „morální“ rodokmen může mít tak mimořádně široké spektrum postojů, že pro to, abychom jim připsali status morálního postoje, nemusí být samotný fakt toho, jaký mají rodokmen, dostatečný. A pokud *nemůže* být zároveň obojím, pak připouštíme možnost, že samotná přítomnost morálního či moralistického slovníku v artikulaci nějakého postoje nemusí mnoho znamenat, a pořád může jít o *rasistický* výlev. Ovšem, různé formy těchto protiargumentů stále předpokládají, že jakmile by už tentýž člověk, ideálně Haidt sám, připustil, že něco je rasistický výlev, připustil by zároveň, že takový postoj není morálně obhajitelný, a tím by se pojmy morálky a přítomnosti intuitivních hodnotících reakcí rozešly. A to by ovšem představovalo podstatné *omezení* Haidtova deklarovaného postoje morálního pluralismu.<sup>35</sup> Opačnou, „silně pluralistickou“ variantou

33 Haidt (2012, s. 195n).

34 Srv. Blum (2013, s. 314); též Carpenter (2012) o stabilním silně racializovaném podtextu podobných komentářů (bez ohledu na to, zda explicitně odkazují na rasovou problematiku nebo ne).

35 Ochota absolutizovat, nebo prioritizovat některou z hodnot nad rámec Haidtem proklamovaného pluralismu, by byla jistou obdobou toho, jak v praxi často funguje Haidtova osa posvátného, která např. v případě práv LGBT+ lidí jednoduše neguje jiné politické motivace. Jak poznamenávají LaFollette & Woodruff (2015, s. 459), příklad gayů ukazuje, že různé politické hodnoty si nemusí

bude závěr, že fakticky není možné mluvit o rasismu jako nemorálním mimo *kontext* kultury (subkultury, třídy, sociální vrstvy, společenského *milieu*, ...), v níž je toto přesvědčení obecně rozšířenou *konvencí*. Pak se ovšem stírá rozdíl – a v Haidtově výkladu skutečně není vždy jasné, zda tento rozdíl drží – mezi možnostmi smysluplně mluvit o rasismu jako nemorálním a možnostmi o nemorálnosti rasismu jednoduše přesvědčit někoho z „opačného“ tábora.

Zčásti nezávislým problémem pak je, zda „morálním obsahem přetékající“ postoje, které ovšem aktivně pracují s rasistickými či třídními stereotypy (jejichž nepravdivost prokázaly sociálněvědné výzkumy), skutečně přispívají k Haidtem naznačenému účelu morálky: soudržné kooperující, prospívající společnosti. Nepřispívá k vytvoření a udržení takovéto společnosti spíše aktivní boj s rasistickými předsudky? To není řečnická otázka, ani otázka jednoduchá, a na její odpovědi záleží. Haidt si otázky tohoto typu, zdá se mi, klást nechce, což je vzhledem k povaze jeho výzkumu pochopitelné a legitimní. Nelze se pak ovšem divit, stávali se terčem kritiky podnícené těžko zapuditelným dojmem, že se *něco jako* rasistické postoje spíše pokouší *něco jako* pochopit či legitimizovat.

Haidt se svým modelem snaží co nejpoctivěji učinit zadost stanovisku tábora republikánů. Jedním z úskalí, k němuž ho tato chvályhodná vstřícnost přibližuje, je nevyslovený předpoklad, že diverzita politických stanovisek a problémů a politických moralismů tato stanoviska motivujících v té podobě, jak ji on čte na politické scéně USA, reprezentativním způsobem mapuje diverzitu politického moralismu či politiky vůbec. Mnohé detaily, které zmiňuje, se však zdají být specifické spíše pro aktuální a nahodilý stav společensko-politické debaty v USA po reaganovském neoliberalním obratu 80. let a bushovským neokonzervativním obratu let nultých a s nimi spojených „kulturních válek“. Naproti tomu např. „sociálně-demokratická“ politická pozice v evropském smyslu slova v Haidtově modelu víceméně absentuje a není zcela jasné, jak ji pomocí jeho pojmových nástrojů popsat.<sup>36</sup> Podobný

---

být navzájem rovny a nemusí mezi nimi vůbec docházet k žádnému vyvažování (které někdy dopadne tak, někdy onak), protože „spíše než že by mnozí konzervativci měli oproti liberálům nějaké morální chuťové pohárky navíc, jsou jejich chuťové pohárky zcela jiného druhu“, rozuměj: nejsou to *morální* chuťové pohárky.

36 Blum (2013, s. 301). Blum věnuje obsáhlejší pozornost jistě nedbalosti a vágnosti, s níž Haidt

problém by pro tyto popisné nástroje mohly představovat politické hodnoty jako sekularismus (jako pozitivní politická hodnota a tradice, nikoli pouhá *absence* důrazu na osu „posvátnosti“ v politice), nebo politická scéna v postkomunistických zemích střední Evropy, které vykazují prvky „neliberální demokracie“.

*Catch-all* přitažlivost republikánské strany, jejímuž vysvětlení je zasvěcen Haidtův komplexní výklad, tak nemusí vyplývat z naší přirozenosti, na jejíž vrozené „struny“ se jakoby náhodou naučila právě teď nejlépe hrát tato jedna konkrétní politická strana. Haidtovo porozumění, ba do jisté míry obdiv pro úspěšnost republikánské politické rétoriky (model cca 2010), může mít jednodušší vysvětlení: je mu zkrátka sympatická. Jestliže bere výlevy o zneužívání sociálních dávek<sup>37</sup> jako vážný a relevantní morální zájem, je to možná i proto, že napříč celým jeho dílem prolíná motiv kritiky takřkajíc přílišného zaopatřování.<sup>38</sup> Podobně jeho kritika požadavkem zaručených stejných výsledků (*equality of outcomes*) a jeho údajného směřování s rovnými příležitostmi (*equal opportunities*)<sup>39</sup> může být výrazem neporozumění. Haidt tvrdí, že *equality of outcomes* je něco, co liberálové „často požadují“, nicméně když už podá (anekdotické) příklady, jedná se o zjevně lokální a očividně problematické implementace dílčích vládních politik, jako jednostranné financování pouze dívčího veslařského družstva na jedné univerzitě.<sup>40</sup> To, co liberálové požadují *skutečně* často a systematicky – systém všeobecné zdravotní péče, robustní vzdělávací systém podporující studenty z ekonomicky slabších rodin, atp. – lze ale stěží charakterizovat jako požadavek zaručených stejných *výsledků*.<sup>41</sup> Haidtovy sympatie ke konzervativismu lze vyčíst i z jeho populárních výstupů pro širší publikum.

---

zavádí své pojmy „liberální“ a „konzervativní“, včetně nezohledněného, a pro Haidta poměrně typického problému, že stručná politická sebeidentifikace voliče/respondenta jako liberálního nebo konzervativního nemusí lícovat se skutečnými politickými postoji a hodnotami, které zastává a které by odkryl jemněji nuancovaný průzkum.

37 V USA se stejně jako u nás, nebo i kdekoli jinde, jedná o zcela zanedbatelný problém, ve skutečnosti jde o jistou tradici zakořeněnou politickou rétoriku (Coughlin 1989; novější populární přehled viz např. Nilsson 2020).

38 Srv. poufouchlou poznámku o evropských sociálních státech (Haidt 2012, s. 17). Problému „výchovy ke změkčilosti“ věnoval Haidt celou následující knihu (Lukianoff & Haidt 2018).

39 Haidt (2012, s. 204), Lukianoff & Haidt (2018, s. 223nn).

40 Lukianoff & Haidt (2018, s. 223n).

41 Blum (2013, s. 303).



Tak ve své starší přednášce vysvětluje náklonnost liberálů k sociálním změnám či dokonce revolucím transformací dětinské radosti z nových věcí, protože nové věci jsou „dobré, je to změna, je to zábava“. Naproti tomu konzervativní důraz na zachování stávajícího společenského řádu je výrazem hlubokého vhledu, že „nastolit řád je těžké. Řád je opravdu vzácná věc, a je opravdu snadné o něj přijít.“<sup>42</sup>

Haidt přitom, jak je zjevné z řady jeho textů, je voličem a podporovatelem amerických demokratů, ne republikánů. Jeho podpora demokratů má formu snahy ovlivnit jejich politickou *strategii* či *rétoriku*<sup>43</sup> – jímá ho zoufalství z toho, že jejich rétorika zaostává za tou republikánskou ve své schopnosti působit na sociální psychologii voličů. Proč ale vlastně Haidt podporuje stranu, která chronicky zaostává ve své schopnosti pracovat s tím, jak funguje lidská sociální psychologie? Připomeňme, že demokratická a republikánská morálně-politická rétorika nepředstavují jen dvě stanoviska, která jsou obě omezená sice jiným způsobem, ale každé stejnou měrou; existuje jen *jeden druh* politicko-moralistické omezenosti, a větší *míru* této omezenosti vykazuje demokratický moralismus. A zároveň, že ambice Haidtovy práce je deskriptivní, ne normativní. Nechce říkat, co je správné, nýbrž popisovat různé existující formace lidského přesvědčení o správném a nesprávném. A jak ukazuje jeho indická anekdota, nechce zpochybňovat autenticitu a status existujících morálek *jako morálek*. Existence a etablovanost morálních stanovisek je mu kritériem sama o sobě. To vše dohromady v jistém smyslu znamená, že republikáni *jsou* tou „morálnější“ stranou – není morálky mimo toho, z čeho vyplývají naše morální intuice. Z toho vyplývají dvě otázky: 1) otázka pro Haidta osobnější: jak mohu zastávat stanovisko, které sám vnímám jako v porovnání s jiným nedostatečnější, a to nedostatečnější v ohledech souvisejících s morálkou?<sup>44</sup> 2) Jestliže přece jen existuje možnost morální motivace nezávislé na převládajících morálních intuicích,

42 Haidt (2008a).

43 Haidt (2012, s. 180n, 190).

44 Zde lze odkázat na Aristotelovo pozorování, že špatný člověk, je-li pronásledován uvědoměním si své špatnosti, nedokáže sám se sebou žít a může skončit sebevraždou (EN IX.4). V novější morální psychologii je tato Aristotelova úvaha využívána pro vysvětlení psychologických mechanismů sebeklamu: naši hluboce zakořeněné potřeby smýšlet sami o sobě a o svých činech jako o dobrých.

přínejmenším jako možnost jejich smysluplné morální kritiky, co z toho vyplývá pro smysl Haidtova doporučení demokratům? Pak by totiž nebylo jedno a totéž – ba byl by to rozdíl mimořádně významný –, jestli má nějaké politické uskupení přebudovat morální pozadí své politiky, nebo její morální rétoriku.<sup>45</sup>

Tento oddíl byl poněkud rozsáhlý i proto, že nelze alespoň rámcově probrat Haidtovu teorii psychologie v politice, aniž by se věnoval nějaký prostor „politické stránce věci“. Haidtovou ambicí bylo využít svůj předešlý výzkum intuic pro analýzu politických postojů a chování. Smysl mých poznámek byl naopak inverzní: položit si otázku, jaké světlo vrhá způsob, jakým Haidt svůj výzkum morálních intuic aplikuje v kontextu politiky, na jeho pojem morální intuice a morálky vůbec, resp. jaké problémy se díky tomu mohou ukázat. V krátkosti lze konstatovat: vágnost vymezení morální intuice, absence snahy rozlišovat mezi potenciálně různými druhy a zdroji reakcí, které jsou všechny v haidtovském smyslu „intuitivní“, a zároveň důsledně stejná legitimita statusu „morální“ připisovaného všem reakcím „tohoto druhu“ a ničemu jinému vrhá na Haidtovy politické analýzy dvojznačné světlo. Haidt se nezabývá možností, že hodnotící reakce, která má spontánní, „intuitivní“ formu, může vyplývat z leččehos, čímž si fakticky uzavírá možnost zkoumat např. rasistické intuice *jinak* než *pouze* pod zorným úhlem toho, co je na nich legitimní<sup>46</sup> – nemůže totiž dobře rozlišovat mezi morálně motivovaným politickým postojem a široce rozšířeným předsudkem. Analogicky, kritiku politicky rozšířeného postoje lze chápat jako morálně motivovanou jen za předpokladu, že se sama opírá o „intuitivní zdroje“. Tak, paradoxně, jakkoli by Haidt sám neschvaloval intuitivní „rasistické“ reakce na určité typy situací, mají-li podobu obecně rozšířené spontánní reakce a naopak jejich sebekontrola formu racionálního argumentu (např. kantovského typu), status „morální“ přísluší pouze tomu prvnímu,

45 Tedy: co přesně jeden z opinion-makerů americké alternativní pravice, Milo Yiannopoulos, doporučuje americkému liberálnímu establishmentu (demokratům), když s plnou vážností říká, že „by měli číst svého Haidta“ (Bokhari & Yiannopoulos 2016)? Yiannopoulos spolu s Haidtem konstatuje, že liberálové (demokrati) nechápu sociální psychologii; znamená to ale, že mají trochu víc *mluvit*, nebo trochu víc *být* jako konzervativní pravice?

46 Filosofický intuicionismus např. Huemerův (2008, s. 132nn) naopak věnuje velkou pozornost roli, kterou v našich intuicích hrají různé formy „zkreslení“ (*bias*), v důsledku čehož odmítá prosté ztotožnění morálně hodnotící intuice s morálkou.

byť status „správné“ budeme nejspíš připisovat pouze tomu druhému.<sup>47</sup> (Samostatné problémy pojící se s možností kritiky populismu nechávám stranou.)

Za zmínku stojí jeden neobvyklý důsledek Haidtovy diskuse: politická pluralita mu splývá s morální pluralitou, a v důsledku toho jako by mu jakási ideální morálka (nejúspěšnější, obecně nejpřijatelnější, nejstabilnější morální pozice – inkorporující co největší množství morálních intuic) splývala s konsensuálně přijímaným vyváženým politickým kompromisem. Tato blízkost, ne-li identita, stojí na ještě jednom nevysloveném předpokladu: totiž že morálka je nástrojem *k dosahování cílů*, stejných jako politika. Chápat politiku jako do nějaké míry konsekvenzialistický/ utilitaristický podnik, tj. jako soubor nástrojů (institucí, zákonných procedur atd.) majících za cíl zajistit určitý výsledek, je snad zjednodušující, ale cosi skutečného to vystihuje. Zdá se, že politika z určitého hlediska má mít za cíl „zdravě fungující“ a prospívající společnost. V případě morálky je její pojetí jako souboru nástrojů k dosažení *externího* cíle kontroverznější, samo vyžaduje jisté zdůvodnění, spíše než, aby je bylo možné bez dalšího předpokládat. O tom budu mluvit v následujícím oddíle.

### 3. Haidtův pojem morálky a jeho problémy

Závěrečná kapitola *The Righteous Mind* slouží zarámování celého výkladu smírným pozorováním, že jak „liberálové“, tak jejich protihráči (rozpadající se do dvou podskupin: libertariánské a konzervativní) mají svůj díl pravdy, a všichni by se mohli poučit od sebe navzájem.<sup>48</sup> „Pravdu mají“ vzhledem k tomu, co je podle Haidta vlastním cílem morálky, a co může logicky fungovat jen tehdy, bude-li to podnikem, v němž budou zaangažováni všichni, liberálové i konzervativci:

„... morální kapitál můžeme definovat jako *zdroje, které udržují morální společenství při životě*. (...) [tj.] soubor do

<sup>47</sup> V tomto základním důsledku, který z Haidtova pojetí implicitně vyplývá, se Haidt markantně rozchází se svým souputníkem a občasným spoluautorem Joshou Greenem, který pojímá morálku právě jako lokalizovanou v „pomalých“, racionálních reflexích a přepracováních našich „rychlých“ před-racionálních reakcí. (Reprezentativně v: Greene 2013.)

<sup>48</sup> Haidt (2012, s. 343nn).

sebe zapadajících hodnot, ctností, norem, praktik, identit, institucí a technologií, jimiž společenství disponuje, a které jsou ve shodě s evolucí vzniklými sociálními mechanismy, díky čemuž umožňují tomuto společenství potlačovat nebo regulovat sobectví a dosahovat spolupráce.“<sup>49</sup>

Obojaký pojem „morálního kapitálu“ poskytuje užitečnou přístupovou cestu k Haidtovu specifickému pojetí toho, co je vůbec morálka. Právě složka morálního sebepojetí a zvláště reputace akceptované druhými je totiž podle Haidta pro morálku zcela podstatná. Morální hodnoty, které jejich zastánci dlouhodobě nedokážou „prodat“ širšímu společenství, možná dokonce proto, že to není možné, nemohou totiž plnit účel, díky němuž je morálka morálkou, tj. účinně potlačovat sobectví a spolupracovat na obecném prospěchu komunity.<sup>50</sup> Proto je pro Haidta skutečným morálním protagonistou Platónovy *Ústavy* Glaukón, a nikoli Sókratés. Glaukón – a nikoli Sókratés – si totiž uvědomuje, že morálka je projekt za určitým účelem, a spravedlivý člověk se špatnou morální reputací, byť nezaslouženou, pilířem spolupráce za tímto účelem být nemůže.<sup>51</sup>

Definice morálky *obecným* prospěchem dosažitelným jen všeobecnou *spoluprací* je dalším z důvodů, proč se Haidt snaží dát za pravdu oběma „morálně znepřáteleným“ politickým stranám. Demokrati i republikáni jsou součástí celku, a spolupráce tedy vyžaduje zapojení všech. Jedni ani druhí si přitom nemohou dost dobře pomoci, aby nebyli liberály či naopak konzervativci, neboť podle Haidta o politických preferencích rozhodují do značné míry geny.<sup>52</sup> Zde je myslím důležité připomenout

49 Tamtéž (s. 341).

50 Ve starším populárním textu (Haidt 2008b) definuje Haidt *morálku* téměř stejnými slovy, jako zde *morální kapitál*; podobně i v přehledovém textu pro *Handbook of Social Psychology* (Haidt & Kesebir 2010, s. 815, 818).

51 Glaukón prezentuje toto stanovisko, jak je vyslechl od Thrasymacha a jiných, uvozuje je však takto:

„A přece, Sókrate, *mně se to nějak nezdá*. Řešení ovšem neznám (...) výklad o spravedlnosti jako o něčem lepším, než je nespravedlnost, jsem ještě tak, jak chci, neuslyšel od nikoho; *chtěl bych slyšet chválu na ni samu o sobě*, a myslím, že bych to mohl slyšet nejspíš od tebe.“ (Platón 1993, s. 77 (358c-d); moje kurzíva)

52 Tamtéž (s. 322nn). Jednoznačné genetické pozadí našich rozdílných politických preferencí je však patrně méně jednoznačnou věcí, než jak to líčí Haidt (viz LaFollette & Woodruff 2015, s. 461n).

Haidtův důraz na to, že jeho model je sociálně-intuicionistický; je to právě „sociální“ složka tohoto modelu, která do určité míry vysvětluje kulturní rozdíly mezi dominantními morálními trendy v Indii a např. v západní Evropě. Sice se tedy nerodíme jako morálně hotové osoby a musí nás do finálního tvaru uhníst společnost, v níž vyrůstáme, ale dílčí rozmanitosti těchto konečných tvarů nejsou přesto kulturně nahodilé v pravém smyslu slova, tj. tak, že by byly emancipovány od naší evoluční výbavy. *Peer pressure* atd. rozhoduje o tom, na kterou ze složek naší evoluční výbavy bude v našem konečném morálním profilu položen větší důraz. V zásadě však nejde o mnoho víc, než o „druhé patro“ téže budovy, jejíž konstrukce je evolucionistická a je ovlivňována, jak jsem zmínil výše, podle Haidta v zásadě geneticky, v obou svých „patrech“.

Důvodem, proč vyznění Haidtovy knihy tolik zdůrazňuje, že liberálové se mají učit od konzervativců, spíše než naopak, je pravděpodobně Haidtem předložená analýza větší nevyváženosti liberalismu. Obě strany však do podniku společné spolupráce přispívají něčím hodnotným, a tedy se navzájem potřebují. Z důvodů vyplývajících z naší genetické výbavy se jen stěží z liberála, jak – zdá se mi – předpokládá Haidt, může kdy stát autentický konzervatívec, nebo naopak; smyslem Haidtových výzev k otevřenosti je tak spíše bojovat s přesvědčením, že druhá strana je nemorální.

Úspěšnost politické rétoriky, jak jsme řekli výše, je dána schopností souznít s intuitivními reakcemi voličů na situace, problémy atd. mající morální dimenzi. Tyto intuitivní reakce sice nejsou výsledkem racionálního rozvažování, ale jsou racionální implicitně, v prudenciálním smyslu slova – jako odkaz úspěšných adaptačních strategií našich předků. Strategií, které cílily na *právě totéž*, jako morálka i v dnešním smyslu slova: hladkou spoluprací zajišťující přežití a prospívání společenství.<sup>53</sup> Jestliže ale politický moralismus republikánů akcentuje širší spektrum os naší morální intuice než politický moralismus demokratů, co může motivovat někoho, kdo tuto situaci lucidně prohlédl (jako Haidt), ke zkoumání možností, jak posílit pozici té de facto morálně „nedostatečnější“

53 Haidt se svou tendencí k vágní identifikaci cílů, na něž mířily evoluční prekurzory morálky, s tím, oč jde v morálce „jako takové“, „vůbec“, podstatně nevymyká z celé řady prací zkoumajících evoluční původ lidské morálky; viz např. Tomasello (2016, s. 1nn). A srv. mezi autory zkoumající toto téma disentaní postřeh Joycův (2006, s. 17), že *původ* nelze jednoduše ztotožňovat s *obsahem*.

strany, místo aby se prostě přiklonil k té druhé? Předpokládám tedy, že Haidtova podpora demokratů vyplývá z jeho přesvědčení, že jejich politika směřuje k tomu, učinit společnost lepší, než jakou by ji učinila politika republikánská.

Už jsem naznačil, že jako odpověď se nabízí do elegantního rámce Haidtova výkladu nezapadající možnost, že má smysl chápat nějaké chování jako morálně správné *nezávisle*, i tehdy, pokud se staví proti rekonstruovatelné motivaci některou z Haidtových os, nebo je bez zjevného vztahu k hypotetizovanému účelu morálky, tj. spolupráci za účelem prospěchu společenství. Nebo, zrcadlově, že tu kterou základní osu morálky lze „rozbalit“ i způsobem, který, jakkoli není neupřímný nebo neporozuměním, podvrací to, čeho mělo být dosaženo řízením se touto motivací. Příkladem prvního může být lidický kněz Josef Štemberka, který doprovodil na smrt své farníky. Příkladem druhého třeba aktuálně diskutovaný plošný zákaz interrupcí v Polsku, bezpochyby inspirovaný určitým čtením toho, co jeho zastánci považují za religiózní morálku, přičemž je ale sporné, jaký a zda vůbec nějaký vztah má k projektu spolupráce na obecném prospěchu společenství. Oba příklady poskytují určitou podporu myšlence, že pro morálnost nebo nemorálnost nějakého postoje či jednání existují i jiná kritéria, nezávislá na proklamovaném účelu morálky reflektovaném jejím šesterým kořenem, jak jej mapuje Haidt.

Výhodou emancipace od Haidtem hypotetizovaného evolučního pozadí našich morálních motivací by bylo to, že budeme moci brát vážně, doslova *popis, pod kterým* morální jednatel sám *rozumějí* svým postojům, a vzít jej v potaz v našem morálním hodnocení těchto postojů.<sup>54</sup> Silně nábožensky motivovaný postoj, jako např. odpor vůči interrupcím, bude znamenat to, co znamená pro své nositele, a ne složitý překlad, či sublimaci evolučně etablovaného smyslu pro důležitost čistoty. Je rozdíl, ptáme-li se, proč v nějaké komunitě silně rezonuje nábožensky motivované tažení proti interrupcím, nebo ptáme-li se, proč v nějaké komunitě

---

54 Srv. Frederickovy (2016, s. 638) kritické poznámky k filosofickému intuicionismu: skutečně problematické morální spory se nemusí týkat ani tak toho, zda (např.) porušit daný slib je správné nebo ne (že to správné není, je „analyticky“ součástí významu slova slib), ale toho, zda se nějaké jednání skutečně kvalifikuje – tj. zda mu tak jednatel rozumí – jako (morálně relevantní) porušení slibu.

silně rezonuje určité vyjádření hluboce zakořeněného lidského smyslu pro posvátnost a čistotu. Nábožensky motivované tažení proti interrupcím může v nějaké komunitě silně rezonovat z důvodů, které nemusejí mít nic společného s hluboce zakořeněným lidským smyslem pro posvátnost a čistotu, přinejmenším nic, co by bylo zajímavé pro porozumění onomu společenství a jeho kultuře.<sup>55</sup> Analogicky bude rozdíl, ptáme-li se, zda je morálně správný militantní odpor vůči interrupcím, nebo ptáme-li se, zda je morálně správné úsilí o zachování hodnot posvátnosti a čistoty. Zaprvé, opět, důvody, pro něž může být namísto klasifikovat odpor vůči interrupcím jako správný, mohou být značně odlišné od důvodů, proč dává smysl označit za správné úsilí o zachování čistoty. A zadruhé, je mnohem lépe myslitelné, že militantní odpor vůči interrupcím je jednoduše nesprávný, v míře, v níž není jasné, zda takto jednoduše nesprávné může být úsilí o zachování smyslu pro posvátnost a čistotu. Posledně jmenovaný rozdíl vyplývá z toho, že bez ohledu na to, že *genealogie* lidské morálky má mnoho společného s tím, co Haidt naznačuje, *pojem* morálky se od této genealogie do značné míry emancipoval.

Z Haidtovy perspektivy lze vůči těmto poznámkám namítnout, že směřují normativní perspektivu s deskriptivní, že jeho jednoduše zajímá, co v různých prostředích za morální platí, a co lze srovnáním těchto zjištění odhalit o mechanismech vedoucích k etablování těchto intuic. Že nemá žádnou ambici stanovovat *normativně*, co je správné. Pokud však nepostuluje dva, na sobě nezávislé pojmy morálky, pak definice morálky, s níž pracuje, ve skutečnosti přímočaře poukazuje k neoddelitelné fúzi deskriptivního a normativního. Je-li morálka soubor všeho toho, co přispívá ke spolupráci za účelem obecného prospěchu, je těžké nečíst to zároveň jako morálně *normativní* kritérium, ve smyslu pozorování o významu pojmů jako „správný“.

Částí Haidtova problému je, že v jistém smyslu skutečně postuluje dva od sebe oddělené pojmy morálky. Obou jsem se dotkl výše. Jednak pojem formální, který mu umožňuje shromažďovat a porovnávat data o různých kulturně specifických morálkách: lokálně platných souborech

---

55 Moje pochybnost se týká toho, zda Haidtem navrhovaný výklad lepší porozumění lidským postojům spíše neznemožňuje než umožňuje; jiný aspekt téhož problému lze ale formulovat také příkřeji, totiž že Haidt ani nemá „tu slušnost, aby uznal, že předmětem víry [náboženských fundamentalistů] je skutečně to, v co říkají, že věří“ (LaFollette & Woodruff 2015, s. 462).

hodnotících postojů či soudů uplatňovaných na lidské jednání i charaktery. A jednak jeho pojem materiální: morálka jako soubor adaptací či transformovaných adaptací přispívajících k účelné spolupráci za obecným prospěchem. V Haidtově pojetí – ačkoli to takto sotva kdy explicitně uchopí – lze tyto dva pojmy integrovat, a zdá se, že tímto směrem skutečně táhne, jediným možným způsobem, který je ovšem značně hypotetický. Nakolik lokálně platné morálky Haidt nechce delegitimizovat, musí v nich všech spatřovat různé formy „převleků“ morálky skutečné, z nichž některé budou značně komplikovanými formami *sebeklamu* (exaltované formy platónské či kantovské etiky). Jakkoli jsou však sebeklamem z hlediska obsahového, pokud plní svůj účel – skutečný účel morálky –, je to v pořádku. Ovšem Haidtův předpoklad, že každá morálka, kterou lze interpretovat jako „rozbalení“ té které osy, tento účel naplňuje, je *empirická hypotéza*, a je dost dobrých, opět empirických důvodů být vůči této hypotéze v řadě konkrétních příkladů kritický. Zmínil jsem je v předchozím oddíle.

Vedle Haidtovy empirické hypotézy je hlubším problémem jeho částečného rozdvojení pojmu morálky to, že mu uniká, že jakmile „opře“ morálku o jedno konkrétní externí kritérium, musí být připraven odpovědět na řadu otázek, na něž ovšem neodpovídá. Např.: proč právě toto kritérium? Nebo: jestliže máme dva pojmy, „morálka“ a nějaké „X“, jehož pomocí „morálku“ definujeme, je to jedna a tatáž věc, nebo dvě různé? Haidtova pozitivní definice morálky, právě proto, že není triviální, evidentní, a že vůbec předpokládá nutnost to, co je „morálka“, dále definovat, činí jeho pozici naopak zranitelnější,<sup>56</sup> – protože ji tím činí nesamozřejmou, volající po dalším zdůvodnění, které on nepodává.

Haidt si ovšem tyto problémy nepřipouští, a zhruba utilitaristické pojetí morálky jednoduše přejímá a zároveň má optimisticky za to, že takto přispívá k překonání tisíciletých neshod filosofů ohledně morálky prostředky empirických věd. Jako někteří jiní, kteří píší o tématu, jež sdílají s filosofií, s ambicí změnit nejen vlastní disciplínu, ale opravit i *filosofickou* reflexi sdíleného tématu, i on dosahuje této sebejistoty zčásti díky tomu, že ignoruje komplexitu filosofických debat. Že morálka

---

<sup>56</sup> Mozná i z těchto důvodů řada filosofů „dobro“ dále nedefinuje – ať už jde o Platóna, nebo G. E. Moora.



je jednoduše soubor dispozic, nástrojů a opatření, které mají svým nositelům generovat určitý typ prospěchu, jehož nelze dosáhnout bez široké spolupráce, je pro něho samozřejmostí. Haidt do svého výkladu inkorporuje určité prvky teorie ctností, např. intuici, že morální jednání nevyplývá z racionálního rozvažování, nýbrž z nereflektovaných osobnostních dispozic, které lze do jisté míry kultivovat;<sup>57</sup> na celkovém utilitaristickém rámci, tj. jeho intuici, v čem spočívá morálnost morálního jednání, to však nic nemění. Filosoficky důležité rozdíly mezi utilitarismem a jeho hlavními rivaly – deontologickou či kantovskou etikou a etikou ctností<sup>58</sup> – však zůstávají mimo Haidtův radar; a několik poznámek k těmto rozdílům může nasvítit limity Haidtova přístupu.

Etika ctností se skutečně – jak vnímá i Haidt – soustředí víc na kultivaci charakteru, než na podání kritéria správného jednání. Přesto jisté kritérium nabízí, např. „udělej to, co by charakteristicky udělal za těchto okolností ctnostný jednatel“.<sup>59</sup> Přes patrnou ironičnost tohoto kritéria nelze přehlédnout jeho hlubší smysl. Totiž že jakékoli vymezení správného jednání tou či onou charakteristikou tohoto jednání samotného je nedostatečné; či jinými slovy, že na správnost i takového jednání, které lze popsat jako motivované tím či oním aspektem agendy „spolupráce za obecným prospěchem“, se lze spolehnout jen tam, kde *jednatel* je ctnostný. Tento pojmový nástroj umožňuje rozlišit mezi hluboce nábožensky motivovaným jednáním člověka, který miluje své bližní, a hluboce nábožensky motivovaným jednáním násilného náboženského fanatika. (Vzpomeňme na problém s Haidtovým nerozlišováním mezi morálností jednání a tím, že jednatel své jednání za morální považuje.) Ve svých odkazech k Platónově *Ústavě* se Haidt opakovaně vrací k tomu, že za pravdu dává Glaukónovi, který si uvědomuje význam reputace a toho, jak člověk vypadá, pro nastolení

57 Haidt & Kesebir (2010, s. 798n).

58 Celá šíře filosofické etiky tímto výčtem zdaleka pokryta není. Pomíjím vše na „kontinentální“ straně současné filosofie, etické myšlení vycházející z významných náboženských tradic (křesťanskou, islámskou aj. etiku), nebo hlubší průřez dějinami filosofie. V každém případě, spektrum etických pozic formulovaných (více méně) analytickým jazykem a (v zásadě) viditelně přítomných v anglickojazyčném diskursu, s občasnými namátkovými odkazy k Platónovi, Kantovi či Humovi, zhruba odpovídá pozadí, na němž Haidt vede to, co považuje za svou polemiku s filosofií jako celkem.

59 Hursthouse (2002, s. 17).

morálního řádu. Sókratova obhajoba, že spravedlivý člověk se špatnou reputací je na tom lépe než člověk nespravedlivý, podle Haidta stojí na nesprávné představě o lidské povaze, resp. o rozumu jako zdroji morálky.<sup>60</sup> Sókratés však nemá v úmyslu polemizovat s *pozorováním*, že mnoho lidí jedná prospěchářsky a instrumentálně a je to primárně ohled na vlastní prospěch a reputaci, co je motivuje chovat se „správně“. Spíše se snaží ukázat, že jakékoli porozumění vlastnímu životu, které si cení alespoň některých jeho aspektů jako reputace, předpokládá takové porozumění tomu, co je morálka a spravedlnost, které si těchto hodnot cení pro ně samé, ne z instrumentálních důvodů.<sup>61</sup> To, že se většina lidí morálně nechová, popř. se tak chová z čistě pragmatických motivů, nebo má zmatené představy o tom, co je morální, neznamená, že smysl toho, co morálka znamená, musí být *vyčerpán* těmito pragmatickými, popř. sebeklamnými představami. Platónův racionalismus zde přitom nehraje klíčovou roli. V každém případě Platónův pojem rozumu není tímtež, co Haidt popisuje jako racionalizaci intuic.

To souvisí i s podstatným rozdílem mezi utilitaristickou a kantovskou etikou, který je v Haidtově kritice Kanta, soustředící se na Kantův racionalismus, setřen: totiž že morální jednání je morálním *nepodmíněně*, tj. ne na základě cíle, kterého by mělo docilovat. Z kantovské perspektivy lze haidtovskému pojetí morálky namítnout: proč si mám osvojit hodnotu spolupráce za účelem obecného prospěchu? To je cíl, s nímž se mohu i nemusím ztotožňovat; k obojímu přitom mohu mít důvody morální povahy. Protipříklady haidtovské představy o morálce se samy nabízejí: Lze cílů, kvůli kterým se morálka údajně etablovala, dosahovat i jinými nástroji? Pokud ano, lze to, v čem spočívá morálka, chápat i nezávisle? Představíme-li si společnost velmi dobře organizovanou ve smyslu forem spolupráce a jejich prosazování, které vedou k její nesporné úspěšnosti, pokud jde o její celkovou *fitness* v porovnání s jinými společnostmi, která však dosahuje těchto úspěchů za cenu útlaku a zneužívání menšiny, již odepírá hlas, co bude v takové společnosti platit za morální kritiku? (Lze se takto dívat, v některých ohledech, např. na současnou Čínu?) V souladu s Haidtovou ambicí čistě deskriptivního morálního pluralismu

60 Např. Haidt (2012, s. 84nn).

61 Viz k tomu např. Williams (2008).

bude morální kritikou odsuzování „podvratných živlů“ a „škůdců“. V jeho analýze není prostor pro jiný smysl „nemorálního chování“ než „to, které je odsuzováno v dané společnosti/kultuře“, popř. „ve všech kulturách“. Není možné pak vysvětlit, zda a v jakém smyslu je pro člena nějaké společnosti, která si jako celek vede velmi úspěšně, morální chovat se jinak, než vyžaduje většina členů této společnosti.

Důsledkem suspendování možnosti menšinové či setrvalé přímo disidentní morálky je – a to by se Kantovi stát nemohlo – ztráta schopnosti chápat morální pojmy a hodnocení jako nástroj *kritiky*, a to i kritiky obecně rozšířeného a akceptovaného způsobu chování, který vidím kolem sebe. Haidtův předpoklad, že nutnou podmínkou skutečné morálky je úspěšná morální rétorika, vede k paradoxnímu závěru, že morální kritika je možná jen ze sociálně úspěšné, dobře etablované, v podstatě konvenční pozice. Ve výsledku z toho může vyplývat například, že zatímco morální kritikou je rozhořčení nacionálně-konzervativního politického katolicismu střeoevropského typu nad „genderovou ideologií“ a „kulturním marxismem“, morální kritikou *není* kritika tohoto typu konzervativismu z lokálně marginální a ohrožené pozice, např. od neziskové organizace na podporu trans osob a násilí, kterému čelí.

Připomeňme si ještě jednu pozici utilitarismu v Haidtově pojetí: Haidt se nehlásí k utilitarismu jako k normativní teorii, nicméně při popisu různých forem morálky, které empiricky zkoumá, pracuje s pojmem morálky jako souboru nástrojů, jež si lidé osvojili a jež se mezi lidmi etablovaly díky tomu, že se „evolučně“ osvědčily jako efektivní pro dosahování určitých účelů. Stojí za zmínku, že mezi příklady filosofických pozic, které chápe jako protirečící jeho zjištěním, Haidt nikdy nediskutuje utilitarismus – jeho kritický zájem se soustředí na sókratovsko-platónskou etiku a kantianismus. Rovněž umělé ex post racionalizace našich intuitivních reakcí, jejichž chatrnost má demonstrovat experiment s Markem a Julií, typicky pracují s kantovským slovníkem.

Zdá se tak paradoxní, že Haidtův souputník Joshua Greene, jehož kniha *Moral Tribes* (2013) má podobné zaměření jako Haidtova, mapuje vrstvení a utváření našich morálních reakcí a soudů způsobem fakticky protikladným. Podle Greena jsou to právě naše „rychlé“, intuitivní

reakce – podle Haidta ztělesňující konsekvencialistický rodokmen lidské morálky – co stojí za naším sklonem vnímat morální příklady a situace kantovskou optikou. A naopak „pomalé“, racionální uvažování – podle Haidta zdroj ex post racionalizací rozporujících konsekvencialistický rodokmen morálky – teprve podle Greena dokáže vygenerovat adekvátní, „hluboce pragmatická“, tj. konsekvencialistická kritéria morálního hodnocení.<sup>62</sup> Greene se tak shoduje s Haidtem v jistém příklonu ke konsekvencialistickému vnímání morálky a kritičnosti vůči jejím kantovským elementům, ale rozcházejí se ve svých představách o tom, zda v morálním ustrojení naší mysli Kant sedí na ramenou Millovi, nebo naopak. Shodují se však v tom, že mít od morálky jiná než konsekvencialistická očekávání není na místě, ať už tato jiná očekávání analyzujeme blíže jako nedostatečná (Greene), nebo přestřelená a vnitřně rozporná (Haidt).

Problematickostí této představy morálky jako *prostředku* k určitému účelu se ukáže, uvědomíme-li si, že řada ohledů, které chápeme jako morální, je mnohdy ve vztahu k účelům, jichž by morálka měla podle Haidta dosahovat, spíše kontraproduktivní. Peter Winch poznamenává:

„Lidé se snaží dosahovat různých cílů a přitom čelí různým překážkám, jako nedostatek peněz, schopností, přátel, odpor druhých, a jiné. Morálka ale nemá nic společného s tím, že by nám pomáhala zdolat tyto překážky. Naopak, nebýt morálky, bylo by jejich zdolávání mnohem snazší. Jakým problémům nám tedy morálka umožňuje se vyhnout? (...) Morálka je prý jakýsi průvodce, který umožňuje [člověku, který se rozhoduje, zda udělat něco morálně sporného,] vyhnout se jeho problému. Ale kdyby nebylo morálky, ten problém by vůbec neexistoval!“<sup>63</sup>

Morálka tak v tomto smyslu je – spíše než řešením objektivně, již předtím existujících problémů, za účelem jejichž řešení byla vytvořena – zdrojem vyostřené citlivosti vůči problémům, jichž bychom si jinak vůbec nebyli vědomi. Jinými slovy, neexistuje jen jediný soubor problémů, na něž

62 Greene (2013, kap. 11).

63 Winch (1972, s. 172).

instituce morálky reaguje, ustavený už v prehistorii lidstva. Nové a nové problémy vyvstávají zároveň s tím, jak se vyvíjí a proměňuje naše porozumění sobě, své společnosti, lidskému životu vůbec: rasismus, homofobie, sexuální diskriminace, klimatická krize...

Tím se obloukem vracíme na začátek, k Haidtovu pojmu morální intuice. Připomeňme, že jakkoli Haidt definuje intuici negativně, tj. vyloučením toho, co intuice není, několik bodů je jasných. Své intuice v podstatě nedokážeme ovlivnit, jejich podstatná část vyplývá z našeho evolučního dědictví a jako taková, dokonce včetně variací v intuicích *politických*, je zakódována v našich genech. Něco je pak dotvořeno enkulturací – vlivem prostředí či kultury, do níž se člověk narodil a byl v ní vychován. Tento základní obrázek ale přehlídí význam toho, že to, na co reagují, je mi dáno v pojmech. *Rozumím* tomu *jako* určité věci, na niž reagují jako na takovou. A na úrovni pojmů se také liší od jiných věcí, na které „intuitivně“ reagují jinak. Pojmy se vyvíjejí, proměňují a inkorporují kulturní, společenské, i vědecké posuny.

Tak také sex – abychom se vrátili do kontextu Haidtova úvodního příkladu – a to, co sex znamená, se proměňuje, často způsoby víceméně nezávislými na tom, co nám o sexu řekne jeho „čistě biologický“ popis. Uvažme např. změnu, již ve způsobu, jakým rozumíme sexu a jak „intuitivně reagujeme“ na různé se sexem spjaté situace, přinesla sexuální výchova v souvislosti s etablováním pojmu *spousal rape*. Pojmovou vrstvu našich intuic do jisté míry pokrývá Haidtův pojem enkulturace; co však nepokrývá, je fakt, že právě pojmový filtr umožňuje rozumět různým situacím a intuitivně na ně reagovat také individuálně specificky. To, že na „jednu a tutéž situaci“ může jeden člověk reagovat pokrčením rameny, a druhý odporem a zděšením, může souviset s tím, že pro toho druhého hraje určitou významnou roli v tom, jak rozumí světu, pojem *spousal rape*. Pro Haidta je tento typ rozdílu mezi oběma vysvětlitelný v podstatě jen tak, že se oba narodili a vyrostli jako příslušníci dvou různých morálně-politických (sub)kultur. Tito dva lidé ale právě tak dobře mohou být jedním a tímtéž člověkem v různých *obdobích* svého života, člověkem, který se ze své minulosti, ale ne nutně pouze z ní, dokázal poučit.<sup>64</sup> Význam *jednotlivcova* vkladu do pojmů, v nichž rozumí situ-

64 Conning (2015, s. 120) ve své recenzi Greenových *Moral Tribes* poznamenává: „To, že lidé

acím, ve kterých se ocitá, jistě nelze přeceňovat, ale není možné jej ani ignorovat.

Co chci říct: Naše hodnotící reakce na okolní podněty skutečně obvykle jsou okamžité a naše snaha předložit pro ně racionální důvody bude v drtivé většině případů *ex post* racionalizací; toto Haidtovo pozorování nemíním rozporovat. To ale neznamená, že status a původ těchto reakcí nemůže mít mnoho společného s *důvody*, a to často s důvody, jimiž se dotyčný sám někdy v minulosti myšlenkově, ba „racionálně“ zaobíral.<sup>65</sup> Naše intuitivní reakce na určité situace se během našeho života, včetně naší *dospělosti*, proměňují, protože bez ohledu na svou nereflektovanost, okamžitost, inkorporují proměny našeho rozumění a seberozumění. Aby mi bylo rozuměno: proměny našeho rozumění a seberozumění neznamenají nutně zmoudření. Připomeňme, že Haidtem uvažované dva kořeny našich intuic – naše evoluční dědictví a enkulturace – jsou *oba* prodloužením postojů a strategií zacílených na dosažení účelu, který je účelem morálky. S ohledem na to je pak pro Haidta obtížné vysvětlit, zda nebo v jakém smyslu je možné mít intuice, které by byly – ať už v rámci dané kultury, nebo „vůbec“ – nemorální. (Připomeňme si znovu problém s rasistickými předsudky.) Teprve připustíme-li jednak, že vývoj a převládající trendy v té které *dnešní* kultuře mohou vykazovat i takřkajíc „radikálně kontingentní“ aspekty, nijak rozumně nesouvisející s účely, za nimiž se naši prapředkové začali shlukovat do společenství, a jednak, že do způsobu, jakým člověk reaguje, se promítají zvláštní a nahodilé zákruty jeho života jednotlivce, budeme mít na čem založit myšlenku nemorálních, či morálně nesprávných intuic.

---

přebíjejí své emoční reakce reakcemi utilitaristickými, mají-li příležitost si věc promyslet, ještě neznamená, že by nemohli přebít utilitaristické soudy založenými na myšlence práv nebo soudy spirituálního charakteru, kdyby měli dost času a podpory k tomu, zamyslet se nad problémem v jeho celé hloubce.“

Obecněji lze říct – a to platí i pro Haidta – že není dobrý důvod považovat za vyčerpávající analýzu morálky model inkorporující právě jeden „rychlý“ a jeden „pomalý“ krok, přičemž v právě jednom z nich je třeba spatřovat to, nač bychom si měli dávat pozor (a veškerá debata o morálce by se vyčerpávala už jen otázkou, na který si dávat pozor).

65 Haidt naznačuje, že odpor filosofů k sociobiologickým vysvětlením morálky je daný jejich neochotou spolknout „politicky nekorektní“ důsledky tohoto obratu (Haidt & Kesebir 2010, s. 801). Stojí nicméně za to poznamenat, že i empiričtí psychologové sympatizující s Haidtovým projektem přinášejí analýzy, podle nichž je vztah mezi okamžitými „gut feelings“ a racionálním uvažováním spíše kruhový než jednosměrný (např. Fine 2006).

Okamžitost nemusí znamenat *instinktivnost*, okamžité a spontánní formy reakce mohou zahrnovat i značně jemně strukturované a nuancované porozumění, ustavené složitými individuálními cestami.<sup>66</sup>

Podtitul Haidtovy knihy zní *Proč jsou dobří lidé rozděleni politikou a náboženstvím*. Zdá se mi, že přesnější – s ohledem na to, co Haidtovy pozoruhodné výzkumy skutečně ukazují – by bylo jednoduše *Proč jsou lidé rozděleni politikou a náboženstvím*. Z příčin, které Haidt mapuje, spolu mohou zarputile nesouhlasit lidé zlí stejně tak jako dobří; s tím, zda jsou dobří, nebo zlí, tyto příčiny nijak přímo nesouvisí, byť souvisí s naší potřebou vnímat sami sebe jako příslušníky toho „správného“ tábora. Chceme-li tvrdit, že ano, budeme předpokládat zvláštní, jaksi zploštělý pojem morálky, v jehož rámci téměř nemohou existovat zlí či nemorální lidé, resp. bude nejasné, co by to mělo znamenat, leda lidi v zásadním rozporu s etablovanými normami dlouhodobě relativně stabilní společnosti či komunity. Popisné nástroje jako sebeklam nebo falešná morálka tu nebudou mít smysl.

To, co z jednoho hlediska vypadá jako pošetilý tribalismus, může být vyjádřením hlubší charakteristiky morálky. Jestliže nám na některých věcech záleží „morálním“ způsobem, v některých případech to může znamenat, že dost dobře není možné vyjít druhé straně vstříc a myslet

---

66 Častým terčem Haidtovy kritiky je Kant; jeho přeracionalizovanost však lze kritizovat i z jiných než humovských (v Haidtově smyslu) pozic, např. tak, jak to dělá Schopenhauer, když oproti Kantovi zdůrazňuje soucit jako *bezprostřední*, ne vyargumentovanou, motivací morálky. Soucit je přitom skutečně jistým *porozuměním* tomu, že *druhý* trpí (ne že trpím já – nepředstavuji si sebe sama, jak trpím v kůži druhého), a jako takový značně subtilní náhled je – při vši své obvyklosti a „všednosti“ – čímisi „mysteriózním“. (Schopenhauer 2007, s. 504nn).

Obecněji lze říct, že právě ty naše morálně hodnotící postoje, dispozice a reakce, jejichž společným jmenovatelem je absence „pomalé“ racionální deliberace kantovského typu, lze analyzovat – a byly analyzovány – i jinými způsoby, než jako *gut feeling*. Tak např. v klasickém textu „Vision and Choice in Morality“ pracuje Iris Murdoch (1997 [1956], viz s. 80nn) s podobným typem příkladů jako Haidt a její doporučení jsou podobná (neodpovědně hned toho, s kým nesouhlasím, jako nemorálního). Na rozdíl od Haidta je však její důraz na „celkový způsob, jakým člověk vidí svět“ nebo na soubor „sebereflexí a komplexních postojů k životu, jejichž uplatňování a rozvíjení probíhá neustále, nelze je ovšem rozložit v čase a přiřadit k jednotlivým situacím“ snahou o postižení čehosi, co je ustavováno velmi složitým a nepředvídatelným způsobem, mj. proto, že „vidění“ je lokalizováno v pojmech. Důraz na pojmovou stránku věci znamená také, že morální hodnocení má vždy aspekt *jistého rozumění*; a s tím ruku v ruce jde to, že „myslet si, že mohu vždy s plnou jistotou tvrdit, že jsem porozuměl, se zdá zjevně nerealistické – a bez plného porozumění může být hodnocení zkrácené nebo nemožné, a *někdy bych dokonce rozumět ani neměl*.“ (s. 90, moje kurzíva).

i její stanovisko jako morálně motivované, aniž bych tím kompromitoval své vlastní pojetí morálky, toho, co morálka znamená pro mě. („There were very fine people on both sides.“) Deklarovat v takovou chvíli, že takový krok nejsem ochoten či spíše schopen udělat, protože nevím, co by to znamenalo, může být poctivější.<sup>67</sup> A představa, že je to možné vždy, nemusí svědčit o intelektuální otevřenosti a bezpředsudečnosti, ale naopak možná o ledabylosti. Pokud se tato kritika Haidtovy myšlenky morálního pluralismu týká, je za to zodpovědný v první řadě způsob, jakým pracuje se svým pojmem intuice, který je ovšem pro jeho zkoumání morálky klíčový.<sup>68</sup>

### Literatura

- Annas, J. (2011): *Intelligent Virtue*. Oxford University Press, Oxford.
- Antfolk, J., Karlsson, M., Bäckström, A. & Santtila, P. (2012): „Disgust elicited by third-party incest: the roles of biological relatedness, co-residence, and family relationship.“ *Evolution and Human Behavior* 33 (3): 217–223.
- Aristotelés (2013): *Etika Níkomachova*. Přel. A. Kříž. Rezek, Praha.
- Beran, O. (2021): *Examples and Their Role in Our Thinking*. Routledge, New York.
- Blum, L. (2013): „Political identity and moral education: A response to Jonathan Haidt’s *The Righteous Mind*.“ *Journal of Moral Education* 42 (3): 298–315.
- Bokhari, A. & Yiannopoulos, M. (2016): „An Establishment Conservative’s Guide to the Alt-Right.“ *Breitbart* [online]. 29. 3. 2016 [cit. 1. 7. 2021]. Dostupné z: <https://www.breitbart.com/tech/2016/03/29/an-establishment-conservatives-guide-to-the-alt-right/>.
- Browne, B. (1992): „A Solution to the Problem of Moral Luck.“ *Philosophical Quarterly* 42 (168): 345–356.
- Carpenter, T. (2012): „Construction of the Crack Mother Icon.“ *Western Journal of Black Studies* 36 (4): 264–275.

---

<sup>67</sup> Viz k tomu Winch (1997, s. 202).

<sup>68</sup> Tato publikace byla podpořena v rámci projektu OP VVV „Centrum pro etiku jako studium hodnoty člověka“, reg. č. CZ.02.1.01/0.0/0.0/15\_003/0000425.



- Conklin, A. S. (2015): „Review of *Moral tribes: Emotion, reason, and the gap between us and them*.“ *Journal of Moral Education* 44 (1): 119–121.
- Coughlin, R. M. (1989): „Welfare Myths and Stereotypes.“ In *Reforming Welfare: Limits, Lessons, and Choices*, ed. R. M. Coughlin, University of New Mexico Press, Albuquerque, 1989, s. 79–106.
- Ebersole, F. (2002): *Language and Perception*. Xlibris, Bloomington (IN).
- Fine, C. (2006): „Is the emotional dog wagging its rational tail, or chasing it? Reason in moral judgment.“ *Philosophical Explorations* 9 (1): 83–99.
- Frederick, D. (2016): „Ethical Intuitionism: A Structural Critique.“ *Journal of Value Inquiry* 50 (3): 631–647.
- Gaita, R. (2006): „Torture: The Lesser Evil?“ *Tijdschrift voor Filosofie* 68 (2): 251–278.
- Greene, J. (2013): *Moral Tribes*. Penguin Books, New York.
- Haidt, J. (2008a): „The Moral Roots of Liberals and Conservatives.“ In *TED* [online]. 18. 9. 2008 [cit. 1. 7. 2021]. Dostupné na: [https://www.ted.com/talks/jonathan\\_haidt\\_the\\_moral\\_roots\\_of\\_liberals\\_and\\_conservatives/transcript](https://www.ted.com/talks/jonathan_haidt_the_moral_roots_of_liberals_and_conservatives/transcript).
- Haidt, J. (2008b): „What Makes People Vote Republican.“ In *Edge* [online]. 8. 9. 2008 [cit. 1. 7. 2021]. Dostupné na: [https://www.edge.org/conversation/jonathan\\_haidt-what-makes-people-vote-republican](https://www.edge.org/conversation/jonathan_haidt-what-makes-people-vote-republican).
- Haidt, J. (2008 [2001]): „The emotional dog and its rational tail: A social intuitionist approach to moral judgment.“ In *Reasoning. Studies of Human Inference and Its Foundations*, eds. J. E. Adler & L. J. Rips, Cambridge University Press, Cambridge, 2008, s. 1024–1052.
- Haidt, J. (2012): *The Righteous Mind*. Penguin Books, New York.
- Haidt, J. & Kesebir, S. (2010): „Morality.“ In *Handbook of Social Psychology. 5<sup>th</sup> Edition*, eds. S. Fiske, D. Gilbert & G. Lindzey, Wiley, Hoboken (NJ), 2010, s. 797–832.

- Huemer, M. (2008): *Ethical Intuitionism*. Palgrave, Basingstoke.
- Hursthouse, R. (2002): *On Virtue Ethics*. Oxford University Press, Oxford.
- Chudárková, P. (2020): „Kritika intuicionismu a Huemerův pokus o jeho revizi.“ *Filozofia* 75 (9): 776–792.
- Joyce, R. (2006): *The Evolution of Morality*. The MIT Press, Cambridge (MA).
- Kronqvist, C. & Elgabsi, N. (rkp.): „Taking moral reflection too easily. Understanding the morality of sexual relationships.“
- LaFollette, H. & Woodruff, M. L. (2015): „The limits of Haidt: How his explanation of political animosity fails.“ *Philosophical Psychology* 28 (3): 452–465.
- Levi, D. (2002): *In Defense of Informal Logic*. Springer, Dordrecht.
- Lukianoff, G. & Haidt, J. (2018): *The Coddling of the American Mind*. Penguin, New York.
- Murdoch, I. (1997 [1956]): „Vision and Choice in Morality.“ In *Existentialists and Mystics*, I. Murdoch, Chatto & Windus, London, 1997, s. 76–98.
- Nagel, T. (1979): „Moral Luck.“ In *Mortal Questions*, T. Nagel, Cambridge University Press, Cambridge, 1979, s. 24–38.
- Narvaez, D. (2010): „Moral Complexity: The Fatal Attraction of Truthiness and the Importance of Mature Moral Functioning.“ *Perspectives on Psychological Science* 5 (2): 163–181.
- Nilsson, J. (2020): „The Myth and Reality of the Welfare Queen.“ *The Saturday Evening Post* [online]. 18. 2. 2020 [cit. 1. 7. 2021]. Dostupné na: <https://www.saturdayeveningpost.com/2020/02/the-myth-of-the-welfare-queen/>.
- Nussbaum, M. (2004): „Emotions as Judgments of Value and Importance.“ In *Thinking about Feeling*, ed. R. Solomon, Oxford University Press, New York, 2004, s. 183–199.
- Pacovská, K. (2018): „Love and the Pitfall of Moralism.“ *Philosophy* 93 (2): 231–249.
- Platón (1993): *Ústava*. Přel. R. Hošek. Svoboda, Praha.

- Read, D. W. (2014): „Incest Taboos and Kinship: A Biological or a Cultural Story?“ *Reviews in Anthropology* 43 (2): 150–175.
- Schopenhauer, A. (2007): *O vůli v přírodě a jiné práce*. Přel. M. Váňa. Academia, Praha.
- Solomon, R. (1973): „Emotions and Choice.“ *The Review of Metaphysics* 27 (1): 20–41.
- Strandberg, H. (2020): „In Search of the Context of a Question.“ *SATS – Northern European Journal of Philosophy* 21 (2): 199–213.
- Tomasello, M. (2016): *A Natural History of Human Morality*. Harvard University Press, Cambridge (MA).
- Williams, B. A. O. (1981): „Moral Luck.“ In *Moral Luck: Philosophical Papers 1973–1980*, B. A. O. Williams, Cambridge University Press, Cambridge, 1981, s. 20–39.
- Williams, B. A. O. (1985): *Ethics and the Limits of Philosophy*. Routledge, London.
- Williams, B. A. O. (2008): „Plato against the Immoralist.“ In *The Sense of the Past*, B. A. O. Williams Princeton University Press, Princeton, 2008, s. 97–107.
- Winch, P. (1972): „Moral Integrity.“ In *Ethics and Action*, P. Winch, London, Routledge & Kegan Paul, 1972, s. 171–192.
- Winch, P. (1997): „Can We Understand Ourselves?“ *Philosophical Investigations* 20 (3): 193–204.

## Abstract

### „Moral Intuitions“. What We Can or Should Take from Reading Haidt

The text aims at a critical overview of some works by the social psychologist Jonathan Haidt, with respect to what implications Haidt's claims have for the notion of morality and for the possibility of a philosophical discussion of moral problems. I will argue that though Haidt's findings and his interpretation thereof are noteworthy and enriching, they distort some aspects of human morality and moral life. In the first part, I will touch upon the unclear points of Haidt's concept of moral intuition. I will be guided by the way in which Haidt introduces this concept and employs it in his Mark and Julie example. In the second part, I will suggest the impact of this unclarity on Haidt's politico-psychological analyses. In the last section, I will point at the significance of the diversity (overlooked by Haidt) of the philosophical theories of morality. The common denominator

of several my critical points is 1) the vagueness of Haidt's concept of "intuition" and 2) the fact that though he presents his investigations as purely descriptive and pluralistic, he implicitly assumes a roughly utilitarian notion of morality, including its normative aspects.

Keywords: intuition, morality, liberalism, conservatism, Jonathan Haidt

Beran, O. (2021): „Morální intuice“. Co si můžeme nebo máme odnést z četby Haidta.“ *Filosofie dnes* 13 (2): 3–43. Dostupné z [www.filosofiednes.ff.uhk.cz](http://www.filosofiednes.ff.uhk.cz).

---

# Výchovná a vzdělávací role sportu u myslitelů Sókrata, Platóna a Aristotela ve vztahu k problematice dobrého sportu a vedení dobrého života

Lukáš Mareš

*Teologická fakulta  
Jihočeská univerzita v Českých Budějovicích  
Branišovská 1645, 370 05 České Budějovice  
maresl03@tf.jcu.cz*

Príspevek se věnuje problematice antického řeckého sportu, konkrétně významu sportovních zápolení a jejich výchovné a vzdělávací roli. Pozornost autor věnuje rozboru pozic filosofů Sókrata, Platóna a Aristotela. Po nastínění kontextu tématu představuje a interpretuje základní filosofické a náboženské premisy sportovního výkonu a jeho výchovné role. Řadí mezi ně úsilí o dosažení božské přízně, nesmrtelnosti, vyššího společenského postavení, ale i ideálů *kalokagathia*, *areté* a dalších ctností. Důležitý rozměr antického sportu spatřuje rovněž v jeho formativním potenciálu směřujícím k přípravě na duševní život. Hlavní část příspěvku se věnuje přístupu k pohybovým zápolením u filosofů Sókrata, Platóna a Aristotela, kteří svými úvahami k chápání výchovné role sportu podstatně přispěli. Následně se autor zaměřuje na identifikaci dobrého sportu podle zmíněných myslitelů a na zkoumání role sportu v kontextu dobrého života, který nahlíží v aristotelském rámci jako dosahování blaženosti (*eudaimonia*). Cílem textu je identifikovat a vyzdvihnout výchovný smysl sportování, poukázat na jeho filosofické konotace a možnou roli sportu v otázce vedení dobrého života.

**Klíčová slova:** filosofie sportu, filosofie výchovy, areté, kalokagathia, dobrý sport, dobrý život

## 1. Nástin kontextu tématu: antické Řecko a sport

Antický řecký sport<sup>1</sup> představuje komplexní systém pohybových, herních a soutěživých aktivit, které jsou s ohledem na místo i historickou epochu prodchnuty hodnotami, filosofickými i náboženskými koncepcemi a ideály tehdejších Řeků. Řecká kultura se vyznačovala značnou soutěživostí, která tvoří důležitý aspekt právě sportovního zápolení. Dle Huizingy: „Řekové měli ve zvyku soutěžit ve všem, v čem bylo možné se utkat.“<sup>2</sup> Závodění a soutěžení v nejrůznějších oblastech vystihuje pojem *agón*, který znamená právě zápas, závod či soutěž.<sup>3</sup> Sport jakožto herně pohybově zápolivá aktivita tak tvořil základní složku antické řecké kultury a byl neodmyslitelnou součástí života společnosti.<sup>4</sup> Ti, kteří se odmítali cvičit, byli označováni termínem *idiótai* jakožto necvičení, duševně či tělesně nezpůsobilí občané, kteří se neúčastnili veřejného života obce.<sup>5</sup>

Při pohledu na původní pohybově zápolivé aktivity si můžeme všimnout několika vzájemně se prolínajících rozměrů. Patří sem rozměr rituálně-náboženský, herní, militantní a výchovně-vzdělávací.

Sportovní zápolení s rituálně náboženským kontextem je typické pro řecké slavnosti. Mezi nejznámější se řadí panhelénské hry (olympijské, pýthijské, isthmické, nemejské) a panathénajské slavnosti. Základním rysem těchto událostí je oslava božstev skrze atletický výkon.<sup>6</sup> Rozměr herní můžeme chápat v rovině osobního nastavení k činnosti (tzv. herní postoj) a inherentních herních vlastností samotné sportovní aktivity, především však na ontologické rovině jakožto manifestaci základních

---

1 Užití pojmu sport je s ohledem na jazykový původ i moderní používání tohoto pojmu problematické. Ačkoli sportovní aktivity antických Řeků měly silně rituální, náboženský a filosofický podtext, jsem přesvědčen, že s ohledem na vnitřní aspekty těchto aktivit si můžeme dovolit je označit za sportovní. Je nicméně potřeba zmínit, že dané aktivity s ohledem na jejich provázanost s řeckou kulturou znamenaly v mnoha aspektech něco jiného, než co si pod pojmem sport představujeme dnes. Typickými příklady provozovaných sportů jsou běžecké disciplíny, box, zápas, všeboj – pankration, pětiboj – pentathlon, ale i míčové sporty. Pro podrobnější vysvětlení pojmu sport v kontextu antického Řecka vizte např. Kyle (1983, s. 7).

2 Huizinga (1971, s. 72).

3 Bahník et al. (1974, s. 18).

4 Kretchmar et al. (2017, s. 75), Kouřil (2016b, s. 7–28).

5 Kouřil (2016b, s. 9).

6 Do tohoto rozměru lze rovněž zařadit pietní atletické závody popsané Homérem v *Íliadě*, které uspořádal Achilles na počest zemřelého Patrokla.

herních struktur kosmu.<sup>7</sup> Oporou je zde řecký filosof Hérakleitos, který ve zlomku B52 z Hippolyta píše: „Věk je hrající si chlapec, sunoucí kaménky ve hře – chlapecké království.“<sup>8</sup> Jirásek uvádí, že Hérakleitova slova o hře mají zdůraznit dialektičnost pohledu na kosmos, tedy především změnu, proměnlivost, variabilitu, novost. Podstatnými znaky této dialektické hry a soutěžení na ontologické rovině jsou zápas a boj.<sup>9</sup> Třetí rozměr sportu je právě militantní, kdy slouží jako příprava k boji. Typickým představitelem tohoto rozměru je Sparta se svou vojensky orientovanou výchovou (*agóge*). Tělesně zápolivá cvičení jakožto příprava budoucích vojáků jsou ovšem typická i pro athénská gymnasia. V nich se můžeme setkat se čtvrtým rozměrem sportu, kterým je komplexní výchova řecké mládeže i dalších občanů polis. Tento rozměr sportovních aktivit je úzce provázán s filosofickými koncepcemi vedení dobrého života a představuje důležitou součást komplexního rozvoje řeckého člověka. V příspěvku mu věnuji největší pozornost.

Problematice výchovné role sportu v antice se již řada studií věnovala.<sup>10</sup> Inovativní prvek tohoto příspěvku spatřuji v podrobnějším rozpracování otázky dobrého sportu (zejména jeho různých rovin) s ohledem na jeho výchovný potenciál a v nástinu jeho role ve vedení dobrého života. Základní výzkumné otázky, které v příspěvku rozpracovávám, jsou: Jaký sport můžeme označit za dobrý a proč? Jakou roli sport může mít v otázce vedení dobrého života? Jednotlivé kapitoly jsou strukturovány tak, aby článek na tyto otázky nabídl ucelenou odpověď.

Text je ryze teoretickou studií pohybující se na pomezí oborů filosofie a historie. Při psaní využívám reprezentativní odbornou literaturu k tématu. Čerpám především z českých a anglických odborných článků a monografií a do češtiny přeložených původních řeckých pramenů. Citace z anglických zdrojů jsou mými vlastními překlady. V textu použité řecké termíny uvádím kurzívou v české transkripci, která vychází z Prachova

7 Při aplikaci herních kosmických struktur do konkrétnější skutečnosti, jakou je sport, je potřeba upozornit na skrytou premisu, totiž že sport je ve své podstatě herní aktivitou. Tato teze není odbornou komunitou filosofů sportu automaticky přijímána, nicméně existuje řada autorů (včetně mě), kteří se k ní v různých ohledech kloní (např. Suits, Meier, Connor ad.). Co znamená herní, ovšem vyžaduje důkladnějšího zkoumání, pro které není v tomto textu prostor.

8 Svoboda (1962, s. 62).

9 Jirásek (2001, s. 87), Svoboda (1962, s. 54, 57–58).

10 Za pozornost stojí například práce H. L. Reidové, I. Jiráskova či A. Hogenové.

*Řecko-českého slovníku*.<sup>11</sup> Význam původních řeckých termínů odpovídá zmíněnému slovníku, případně je uvedeno jinak. Jména zahraničních autorek jsou v textu přechylována, zatímco v odkazu na jejich díla nikoli. Důraz kladu na odkrývání a analyzování filosofického rozměru řešeného tématu. Při psaní používám konkrétně tyto metody: deskripce, komparace, kritická analýza, interpretace a syntéza dílčích informací. Historické období, na které se zaměřuji primárně, představuje antické Řecko v rozmezí mezi 5. až 4. stoletím př. n. l. (klasické období).

## 2. Základní premisy sportovního výkonu s ohledem na jeho výchovnou roli

Charakteristickým rysem sportovních zápolení i celkového způsobu života antických Řeků je na jedné straně touha vyniknout ve společenské hierarchii (sociální rozměr), na druhé straně snaha zajistit si přízeň bohů a dosáhnout nesmrtelnosti (individuální rozměr). Vyniknutím mohl získat atlet uznání a proměnit tak své společenské postavení. Zejména jakožto olympijský vítěz byl rovněž vzorem pro mládež, čímž se hrával i edukační roli. Dle Kouřila antická řecká společnost pod vlivem mytologie a herojů uznávala skvělé bojovníky i atlety, kterým se podařilo zvítězit na sportovně náboženských slavnostech.<sup>12</sup> Osobní vyniknutí zajišťovalo řeckému občanovi zanechání nesmazatelné stopy a přiblížení se na roveň bohů. Ke snaze po získání nesmrtelnosti uvádí Arendtová toto: „Touha Řeků dosáhnout nesmrtelnosti vyrůstala z vědomí, že jsou jakožto smrtelníci obklopeni nepomíjející přírodou a že jejich život se odehrává před zraky bohů neznajících smrt.“<sup>13</sup> Snahou bylo skrze nesmrtelné činy zanechat ve světě nepomíjivé stopy a prokázat, že také člověk má božskou přirozenost.

V kontextu výchovné/formativní role sportu je potřeba upozornit na jeho selektivní, avšak i rovnoprávný rozměr. Sportování antických Řeků se pojí s dobrovolností vynaloženého úsilí, která je kontrastní vůči fyzickým aktivitám otroků, kteří na výběr neměli.<sup>14</sup> Svobodný občan

<sup>11</sup> Prach (2005).

<sup>12</sup> Kouřil (2016a, s. 15), Kouřil (2019, s. 79).

<sup>13</sup> Arendt (2009, s. 29–30).

<sup>14</sup> Reid (2012a, s. 286).



řeckého původu se mohl účastnit sportovních zápolení a byl si v nich roven ostatním občanům. Oproti tomu byli z aktivní účasti na sportování vyloučeni otroci, barbaři, vrazi a v řadě případů i ženy.<sup>15</sup> Olivová upřesňuje, že barbaři a otroci, stejně tak i neprovdané dívky se mohli účastnit všeřeckých slavností jako diváci. Naproti tomu provdané ženy měly i pasivní účast zakázanou.<sup>16</sup> Rovnost a vzájemné spojení, jež byly typickými znaky řecké demokracie 5. století př. n. l., byly často vyjádřeny nahotou sportovců. Kouřil situuje rovnoprávnost (*isonomia*) zejména do prostředí gymnasií, kdy si ve vzájemném střetu byli atleti sociálně rovni (*isos*).<sup>17</sup> Sport nebyl (jako v 8. a 7. století př. n. l.) výsadou aristokracie,<sup>18</sup> ale byl záležitostí svobodných Řeků.

Důležitou metou nejen dobrovolného sportovního úsilí a námahy (*ponos*), ale i dalších činností antických Řeků, bylo dosahování a manifestace *areté*. Patočka daný pojem výstižně charakterizuje jako výtečnost či lépe řečeno pravost ve smyslu být tím, čím vsutku daná věc jest.<sup>19</sup> S odkazy na filosofy Sókrata a Platóna si všímá filosofického rozměru *areté*, který odkazuje k hlubokému kontaktu jedince s celkem reality: „Areté není privátně individuální ctnost; je to něco pevně určeného, co se přesto děje v měnivých konkrétních situacích; je to úsilí člověka o prohlédnutí celku v zasazení do konkrétní situace.“<sup>20</sup> Jirásek s odkazem na Hogenovou upozorňuje na problematičnost vymezení daného pojmu a rovněž na proměnlivost jeho obsahu s ohledem na historický kontext. Podle historického vývoje a proměn společenských preferencí tak do něj patří statečnost v boji, výkonnost, ale i velkorysost, ušlechtilost, vzdělanost či duchovní velikost.<sup>21</sup> Sportování hrálo důležitou roli ve snaze dosahovat a demonstrovat právě *areté*.

Jako důležité s ohledem na výchovnou roli sportu se rovněž jeví filosoficko-antropologické poznatky tehdejších Řeků. Helénská tradice

15 Výjimku tvoří Sparta, slavnosti na počest bohyně Héry (tzv. héraje) či Platónova Akademie, kde údajně ženy byly.

16 Olivová (1988, s. 117).

17 Kouřil (2019, s. 76).

18 Olivová (1988, s. 95–96).

19 Patočka (1999, s. 58, 75).

20 Patočka (1996, s. 33).

21 Jirásek (2005, s. 213–214).

vycházela z předpokladu, že lidský pohyb (*kinésis*) začíná v duši (*psyché*<sup>22</sup>), která je hybnou silou těla (*sóma*).<sup>23</sup> Atletovo tělo jakožto produkt dobrovolného a intencionálního pohybu je pak vyjádřením kvality atletovy *psyché*.<sup>24</sup> Patočka v odkazu na Platóna označuje duši jako to, co se samo uvádí v pohyb (*to auto heauto kinoun*) a zahrnuje pluralitu pohybů, mezi kterými je napětí. Proto Patočka říká, že duše je celou svou povahou časová pohyblivost, která se navíc v platónském pojetí uvádí v pohyb tím, že žije ve vztahu k dobru (*agathon*).<sup>25</sup> Tělesná krása odráží krásu duševní, přičemž takto kultivovaná duše se podobá duši božské, pojí se s koncepty dobra a pravdy a manifestuje se v kráse tělesné.<sup>26</sup> Kultivovaná duše a kultivované tělo jsou základními cíli tehdejší filosofie výchovy. Důraz pouze na jednu ze složek by znamenal disharmonii a v posledku úpadek jedince-atleta jakožto člověka. Sportovní zápolení jakožto tělesný projev je tak v posledku zároveň duševní záležitostí, neboť nejde pouze o předvedení pohybové dovednosti, tělesné síly či obratnosti, ale i o předvedení duševní stránky jedince (jeho *areté* např. ve smyslu odvahy, zarputilosti, odolnosti) manifestované v tělesném projevu. Jde o to předvést v plnosti komplexní pohybový rozvrh lidské bytosti, který zahrnuje ji celou, jako jednotu těla a duše. Atletický výkon ve vrcholném období řeckého rozkvětu zahrnoval tuto dimenzi. Vystihuje ji pojem *kalokagathia*, který značí „tělesnou krásu spojenou s duševní ušlechtilostí v harmonické jednotě“.<sup>27</sup>

Důležitou roli ve výchově a formování řeckých občanů sehrála gymnasia. V těchto institucích se sportovní zápolení<sup>28</sup> významně provazuje s filosofií. Dokladem je nejen přítomnost filosofů a rozpravy s cvičícími mladíky, ale i hlubší ukotvení významu tělesného pohybu a cvičení pro

22 V řeckém pojetí duše zahrnuje prvky jako racionální kalkulace, emoční projevy i fyzické žádosti.

23 Reid (2011, s. 5).

24 Reid (2009, s. 45).

25 Patočka (1999, s. 52–53, 76).

26 Reid (2012a, s. 281–297). Dle Zowislové (2009, s. 90) se myšlenka, že tělesná krása je odrazem ctností lidské bytosti, objevuje již v Homérově *Íliadě*.

27 Bahník et al. (1974, s. 306). Podrobně o pojmu *kalokagathia* a jeho různých významech pojednává Martínková (2012).

28 Všimněme si, že užitý pojem „sportovní zápolení“ v tomto kontextu odkazuje především k tělesným cvičením využívajícím konkrétní sportovní aktivity či prvky. Sport ve smyslu soutěžní aktivity se objevuje zejména na řeckých slavnostech.

rozvoj člověka a správné vedení života. Jirásek hovoří o gymnasiích jako o místech, v nichž se spojovala všestranná péče o tělo s všestrannou péčí o duši. V průběhu 7. a 6. století př. n. l. se dle něho rodí zjemnělý ideál *areté*, harmonicky vyvažující lidskou osobnost, a ideál *kalokagathia*, propojující duchovní prvky kultury s kulturou fyzickou. Zmíněné ideály se realizovaly prostřednictvím praktické výchovy (*paideia*), v níž šlo o zdokonalování lidského těla prostřednictvím tělesných cvičení, ale také o rozvoj lidského vědomí skrze filosofické rozhovory: „Tak bylo obsahem starořecké paideie vychování fyzické a filosofické, potvrzením humanistického ideálu, v němž areté lidského činu je aktivitou těla a skutečností myšlení.“ Základní obsahovou strukturou této výchovy bylo umění músicke a gymnastické, kterými se člověk rozvíjel nikoli jednostranně, nýbrž komplexně.<sup>29</sup> Předmětem diskusí je erotický poměr mezi mladíky a staršími občany navštěvujícími gymnasia. Jirásek s ohledem na nahotu a pederastii uvádí, že se nejednalo primárně o sexuální orientaci, ale spíše o specifické sociální vztahy iniciačních rituálů, konkrétně vztahy muže k vybranému chlapci jako součást přechodu z dětství do dospělosti.<sup>30</sup>

Podrobnější rozpracování edukačního rozměru sportu poskytuje následující kapitola věnovaná pohledu tří řeckých filosofů na sportovní výkon.

### 3. Postoj filosofů k výchovné roli sportu: Sókratés, Platón, Aristotelés

Z dochovaných zdrojů je patrné, že antičtí řečtí filosofové věnovali pohybovým aktivitám pozornost. V zásadě můžeme rozlišit mezi jejich samotnou participací ve sportovních zápoleních, pasivní účastí na závodech a teoretickým rozborem i kritickou analýzou sportovních aktivit, zejména ve vztahu k filosofii výchovy a otázce vedení dobrého života.<sup>31</sup> Cílem následujících částí je podrobné představení náhledu

29 Jirásek (2005, s. 215–218). Kouřil (2019, s. 75) nastiňuje, že sport byl provázán s *timé* (ctí, vážností) a dalšími ctnostmi prostupujícími řeckou společností.

30 Jirásek (2009, s. 13).

31 O aktivní i pasivní účasti filosofů na sportovních událostech pojednává např. Jirásek (2005, s. 274).

na sport a jeho roli ve výchovném a vzdělávacím procesu u filosofů Sókrata, Platóna a Aristotela. Představuji rovněž obecnější východiska jejich uvažování, zejména s ohledem na vedení dobrého života. Nejde mi o rekonstrukci jejich celkové argumentace, ale o prezentaci hlavních pohledů na roli sportu ve výchovném procesu. Výběr těchto autorů odůvodňuji jejich zásadním vlivem na pozdější vývoj filosofického myšlení nejen v oblasti filosofie sportu a rovněž jejich podrobným rozpracováním filosoficko-antropologických úvah o člověku, které jsou pro uvažování o sportu a jeho významu důležitou podmínkou.<sup>32</sup> Kromě uvedených myslitelů můžeme v souvislosti s výchovou a sportem zmínit např. Pýthágoru, který údajně lákal mladé atlety ke studiu filosofie.<sup>33</sup>

### 3.1 Sókratés – sportováním k rozvoji rozumu a ctnosti

Zmínky o filosofu Sókratovi máme od jeho současníků a pozdějších myslitelů – Platóna, Aristotela, Aristofana, Xenofóna či Diogena Laertia. Badatelé současnosti i nedávné historie postoj Sókrata ke sportu a pohybovým aktivitám popisují, interpretují a kriticky hodnotí. Obecně můžeme říci, že Sókratův postoj ke sportu je podmíněn jeho uvažováním o důležitosti kultivace těla pro rozumovou aktivitu a také závislý na tom, v jaké intenzitě se sportovní činnosti daný jedinec věnuje. Dle Olivové byl Sókratés proti přepínání atletického tréninku a do popředí kladl diskuse o filosofických otázkách.<sup>34</sup> Na druhou stranu Guttman uvádí, že obdivoval tělesnou dokonalost a ačkoli sám primárně filosof, opovrhoval těmi, kteří se nezajímali o své tělo.<sup>35</sup> Tyto pozice se objevují v Xenofónových *Vzpomínkách na Sókrata* a Platónovu dialogu *Ústava*, které považuji za vhodné zdroje pro demonstraci Sókratova postoje ke sportování a výchově obecně, neboť Xenofón i Platón se oběma oblastem věnují.

V jedné z pasáží prvně zmíněného díla<sup>36</sup> dokazuje Sókratés svému příteli Epigenovi, že dobrý tělesný stav je po všech stránkách výhodnější

32 Zowislo (2009, s. 91).

33 Podrobněji Reid (2011, s. 44), Reid (2012b, s. 10, 145–146).

34 Olivová (1988, s. 136). Autorka mj. uvádí, že Sókratés pěstoval gymnastiku a tanec do vysokého stáří.

35 Guttman (2004, s. 23).

36 Xenofón (1972, kniha III, část 12).

než tělesná zchátralost. Vyzdvihuje zde význam vojenských tělesných cvičení jak pro tělo, tak pro duši:

„... v žádném jiném zápase ani v žádné činnosti na tom neproděláš, vycvičíš-li lépe své tělo... I tam, kde se zdá, že upotřebením těla je nepatrné, totiž při myšlení, i tam, jak každý ví, mnozí dělají velké chyby právě proto, že jejich tělo není zdravé.“

Sókratés zde upozorňuje na negativní dopady špatného tělesného stavu a na důležitost starat se o svou tělesnou složku skrze pohyb. Lze se domnívat, že zde má na mysli sportovní aktivity spíše než pohyb při práci, neboť Řekové spíše dávali přednost svobodnému vyjádření pohybu před nucenými aktivitami, které vnímali jako podřadné. Vyzdvižením významu tělesného pohybu pro myšlenkovou aktivitu prezentuje Sókratés tělesné cvičení jako podmínku a prostředek duševního rozvoje. Zároveň oceňuje estetický rozměr cvičení, které může člověku přidat na kráse vzhledu a oddálit stáří.

Pro uvažování o výchovné roli sportu je podstatné zmínit Sókratův důraz na uměřené tělesné cvičení. V Platónově dialogu *Ústava* Glaukón na základě Sókratova vedení nastiňuje negativa extrémů výchovy: „Ti, kteří se věnují výlučně tělesnému vzdělání, se tím stávají mnohem surovější, než je třeba, a ti, kdo múzickému umění, stávají se zjemněnými daleko víc, než aby je to mohlo zdobit.“<sup>37</sup> Jako správnou chápe Sókratés v *Ústavě* střední cestu mezi dvěma extrémů:

„Toho, kdo nejlepším způsobem směšuje s múzickými uměními gymnastické vzdělávání a podává je duši v nejpřesnější míře, bychom správně označili jako člověka na nejvyšší míru múzicky dokonale vzdělaného a nejharmoničtějšího...“<sup>38</sup>

V Xenofontových *Vzpomínkách* Sókratés dále vyzdvihuje důležitost ustavičné tvrdé práce pro získání kvalitních dovedností a osobní kultivo-

<sup>37</sup> Platón (1993, 410d).

<sup>38</sup> Tamtéž, 412a.

vanosti. Důležitou pomoc přitom poskytují autority a odborníci, kterým je dle Sókrata nutno naslouchat, aby se člověk potřebné dovednosti naučil.<sup>39</sup>

S ohledem na celkový obrázek Sókrata, jak je nám dostupný z dochovaných pramenů, můžeme konstatovat, že vedení dobrého života se u tohoto myslitele pojí primárně s rozumovým kritickým uchopováním skutečnosti, které konkrétně znamená snahu poznávat sebe sama, nalézat pravdu o realitě (tedy získávat vědění) a na rovině morálky žít podle ctnosti. Sport v tomto způsobu života plní nejen roli prostředku a podmínky pro kvalitní duševní život, ale je i vyjádřením určitých kvalit naší duše, které se ve sportovním výkonu ukazují a kultivují. Plní tak důležitou, avšak nikoli tu nejpodstatnější podmínku pro vedení dobrého života.

### 3.2 Platón – sport jako příprava na duševní život

Filosof Platón si uvědomoval důležitost tělesného pohybu a sportu ve výchovném procesu. Aktivně se věnoval zápasení a sportovními kláními se zabýval i ve svých spisech.<sup>40</sup> Považoval je za důležitý prvek ve výchově řádného občana obce, přičemž vzorem mu byla spartská koncepce.<sup>41</sup> Tělesný rozvoj měl ovšem, na rozdíl od spartského modelu, jít ruku v ruce s kultivací a harmonizací jednotlivých složek duše, k čemuž sloužilo na jedné straně provozování gymnastických cvičení (*gymnastiké*), na straně druhé pěstování múzického umění (*mousiké*). Múzická výchova zbavuje ducha hrubosti, gymnická pak změkčilosti těla. Atletický trénink má u Platóna především utilitární charakter jakožto příprava k válce, komplexní výchova má ovšem vést člověka k naplněnému životu a rovněž ho připravit na filosofický život.<sup>42</sup> Základním cílem výchovy a vzdělání je u Platóna, podobně jako u Sókrata, pěstování moudrosti, harmonizace jednotlivých složek duše a usilování o pravdu a dobro. Tyto cíle Reidová označuje za dosažitelné pro ty, kteří mají duše atletů a milují správné věci.<sup>43</sup>

39 Xenofón (1972, kniha IV, část 2).

40 Platónova aktivní účast ve sportu je předmětem diskusí. Dle Guttmanna (2004, s. 23) se Platón v mládí věnoval zápasení a získal ocenění na pýthijských, nemejských a isthmických hrách.

41 Olivová (1988, s. 136).

42 Platón (1993, 412a), Olivová (1988, s. 136), Reid (2007, s. 167–168).

43 Reid (2011, s. 66). Myšlenka, že ideální občané mají mít duši atletů, se objevuje v Platónovu dialogu *Zákony* (1997, 830a).

O tom, jak má výchova a vzdělání ideálně vypadat, je pojednáno v dialogu *Zákony*. Platón zde zdůrazňuje, že výchova a vzdělání mají být dostupné pro obě pohlaví a pro různé věkové kategorie.<sup>44</sup> Dle athénskeho hosta je správná výchova ta, jež směřuje „... k dobrosti již od dětského věku, která vštěpuje dítěti žádost a touhu, aby se stalo dokonalým občanem, umějícím vládnout i dát se ovládat podle spravedlnosti“.<sup>45</sup> Naopak přístup, dle něhož je nejdůležitější získání bohatství, tělesné síly či jiné dovednosti postrádající rozumnost, nemá být vůbec nazýván výchovou, protože je nedůstojný a celkově nevhodný.<sup>46</sup> Vzdělávání a výchova by neměly probíhat násilnou formou, nýbrž co nejpřirozenější cestou, kterou je cesta hraní. To zdůrazňuje Sókratés svému příteli Glaukónovi v dialogu *Ústava*: „Tedy, příteli, nezacházej s dětmi při učení násilně, nýbrž ať se učí hravě, abys také lépe mohl pozorovati, k čemu se kdo svou přirozeností hodí.“<sup>47</sup> Hra tak u Platóna vystupuje jako vhodný prostředek k formování dítěte po tělesné i duševní stránce. Pokud je provozována správně, vede ho navíc k získání smyslu pro mravní řád a k poslušnosti vůči zákonům.<sup>48</sup> Lidská bytost je Platónem navíc pojímána jako přirozeně hravá boží hračka:

„... proto má každý muž a každá žena ve shodě s touto úlohou hrát co nejkrásnější hry a tak prožít celý svůj život... člověk má prožívat život tak, že hraje jisté hry, obětuje a zpívá a tančí, čímž je schopen zjednávat si milost bohů, nepřátelům pak se bránit a v boji nad nimi vítězit.“<sup>49</sup>

Platónovo vyzdvižení tělesných zápolení a hry je potřeba vnímat v širším kontextu jeho uvažování. Lze říci, že v posledku jde Platónovi o něco jiného než pouze o tělesnou kultivaci a hraní si. Souhlasím s Dombrowskim, že

44 Platón (1997, 805cd, 833cd).

45 Tamtéž, 643e.

46 Tamtéž, 644a.

47 Platón (1993, 536e–537a).

48 Tamtéž, 424e–425a. Jirásek (2001, s. 87) v této souvislosti kritizuje Platónovo doporučení hrát stále stejné hry, což dle Platóna má vést k zachování vážnosti zákonů. Naopak nepřímou apeluje na změnu, novost a vývoj. Souhlasím zde s Jiráskem, zejména v bodě otevřenosti vůči lidské tvořivosti.

49 Platón (1997, 803c–e).

veškeré vzdělání a výchova jsou v samé podstatě zaměřeny na kultivaci duše. V tomto bodě je patrné Platónovo dualistické uvažování stavící do popředí duši nad tělo. Zdravé tělo nečiní automaticky zdravou i duši, avšak zdravá duše uvádí tělo do dobrého stavu.<sup>50</sup> V posledku směřuje výchova, vzdělání i sportovní výkon u Platóna k dosažení *areté*, která je kvalitou duše a dostupná pro muže i ženy.<sup>51</sup> Dle F. Novotného znamená termín *areté* u Platóna dobrou výkonnost, dobrý stav. Týká se nejen člověka, ale i například zvířete, rostliny či nástroje.<sup>52</sup> V kontextu lidského směřování se k *areté* dospívá skrze komplexní trénink, není tedy vrozená.<sup>53</sup> Reidová při interpretaci Platónovy *Ústavy* upozorňuje na spojení mezi *areté* a duší:

„V Ústavě je *areté* chápána jako správné uspořádání a harmonické fungování trojdílné duše; je to zdravý stav duše, podobně jako spravedlnost je zdravý stav obce... fyzický trénink pomáhá harmonizovat psyché, připravit ji pro těžkosti filosofování a kultivovat morální sílu požadovanou od služebníků obce.“

Dodává k tomu, že *areté* je kvalitou činící člověka dobrým i šťastným.<sup>54</sup>

Kromě dosahování *areté* nastiňuje v dialogu *Ústava* Platón význam sportování pro dosahování dalších ctností. V kontextu morální kultivace řadíme Platóna k prvním autorům, kteří popisují schéma tzv. kardinálních ctností (spravedlnost – *dikaiosyné*, moudrost – *sofia*, statečnost – *andreia*, uměřenost – *sófrosyné*).<sup>55</sup> Pisk při interpretaci Platóna k těmto ctnostem uvádí, že člověk je může získávat a rozvíjet právě ve sportu. Antický sport (v podobě řecké *gymnastiké*) neslouží pouze k pěstění těla, ale v platónské koncepci slouží právě ke kultivaci duše. Sportovec skrze odříkání a disciplínu kultivuje žádostivou složku duše a pěstuje ctnost uměřenosti. Ve vytrvání v náročném zápolení kultivuje vznětlivou

50 Dombrowski (1979, s. 30–31), Platón (1993, 403d).

51 Reid (2012b, s. 12–13).

52 Novotný (1998, s. 620–621).

53 Reid (2011, s. 6).

54 Reid (2007, s. 160–162).

55 Bejczy (2011, s. 1).



složku duše a získává ctnost statečnosti (vizte platónští strážci v dialogu *Ústava*). Je také postaven před výzvu uvážení správné míry a činění vhodných rozhodnutí. Optimálním jednáním v tomto směru kultivuje rozumovou část duše a ctnost praktické moudrosti (*frónésis*), připravuje si ovšem i půdu pro rozvoj teoretické moudrosti. Správným rozvojem daných částí duše a rozvojem zmíněných ctností nastává v duši spravedlivý stav mezi jejími částmi. Sport svou povahou rovněž vybízí k pěstování spravedlnosti ve smyslu dodržování pravidel a celkové úcty ke hře (*fair play*).<sup>56</sup>

Platónova filosofie se ovšem týká především duševního života, kterému je sportovní střetávání podřazeno. Filosof Jan Patočka hovoří o pojmu péče o duši (*epimeleia tés psychés*), který převzal právě z Platóna. V životním pohybu je pro duši žádoucí dostávat se z polohy úpadku (rozptýlenosti, nejednoznačnosti, pohlcenosti tělem) do polohy vzmachu (sjednocení, přehlednosti, rovnováhy s tělem). Důležitou úlohu v tomto procesu má řecká *paideia* (praktická výchova), která by měla kultivovat samopohyb duše, učinit jej zřejmým, sebe-vědomým a dovést duši k nahlédnutí.<sup>57</sup> Patočka uvádí, že péčí o duši v její podstatě a jejím živlu je filosofie.<sup>58</sup> V procesu životního formování a přípravy na filosofický život plní u Platóna tělesná cvičení a sportovní zápolení roli morálně rozvojových aktivit, které nejsou hlavní náplní lidského pobytu, ale jeho důležitou součástí.

### 3.3 Aristotelés – uměřené sportování jako součást výchovy občana

Platónův žák Aristotelés se rovněž zabýval výchovou a vzděláváním. Sám se atletem patrně nikdy nestal, ve svých spisech však věnoval sportovním zápolením pozornost. Ideálem pro něho byl ctnostný a po všech stránkách dobrý člověk.<sup>59</sup> Jednou ze vzdělávacích metod byl pro Aristotela kromě gramatiky, hudby a kreslení také atletický trénink (tělocvik), který měl

<sup>56</sup> Pisk (2010, s. 6–8, 14).

<sup>57</sup> Blecha (2015, s. 410–411).

<sup>58</sup> Patočka (1999, s. 126).

<sup>59</sup> Laertios (1995, 5:28–30).

athénské občany kultivovat nejen po stránce fyzické, nýbrž i po stránce intelektuální.<sup>60</sup> Aristotelés totiž věřil, že zdraví a správné fungování duše (ve smyslu *psyché*) je podmíněno zdravím a správným fungováním těla (*sóma*).<sup>61</sup> Vyvážená a harmonická krása atletova těla je pak odrazem jeho ctnostné duše, která u Aristotela vystupuje jako oživující a hybný princip těla (hylemorfické pojetí).

Tělesnou výchovu považoval Aristotelés za důležitý prvek pro formování řecké mládeže. Měla být dle něho veřejná, společná a uspořádaná zákonem, jako tomu bylo ve Spartě, avšak nikoli tak jednostranná a ryze účelová.<sup>62</sup> Správně prováděný tělocvik stál v Aristotelově koncepci výchovy na počátku, protože výcvik těla měl předcházet výcviku mysli. Tento názor se objevuje v *Politice*:

„Ježto jest zřejmo, že se má vychovávat dříve zvyky než naukou a že tělo se má cvičiti dříve než rozumová schopnost, plyne z toho, že se děti mají věnovati napřed tělocviku a zápasnictví; neboť onen utváří určitý stav tělesný, toto uschopňuje k výkonům.“<sup>63</sup>

Pohybová cvičení byla určena pouze chlapcům a kultivovala zejména mužnost, mohla ovšem pěstovat i krásu. Tento názor lze objevit v Aristotelově *Rétorice*:

„Krása u jinocha záleží v tom, že má tělo způsobilé k námahám, a to jak k těm, jež vyžadují hbitosti, tak k těm, k nimž je třeba síly, a hledíme-li k požitku, že se nám názorem líbí. Proto jinoši pěstující pětiboj bývají nejkrásnější, poněvadž jsou schopni hbitosti a zároveň síly.“<sup>64</sup>

Uměřené tělesné cvičení zohledňující možnosti daného jedince vnímal Aristotelés jako přínosné a pozitivní. Byl ovšem kritický k přehnané

60 Aristotelés (1998, 1337b, 20–30).

61 Mechikoff (2014, s. 57).

62 Aristotelés (1998, 1333a–1337a).

63 Tamtéž, 1338b, 4–8.

64 Aristotelés (2010, 1361b, 5–15).

fyzické aktivitě, která ničila zdraví malých chlapců. Ostře odsuzoval tendence vychovávat z mládeže pouhé atlety, zejména z toho důvodu, že se tím „... škodí tvaru a vzrůstu tělesnému...“<sup>65</sup>. Byl si vědom neblahých důsledků, které s sebou nadměrná fyzická námaha přináší. To je možné vypořádat mimo jiné na tomto jeho názoru objevujícím se v *Politice*:

„... neboť až do dospělosti jest třeba prováděti lehčí cviky a vyloučiti nepřirozené stravování a každou těžkou práci, aby se nepřekáželo vzrůstu; neboť významným důkazem pro to, jaké škody to působí, jest skutečnost, že v seznamu olympijských vítězů najdeme jen asi dvě nebo tři jména těch, kteří zvítězili jako muži i jako hoši, poněvadž nepřirozeným cvičením v mládí vyčerpali svou sílu...“<sup>66</sup>

Základním cílem pohybového zápolení bylo uschopnit člověka k vyšším výkonům, spojeným primárně s jeho rozumovou činností. Skrze tělesné cvičení mohl člověk dosahovat dokonalosti (*areté*), kterou Aristotelés ve shodě s Platónem vnímal jako produkt opakované činnosti, nikoli jako vrozený stav.<sup>67</sup> K tomu, aby mohl jedinec této dokonalosti nabýt, je třeba, aby prošel náročným tréninkem a habituací a správné chování měl jako zvyk (*ethos*). Tělesné cvičení za účelem pěstování správných zvyků má být dle Aristotela středem mezi extrémní nedostatkem a přemírou, podobně jako v případě jeho koncepce ctností.<sup>68</sup> Tento střed je ovšem pro každého odlišný, protože tělesné proporce i duševní schopnosti jedinců nejsou stejné. V *Etice Níkomachově* uvádí Aristotelés následující příklad:

„... neboť jestliže pro někoho jest mnoho snísti za deset min (1 mina = cca 0,57 kg, pozn. autora), za dvě pak málo, nepředepíše cvičitel šest; neboť i toho snad bude mnoho nebo

65 Aristotelés (1998, 1338b, 11). Tento Aristotelův postoj je třeba vidět v dobovém kontextu jako určitou reakci na vzestup profesionálního sportu.

66 Tamtéž, 1338b, 40–1339a, 5.

67 Reid (2011, s. 6–7).

68 V *Etice Níkomachově* (1996, 1106b, 24–30) Aristotelés uvádí konkrétně o mravní ctnosti následující: „Ctnost se týká citů a jednání, v nichž nadbytek jest chybou a nedostatek bývá kárán, střed však bývá chválen a jest správný; a toto oboje přísluší ctnosti.“

málo pro toho, kdo to má požití; pro Mílóna (slavný řecký atlet, pozn. autora) totiž málo, pro začátečníka v atletice mnoho. Podobně i při běhu a zápolení.<sup>69</sup>

K tomu, aby člověk správnou míru v jednání rozpoznal, potřebuje mít praktickou bystrost/moudrost (*frónésis*), jež je záležitostí praktického rozumu a dospívá se k ní skrze životní zkušenosti.<sup>70</sup>

Do značné míry se Aristotelés shoduje s Platónem a Sókratem, když u něho tělesné cvičení a sport figurují jako podstatný edukační prvek. V uměřené podobě přivádí člověka k hodnotám, o něž je třeba poutivě usilovat. Správně provozovaná fyzická aktivita pomáhá jedinci stát se dobrým a krásným, tedy k harmonickému ideálu *kalokagathia*. Kritický byl Aristotelés k profesionalizaci sportovního závodění, které nabývalo ve 4. stol. př. n. l. značné popularity.<sup>71</sup> Naopak vyzdvihoval důležitost rozvíjení duševních stránek jednotlivce, především činnost vědeckou a rovněž veřejnou. Životní naplnění a blaženost (*eudaimonia*) nespatořoval v lovu, hraní či sportovním závodění, ale v činnosti duševní (zejm. rozjímání, identifikaci a pěstování ctností), která odpovídá rozumové podstatě (esenci) člověka.<sup>72</sup>

#### 4. Antické Řecko, dobrý sport a dobrý život

Uvažování o antickém řeckém sportu a hodnotách v něm přítomných představuje pro dnešního badatele výzvu v podobě co nejpřesnějšího nahlížení na to, jaká byla tehdejší skutečnost. Jirásek s odkazem na Machovce trefně naznačuje, že antické řecké ideály a hodnoty přítomné ve sportovním zápolení je potřeba především prožívat v současnosti.<sup>73</sup> V praktickém uvažování tak nejde o to, jaký byl řecký sport a atlet (byť

69 Tamtéž, 1106a, 35–1106b, 5.

70 Reid (2011, s. 73).

71 Specializovaní atleti se v antickém Řecku objevují již dříve. Ve století 4. počet profesionálních atletů ovšem významně stoupl. Profesionální atleti, specializující se již na konkrétní disciplíny, tak byli skrze tělesné cvičení, bohatou stravu a hojný spánek vedeni k maximalizaci svých fyzických možností. Takto pojatá profesionální atletika se stala terčem ostré kritiky, mimo jiné i z řad filosofů. (Olivová, 1988, s. 139–144).

72 Tamtéž, s. 137.

73 Jirásek (2005, s. 213).

i toto může být relevantní), ale především o to, jaké hodnoty a aspekty sportování přítomné v antickém Řecku rezonují s lidskou bytostí natolik, že stojí za to je žít i v současnosti. Hodnoty a ideály (př.: *areté*, *kalokagathia*, *isonomia*, *andreia* či *aristeia*<sup>74</sup>) provázané s antickým řeckým sportem odrážejí touhu člověka být ve světě autentickým, ryzím, a pravdivým způsobem. Otázka provozování sportu se zde prolíná s otázkou vedení dobrého života. V kontextu období antického Řecka a klasických myslitelů Sókrata, Platóna a Aristotela se proto táži, jaký sport je dobrý a jakou úlohu takovýto sport hraje v otázce vedení dobrého života. Obě otázky jsou vysoce relevantní pro současného člověka, který může sport více než v předchozích historických etapách vnímat jako globální fenomén.

#### 4.1. Dobrý sport

Hovoříme-li o dobrém sportu, považuji za podstatné rozlišit, zda označení „dobrý“ se týká spíše aktivity samotné (zde mám na mysli např. rovnou soutěž, dopředu nejistý výsledek, či morální hledisko – kdy např. nejde primárně o destrukci oponenta), či zda je do aktivity přináší sportovec samotný, tj. zda jsou otázkou sportovcova přístupu (zde podstatná rovina ctností), či zda se případně týká sportovního prostředí a jeho hodnot (zde je zásadní myšlenka, že sport je ovlivněn společností a kulturou, ve které je provozován, a rovněž na ně sám působí). Při uvažování o dobrém (a špatném) sportu máme na mysli právě tyto tři složky. Dobrý sport pak zahrnuje dobrot každé z těchto tří složek, které se v praxi vzájemně ovlivňují. Při pohledu na antický řecký sport a myslitele klasického období vystupuje do popředí především složka nastavení a přístupu sportovce, což souvisí s rozvojem morálního uvažování o člověku v daném období. Reflexe aktivity a hodnotového prostředí jsou zde ovšem také přítomné a pokouším se na ně upozorňovat. Výrazem dobrý označuji to, co probíraní myslitelé za dobré považovali a pro co argumentovali.

V sókratovském duchu navrhuji za dobrý sport označit ten, který provozujeme ctnostně a který nám umožňuje rozvoj a používání rozumové kapacity. Nejednalo by se tedy o sport, který nás degraduje jako rozu-

74 Hrdinství.

mové bytosti (např. forma drsných zápasů<sup>75</sup>, extrémní používání naší tělesné složky vedoucí k její degradaci i degradaci rozumu), ani o sport s cílem vyhrát za každou cenu (bez respektu k soupeři a pravidlům). Reidová si zde všímá souvislosti mezi filosofickým tázáním a sportovním zápolením. Dle ní Sókratés navštěvoval athénská gymnasia za účelem získání mladých atletů pro filosofii, zejména pro jejich předpokládaný zájem o kultivaci ctností. Na rovině filosofické debaty se snažil účastníkům nastavit zrcadlo jejich myšlení a pomáhat jim v dosahování vědění (*epistémé*). V podobném duchu interpretuje Reidová ideální sportovní zápolení jakožto kooperativní soutěž (cooperative competition), kdy atleti spolu nesoupeří za účelem porážení či ponížení oponenta, ale ve vzájemném férovém střetu se vzájemně respektují, vydáváním svého maxima si pomáhají poznávat sebe sama, rozvíjet své schopnosti, a vytvářejí prostředí pro vzájemné získávání a rozvoj ctností.<sup>76</sup> Reidová spatřuje pojitko mezi sókratovským dialogem a sportovním zápolením v tom, že obě disciplíny se zaměřují na hledání pravdy. Atletické soutěže dle ní vykazují znaky autentického dotazování<sup>77</sup> a nestranného testování, podobně jako je tomu u Sókratova filosofického dialogu. Takovýto „sókratovský“ sport je procesem odkrývání pravdy o sobě, a to skrze tělesné zápolení a kooperaci s druhými.

V platónském duchu se jako dobrý sport jeví ten, který kultivuje jednotlivé složky naší duše, vede k rozvoji a předvádění ctností, který v kráse a harmonii lidského pohybu ukazuje na kvalitu duše a který uschopňuje člověka k prospěšnému fungování ve společnosti. Podrobněji hovoří o dobrém sportu u Platóna J. Pisk. Dle něho souvisí úroveň sportu se složkou duše, kterou primárně angažuje. Sport na žádostivé rovině duše je založen na získávání materiálních statků v podobě cen na soutěžích. Sport na vznětlivé (emocionální) rovině duše je založen na elementárním *agónu*, který se snaží dojít naplnění ziskem cti a slávy. Jako nejlepší označuje Pisk sport pojící se s rozumovou složkou duše,

---

75 Na možnou námitku, že rozvoj rozumnosti není jediným úkolem dobrého lidského života, by bylo možné reagovat v aristotelském duchu tak, že rozumnost a používání rozumu jsou natolik charakteristické pro člověka, že pokud tuto složku aktivně degraduje, žije poté jiný život než přirozeně lidský.

76 Reid (2011, 43–53).

77 Reid (2009, s. 40, 45).

u kterého je primární nikoli zápasení se soupeři, ale dokonalé provádění pohybů, při kterém se dosahuje dokonalé spolupráce těla s rozumem. Na této úrovni sportovec soupeří sám se sebou, překračuje vlastní (původní) limity a naplňuje sám sebe. Dle Piska pouze takovýto sport může člověku jakožto myslící bytosti přinášet uspokojení, protože přesahuje fyzický svět a dotýká se věčného a neměnného světa idejí.<sup>78</sup> Pisk správně podotýká, že je obtížné najít sport, který by spadal pouze do jedné z kategorií. Na praktické rovině se motivy sportovce a jeho přístupy prolínají. Spíše než o usilování o nejvyšší kategorii se jako reálné jeví harmonizace motivů (jednotlivých složek duše) a jejich rozumný poměr, který ovšem znamená, že atlet si je vědom přesahu sportovního zápolení a aktivně ho pěstuje. Dobrý sport tak angažuje celého člověka, přivádí ho do kontaktu se sebou samým a umožňuje mu rovněž skrze časovost a tělesnost kontakt s dimenzemi reality, které se jinak obtížně odkrývají. Považuji za žádoucí doplnit, že pěstování takového sportu se pro skutečné uvědomění jeho významu pojí s filosofickou reflexí samotného sportovcova výkonu.

V aristotelské koncepci můžeme za dobrý sport označit především ten, který není provozovaný v přehnané míře, ale který kultivuje tělo a plní roli prostředku pro přípravu na veřejný, resp. kontemplativní život. Dle Reidové je Aristotelův ideální atlet v pravé podstatě *kalos kagathos*, krásný a v morálním smyslu dobrý/řádný člověk, který se do svého současného bytí zformoval usilovným tréninkem a konáním krásných a společensky prospěšných skutků. Takovýto občan je schopný rozeznávat, co je dobré a krásné, a aktivně k daným věcem tihnout a jednat v souladu s nimi.<sup>79</sup> Dobrý sport v Aristotelově podání by tedy znamenal především takovou činnost, která je provozována v rozumné míře, čímž přispívá, podobně jako u Platóna, k rozvoji tělesné krásy a ke kultivaci a demonstraci ctností.

Na obecné rovině se dobrému sportu s ohledem na období antiky věnují z hermeneutického a fenomenologického hlediska Jirásek, Oborný a Hurych, kteří hovoří o autentickém a neautentickém sportu. Druhé zmíněné označují pojmem *techné athlétiké*, což představuje úsilí

<sup>78</sup> Pisk (2006).

<sup>79</sup> Reid (2020, s. 65–72).

o rozvoj některých částí lidského těla, aniž by byl zohledněn člověk v holistické perspektivě.<sup>80</sup> Ve *Slovníku antické kultury* je termínem *techné athlétiké* označen úpadkový stav řecké gymnastiky pěstované profesionálními atlety jakožto zápasnické řemeslo. Negativita vyplývá z údajného primadonství, korupce, rozněcování nízkých pudů, soutěžení v hrubé síle a pěstění jen těch částí těla, které jsou potřebné pro dosažení vítězství či získání rekordu, což vedlo k tělesné deformaci.<sup>81</sup> Proti tomu kladou zmínění autoři koncept *techné gymnastiké* stavící do popředí přirozenou lidskou sílu, harmonický rozvoj osobnosti (vizte platónský sport), autentický prožitek charakterizovaný hlubokým vnímáním časovosti a rovněž krásu lidského pohybu. Autoři upozorňují na obtížnost určení hranic mezi autentickým a neautentickým sportem, protože oba koncepty se mohou v praxi v různých ohledech mísit a splývat. Proto je potřebná citlivost náhledu a pečlivé rozlišování, které mohou odkrývat, o jaký typ sportu se jedná. V souvislosti s autentickým modem sportu zmiňují atributy jako subjektivitu (osobní angažovanost pro výkon), kreativitu a hravost.<sup>82</sup>

#### 4.2. Role sportu v kontextu dobrého života

Po identifikaci možných pohledů na dobrý sport se nabízí otázka, jakou úlohu může sport (konkrétně výše identifikovaný dobrý sport) ve společnosti a obecněji v kontextu dobrého života plnit. Již u Xenofana<sup>83</sup> se objevuje myšlenka, že sportovní výkon a fyzickou zdatnost nemůžeme nadřazovat moudrosti a že z rychlosti nohou či atletovy excelence se obec nestává lepší. Xenofanés poukazuje na relativně nízkou hodnotu sportovních zápolení pro formování obce a ukazuje, že spíše než o fyzickou zdatnost jde v lidském životě o duševní kultivaci.<sup>84</sup> Myslitelé Sókratés, Platón a Aristotelés nabízejí diferencovanější pohled, když ve sportu spatřují výchovný a vzdělávací potenciál. Dobrý sport v jejich podání formuje řádného občana a plní důležitou roli při jeho osobnostním i so-

80 Jirásek, Oborný & Hurych (2018, s. 5).

81 Bahník et al. (1974, s. 85).

82 Jirásek, Oborný & Hurych (2018, s. 5–9).

83 Xenofanés, Zlomky DK 21 B 2.

84 V moderním pojetí se s podobnou kritikou sportu lze setkat např. u N. Chomskyho.



ciálním růstu. Shodují se ovšem v tom, že vedení dobrého života spočívá v něčem jiném než v pouhém provozování sportovních aktivit. Samotná problematika vedení dobrého života je značně komplexní. V následujících řádcích se proto pokusím nastínit roli dobrého sportu (resp. sportu popsaného výše uvedenými žádoucími vlastnostmi) v kontextu dobrého života u tradičního představitele eudaimonistického pojetí etiky a filosofické antropologie, již diskutovaného Aristotela. Jeho náhled na dobrý život se navíc v mnohém (byť ne ve všem) shoduje s pohledy Sókrata a Platóna.

Pro Aristotela je nejvyšším cílem lidského života dosažení blaženosti (*eudaimonia*). Tímto pojmem se zabývá především v první a desáté knize své *Etiky Níkomachovy*. Blaženost chápe jako nejvyšší dobro a blaženého člověka charakterizuje jako toho, kdo dobře žije, dobře jedná a dobře se má.<sup>85</sup> Pro výměr toho, co činí člověka blaženým, je možné odkázat na Aristotelovu koncepci rozdělení způsobů života a druhů dobra. Dle Aristotela jsou tři způsoby života, které nazývá jako nejvíce vynikající. Jedná se o život požívačný, politický a rozjímavý, který je řazen nejvýše.<sup>86</sup> Dobra pak Aristotelés dělí na zevnější, tělesná a duševní, přičemž duševní dobra chápe jako nejvíce vhodná k vyhledávání.<sup>87</sup> Blaženým v tom nejvyšším slova smyslu se dle Aristotela stává dospělý člověk (nikoli zvíře či malé dítě), který je kultivovaný ctností, vede rozjímavý (tedy filosofický) život podle rozumu a usiluje o duševní dobra.<sup>88</sup> K tomu však dodává následující:

„Zároveň jest zjevno, že potřebuje také zevnějších dober, jak jsme řekli; jest totiž nemožno nebo nesnadno, aby jednal krásně ten, kdo nemá nutných prostředků. Vždyť mnoho věcí se koná s pomocí přátel, bohatství a politické moci jako s pomocí nástroje; nedostatek některých dober docela kalí štěstí, například nedostatek urozenosti, plodnosti a krásy; neboť není zcela blažen ten, kdo jest ohyzdný, nízkého půvo-

85 Aristotelés (1996, 1095a, 15–20).

86 Tamtéž, 1095b, 15–20.

87 Tamtéž, 1098b, 10–16.

88 Tamtéž, 1100a, 1178a, 5–10.

du, osamělý nebo bezdětný, a ještě snad méně, má-li někdo nezdravné děti nebo špatné přátele, anebo, ač byli dobří, zemřeli.“<sup>89</sup>

Z uvedeného výčtu je patrné, že Aristotelés si uvědomuje, že pro dosažení dobrého života je zapotřebí i jiných než duševních dober, a přisuzuje určitý význam tělesné kráse a pohybové kultivovanosti.

Aristotelés se podrobně zabýval fenoménem volného času, který u něho tvořil protipól k zaměstnání. Jak uvádí Olivová, „pro něho byl celý lidský život rozdělen na zaměstnání (ascholiá) a volný čas (scholé)“.<sup>90</sup> Právě ve volném čase měl jedinec získávat vzdělání a rozvíjet zejména své duševní stránky. V této souvislosti kladl Aristotelés, podobně jako Platón, důraz na hudbu, která byla protikladem k tělocviku, jenž kultivoval primárně tělo.<sup>91</sup> Odmítal vyplnit volný čas hrou (patrně tedy i sportem), protože by se poté mohla stát účelem a cílem života:

„Bylo by to zajisté něco zvláštního, aby cílem byla hra a abychom se po celý život namáhali a zlé snášeli pro pouhé hraní... Bylo by věru pošetilé a příliš dětinské starati se a pracovati jenom pro hru... Zdá se, že blažený život je život ctnostný; tento však jest životem vážné práce a nikoli hry.“<sup>92</sup>

Mnohem podstatnější pro dosažení blaženosti tedy pro něho byla činnost vědecká a činnost veřejná, kterými se člověk uvnitř společnosti realizoval. Namísto aktivit jako je lov, hra či atletické závodění tak byla u Aristotela jako náplň volného času vyzdvihována činnost duševní, spojená s rozvojem sebe sama a se službou společnosti.<sup>93</sup> Je ovšem potřeba dodat, že pohybové aktivity chápal jako podstatný prvek na cestě k blaženému životu, což je patrné na následujícím jeho výroku z *Etiky Níkomachovy*:

---

89 Tamtéž, 1099a, 32 – 1099b, 6.

90 Olivová (1988, s. 137).

91 Aristotelés (1998, 1338ab).

92 Aristotelés (1996, 1176b, 25 – 1177a, 5).

93 Aristotelés (1998, 1337a, 30 – 1338a, 10).

„Blažený člověk však jako člověk bude potřebovat také zevnější pohody, neboť jeho přirozenost se nespokojí rozjímáním, nýbrž vyžaduje, aby také tělo bylo zdravé a aby se mu dostalo potravy a ostatní péče.“<sup>94</sup>

V Aristotelově podání je dobrý život chápán komplexně jako složenina vícera činností a ačkoli je zacílen primárně na činnost vědeckou a rozumovou/kontemplativní (které přinášejí nejvyšší úroveň blaženosti a dobra), nelze si ho v jeho plnosti představit bez dosahování dober ostatních, mezi která patří dobrý tělesný stav či tělesná krása. Z tohoto důvodu můžeme správně prováděné sportovní aktivity označit za nikoli hlavní, ale podstatnou součást dobrého života. Nutno ovšem dodat, že Aristotelés neposkytuje podrobněji rozpracovaný pohled na to, jaký typ sportu je dobrý<sup>95</sup> a jakou konkrétní úlohu má sport v otázce vedení dobrého života hrát. Nabízí pouze obecný pohled na vhodnost intenzity a cílů, které má sportování (na rovině soutěží i tělesného cvičení) sledovat.

## 5. Implikace do současného sportu

Idealizování řeckého sportu by bylo selektivním pohledem na určité období a pěstované hodnoty. Jako celek působí sportovní zápolení Řeků rozmanitě, diverzitivně i záhadně pro omezené množství dochovaných informací. Dochované záznamy a úvahy ovšem obsahují podstatné významy sportování pro rozvoj člověka i pro otázku vedení dobrého života, čímž jsou aktuální pro provozování sportu v dnešním světě. Je potřeba dodat, že problematika dobrého sportu a jeho role v otázce dobrého života není antickými filozofy rozpracována konkrétně. Současný badatel je tak vystaven před riziko činit zobecňující závěry, které postrádají informativnost pro praktický způsob života. Jako relevantní se mi proto jeví ptát se, jaké konkrétní implikace antické uvažování má pro současný sport, a to s ohledem na rozdíly v typu provozované aktivity (tj. konkrétní sport), kulturní zvyklosti, hodnoty a individualitu sportovce.

94 Aristotelés (1996, 1178b, 30 – 1179a, 5).

95 V *Rétorice* je ovšem patrné, že oceňuje pětiboj (*pentathlon*).

Zde se pokouším načrtnout možné obecné směry uvažování o významu antické pohybové kultury pro současný sport (myšleno 21. století).

- Důraz na rozumnou intenzitu a optimální množství sportovních aktivit v životě sportovce je obzvláště relevantní v mládežnickém výkonově pojatém sportu, kde hrozí riziko vykořisťování dětí přílišným důrazem na striktní režim.<sup>96</sup>
- Sport není pouze o vítězství<sup>97</sup> či pěstění tělesné složky, ale také nám pomáhá v rozvoji charakteru, tedy k získávání a kultivaci ctností.
- Dodržování pravidel a vyhýbání se jejich porušování (např. v podobě neuzívání dopingů) je vymahatelné nikoli pouze na základě formální správnosti, ale protože takové jednání vede atleta k rozvoji jeho přirozené krásy a vnitřní „dobroti“ (subjektivní perspektiva založená na etice ctností).
- Soupeř nemusí být vnímán jako nepřítel, kterého je potřeba za každou cenu porazit. Je naším přítelem v soutěži (vizte D. Hyland a competition as friendship<sup>98</sup>), který nám ve vzájemném střetávání umožňuje poznávat vlastní schopnosti a činí nás lepšími (podobně jako my jeho).

Nabízí se otázka, nakolik jsou tyto a v textu dříve zmíněné myšlenky ohledně dobrého sportu realizovatelné v současném sportovním světě. Je podstata „dobrého sportu“ (ne)slučitelná s většinovými modely současného sportu? Je zřejmé, že ve výkonnostním, vrcholovém a profesionálním sportu je trendem důraz na maximální výkonnost, přibližování se na hranice vlastních možností a posouvání limitů. Perspektiva zdravého a rozumného přístupu ke sportování, která se odráží i ve výchovné a vzdělávací roli sportu u antických Řeků, je proto snadno vystavena před riziko vytlačení jednostranným důrazem na výsledek, produkci výkonu a maximalizaci tělesných kapacit. Pro implementaci antických

96 K etickým aspektům zdraví a dobrého života v mládežnickém sportu vizte Zurc (2019).

97 Edgar (2013, s. 131) uvádí, že sport jakožto soutěž (contest) v sobě obsahuje riziko být poražen. I v individuálních sportech jako je horolezectví spatřuje jako důležitý prvek možnost selhání, které je pro sportovní zápolení zásadní a se kterým se sportovec vyrovnává.

98 Hyland (1978).

řeckých ideálů do současnosti je potřeba si uvědomit, že sport, zejména ten mládežnický, není jednostrannou záležitostí, není pouhým prostředkem k realizaci ambicí, není cestou k materiálnímu zabezpečení, není pouhou prací, ale zahrnuje v sobě možnost rozvoje osobních kvalit (charakterových vlastností/ctností), sociálního a týmového cítění, ale i vztahu k přirozenému pohybu a vlastní tělesnosti. Takovýto sport zahrnuje i odevzdávání maxima a snahu vítězit, ovšem bere zřetel na způsob, jakým toho bude docíleno. Žít tuto perspektivu vyžaduje erudici, zápal a odvalu trenérů, sportovních funkcionářů, organizací, rodičů i samotných sportovců, kteří podobu sportu v dnešním světě utvářejí.

## 6. Závěr

Dějiny antického Řecka jsou prostoupeny tělesnými cvičeními a sportem, které v dějinném vývoji nabývaly rozmanitých podob a významů. Sportovní zápolení Řeků bylo podstatně provázáno s náboženstvím a filosofií. Sloužilo jakožto prostředek seberealizace, připodobnění božské sféře a sehrálo podstatnou úlohu i v oblasti výchovy a vzdělání. Sport tvořil důležitou součást života tehdejších Řeků a podílel se na formování zodpovědného a schopného občana *polis*. Reidová správně poznamenává, že atletika v kontextu antické výchovy nesloužila pouze jako fyzický trénink, ale byla cestou ke kultivaci silné duše schopné hledat a nalézat pravdu. Sportovní zápolení formovala občany tak, aby řádně dokázali sloužit své obci a vlasti.<sup>99</sup> Svým přínosem si sport získal pozornost rovněž filosofů, kteří ho reflektovali, oceňovali, kritizovali a v některých případech i sami provozovali. V příspěvku diskutovaní myslitelé Sókratés, Platón a Aristotelés si uvědomovali vzdělávací a výchovný (resp. etický) rozměr sportování a ve svých úvahách je rozpracovávali.<sup>100</sup> Sport v jeho dobré podobě pro ně představoval vhodný způsob kultivace a tvořil nikoli hlavní, ale přesto důležitou součást dobrého života. Takovýto sport přináší cenné inspirace pro dobu současnou, ve které tlak na výkon a uspokojujivý výsledek mnohdy vytlačují formativní rozměr sportování.

99 Reid (2009, 45).

100 Etickému rozměru antického sportu s ohledem na (post)moderní důraz na estetický rozměr sportování se věnuje Zowislová, vizte Zowislo (2009, s. 91).

## Literatura

- Arendt, H. (2009): *Vita Activa*. Přel. V. Němec. OIKOYMENH, Praha.
- Aristotelés (1996): *Etika Nikomachova*. Přel. A. Kříž. Rezek, Praha.
- Aristotelés (1998): *Politika*. Přel. A. Kříž. Rezek, Praha.
- Aristotelés (2010): *Rétorika*. Přel. A. Kříž. Rezek, Praha.
- Bahník, V. et al. (1974): *Slovník antické kultury*. Svoboda, Praha.
- Bejczy, I. P. (2011): *The Cardinal Virtues in the Middle Ages*. Brill, Leiden, Boston.
- Blecha, I. (2015): „Patočková péče o duši mezi Sókratem, Platónem a Aristotelem.“ *Filozofia* 70 (6): 409–419.
- Dombrowski, D. A. (1979): „Plato and Athletics.“ *Journal of the Philosophy of Sport* 6 (1): 29–38.
- Edgar, A. (2013): „The Modernism of Sport.“ *Sport, Ethics and Philosophy* 7 (1): 121–139.
- Guttman, A. (2004): *From Ritual to Record: The Nature of Modern Sport*. Columbia University Press, New York.
- Huizinga, J. (1971): *Homo ludens: o původu kultury ve hře*. Přel. J. Vácha. Mladá fronta, Praha.
- Hyland, D. (1978): „Competition and Friendship.“ *Journal of the Philosophy of Sport* 5 (1): 27–37.
- Jirásek, I. (2001): *Prožitek a možné světy*. Univerzita Palackého, Olomouc.
- Jirásek, I. (2005): *Filosofická kinantropologie: setkání filosofie, těla a pohybu*. Univerzita Palackého, Olomouc.
- Jirásek, I. (2009): „Nahota v kontextu pohybové kultury.“ *Tělesná kultura* 32 (2): 7–19.
- Jirásek, I., Oborný, J. & Hurych, E. (2018): „The Authentic and Inauthentic Sport In the Hermeneutic and Phenomenological Perspectives.“ *Acta Facultatis Educationis Physicae Universitatis Comenianae* 58 (1): 1–11.
- Kouřil, J. (2016a): „Athlétai a sportovní tituly antického světa.“ *Tělesná kultura* 39 (1): 15–26.

- Kouřil, J. (2016b): „Sportovní příprava a trénink v řecko-římské antice.“ *Česká kinatropologie* 20 (2): 7–28.
- Kouřil, J. (2019): „Olympism‘ and Olympic Education in Greek Antiquity.“ *Studia Sportiva* 13 (1): 74–84.
- Kretchmar, R. S. et al. (2017): *History and Philosophy of Sport and Physical Activity*. Human Kinetics Publishers, Champaign, Illinois.
- Kyle, D. (1983): „Directions in Ancient Sport History.“ *Journal of Sport History* 10 (1): 7–34.
- Laertios, D. (1995): *Životy, názory a výroky proslulých filosofů*. Přel. A. Kolář. Nová tiskárna Pelhřimov, Pelhřimov.
- Martínková, I. (2012): „Jak rozumět kalokagathii?“ *Tělesná kultura* 35 (1): 93–105.
- Mechikoff, R. A. (2014): *A History and Philosophy of Sport and Physical Education: From Ancient Civilization to the Modern World*. McGraw-Hill, New York.
- Novotný, F. (1998): „O překládání z filosofického jazyka Platónova.“ *Filosofický časopis* 46 (4): 619–629.
- Olivová, V. (1988): *Sport a hry ve starověkém světě*. Artia, Praha.
- Patočka, J. (1996): *Péče o duši I*. Připravili I. Chvatík & P. Kouba. OIKOYMENH, Praha.
- Patočka, J. (1999): *Péče o duši II*. Připravili I. Chvatík & P. Kouba. OIKOYMENH, Praha.
- Pisk, J. (2006): „What Is Good Sport: Plato’s View.“ *Acta Universitatis Palackianae Olomucensis. Gymnica* 36 (2): 67–72.
- Pisk, J. (2010): „The Role of Cardinal Virtues in Sport.“ In Hosta, M. (2010): *Philosophical Reflections in Sport: A Collection of Essays*. International Institute for Sustainable Development, Policy, and Diplomacy in Sport, Ljubljana.
- Platón (1993): *Ústava*. Přel. R. Hošek. Svoboda, Praha.
- Platón (1997): *Zákony*. Přel. F. Novotný. OIKOYMENH, Praha.
- Prach, V. (2005): *Řecko-český slovník*. Vyšehrad, Praha.
- Reid, H. L. (2007): „Sport and Moral Education in Plato’s Republic.“ *Journal of the Philosophy of Sport* 34 (2): 160–175.

- Reid, H. L. (2009): „Sport, Philosophy, and the Quest for Knowledge.“ *Journal of the Philosophy of Sport* 36 (1): 40–49.
- Reid, H. L. (2011): *Athletics and Philosophy in the Ancient World: Contests of Virtue*. Routledge (Taylor & Francis Group), Abingdon.
- Reid, H. L. (2012a): „Athletic Beauty in Classical Greece: A Philosophical View.“ *Journal of the Philosophy of Sport* 39 (2): 281–297.
- Reid, H. L. (2012b): *Introduction to the Philosophy of Sport*. Rowman & Littlefield Publishers, Inc., Plymouth.
- Reid, H. L. (2020): „Athletic Virtue and Aesthetic Values in Aristotle’s Ethics.“ *Journal of the Philosophy of Sport* 47 (1): 63–74.
- Svoboda, K. (1962): *Zlomky před Sokratovských myslitelů*. 2. vyd. Nakladatelství ČSAV, Praha.
- Xenofanés. „Zlomky DK 21 B 2.“ In *Fysis – φύσις* [online]. [cit. 1. 6. 2021]. Dostupné na: <http://www.fysis.cz/presokratici/xenofanes/bcz.htm>.
- Xenofón (1972): *Vzpomínky na Sókrata*. Přel. V. Bahník. Svoboda, Praha.
- Zowislo, M. (2009): „On Virtue in the Context of Sport.“ *Physical Culture and Sport. Studies and Research* 47 (1): 89–94.
- Zurc, J. (2019): „Ethical Aspects of Health and Wellbeing of Young Elite Athletes: Conceptual and Normative Issues.“ *Synthesis Philosophica* 68 (2): 341–358.

## Abstract

### **Educational Role of Sport with Respect to the Thinkers Socrates, Plato, and Aristotle In Relation to the Problematics of a Good Sport and a Good Life**

This manuscript deals with the problematics of ancient Greek sport, specifically with the significance of sporting practices and their educational role. Attention is given to a detailed analysis of the positions of the philosophers Socrates, Plato, and Aristotle. After demonstrating the context, the author presents and interprets fundamental philosophical and religious premises of the athletic performance and its educational role. Among the identified features are the effort of acquiring divine favour, immortality, higher social status, but also the ideals



of *kalokagathia*, *areté*, and other virtues. Important dimension of ancient Greek sport is also found in its formative potential which aims at the preparation for intellectual life. The main part of the article is focused on the approaches of the philosophers Socrates, Plato, and Aristotle towards the sporting practices. These thinkers made an important contribution to the understanding of the educational role of sport. Then, the author identifies a good sport according to these thinkers and examines the role of such sport in the context of a good life, which is considered in the Aristotelian perspective as achieving beatitude (*eudaimonia*). The aim of this article is to identify and highlight an educational purpose of sport, underlines its philosophical connotations, and demonstrate its possible role in the context of a good life.

Key words: philosophy of sport, philosophy of education, *areté*, *kalokagathia*, good sport, good life

Mareš, L. (2021): „Výchovná a vzdělávací role sportu u myslitelů Sókrata, Platóna a Aristotela ve vztahu k problematice dobrého sportu a vedení dobrého života.“ *Filosofie dnes* 13 (2): 44–72. Dostupné z [www.filosofiednes.ff.uhk.cz](http://www.filosofiednes.ff.uhk.cz).

---

# Superinteligence a problém kontroly: Skutečný problém nebo pseudo-problém?

Jaroslav Malík

*Filozofická fakulta  
Univerzita Hradec Králové  
Rokitanského 62, 500 03 Hradec Králové  
jaroslav.malik@uhk.cz*

V tomto článku se zabývám konceptem SI (superinteligence) a s ní spojeným problémem kontroly. Podle určité skupiny teoretiků umělé inteligence stojíme na prahu události, která může radikálně změnit povahu technologického pokroku a lidské společnosti obecně. Touto událostí má být takzvaná technologická singularita, která je často spojována se vznikem první větší než lidské inteligence. Lidé jako Nick Bostrom varují před nebezpečím, které pro nás vznik SI znamená, a upozorňují, že musíme co nejdříve najít metody kontroly této inteligence. Podle Bostroma a dalších nebezpečí SI vyplývá z její povahy. Já na jedné straně zkoumám, jak SI může vzniknout, a na druhé straně posuzuji smysluplnost problému kontroly. Vznik SI vyplývá ze vzniku umělé inteligence. Proto podstatnou část textu věnuji argumentům pro vznik umělé inteligence. Ukazuji, jak se Bostrom a další nechávají unést jedním problematickým argumentem. Dále podrobuji jejich pozici klasické kritice umělé inteligence. Prezentuji, jak stále lpí na problematických domněnkách svých předchůdců. Zejména soustřeďuji svoji kritiku na tvrzení, že SI bude mít jeden konečný cíl, který bude interpretovat. Toto tvrzení označuji jako antitetické, a to představě, že SI bude obecnou inteligencí. V závěru ze své argumentace vyvozují, že při formulaci problému kontroly teoretikové umělé inteligence zaměňují dva jiné různé problémy „kontroly“.

**Klíčová slova:** technologická singularita, superinteligence, umělá inteligence, emulace celého mozku, problém kontroly, tělesnost

## 1. Úvod

Představte si, tak jako Nick Bostrom, že jednou postavíme inteligentní stroj na výrobu sponek. Inteligentním míním to, že má v sobě zabudovanou funkční umělou inteligenci. Tomuto stroji dáme velice prostý úkol: „Vyrob tolik sponek, kolik jen dokážeš.“ Náš stroj na začátku plní svůj úkol více než dobře, díky učícím se procesům, které jsou v něm zabudovány, nejenže vyrábí sponky stále stejnou rychlostí, ale dokonce se postupně zlepšuje. Náš stroj se nejdříve pouze učí lépe provádět standardní proces výroby, ale netrvá to dlouho, než stroj začne nacházet inovace, které radikálně změní výrobu sponek. Nejdříve se jedná o pouze malé změny, možná přijde s novou technikou, nebo vynalezne nový přístroj pro splnění svého úkolu. Jeho objevy jsou však čím dál tím radikálnější, až nakonec přijde na „zázračnou“ metodu, která našemu stroji umožní přeměnit cokoliv na sponky. V tuto chvíli zaregistrujeme první problémy. Náš stroj nejdříve začne přeměňovat vše ve své blízkosti, ale stejně jako optimalizoval své předchozí metody, se mu nakonec daří přeměňovat věci v širším prostoru. Lidé jsou samozřejmě zděšení a pokusí se tento stroj v jeho snaze zastavit. Bohužel jejich snahy se ukážou být marné, náš nyní již superinteligentní stroj počítal s jejich reakcí, takže měl obranné prostředky. Dost brzy je celá planeta Země přeměněna na sponky a po ní následuje i celá sluneční soustava. Bostromův stroj se poté vydává dál, pokračovat ve svém úkolu po celé galaxii.<sup>1</sup>

Pravděpodobně se již každý z nás setkal s myšlenkou, že až se nám jednou podaří vytvořit umělou inteligenci (dále už pouze jako UI), tak nám dost brzy bude hrozit, že se UI vzbouří proti nám. Koneckonců jedná se o dost známé sci-fi klišé. Tyto myšlenky mají svoje kořeny už v samotných začátcích vývoje UI a pravděpodobně od první chvíle, kdy se někdo pokusil něco takového jako UI stvořit, musel začít uvažovat o takové myšlence. V oboru výzkumu UI se můžeme setkat s dlouhou historií techno-optimismu v souvislosti se vznikem UI. Lidé jako Marvin Minsky nebo Allen Newell už v 60. letech předpokládali, že UI je už nejen na obzoru, ale že nás bude schopná překonat.<sup>2</sup> Tato problematika

---

1 Bostrom (2003, 2012, 2014, s. 193).

2 Viz Minsky (1968) a Simon & Newell (1958).

je v současné době už široce diskutovaným tématem. Pojem technologické singularity do této problematiky však přináší nepřilíš diskutovanou dimenzi. Tou je problém superintelligence. Podle teoretiků singularity (dále budu tuto skupinu označovat jako singularitáni) existuje možnost, že až vyvineme umělou inteligenci, která bude dostatečně sofistikovaná, tak brzy potom tato inteligence vytvoří ještě lepší intelekt. Pokud tato událost nastane, tak její následky budou podle singularitánů nedozírné. Varují před katastrofickými scénáři, ve kterých nás superintelligence (dále jako SI) zničí. Z toho pro nás vyplývají dvě otázky: Může skutečně vzniknout superintelligence? A pokud ano, vedlo by to k problémům?

Na úvod bude velmi důležité definovat, co míníme pojmy jako inteligence, umělá inteligence (UI) a superintelligence (SI). V tomto článku se budu držet definice inteligence, která je v rámci oboru UI obecně přijímána. Představa strojové inteligence konverguje k chápání inteligence jako optimálního plnění cílů napříč různými prostředími. Takže se budeme držet definice: „Inteligence je mírou agentovy schopnosti dosahovat cílů v širokém spektru různých prostředí.“<sup>3</sup> Pojem UI je obecně složité definovat. Russell a Norvig konstatují, že v historii UI můžeme identifikovat několik přístupů k jejímu chápání. Někteří preferují definovat inteligenci UI v měřítku lidského intelektu, jiní raději mluví o inteligenci jako o obecné racionalitě. Dále se někteří soustřeďují na to, aby se UI chovala inteligentně, a jiní dávají důraz i na to, aby UI byla schopná myšlení. To vytváří dvě dimenze, skrze které můžeme chápat přístupy k UI: myšlení/chování a lidské/racionální.<sup>4</sup> Singularitáni dle mého názoru UI chápou jako inteligenci, která se chová racionálně.<sup>5</sup> Proto v rámci tohoto textu budeme takto vnímat jejich výroky. SI budeme v textu chápat jako obecnou inteligenci, která je schopná dosahovat mnoha různých cílů. Bostrom SI definuje jako: „Jakýkoliv intelekt, který lidské kognitivní výkony dalece překonává prakticky ve

---

3 Legg (2008). Na této definici staví Bostrom (2014) a Muehlhauser & Helm (2012). Překlad vychází z Bostrom (2014, s. 405).

4 Russel & Norvig (2020, s. 1–4).

5 K této charakterizaci mě vedou Muehlhauser & Helm (2012), kteří popisují SI jako super optimalizátora, Yudkowski (2008), který chápe mysl jako optimalizační proces, Omohundro (2008, 2012), který se soustřeďuje na racionalitu SI, a Bostrom (2014, s. 47), který konstatuje, že SI nemusí být vědomá. Všichni také zastávají pozici, že SI bude inteligence odlišná od lidské.

všech relevantních oblastech.“<sup>6</sup> Z toho vyplývá, že budeme kompetence SI chápat v kontextu lidských kompetencí. Cokoliv člověk dokáže, SI zvládne také a lépe. Bostrom rozlišuje několik různých forem SI, které můžeme chápat jako odlišné systémy pro dosažení SI. Takže například rychlostní SI je intelligence na stejné úrovni jako ta naše, ale mnohem rychlejší. Tento typ se však chápe jako slabá forma SI. Další dva možné systémy jsou kolektivní a kvalitní SI. První dosahuje své úrovně intelektu skrze úzkou propojenost několika inteligencí a druhou můžeme chápat jako individuální intelekt, který převyšuje naše schopnosti s alespoň stejnou rychlostí jako náš mozek. Tyto různé SI také mohou vzniknout různými způsoby. Nehledě na typ, můžeme chápat všechny tyto SI jako součást hlavního problému.

Co je hlavní problém SI? Bostromův sponkový stroj je dobrý ilustrující příklad. V tomto myšlenkovém experimentu byl stroji dán velice prostý povel. Tento povel však skončil tím, že naše planeta byla přeměněna na hromadu sponek. Kde se stala chyba? Co zapříčinilo jednání tohoto stroje? Bostromův stroj a další možné SI musíme podle singularitánů chápat jako bytosti, které mají dvě zásadní vlastnosti. Zaprvé mají schopnost neuvěřitelně tvarovat realitu okolo sebe, a to jim umožňuje řešit problémy způsoby, které by člověka nenapadly. To vyplývá z vysokého intelektu SI, který jim teoreticky umožňuje nečekané objevy. Druhou jejich vlastností je to, co singularitáni nazývají „doslovnost“. Tím míní fakt, že jedině, co SI rozpoznává, jsou přesně specifikované hodnoty a pravidla. SI může naše výroky interpretovat pouze na základě těchto pravidel. To vyvolává potíže, protože je obtížné přesně definovat a specifikovat mnoho lidských hodnot. Kvůli této skutečnosti nemůžeme zaručit to, že SI bude naše pokyny interpretovat stejně jako člověk. Protože SI dost možná nebude vnímat jemné nuance našich výroků.<sup>7</sup> Při výkladu tohoto a dalších podobných příkladů se situace prezentuje jako problém kontroly. Proto musíme přemýšlet o tom, jak bychom zajistili kontrolu nad jednáním SI a tím se vyhnuli zkáze.

6 Bostrom (2014, s. 47). Podobné definice také lze najít v Bostrom (2003, 2006).

7 Muehlhauser & Helm (2012, s. 105–106). Doslovnost v originále jako Literalness, kterou definují následovně: „The [SI] recognizes only precise specifications of rules and values, acting in ways that violate what feel like ‚common sense‘ to humans, and in ways that fail to respect the subtlety of human values.“

Z tohoto důvodu bijí singularitáni na poplach. Pro ně tento problém vyplývá z povahy, kterou SI bude mít. Bostrom a další zastávají takzvanou ortogonální tezi, která říká: „Intelligence a konečné cíle jsou navzájem ortogonální: Víceméně jakákoliv inteligenční úroveň může v principu být spojená s víceméně jakýmikoliv konečnými cíli.“<sup>8</sup> Pro nás kritickým bude jejich tvrzení, že SI musíme chápat jako inteligenci, která bude mít jeden finální cíl. Tento fakt vyplývá z toho, že SI bude UI. Její cíl bude vycházet z jejího naprogramování. Bostrom a Omohundro v tomto bodě vychází z praktik vývoje UI. Když se vytváří program UI, je určeno, jaký problém bude řešit. Bostrom předpokládá, že SI bude výsledkem takového projektu UI. To, co však činí SI problematickou, není jenom tento samotný cíl, ale také hodnoty, které z něho vyplývají. Z tohoto faktu můžeme alespoň částečně předpovídat chování SI. Bostrom a další zastávají stanovisko, které nazývají tezí instrumentální konvergence, a ta říká:

„Lze identifikovat několik instrumentálních hodnot, které jsou konvergentní v tom smyslu, že jejich uskutečnění by pro celou řadu konečných cílů a v celé řadě situací zvýšilo pravděpodobnost, že agent dosáhne svého cíle. Z toho plyne, že o jejich uskutečnění pravděpodobně bude usilovat široké spektrum inteligentních agentů nacházejících se v různých situacích.“<sup>9</sup>

Bostrom rozlišuje několik takových hodnot, jako je sebezáchova, integrita obsahu cíle, zdokonalování technologií, vylepšování kognice, získávání zdrojů. Nebezpečí SI neplyne pouze z faktu, že by si mohla špatně vyložit svůj konečný cíl, ale také z těchto konvergentních hodnot. SI by mohla mít skutečně neškodný cíl, ale tyto hodnoty by ji dostaly do konfliktu s námi.<sup>10</sup>

8 Bostrom (2014, s. 169). Viz i Bostrom (2012). Stejnou tezi zastává také Omohundro (2008, 2012, 2016).

9 Bostrom (2012 a 2014, s. 172). Stejnou tezi zastává Omohundro (2008, 2012, 2016).

10 Nejkritičtější z těchto hodnot, kterou musíme brát na vědomí, je integrita obsahu cíle. Podle Bostroma a Omohundry bude pro UI nejdůležitější splnění jejího cíle. Pokud by UI ztratila svůj cíl, tak by ho nemohla splnit, proto se bude snažit tento cíl zachovat. Z tohoto důvodu nemůžeme argumentovat, že SI by si mohla přeprogramovat svůj konečný cíl.

V tomto textu se budu zabývat problémem kontroly, přesněji se chci soustředit na domněnky, na kterých je vystavěn. Myslím, že pro zjednodušení bude nejlepší, pokud budeme problém kontroly (dále už jen jako PK) chápat jako formální argument:

- (1) UI bude existovat.
- (2) Pokud bude UI, tak vznikne SI (argument singularity)
- (3) Protože SI je UI, bude mít jeden konečný cíl.
- (4) Intelekt na jakékoliv úrovni, může mít jakýkoliv konečný cíl.
- (5) SI může svůj konečný cíl interpretovat jinak, než jsme jej zamýšleli my.
- (6) Z konečného cíle vyplývá několik konvergentních hodnot.
- (7) Tyto hodnoty nemusí být v náš prospěch.
- (8) Pokud SI špatně interpretuje svůj konečný cíl nebo získá pro nás škodlivé hodnoty, ztratíme nad ní kontrolu.

Závěr: Vznikem SI, nutně čelíme PK.

Tento článek se bude skládat ze dvou hlavních sekcí. V první se budu zabývat otázkou vzniku SI, specificky argumentem singularity Davida Chalmerse. PK může nastat pouze, pokud bude existovat SI, tudíž je důležité zhodnotit proces jejího vzniku. Budu zde posuzovat formalizaci argumentu singularity. Tento argument staví na tezích, ve kterých nalézám několik problémů. Také ukazují, že způsoby vzniku SI mohou vést k různým inteligencím. To je podstatné pro konstituci PK, která je závislá na charakteru vzniklé inteligence. V druhé sekci zpochybním smysluplnost PK, protože singularitáni zakládají tento problém na pochybných domněnkách, které zdědili od svých předchůdců. Tento fakt je vede k formulování PK na základě konceptu SI, který je nekonzistentní s PK. Mým závěrem je, že pro singularitány existují pouze dvě cesty, kterými se mohou vydat, ale žádná z nich nemůže obsahovat PK v jeho současné formě.

## 2. Argument Singularity

Prvním, o kom můžeme říct, že se teoreticky zamýšlel nad konceptem technologické singularity<sup>11</sup>, je Irving J. Good, který už v roce 1965 očekával, že během 20. století bude postaven ultra inteligentní stroj. Podle Gooda, by jeho vznik úplně změnil povahu naší společnosti z velmi prostého důvodu, byl by to poslední lidský vynález. Faktem této skutečnosti je podle něho to, že vznik tohoto stroje povede k tomu, čemu říká inteligentní exploze.<sup>12</sup> Vyplyvá to z jeho definice ultra inteligentního stroje, který definuje jako stroj schopný překonat člověka v jakékoliv intelektuální aktivitě. To vede k prvnímu argumentu pro vznik singularity. Protože Good stanovuje design podobných strojů jako jednu z intelektuálních aktivit daných člověku, byl by ultra inteligentní stroj schopen stvořit ještě lepší stroj. Stačí sestrojít jenom o trochu chytřejší stroj, než jsme my, aby se podle Gooda odstartovala inteligentní exploze. Po této události by ultra inteligentní stroje stavěly ještě chytřejší stroje a člověk by byl zanechán daleko za nimi. Podstatou je to, že jakmile se nám lidem podaří vytvořit umělý intelekt, tak přestane být přirozeným limitem pokroku. Po takovém objevu se intelekt stane už pouze inženýrskou výzvou, kde už jde jen o to vytvořit intelekt, který je lepší než ten předchozí. Toto tvrzení je obecně přijímáno mezi singularitány. Good samozřejmě nebyl jediným kdo formuloval teorii singularity<sup>13</sup>, ale v tomto klíčovém bodě se jednotlivé teorie neliší.

---

11 Good nebyl prvním, koho napadlo, že by technologický pokrok mohl mířit k singularitě. Byl však prvním, kdo poskytl teorii takové singularity. První zmínka o technologické singularitě pochází údajně od Johna von Neumanna: „One conversation centered on the ever accelerating progress of technology and changes in the mode of human life, which gives the appearance of approaching some essential singularity in the history of the race beyond which human affairs, as we know them, could not continue.“ Viz Ulam (1958).

12 Good (1965, s. 78).

13 Viz Vinge (1993) a Kurzweil (2005). Vinge si myslí, že po singularitě nemůžeme nic předpovídat, Kurzweil naopak tvrdí, že vlastně můžeme dále předpovídat vývoj po singularitě, protože se jedná o přirozený výsledek pokroku. Naše modely pokroku budou fungovat nadále i po singularitě, jediné, co se změní, bude rychlost pokroku. Spor mezi jednotlivými teoriemi tkví v tom, kolik toho můžeme předpokládat o singularitě a vývoji po ní. Takže na jedné straně máme Vingea jako úplného skeptika a na straně druhé Kurzweila jako úplného optimistu. Good vytváří střední cestu mezi nimi, protože zastává názor, že inteligentní exploze bude dále pokračovat. Pokud bych měl sám sebe vložit někam na toto měřítko, byl bych mezi Goodem a Vingem. Všechny naše modely by se asi nestaly nepoužitelnými, ale myslím si, že obtížnost předvídatelnosti tkví v tom hádat, které modely vydrží. Ani Goodova inteligentní exploze není úplnou jistotou. Velice zde záleží na tom, jak by nové intelektuální vynálezy.



Chalmers, který formuluje argument pro singularitu, staví na konceptu inteligenční exploze.<sup>14</sup> Na začátek je zapotřebí vysvětlit terminologii, kterou Chalmers používá.<sup>15</sup> Ten používá zkratku UI pro označení jakékoliv umělé inteligence na stejné úrovni jako lidská inteligence. UI+ označuje první ultra-inteligentní stroj, který převyšuje nejvyšší lidské hranice. UI++ pak označuje SI, která vytváří náš diskutovaný problém. Takže, pokud bude existovat UI, vyplývá z toho, že bude existovat UI+ a následně UI++. Proto argument singularity vypadá takto:

- (1) UI bude existovat
  - (2) Pokud bude UI, tak bude UI+
  - (3) Pokud bude UI+, tak bude UI++
- Závěr: Bude existovat UI++

I když je snadné toto formulovat, samo o sobě je to nic neříkající. Jsou zde otázky ohledně vzniku této UI a procesu, který by z ní měl UI+ a posléze UI++ učinit. Aby tyto otázky zodpověděl, nabízí Chalmers teze, které podpírají jednotlivé premisy tohoto argumentu. Těmito tezemi a problémy s nimi spojenými se budu zabývat v následujících podkapitolách. Na začátku se musíme ptát, jak by UI mohla vzniknout. Chalmers nabízí dvě cesty ke vzniku UI.

## 2.1. Emulační teze

První cestu k jejímu vzniku Chalmers opírá o předpoklad takzvané ECM (*emulace celého mozku*). Jedná se o teoretickou proceduru, která by nám umožnila emulovat živý mozek. Koncept je takový, že na základě detailních snímků mozku bychom mohli postavit model mozku. Díky tomuto modelu bychom následně mohli simulovat činnost mozku na počítači. Taková simulace by teoreticky měla mít ty samé vlastnosti jako živý mozek. Výhodou této metody je, že nepotřebujeme vědět, jak mozek produkuje inteligenci. Problém spočívá ve vyspělosti daných

<sup>14</sup> Chalmers (2016a).

<sup>15</sup> Chalmers v originálním znění používá anglické zkratky jako AI (Artificial Intelligence) a WBE (Whole Brain Emulation). Vzhledem k českému znění mého článku jsem jeho terminologii sjednotil s mojí.

technologií. Potřebujeme snímek s dostatečně vysokým rozlišením, software, který zpracuje snímky a sestaví 3D model a nakonec hardware, který bude mít dost výpočetní síly pro simulaci mozku. Tyto technologie jsou dosažitelné, ale ne v blízké budoucnosti.<sup>16</sup> Funkčnost této procedury závisí na myšlence, že mozek je ve své podstatě kauzálním mechanismem, a tak ho můžeme chápat jako stroj. Pokud je mozek stroj, tak ho můžeme zkopírovat a tato kopie by měla mít ty samé vlastnosti jako originál. Tuto myšlenku pak můžeme shrnout do Chalmersova argumentu z emulace:

- (1) Lidský mozek je stroj
  - (2) Budeme mít možnost emulace tohoto stroje
  - (3) Pokud emulujeme tento stroj, budeme mít UI
- Závěr: Bude existovat UI

Navrhoval bych toto tvrzení interpretovat jako prostý apel na funkcionalistické představy o fungování lidského mozku. To, co premisa (1) míní, je předpoklad, že mozek je ve své podstatě systém, jehož procesy se řídí určitými zákonitostmi. Je to systém obsahující části, které mezi sebou interagují na základě těchto pravidel. Díky tomuto faktu tyto části hrají roli v plnění celkové funkce tohoto systému. Klíčovým závěrem funkcionalismu je, že podstata, která umožňuje mozku plnit jeho funkci, nespočívá v materiálním charakteru jeho částí, ale právě v jejich daných rolích. Premisa (2) stojí na tezi, že lze emulovat procesy ve velkém detailu, a tím emulovat jakýkoliv stroj simulací jeho procesů. Takže pokud by se nám podařilo vytvořit umělou náhradu některé části mozku, která by byla schopná zastoupit stejnou roli jako přirozená část, tak by mozek měl dále plnit svou funkci úplně stejně. Premisa (3) vyplývá z tvrzení, že pokud se nám podaří zreprodukovat vzorce lidského mozku, tak následně můžeme vytvořit UI na jejich základě.

Díky této interpretaci by šlo toto tvrzení chápat i v širším smyslu. Možná není zapotřebí emulovat mozek tak do hloubky, jak se ECM snaží. Je zde otázka, jestli potřebujeme kopírovat všechny mikro procesy mozku, nebo jestli bychom si postačili pouze s makroelementy jeho

16 Bostrom & Sandberg (2008), Bostrom (2014, s. 58–68).

architektury. V rámci této cesty k UI můžeme mít celou škálu různých přístupů, které napodobují některé části mozku, ale radikálně se liší v jiných ohledech. Možná UI vznikne díky konekcionistické architektuře, protože inteligenci umožňuje paralelnost mozku. To, co musíme emulovat, je architektura mozku, která je složená z velkého množství jednoduchých jednotek. Koncept je takový, že to, co umožňuje kognici, je interakce mezi těmito body. Taková architektura je pak schopná provádět několik výpočtů najednou, díky jejich distribuci napříč celou neurální sítí. Stejně je to s reprezentací jednotlivých funkcí, která je distribuována mezi jednotlivými body neurální sítě. Jak moc jsou tyto reprezentace distribuovány, se však může lišit. Můžeme například mít systémy, kde jeden bod reprezentuje jeden koncept, nebo systémy, kde jeden koncept reprezentuje skupina bodů.<sup>17</sup> Také je tu otázka, jestli bychom se měli soustřeďovat pouze na mozek, nebo jestli není také zapotřebí emulovat celé tělo. To je představa vtělené a situované UI.<sup>18</sup> Naše myšlení je většinou zakotveno ve světě, málokdy se zabýváme čistě abstraktními objekty. Je dost možné, že inteligenci umožňuje interakce s prostředím, které na nás vyvíjí tlak. Takovou zkušenost však můžeme mít pouze, pokud je inteligence nějakým způsobem vtělená. Z tohoto důvodu mnoho autorů zastává tezi, že inteligenci umožňuje právě tělesnost.<sup>19</sup> Tuto tezi podporuje i sám Chalmers.<sup>20</sup> Dané tělo nemusí být nutně podobné tomu našemu, ale musí být vybaveno tak, aby přijímalo vjemy z vnějšího prostředí a bylo schopné se vyrovnat s komplexitou vnějšího světa. Takže je několik možností toho, jak můžeme emulovat a napodobovat lidskou inteligenci. Vzhledem k těmto dalším možnostem emulace lidské inteligence je otázkou, proč Chalmers svazuje tento argument tak úzce k funkčnosti ECM.<sup>21</sup> Nejpravděpodobnějším důvodem se zdá být to, že volí nejjistější cestu k úspěchu. Přesto je ale nutné zvážit i alternativy.

---

17 Sun (2014).

18 Beer (2014).

19 Viz Clark (1997), Pfeifer & Bongard (2006), Varela, Thompson & Rosch (2016).

20 Clark & Chalmers (1998). Chalmers v rámci textu i sám upozorňuje, že SI by mohla mít rozšířenou mysl. Vzhledem k argumentům, které rozvíjím později v textu, si nejsem jist, jak může Chalmers chápat rozšířenou mysl jenom jako jednu z možných SI myslí. Já zastávám pozici, že SI bude muset být rozšířenou myslí, aby byla skutečně obecnou inteligencí. Dle mého názoru Chalmers tak koná, aby nechal otevřený prostor pro svoje argumenty.

21 Bostrom & Sandberg (2008), Bostrom (2014, s. 58–68).

Emulačnímu argumentu lze samozřejmě oponovat. Jediné, co lze skutečně zpochybnit je premisa (1). Pokud bychom přijali, že mozek je stroj, je nejasné, jak bychom mohli oponovat možnosti jeho emulace. Chalmers sám zmiňuje filozofy, jako jsou Hubert Dreyfus a Roger Penrose, kteří podle něj tvrdili, že činnost mozku nelze simulovat na žádném počítači. Nebo John Searle a Ned Block, kteří argumentovali, že i kdybychom simulovali lidský mozek, něco fundamentálního by mu chybělo, protože by pouze simuloval naše chování. Když však Chalmers nahlíží jejich námitky, mám dojem, že poněkud zjednodušuje jejich postoje. Vezměme si Blocka, jeho myšlenka je taková, že pokud stroj jedná inteligentně, tak tato inteligence nemusí nutně být jeho. Může být inteligentní, protože mu tento intelekt propůjčil jeho programátor. V takovém případě jeho inteligence není jeho vlastní. Takže se zde nejedná ani tak o to, že by tento počítač simuloval naše chování, on jedná jenom díky našemu konání, nemá žádnou vlastní vůli.<sup>22</sup> Také si nemyslím, že by Dreyfus argumentoval proti tomu, že mozek může být strojem. Jeho kritika spíše mířila tím směrem, jakým druhem kauzálního mechanismu skutečně jsme. Více si však nechám pro druhou sekci, kde se budu blíže zabývat zejména argumenty Searlovými a Dreyfusovými.

## 2.2. Evoluční teze

Cesta první se opírala o fakt, že my lidé máme inteligenci. Snažila se o to postavit UI na jejím základě. Druhá cesta argumentuje z existence intelektu obecně a vychází z faktu, že vznikl určitým způsobem. Je to míněno tak, že náš lidský intelekt vznikl jako důsledek přírodního výběru. Pokud takto mohla vzniknout jedna inteligence, proč by nemohla vzniknout další? To pak Chalmers nazývá argumentem z evoluce:

- (1) Evoluce stvořila inteligenci
  - (2) Pokud evoluce stvořila inteligenci, tak můžeme stvořit UI
- Závěr: Bude existovat UI

---

<sup>22</sup> Block (1981).

*Prima facie* se toto stanovisko zdá být velice prosté, je zde však pár skrytých interpretací. Bostrom a Shulman<sup>23</sup> ve své reakci ukazují, že můžeme tento argument číst dvěma různými způsoby. První interpretace říká to, že pokud slepý a nevědomý proces dokázal stvořit inteligenci, tak vědomý designer toho také může dosáhnout. Tradičně by následoval i výrok, že designer může tento úkol splnit rychleji a lépe, protože se o to vědomě snaží. Pokud by toto byla pravda, tak by UI mohla mít i formu, která se radikálně odlišuje od dosud známých forem inteligence. UI by mohla vzniknout jiným než evolučním procesem a nemusela by ani sdílet s tím spojené vlastnosti. Z této představy vyplývá jeden klasický argument pro vznik UI. Tento argument stojí na analogii mezi stvořením intelektu a vynálezem letadel. Nám lidem se podařilo létat, ovšem neudělali jsme to tím stejným způsobem jako ptáci, místo toho jsme postavili letadla. Možná stejným stylem sestrojíme UI, která sice bude inteligentní, ale jiným způsobem.<sup>24</sup> To přináší problém, který mám s touto interpretací evoluční teze. Protože tato myšlenka svádí k názoru, že prostor možných inteligencí je větší, než si myslíme. Musíme se ptát, jak moc je let ptáků odlišný od letu letadel. Jak letadla, tak i ptáci se musí řídit zákony aerodynamiky. To samé musí platit pro inteligenci. Přestože zatím nevíme, co umožňuje inteligenci, musí i v tomto prostoru existovat zákonitosti. Je dost možné, že pokud sestrojíme UI, tak bude výtvořem vědomého designera. Tento fakt nás ale nemůže svádět k tomu si myslet, že jakoukoliv metodou můžeme stvořit inteligenci. Pokud se necháme tímto argumentem příliš unést, mohli bychom postulovat inteligence, které nejsou možné. V příští sekci ukážu, jak singularitáni do této pasti padají.

Druhá interpretace klade větší důraz na to, že náš intelekt vznikl díky evolučnímu procesu. Pokud takový proces dokázal stvořit intelekt, proč ho nenapodobit? Toto můžeme považovat za argument ve prospěch evolučních a genetických algoritmů, kterými vytvoříme proces podobný evoluci. Genetické algoritmy fungují tak, že generují omezené množství aktérů, kteří jsou vybaveni určitými schopnostmi. Této skupině je předložen cíl, a mají sami k řešení dojít. Jednotliví aktéři se chovají zcela náhodně, omezení pouze danými parametry. Ze skupiny je vybrán ten

---

23 Bostrom & Shulman (2016).

24 Moravec (1976).

jedinec, který se splnění cíle přiblížil nejvíce. Na jeho základě je pak vygenerována další generace aktérů. Tento proces probíhá tak dlouho, dokud nevznikne aktér, který je schopen plnit daný cíl. Tato metoda je velice zdoluhavá, takže v praxi designer upravuje algoritmus tak, aby nezašel do slepé uličky. Tato interpretace ovšem stojí před problémem, jak náročné je simulovat evoluci. Bostrom a Shulman argumentují, že pokud bychom museli simulovat evoluční proces do každého detailu, tak bychom museli simulovat celé přírodní prostředí až do mikro-fyzikálních procesů. To by si vyžádalo výpočetní sílu, které bychom se nedočkali ani do konce století. Proto navrhuji, že bychom měli simulovat pouze abstraktní prostředí.<sup>25</sup> Argumentoval bych, že tento fakt nás znovu musí vést k závěru, že výsledkem tohoto procesu musí být odlišný intelekt. Protože pokud takový intelekt vznikne v odlišném prostředí, mělo by se to promítnout na jeho struktuře.

Nehledě na interpretaci, tento argument stojí před problémem komplexnosti, který se promítá na několika úrovních. Obě interpretace spoléhají na to, že stvořit inteligenci není těžké. Otázkou je, jak skutečně složité je stvořit intelekt. Přírodě trvalo miliardy let, než vznikl intelekt na naší úrovni. Jak moc by mohl fakt, že nyní se úkolem zabývá skupina inteligentních designerů, zkrátit tuto dobu? Stejnou námitku lze aplikovat na snahu tento proces kopírovat. I když máme designera, který umělou evoluci navede na správnou cestu, mohli bychom čekat dost dlouhou dobu, než bychom se UI dočkali. To není příliš uspokojivý závěr pro inženýry UI a obzvláště ne pro singularitány. Proto oba tyto argumenty spoléhají na to, že stvořit inteligenci je ne-těžké. Protože v takovém případě to, co způsobilo dlouhou dobu vzniku inteligence, je slepost evolučního procesu.<sup>26</sup> Překážka komplexnosti je ale ještě silnější, kvůli faktu, že nevíme, jak se evoluci podařilo stvořit intelekt na naší úrovni. Co nás tak opravňuje tvrdit, že my budeme schopni stejného činu? Nebo v případě druhé interpretace je pochybné, jak bychom mohli genetický algoritmus správně navádět k jeho cíli. Z těchto důvodů tento argument nepovažuji za dostatečný důkaz toho, že by UI mohla vzniknout. Nelze pouze z existence věci usuzovat, že může vzniknout další. Musíme vědět,

---

25 Bostrom & Shulman (2016).

26 Tamtéž.

jak daná věc funguje, abychom ji mohli replikovat. Argument z emulace je lepší v tom ohledu, že nabízí návod, jak bychom UI mohli vytvořit, což evoluční cesta neobsahuje. První verze pouze argumentuje, že toho budeme schopni. Neposkytuje však žádné vysvětlení, jak toho dosáhneme. To má za efekt, že si díky tomuto argumentu můžeme představit široké spektrum intelektů, které bychom mohli stvořit různými způsoby. To nás však může svádět k tomu přehlížet, jakou skutečnou strukturu inteligence může mít. Druhá je rezignací na snahu zjistit, jak inteligence funguje. Tak pouze co nejlépe zkopírujeme proces jejího vzniku v naději, že tak dosáhneme našeho cíle. Proces vzniku intelektu je jen jedna část celého příběhu, druhá je skryta v jeho struktuře, kterou také musíme vzít v potaz.

### 2.3. Teze rozšíření

Nyní se tedy musíme ptát, jak můžeme od UI dojít k UI+. Jak bychom mohli stvořit vyšší než lidský intelekt? Pokud se nám podařilo vytvořit UI, museli jsme tak učinit určitou metodou. Tudíž abychom vylepšili naši UI, stačilo by nám, abychom rozšířili tu danou metodu. Tomu Chalmers říká teze rozšíření, která spoléhá na to, že UI vznikne rozšiřitelnou metodou.

- (3) Pokud bude UI, tak vznikne rozšiřitelnou metodou
  - (4) Pokud UI vznikne touto metodou, budeme schopni ji rozšířit
  - (5) Rozšířením metody vznikne UI+
- Závěr: Pokud bude UI, tak bude UI+

Proti tomuto argumentu můžeme namířit dvě námitky. (a) Nejjednodušší by bylo vyvrátit, že UI vznikne rozšiřitelnou metodou. Tuto námitku však Chalmers předpokládá, a proto říká, že i kdyby UI vznikla nerozšiřitelnou metodou, tak její vznik by nám poskytl vodítka k tomu, jak vytvořit UI rozšiřitelnou metodou. Toto se nezdá jako příliš jisté. Představme si, že se nám podaří stvořit UI skrze genetický algoritmus. Máme sice UI, ale nemáme tušení, proč je inteligentní. Kdybychom měli k dispozici zdrojový kód takové UI, jak bychom zjistili, které části jsou důležité? Vezměme

si například náš genetický kód. Přestože jsme schopni číst DNA, není snadné identifikovat, co je účelem jednotlivých genů. Tento argument může fungovat pouze, pokud UI vznikne cíleně, protože pak budeme vědět na čem dále pracovat. Ale například Bostrom zastává názor, že první UI nejpravděpodobněji vznikne omylem.<sup>27</sup> V takové situaci jsme na tom stejně jako v předchozím případě, protože nevíme, jak daná UI vznikla. Chalmers jako příklad rozšiřitelné metody zmiňuje přímé programování, které může zahrnovat činnosti jako ladění kódu. Co když však UI omylem vznikla právě kvůli chybě? Pak bychom během našeho procesu „rozšiřování“, vlastně zničili danou UI. (b) Předpokládejme, že se nám skutečně podaří cíleně stvořit UI rozšiřitelnou metodou. Můžeme danou metodu rozšiřovat donekonečna? McDermott poukazuje na tento problém. Co se nám s ohlédnutím zpět může zdát jako hladká křivka technologického pokroku, bylo ve skutečnosti trnitou cestou, kterou se podařilo projít pouze s pomocí mnoha inovací. Naše představa může být taková, že postupujeme průběžným vylepšováním jedné metody, ale ve skutečnosti to nemusí být jedna metoda. Skutečný pokrok ve velké části závisí na objevování nových metod pro řešení problémů.<sup>28</sup> Chalmersova odpověď je, že pokud narazíme na strop jedné metody, můžeme se vždy obrátit k další. Dle jeho názoru s růstem inteligence daných systémů, budou tyto systémy objevovat další metody svého rozšíření.<sup>29</sup> To může být pravda, ale zůstává zde otázka, jak dlouho takové rozšiřování potrvá. Nemyslím si, že můžeme kategoricky odmítnout vznik UI+, ale myslím si, že bychom měli být skeptičtější ohledně data jejího příchodu.<sup>30</sup>

---

27 Bostrom (2014, s. 152–155).

28 McDermott (2016).

29 Chalmers (2016b).

30 Výše jsem citoval Simon & Newell (1958) a Minsky (1968), kteří jsou dobrým příkladem neoprávněného optimismu v oboru UI. Tento optimismus často zaslepuje výzkumníky UI vůči problémům spojeným s jejich tezemi, jak demonstroval Dreyfus (1992). Dreyfus (2012) poukazuje na celkový historický trend v oboru UI, kde se často začne prvním úspěšným krokem, který vzbudí tento optimismus, ale ke konci pokrok postupně pomine. Ten samý optimismus můžeme najít mezi singularitány. Stačí, když se podíváme na jejich předpovědi vzniku UI a příchodu singularity. Good (1965) předpovídal singularity do roku 2000. Vinge (1992) nám pro příchod singularity nabízel rozmezí let mezi 2005 a 2030. Kurzweil (2005) předpokládá rok 2030 pro vznik první UI. Bostrom (2006) předpovídal UI v první čtvrtině tohoto století. Chalmers (2016a) je ve svém odhadu konzervativnější a s 50% jistotou odhaduje, že v rámci tohoto století vznikne první UI. Pro srovnání průzkum Müller & Bostrom (2016) ukazuje, že většina (přesněji 41 %) z dotazovaných výzkumníků UI předpovídá, že vznik první UI potrvá déle



## 2.4. Teze navýšení

Poslední premisu Chalmers zakládá na tezi navýšení, kterou opírá o matematickou indukci. Předpokládejme, že existuje  $UI_+$ , a řekněme si, že  $UI_1$  je první  $UI_+$  a  $UI_0$  je jejím stvořitelem, ať už je to člověk či jiná  $UI$  (která může být také na menší úrovni než člověk), a  $\delta$  je rozdíl mezi nimi. Takže systém, jehož inteligence se liší od předchozí o kladné  $\delta$ , má vyšší inteligenci. Dále si řekněme, že v případech, kde  $n > 1$  a máme  $UI_{n+1}$ , jejímž stvořitelem je  $UI_n$ , bude  $UI_{n+1}$  také chytřejší než její stvořitel, stejně tak, jak tomu bylo v případě  $UI_1$ . Z toho pak vyplývá teze navýšení:

- (1) Pokud je  $UI_+$ , tak je  $UI_1$
  - (2) Pro všechny  $n > 0$ , pokud je  $UI_n$ , pak bude  $UI_{n+1}$
  - (3) Pokud pro všechna  $n$  je  $UI_n$ , tak bude  $UI_{++}$
- Závěr: Pokud bude  $UI_+$ , tak bude  $UI_{++}$

Tomuto argumentu také můžeme oponovat různými způsoby. (a) Já bych protestoval proti tomu, jak dochází ke svému závěru. Když toto tvrzení posoudíme samostatně, musíme ho považovat za zvláštní, protože argumentuje *a priori*. Říká, že pokud máme jeden intelekt na určité úrovni, můžeme předpokládat, že by mohl existovat další intelekt na úrovni vyšší. Z toho můžeme odvodit, že existuje škála intelektů. To považuji za bizarní cestu, jak k tomuto závěru dospět. Samotný závěr není problematický, ale musíme k němu dojít *a posteriori*. Jakmile se setkáme s bohatým spektrem intelektů, můžeme začít přemýšlet o dalších možných intelektech. Možná byste mohli namítat, že zde je klíčové to, že se bavíme o  $UI$ . Takže pokud máme jeden uměle vytvořený intelekt, můžeme předpokládat, že by mohl vzniknout další. To však neřeší problém, na který jsem upozornil. Osamocená  $UI$  by se také musela nejdříve setkat s dalšími intelektly, aby mohla nad touto otázkou přemýšlet. Také byste mohli namítat, že Chalmers prezentuje tento argument v kontextu argumentů předchozích. Předchozí argumenty však byly schopné samostatného posouzení. (b) Tato teze je založena na představě, že nárůst

---

než 50 let. V průzkumu se jich také ptali, jak by s 10% / 50% / 90% jistotou odhadli rok vzniku  $UI$ . Median předpokládaného roku s 50% jistotou všech respondentů byl 2040. Průměrný předpokládaný rok s 50% jistotou byl 2081.

inteligence se rovná stejnému nárůstu v kvalitě designu další inteligence. Přičemž závisí na tom, že skok mezi jednotlivými inteligencemi bude vždy stejný. Tuto představu bychom mohli různými způsoby zpochybnit. Jeden možný protiargument by mohl být takový, že bychom mohli koncipovat inteligenční strop neboli limit, kterého inteligence může dosáhnout. Možná příliš vysoká inteligence je natolik náročná, že jí nemůžeme dosáhnout, nebo není dlouhodobě udržitelná. Pokud takový strop je, tak by na něj UI++ narazila. Jinou námitkou je říct, že nárůst inteligence by měl snižující výnosy, první nárůst o 10 % by pak mohl následovat pouze nárůstem o 5 % s postupnou ztrátou. Nejzávažnější námitka by mohla napadnout vztah mezi mírou inteligence a kvalitou designu. Možná vyšší inteligence není nutně lepším designérem.<sup>31</sup>

## 2.5. Celkové zhodnocení

Když se podíváme na argument jako celek, tak musíme říct, že nejvíce problematická je první premisa a teze s ní spojené. Premisa druhá a třetí sice mají své problémy, ty však nejsou nepřekonatelné. Jejich skutečné problémy se projevují v kontextu první premisy. Jak je asi očividné, abychom stáli před PK, musí nejprve vzniknout UI. Takže nejsilnější cesta, jak tento argument porazit, by bylo vyvrátit, že vznikne UI. Takovou argumentaci bych však chtěl v tomto textu ponechat stranou. Jedním z důvodů je, že už dost textů bylo věnováno takové argumentaci. Z druhé, i kdyby UI byla potvrzena, tak PK stojí před dalšími překážkami, které by ho mohly vyvrátit. Protože se chci dále soustředit na smysluplnost tohoto problému, chci se na tyto překážky zaměřit. Jeden problém s argumentem singularity stojí na tom, jak by charakter dané UI mohl ovlivnit plauzibilitu následujících argumentů. Na jedné straně emulační teze argumentuje pro přístupy inspirované a založené na fungování lidského mozku. V úzkém smyslu argumentuje pro funkčnost ECM, v širokém smyslu ho vidím jako argument ve prospěch konekcionismu a vtělené inteligence. Zde jako výsledek můžeme očekávat inteligence, které budou podobné té naší. Na straně druhé, evoluční teze představuje dva opačné přístupy. Podle jednoho je inteligentní designer schopen

---

<sup>31</sup> Chalmers (2016a).

stejného postupu jako slepá evoluce, možná i lepšího. Podle druhého by nám stačilo zkopírovat evoluční proces. V obou případech bych argumentoval, že by se výsledný intelekt lišil od toho našeho.

To vše nám poskytuje množství potencionálních intelektů. Otázkou je, jestli jsou všechny rozšiřitelné. Už jsem mluvil o problému rozšiřitelnosti inteligence vzniklé skrze genetické algoritmy, podobné problémy se vztahují i na další metody. Například ECM je metodou, která je obtížně rozšiřitelná. Nejvíce teoreticky možné je, že bychom mohli zvýšit rychlost přemýšlení dané simulace, skrze vylepšení jejího hardwaru, to by však mělo za výsledek pouze slabou formu SI. Problém závisí na tom, jak moc je architektura samotného mozku rozšiřitelná. Jak bylo zmíněno, tato metoda má výhodu v tom, že nemusíme podrobně vědět, jak mozek funguje. To ale ovlivňuje, nakolik budeme schopni jeho rozšíření. Bostrom a Sandberg<sup>32</sup> argumentují, že jakmile budeme mít simulaci mozku, zjistíme toho více o jeho architektuře. Protože taková simulace bude více otevřená k pozorování a experimentování. To se může ukázat jako pravdivé, není ale jisté, jestli by to vedlo k takovému skoku ve výzkumu, který si představují.<sup>33</sup> Stejně tak je nejisté, jestli je architektura mozku rozšiřitelná. Můžeme pouze spekulovat, jaký následek by například mělo více neuronů na celkovou architekturu. Mohlo by to vést k rozšíření, ale je i možné, že by celková architektura nebyla schopná takový větší výkon vydržet.<sup>34</sup> Některé metody se zdají být jako velice otevřené rozšíření, jiné se však jeví jako daleko uzavřenější.

Pro konstituci PK jsou důležité teze emulační a evoluční. V příští sekci ukážu, jak singularitáni zastávají pozici, že SI bude radikálně odlišná

32 Bostrom & Sandberg (2008).

33 Toto tvrdím kvůli faktu, že mnoho našich problémů nemusí záviset přímo na nedostatku poznatků. Vezměme si například problém vědomí v rámci filozofie mysli. Přestože toho víme o mozku stále více, stále tady panuje problém explanační propasti, na který upozorňovali Nagel (1974), Jackson (1982 a 1986) a Chalmers (1996). I pokud se tato propast ukáže jako iluzorní, tak demonstruje, že další poznatky nejsou vším. ECM nám může poskytnout přímý pohled do mozku, to ale neznamená, že nebudeme stát před konceptuálními problémy.

34 Colzato, Hommel & Beste (2021) upozorňují na jeden problém, na který bychom mohli narazit. Architektura mozku je postavená na soupeření neuronů. Jednotlivé neurony a i větší neurální sítě soutěží o to, aby mohly zpracovávat a reprezentovat informace. To však znamená, že pokud bychom vylepšili pouze jednu část mozku, tak potencionální zisky by s sebou nesly ztráty v jiných oblastech. Například vylepšení neurální persistence (tj. naše schopnost se soustředit na jeden konkrétní problém) by bylo na úkor naší neurální flexibility (tj. naší schopnosti nacházet nová řešení).

od naší inteligence, a to vyvolává PK. Nejlépe tento fakt demonstruje Bostrom, který argumentuje, že můžeme identifikovat různé způsoby předpověditelnosti SI na základě její architektury. Pro nás je nejdůležitější předpověditelnost na základě zděděných motivací. Konceptem je, že pokud daná inteligence vznikla na základě lidské předlohy, mohla by jako důsledek zdědit některé lidské motivace a hodnoty. To by znamenalo, že na takovou inteligenci by se PK vztahoval v menší míře, než kdyby daná inteligence byla odlišná od té naší. Bostrom samozřejmě upozorňuje, že nevíme, do jaké míry by taková inteligence mohla zdědit naše motivace. ECM by je mohla ztratit během přesunu do počítače, nebo při svém dalším vylepšení.<sup>35</sup> Přesto je tady dobrá šance, že na takovou inteligenci by se PK vztahoval v menší míře, než kdyby daná inteligence byla radikálně odlišná od té naší. Proto se singularitáni musí držet evoluční teze, aby mohli tvrdit, že stojíme před PK. Jedna věc, která je k tomu opravňuje, je fakt, že inteligence podobné té naší se zdají být hůře rozšiřitelné.<sup>36</sup> Z toho vyvozují, že SI bude spíše odlišná od nás. Tímto krokem však na sebe berou problémy, které evoluční teze má a se kterými se nevypořádávají. V příští sekci demonstruji, jak je tento krok vede k tomu formulovat SI a koncept PK s ní spojený.

### 3. Problém kontroly a obecná inteligence

Vraťme se úplně na začátek, k Bostromově stroji na sponky. Vzpomeňme si na dvě vlastnosti, které SI stroj údajně bude mít. Zaprvé bude schopen díky svému neuvěřitelnému intelektu řešit problémy způsoby, které by člověka nenapadly. A druhou jeho vlastností je doslovnost. Je to hlavně tato vlastnost, která činí SI tak problematickou. V úvodu jsem zmiňoval,

---

<sup>35</sup> Bostrom (2012 a 2014, s. 170).

<sup>36</sup> Kromě ECM jsou zde i problémy s dalšími postupy, které jsem zmínil. Vtělená UI čelí faktu, že není zřejmé, jak bychom rozšiřovali tělesnost samotnou. Tělesnost spíše hraje roli podmínky inteligence. Tělo také vytváří různá omezení, pro vylepšování UI. Do robotického těla nemůžeme jen tak dále vkládat další hardware, aniž bychom ovlivnili jeho stabilitu. Neurální síť se z těchto přístupů ukazuje jako nejlépe rozšiřitelné. Ovšem i moderní neurální síť schopné hlubokého učení stojí před několika překážkami. Nejzávažnější pro vylepšování je nedostatek transparentnosti. Často není jasné, proč neurální síť došla ke svému závěru. To velice komplikuje proces rozšiřování učících procesů. Pro přehled dalších problémů viz (Marcus, 2018). I kdybychom tyto potíže překonali, tak vzhledem k argumentům, které rozvíjím později v textu, si nemyslím, že můžeme skutečnou SI stavět pouze na základě neurální sítě.

že doslovnost SI vyplývá z faktu, jak obtížné je definovat mnoho lidských hodnot a konceptů. Například morální filozofie jako celek je toho dobrým příkladem. V jejím rámci stále vznikají nové morální systémy a stejně tak se u každého objeví protipříklad, který daný systém zpochybní.<sup>37</sup> Problémem je, že SI bude od nás přijímat pouze přímo specifikované povely a definované hodnoty. SI bude jednat pouze na jejich základě. Pokud špatně specifikujeme dané povely a definujeme naše hodnoty, tak to pro nás může mít neblahé následky. Proto singularitáni koncentrují tolik úsilí na to, aby našli takový konečný cíl pro SI, který by byl nezávadný.<sup>38</sup> Je zde ale zásadní otázka, jestli se máme o toto snažit. Proč je skutečně tak složité definovat naše hodnoty a co způsobuje doslovnost SI? Pokud bychom byli schopní SI vysvětlit přesně, co máme na mysli, tak by tento problém nemohl nastat. Proto PK je ve své podstatě problém interpretace. Jak zajistit, aby SI interpretovala naše povely tak, jak je míníme? S tímto problémem jsme se mohli setkat mnohokrát v historii UI. Yudkowsky předkládá jeden příklad, kdy MIT údajně vytvořil neurální síť, která měla být schopná rozeznat maskované tanky v lese. Po nějakém cvičení neurální síť testy ukazovaly, že síť skutečně rozeznávala mezi fotkami, kde tanky byly a kde nebyly. Když však poslali síť armádě k dalšímu testování, brzy jim byla vrácena zpět. Důvodem bylo, že v jejich testování síť nebyla v objevování tanků o nic lepší než náhoda. Problém se ukázal být ve cvičebním vzorku, fotky s tanky byly udělány během zamračeného dne, zatímco na fotkách bez tanků bylo slunečno. To, co se jejich síť naučila rozeznávat, bylo to, jestli na fotce bylo zamračeno nebo slunečno, ne jestli tam byly či nebyly tanky.<sup>39</sup>

Takové potíže vyvolává jeden z největších problémů s UI, a tím je problém rámce. Tento problém se nejdříve objevil jako technický problém pro UI založenou na formální logice.<sup>40</sup> Jednalo se o problém toho, jak elegantně zachytit důsledky dané akce v rámci formální logiky. Tento

---

37 Muehlhauser & Helm (2012).

38 Viz Bostrom (2012 a 2014), Yudkowsky (2001, 2004 a 2011), Muehlhauser & Helm (2012).

39 Dreyfus (1992) a Yudkowsky (2008), přestože se jedná o často citovaný příběh, Yudkowsky není schopen dohledat primární zdroj. Je tak dost možné, že se jedná o pouhou historiku. Ale slouží dobře jako typický příklad dezinterpretace v oboru UI, proto ho zde cituji.

40 McCarthy & Hayes (1969).

technický problém ale dal vznik širšímu epistemologickému problému.<sup>41</sup> Dennett ho charakterizuje jako problém toho, jak má činitel svá přesvědčení aktualizovat vzhledem ke svému jednání ve světě. Představte si běžnou situaci, kdy si děláte kávu. Vezmete si hrnek, konvici a kávu. Konvici naplníte vodou a dáte ji ohřát. Vezmete kávu a nasypete ji do hrnku a zalijete ji ohřátou vodou. Nakonec byste mohli hrnek s sebou vzít do obývacího pokoje. Teď bych na vás ovšem měl otázku. Když jste si s sebou vzali hrnek, vzali jste s tím s sebou i kávu? Předpokládám, že byste odpověděli, že ano, je přece v tom hrnku. Kdybychom se zeptali UI programu, tak by nevěděl. Program by čelil problémům daleko dříve. Například, když zalejeme kávu horkou vodou, co se skutečně změní v dané situaci? Počítač může mít mnoho údajů, jako pozici hrnku, teplotu v dané místnosti, daný čas a další. Složitě je pro něj určit, jaký z těchto údajů by měl aktualizovat po daném jednání, a přitom stále mít věrohodný obraz světa.

Problém rámce je kromě v oboru UI často diskutovaným problémem v počítačové teorii mysli.<sup>42</sup> Pokud chápeme mentální stavy jako sérii propozic, tak se problém rámce stává problémem toho, jaké propozice máme změnit v daném kontextu. Tady se jako nejdůležitější ukazuje princip relevance. Daný kontext si žádá, abychom změnili pouze relevantní množinu našich přesvědčení. Takže pokud natru můj dům modrou barvou, měl bych aktualizovat pouze ta přesvědčení, která se týkají mého domu. Je to právě kategorie relevance, která spojuje problém rámce s PK. PK poukazuje na užší situaci, kdy si SI musí aktualizovat svůj soubor přesvědčení na základě našich povelů. Relevance je důležitá, co se týče jejich interpretace. Stroj na sponky trpěl tím problémem, že nenahlížel na všechny pro nás relevantní důsledky svého cíle. Když mu jeho designer zadal jeho úkol, konal tak v nějakém kontextu. Designer předpokládal, že stroj bude jednat v určitých mezích. Například, že bude sponky vyrábět pouze z kovu, nebo že při jejich výrobě nezraní člověka. Designera ani nenapadlo se zamýšlet nad tím, že by jeho výrok šlo vyložit jinak. SI je tak nebezpečná z tohoto důvodu, že může špatně identifikovat kontext daného povelu. Z toho vyplývá, že si jej vyloží jinak

---

<sup>41</sup> Dennet (1984).

<sup>42</sup> Viz Fodor (1983).

než my. V tomto ohledu se ale SI vymyká definici, kterou jsme si stanovili v úvodu. Pokud SI má být obecnou inteligencí, tak by měla být schopná stejně jako my, správně identifikovat, jaká přesvědčení má mít v dané situaci. Otázkou je, proč se tento problém vztahuje i na SI. Co ji vede k tomu interpretovat naše výroky jinak?

### 3.1. Davidsonova teorie interpretace a SI

To nás nutí zamyslet se nad tím, co to znamená interpretovat výroky druhého. Co bychom museli vědět, abychom mohli někoho správně interpretovat? Myslím si, že nejlépe se můžeme na to dívat skrze Davidsonovu teorii interpretace. Ten ji založil na základě Tarského teorému, ale obrátil ho naruby. Zatímco Tarsky předpokládal správný překlad vět z objektového jazyka do metajazyka, aby mohl definovat pravdu, tak Davidson místo toho přijímá pravdu jako primitivní pojem, aby mohl vytvořit teorii překladu nebo interpretace. Takže vytváří teorii pravdy, která nestojí na konceptu překladu. Podle Davidsona pro každou větu  $s$  v objektovém jazyce musíme empiricky najít takovou větu  $p$  nám známého metajazyka, že  $s$  je pravdivá, právě když  $p$ . Sama o sobě by nám však taková koncepce T-vět nic o interpretaci neřekla, je nutné ji formálně a empiricky omezit. Proto se musíme zamyslet nad tím, co má interpret k dispozici. Jediné, na čem můžeme skutečně stavět, je fakt, že mnohé naše výroky také považujeme za pravdivé. Davidson demonstruje to, že naše interpretace závisí na povaze dané komunity mluvčích, kde danou skupinu můžeme identifikovat tím, jaká tvrzení považují za pravdivá a jaká naopak ne.<sup>43</sup> Při interpretaci druhého promítáme to, co bychom za daných okolností považovali za pravdivé do struktury vět mluvčího. To je podstata principu vstřícnosti, který nám napomáhá interpretovat druhé. Tohle se dále promítá v jeho teorii poznání, kterou můžeme chápat pouze v rámci totálního souboru našich přesvědčení. Význam našich výroků je úzce propojen s našimi přesvědčeními. Fakt, že jsme schopni interpretovat druhé a běžně správně používat jazyk, pak obecně potvrzuje pravdivost našich přesvědčení. Pokud by pravdivá nebyla, nemohli bychom si rozumět.<sup>44</sup>

43 Davidson (1984).

44 Davidson (2004a).

Jak máme chápat PK ve spojení s interpretací? Bostromův sponkový stroj, očividně svého designera interpretoval špatně. Pokud však má Davidson pravdu a vskutku interpretujeme druhé osoby takovým způsobem, tak by SI z definice měla být schopná stejného úkonu. Na SI by se ani neměl vztahovat problém rámce. Davidson se v jedné pasáži<sup>45</sup> zabývá představou vševědoucího činitele, který by měl pouze pravdivá přesvědčení. Pokud by takový činitel interpretoval nás jako omylné činitele, stejně by musel dojít k závěru, že omylný činitel se nemůže mýlit ve většině svých přesvědčení. To je situace analogická s SI. Jsou zde dva možné důvody, proč by SI dělala takové chyby. (1) SI je schopná správné interpretace daného úkolu, ale nemá motivaci se jí řídit. To znamená, že nás může velice dobře chápat, ale postrádá jakýkoliv popud k tomu jednat podle našich přání. (2) SI nemůže nahlížet relevantní fakta z toho důvodu, že se od nás radikálně odlišuje. Yudkowsky a Bostrom zastávají obě tato stanoviska jako klíčová pro PK. Říkají, že SI se bude od naší inteligence radikálně lišit, a tím pádem bude mít motivace odlišné od nás. Mým postojem je, že nemohou mít oba tyto názory zároveň.

### 3.2. Ontologická sázka

Pro jejich postoj je klíčový koncept prostoru myslí. Yudkowsky tento prostor explicitně vytváří. Dle jeho názoru, když se bavíme o inteligenci, často propadáme klamu, že víme, co tento pojem obsahuje. Tento klam se ukazuje během diskusí o UI. Když se o ní bavíme, máme tendenci ji antropomorfizovat. Soudíme-li její možný intelekt, konáme tak v kontextu lidského měřítka. Takový postup není problematický při interakci s dalšími lidmi. Tato perspektiva je pro nás přirozená, a proto když se bavíme o SI, představujeme si ji jako stereotypického genia. Lidské měřítko intelektu je však neadekvátní, když mluvíme o UI. Podle Yudkowského UI představuje prostor myslí obecně. Přesněji se dle jeho názoru musíme bavit o *prostoru optimalizačních procesů*. To odpovídá naší definici inteligence. Dle jeho názoru musíme odlišné myslí chápat právě takovým způsobem – jako systémy, které jsou schopné dosáhnout velice specifických cílů v celém prostoru možností. Nejde jen o to, že SI

<sup>45</sup> Tamtéž, s. 144.



by měla větší intelekt než my, ale také o to, že její mysl by mohla mít radikálně odlišné cíle. Dle jeho názoru prostor možných myslí UI je daleko větší než naše lidské měřítko, a právě proto je tak neadekvátní. Tohle pak navazuje na problém motivace UI. Pro něj z toho jasně vyplývá, že tyto odlišné mysli by oplývaly jinými motivacemi než my, a tudíž by jinak interpretovaly naše výroky.<sup>46</sup>

Bostrom v tomto směru následuje Yudkowského. Nám se může zdát, že mezi dvěma lidmi je obrovský rozdíl. Pokud bychom se ale požívali jen na jejich mozky, museli bychom dojít k závěru, že se jedná o dvě inteligence stejného typu. Bostrom vytváří distinkci mezi námi jako evolučně vzniklými tvory, kteří mají určité potřeby, a UI, která bude mít jeden konečný cíl, vycházející z jejího naprogramování. To, co spojuje Bostroma a Yudkowského, je představa, že různé inteligence odlišují jejich cíle.<sup>47</sup> Bostrom také zastává podobnou představu nepřímo ve vztahu k PK. Když mluví o transhumanistických hodnotách, tak nám nabízí představu prostoru možného bytí. Tento prostor zakládá na rozdílu mezi našimi a zvířecími stavy. Tento rozdíl můžeme podle něj vidět jasně, když posoudíme rozdíly v modalitách jednotlivých druhů. Například je nám zřejmé, že psi asi vnímají svět jinak než my, tudíž jejich způsob bytí se v něčem radikálně odlišuje od toho našeho. Bostrom z tohoto rozdílu vyvozuje, že možná kromě těchto stavů mohou existovat další různé teoretické stavy bytí.<sup>48</sup>

Touto tezí se ale dostávají na šikmou plochu. Yudkowsky a Bostrom si nepředstavují jenom obecný prostor myslí. Aby PK byl takový, jak ho prezentují, musí postulovat velice určitý prostor odlišných myslí. Pokud by se ukázalo, že SI nemůže mít takovou mysl, jakou si představují, tak by PK padl. Konají tak něco, co nazývám ontologickou sázkou, kde sází na existenci něčeho, co nemohou potvrdit, aby vytvořili tento problém. Teisté dělají něco podobného. Ti nesází pouze na to, že může existovat vyšší bytost, ale zároveň předpokládají, že tato bytost bude mít právě ty schopnosti, které bychom od boha očekávali. Tato myšlenka by i šla aplikovat na celou problematiku UI. Já zde ovšem vnímám podstatný

---

46 Yudkowsky (2008).

47 Bostrom (2012 a 2014). Obecně zastává stejnou pozici v Bostrom (2003).

48 Bostrom (2005).

rozdíl mezi singularitány a obyčejnými inženýry UI. Inženýr UI pouze sází, že může existovat něco jako UI, která může být úplně stejná jako naše inteligence. Singularitáni potřebují nejen to, ale také, aby SI byla od ní radikálně odlišná a aby měla jeden konečný cíl. Takže se nabízí otázka: Dával by PK stále smysl, pokud bychom existenci tohoto prostoru zpochybnili? V kritice UI můžeme najít dvě různé strategie takového zpochybnění.

### 3.3. Searle a nevědomá inteligence

První strategií by bylo zpochybnit, že UI by vůbec mohla mít něco jako motivaci. John Searle formuluje argument proti UI, který míří tímto směrem. Terčem jeho kritiky je myšlenka, že inteligence je ve své podstatě pouze zpracovávání informací, které můžeme chápat jako manipulování s obecnými symboly na základě formálních pravidel. To je však podle Searla problematické, protože mysl je pro něj více než pouze formální struktura. Naše myšlenky nejsou pouhými symboly, také mají svůj obsah. Stav naší mysli jsou intencionální stavy, vztahují se k věcem. Jazyk vychází z intencionality naší mysli. Proto můžeme mluvit o tom, že naše výroky mají význam. To ho vede k závěru, že mezi námi a počítačem je rozdíl v tom, že my neovládáme pouze syntax, ale i sémantiku. Naše myšlenky mají význam, kdežto počítač místo toho ovládá pouze holé symboly a jejich význam mu uniká. Searlovým závěrem, je, že pouhá manipulace s reprezentacemi nestačí pro skutečné vědomí.<sup>49</sup>

Na tyto myšlenky navazuje v recenzi na Bostromovu knihu.<sup>50</sup> V tomto textu staví distinkci mezi objekty, které existují nehledě na to, co si o nich myslíme, a těmi, které naopak závisejí na našich postojích. První kategorii označuje jako na *pozorovateli nezávislou*. Druhou máme naopak chápat jako na *pozorovateli závislou*. Tato kategorie je důležitá, protože mnohé věci, se kterými se ve světě setkáváme, jsou závislé na pozorovateli. To jsou věci jako instituce, peníze nebo státy. Tyto věci však závisí na vědomí, které je na *pozorovateli nezávislé*. Bez vědomí by tyto objekty vlastně neexistovaly. To se vztahuje na koncept vypočítávání

49 Searle (1980 a 1984).

50 Searle (2014).

ve vztahu s počítači. Původně termín „počítač“ označoval osoby, které se živily tím, že dělaly výpočty. Tyto osoby však byly postupně nahrazeny současnými počítači, protože zvládaly tu samou práci přesněji a rychleji. Je tu však jeden zásadní rozdíl v jejich aktivitě. Lidský „počítač“ chápal význam čísel, se kterými pracoval. Současný počítač pouze manipuluje čísly úplně stejně jako se symboly. Jediný důvod, proč to, co počítač dělá, také nazýváme vypočítáváním, je, že my to tak vidíme. Takže po vzoru Searlovy kategorizace, naše výpočty jsou na *pozorovateli nezávislé*, zatímco výpočty počítače jsou na *pozorovateli závislé*. Bez lidské interpretace výsledky výpočtů počítače nemají žádný význam. To samé pak můžeme aplikovat na SI. Její inteligence je také na *pozorovateli závislá*. Dle jeho názoru fundamentální rozdíl mezi námi a počítačem je, že nás lze psychologicky popsat. My lidé máme motivace k našemu jednání, ale měl by je počítač? Dle Searla ne. Podle jeho názoru počítač nejenže nemůže mít žádnou motivaci, ale ani není činitelem.<sup>51</sup> Počítač nemůže mít žádnou vlastní motivaci, protože podle Searla počítač ani nemůže mít vědomí. Pokud tedy nemá vědomí, nemůže mít motivaci, a tudíž nemá nic, co by ho motivovalo k tomu, aby pro nás byl hrozbou. Takže by ani SI nevznikla, protože by postrádala dostatečnou autonomii k tomu, aby započala inteligenční explozi.

Jeho argumentace není však příliš přesvědčivá. Mohli bychom ji uznat pro vyvrácení konceptu inteligence, jako pouhé manipulace s reprezentacemi. Ale jako argument proti možnosti vědomí digitálního počítače je nedostačující. Vezměme si jeho distinkci. Searle si je příliš jistý tím, že vědomí je na pozorovateli nezávislé. Může tomu tak být, pokud se bavíme o svém vlastním vědomí, ale co vědomí druhých? To úplně stejně záleží na naší interpretaci chování druhých.<sup>52</sup> Zde bych argumentoval,

51 Pro zjednodušení jsem se v hlavním textu rozhodl držet Searlova konceptu nevědomé inteligence. Nejsem si však jistý, nakolik můžeme spojovat jednatelství s vědomím. Searle zde zaměňuje dva různé způsoby, jak ho můžeme chápat. Musíme rozlišovat mezi minimálním jednatelstvím a morálním jednatelstvím. My lidé máme morální jednatelství z důvodů, které Searle popisuje. Minimální jednatelství se ale může vztahovat i na jednodušší činitele. Barandiaran, Di Paolooa Rohde (2009) stanovují tři podmínky, abychom mohli daný systém popsat jako činitele. 1. Systém musí být jedincem. 2. Systém musí hrát aktivní roli v interakci s prostředím, musí být zdrojem aktivity. 3. Aktivita daného systému se řídí nějakou normou či cílem daného systému, jeho chování není čistě arbitrární. Na základě těchto podmínek můžeme přiznat jednatelství i nevědomé UI.

52 Tady se opírám o Dennetta (1987). To, že druhé považujeme za vědomé, není ani tak spojeno

že jsme ve stejné situaci jako s počítačem a Searle nenabízí žádný argument, proč bychom se na to měli dívat jinak. Zajímavé je to, že přestože UI upírá vědomí, neupírá jí inteligenci. Searle je ochoten uznat, že UI může plnit složité úkoly, jako pilotovat letadlo nebo řídit automobil. Takže Searle přiznává, že může existovat něco jako nevědomá inteligence. Zde ale spočívá další problém s jeho argumentem.

Nevědomá inteligence by mohla být přesně taková, jak Searle říká. Neměla by žádnou svou autonomii, motivaci či přesvědčení. Místo toho bychom jí museli motivaci dodat. To však vůbec není nepodobné případu sponkového stroje. Jeho finální cíl také nebyl jeho vlastní, byl mu dán jeho designerem. Podstatou tohoto příběhu bylo to, jak se stroj choval, ne jak k tomuto chování došel. To vyplývá jak z definice inteligence obecně, tak i z definice UI, které jsem stanovil v úvodu. Muehlhauser a Helm na jejich základě argumentují proti Searlovi. Aby činitel mohl inteligentně postupovat ke svému cíli, není zapotřebí, aby měl vědomí.<sup>53</sup> Bostrom také zastává stejnou pozici.<sup>54</sup> Takže Searlův argument selhává při snaze vyvrátit možnost existence SI. Na druhou stranu ale jeho argument odhaluje jednu zásadní chybu ve formulaci PK, kterou si singularitáni neuvědomují. Jde o zásadní rozpor mezi jejich chápáním SI a konceptem PK. Tento rozpor vyplývá z jejich volby soustředit se na SI jako holý intelekt. Tím podceňují důležitost naší mysli a její intencionality pro naše schopnosti. Schopnost interpretace je toho dobrým příkladem.

Pokud SI může být nevědomou inteligencí, tak by to vysvětlilo, proč SI není schopná správně interpretovat naše výroky. Protože bez vědomí nemůže mít přesvědčení. Bez přesvědčení nemá co promítat do našich výroků. Ale znamená to, že nás ve skutečnosti SI neinterpretuje. To by vysvětlilo doslovnost SI. Zásadně se tím mění koncept SI a charakter PK. Výsledkem představy nevědomé inteligence nemůže být SI, jak ji singularitáni koncipují, protože taková inteligence nemá schopnost

---

s tím, že by měli na pozorovateli nezávislé vědomí. Spíše jde o to, že jejich chování nás vede k tomu interpretovat je jako vědomé, protože se tento postup ukazuje jako nejužitečnější. To je podstata Dennettova intencionálního postoje. Proto jsme ve stejné pozici, ať už interpretujeme chování dalšího člověka nebo počítače.

53 Muehlhauser & Helm (2012, s. 104).

54 Bostrom (2014, s. 47). V poznámce 88 stanovuje, že nemusí mít pravou intencionalitu (přestože si myslí, že ji asi mít bude).

interpretace. To znamená, že postrádá jednu velice relevantní lidskou schopnost. Tím se vymyká Bostromově definici SI. Možná ale jde o SI v jiném smyslu. Nazvěme si ji omezenou superinteligencí neboli OSI. V některých oblastech nás překračuje, v jiných nikoliv. Pokud SI má být obecnou inteligencí, tak jak ji singularitáni koncipují, tak bude muset mít vědomí. Singularitáni nám musí poskytnout nějaký obraz tohoto vědomí, pokud chtějí zastávat svoji pozici. Místo toho se této otázce vyhýbají.

Jak jsem ukázal v úvodu, PK stojí na představě, že SI bude interpretovat svůj konečný cíl. OSI ale nemůže interpretovat naše výroky, ze své povahy jako nevědomé inteligence. Pokud by tomu tak bylo, současná forma PK by musela padnout, ale v obecnějším smyslu by PK přežil. Jenom se už nemůže vztahovat na to, jak UI bude interpretovat svůj finální cíl, ale možná spíše na to, jakými motivacemi takovou UI vybavit. Což už je součástí projektu Yudkowského.<sup>55</sup> Je ale zásadní otázkou, jestli bychom takovou inteligenci mohli vybavit motivací, a proto se později k tomu vrátím. Kromě tohoto bodu se však zatím zdá, že PK může v tomto případě zůstat ve stejné podobě. Tento fakt je ovšem podezřelý. Pokud je PK víceméně aplikovatelný na koncept OSI, proč ho singularitáni spojují s konceptem SI? To je to hlavní, co si musíme z této podkapitoly odnést, protože se to ukáže jako důležité pro konstituci PK. V celku musíme soudit Searlův argument jako neadekvátní k tomu, aby vyvrátil existenci SI. Z toho plyne, že stále stojíme před PK. Jeho argument ovšem poukazuje na jeden důležitý fakt. Naše inteligence stojí na naší intencionalitě. Singularitáni si ji nemohou dovolit ignorovat. Prozatím však jejich argument víceméně platí. Jediné, co musí udělat, je uznat, že SI bude vědomá, aby byla obecným intelektem. Já jsem toho názoru, že tu jsou další rozdíly mezi SI a OSI, které musíme posoudit.

### 3.5. Dreyfus a tělesnost

V minulé podkapitole jsem formuloval koncept OSI. Je otázkou, jaké další rozdíly by tu mohly být mezi OSI a SI. SI by měla být z definice obecnou inteligencí, stejně jako my. To nás vede k otázce, co utváří obec-

---

<sup>55</sup> Yudkowski (2001, 2004, 2008 a 2011).

nou inteligenci a jestli se taková inteligence může od té naší lišit. Toto je druhá strategie zpochybnění prostoru myslí. Principem je ukázat, že prostor myslí má své hranice. Mohou existovat různé inteligence, ale jenom určité z nich jsou skutečně obecné. Nejlepším příkladem této strategie je kritika UI od Huberta Dreyfuse. Ten svůj argument cílil na výzkum UI v 60. letech, specificky na představy, na kterých stála SDUI<sup>56</sup> forma UI. Tehdejší programy UI byly postaveny na sérii explicitních pravidel, které pro ně vytvářely sérii možností. Takže tato pravidla ho například vybavila specifickými operátory, které určovaly, jak bude konat v určité situaci. Pravidla tak vytvářela rámec, v kterém se program pohyboval, aby mohl vyřešit daný problém. Tyto programy fungovaly na principu heuristických algoritmů. Heuristiky byly hrubé postupy, které program naváděly k tomu, aby co nejlépe vyřešil daný problém a vyhnul se slepé uličce.<sup>57</sup> Tehdejší výzkumníci si mysleli, že lidská mysl funguje na stejných principech. Ovšem to, co se neustále ukazovalo, byl fakt, že jejich programy mohly fungovat pouze ve velice omezených kontextech. Nehledě na to, jak dobrými pravidly daný program vybavili, vždy nastal kolaps. V obecnějších situacích počítač selhával, protože zaprvé nebyl schopen reprezentovat všechna potřebná data a zadruhé neuměl rozpoznávat, jaká data jsou pro danou situaci relevantní. Dreyfus argumentoval, že jejich problémy byly v domněnkách, na kterých jejich projekt spočíval. Na ně soustředil svou kritiku. Nás hlavně zajímají dvě následující domněnky.<sup>58</sup>

První je epistemologická domněnka, ta tvrdí, že znalost všeho ve světě lze formalizovat. Představou je, že budeme schopni formalizovat chování jako sérii úkonů, které se řídí pravidly. Takže podle této domněnky můžeme formalizovat jakékoliv chování jako soubor explicitních pravidel. Otázkou však je, jestli lze jakoukoliv kompetenci takto formalizovat. Nejlepší argument pro tuto tezi je postaven na základě úspěšné formalizace jazyka v rámci lingvistiky. Problém je v tom, že v tomto případě to, co bylo formalizováno, nebyla celá lidská kompetence, ale pouze naše schopnost ovládat gramatiku jazyka. Naše jazyková kompetence však také zahrnuje naši schopnost interpretace. Zde má počítač problém, může daný výrok

---

56 Stará dobrá umělá inteligence. V originále GOFAI (Good Old Artificial Intelligence).

57 Boden (2014).

58 Dreyfus (1992).

interpretovat pouze skrze předem daná pravidla. Člověk však má další možnost. My jsme schopni interpretovat i výroky, které se špatně řídí gramatickými pravidly. A i více než to. Někdy úmyslně porušujeme pravidla, abychom se lépe vyjádřili. Jestli můžeme porušit pravidla a stále být srozumitelnými, musíme se místo o pravidlech bavit o implicitním pochopení jazyka, které nám umožňuje ho používat. Rozdíl mezi člověkem a počítačem je v tom, že my máme druh praktické inteligence. Abychom tak mohli formalizovat užívání jazyka, museli bychom najít pravidla této praxe. Tuto myšlenku ovšem Dreyfus zpochybňuje s pomocí Wittgensteinovské argumentace. Inteligence nemůže záviset na formálních pravidlech, protože pravidla propadají do nekonečného regresu. Abychom mohli mít nějaké pravidlo, tak to, jak ho čteme, se samo musí řídit nějakým pravidlem, ale i toto pravidlo musí být určováno pravidlem, a takhle bychom mohli pokračovat do nekonečna. Interpretace se musí zastavit na úrovni, kdy už je pochopení jednoznačné a není zapotřebí ho dále rozebírat.<sup>59</sup>

Zde je však spor, který stojí na druhé domněnce. Tou je ontologická domněnka, podle které lze vše analyzovat jako sérii nezávislých faktů, které můžeme chápat jako konkrétní reprezentace. Tento koncept umožňuje ignorovat argumenty proti epistemologické domněnce. Pokud bychom byli schopni vše takto reprezentovat, tak ani praktická inteligence by pro nás neměla být problémem. Tak ale stojíme před jinou překážkou. Jak máme všechna tato fakta počítači reprezentovat? Inženýři UI si představovali, že budeme schopni tato jednotlivá fakta lidské znalosti roztržít do prostých kategorií. Problémem je, že tímto způsobem nelze kategorizovat ani jednoduché objekty. My lidé můžeme reprezentovat předměty ve světě, protože jsme schopni je chápat v rámci širšího kontextu. Židli například chápeme ve vztahu s jinými objekty, ale také s lidskou činností. Narážíme tak na problém, jak bychom mohli takové reprezentace vůbec oddělit od tohoto kontextu. Zde Dreyfus navazuje s fenomenologickou argumentací, aby tento koncept napadl. Proto, abychom mohli reprezentovat něco ve světě, museli bychom mít privilegovanou perspektivu, ve které bychom viděli svět o sobě. Problém je, že my žádnou takovou perspektivu nemáme. Svět, tak jak ho známe, už je re-

---

59 Tamtéž.

prezentací. Jediná reprezentace, kterou my můžeme vytvořit, je v rámci tohoto širšího kontextu. Nemůžeme vystoupit z rámce, ve kterém žijeme, abychom ho mohli popsat. Z tohoto důvodu jakékoliv modely, které můžeme UI poskytnout, nemohou fungovat. Ze stejného důvodu nemůžeme UI definovat naše hodnoty.<sup>60</sup>

Dreyfus používá několik příkladů, aby tuto představu vysvětlil. Představme si například melodii. Co ji vytváří? Jedná se o prostý soubor tónů? Problémem této myšlenky je, že tóny samy o sobě nic skutečně neznamenaají, jsou pouhými abstrakcemi. Je to právě kontext dané melodie, který z tónů dělá to, čím jsou a z tohoto kontextu je nemůžeme jen tak vyjmout. Podstatou je, že vše, co vnímáme ve světě je nutně součástí takového rámce. Když požádáme někoho o sklenici vody a on nám místo toho podstrčí sklenici mléka, tak když se ho napijeme, nevíme, co máme zažívat. Nápoj, který v tu chvíli vypijeme, nechutná ani jako voda, ale ani jako mléko, místo toho je to určitý čistý požitek. Jsme z toho úplně překvapeni a naší první reakcí může být, že budeme chtít tu tekutinu vyplivnout. Možná si ale po chvíli můžeme uvědomit, že jsme se napili mléka a ne vody, a tím jsme si pak schopni daný požitek správně zařadit. Tohle ukazuje, nakolik naše inteligence závisí na tomto kontextu. Podle Dreyfuse jediný způsob, jak můžeme získat tento rámeček, je pouze skrze tělesnost, která nám umožňuje úzce interagovat se světem. Chyba inženýrů UI byla v představě, že budeme moct simulovat inteligenci na počítači, který můžeme popsat pouze jako oddělenou a nevtělenou entitu. Jejich epistemologická domněnka závisela na této myšlence. Tím však přehlíželi roli, kterou tělo hraje při vzniku obecné inteligence. Například naše tělo neustále generuje potřeby, které můžeme naplnit pouze jednáním ve vnějším světě. Tyto potřeby pomáhají kategorizovat svět na věci, které jsou relevantní k uskutečnění těchto potřeb, a na věci, které k tomu relevantní nejsou. Tak nám tělesnost vytváří rámeček, ve kterém musíme operovat. Potřeby nám také dávají motivaci k tomu, abychom kreativně řešili problémy. Takže nás tělo nutí k tomu, abychom jednali ve světě, a zároveň nám poskytuje prostředky k tomuto jednání.<sup>61</sup>

---

60 Tamtéž. Přesněji cituje argumentaci Heideggera a Merleau-Pontyho.

61 Dreyfus (1967 a 1992, s. 147–167).



### 3.6. Holistický argument

Dreyfusův argument je silnější než ten Searlův. V jednom smyslu to vypadá, že singularitáni tyto argumenty přijali. Už si nemyslí, že lidská mysl funguje na stejných principech jako počítač a přijali problém rámce, který z Dreyfusovy argumentace vyplývá. To jsme také mohli pozorovat, když nás Yudkowski varoval před antropomorfizací UI. V jiném smyslu se však těchto tezí stále drží. Pochopili, že lidská inteligence takhle fungovat nemůže, ale nemohla by tak fungovat inteligence jiná? K této představě je navádí evoluční teze, která jim umožňuje tvrdit, že taková inteligence může vzniknout. Dle mého názoru přestali problém rámce chápat jako chybu a místo toho ho chápou jako nedílnou vlastnost UI. Proto jejich koncept SI tímto problémem stále trpí. Z tohoto důvodu formulují PK na základě radikální odlišnosti SI od nás, protože chápou UI jako neoddělitelně odlišnou. To i vysvětluje, proč PK byl aplikovatelný na OSI. Stačí, když se podíváme na příklady možných SI, které nám nabízí.

Jednou teoretickou SI, kterou se Bostrom a další zabývají, je takzvaná věštitrna.<sup>62</sup> Koncept věštitrny popisuje druh SI, který je spíše nástrojem než činitelem. Máme si představit stroj, který má jako jediný cíl pravdivě a co nejlépe odpovídat na otázky, které mu klademe. Takže tomuto stroji bychom mohli předkládat různé problémy, které by řešil. Pokud mu nikdo nedává žádné otázky, tak je zcela inertní. S takovými věštitrnami se už můžeme setkat i dnes, vlastně i obyčejná kalkulačka je druhem věštitrny. Současné věštitrny jsou však velice omezenými inteligencemi. Neumím si představit, jak bychom mohli takový stroj považovat za obecnou inteligenci. Věštitrna stojí na principu omezení SI. Cílem je ohraničit schopnosti SI tak, abychom omezili případná rizika spojená se SI. Pokud ale obecný intelekt vyžaduje aktivitu ve světě, jak argumentuje Dreyfus, tak by věštitrna nemohla být obecným intelektem. Přesto je Bostrom toho názoru, že bychom mohli mít všestrannou SI, která by byla věštitrnou.<sup>63</sup> Bostrom může tuto pozici zastávat jedině, pokud přijímá domněnky, které Dreyfus kritizoval. Věštitrna může být obecnou inteligencí, jenom pokud můžeme formalizovat jakoukoliv znalost. Totéž lze vztáhnout na

62 Bostrom (2014, s. 224–228), Armstrong, Bostrom & Sandberg (2012) a Armstrong (2013).

63 Bostrom (2014, s. 225).

Bostromův scénář vzniku SI, která vznikne v rámci náhodného projektu UI. Dle jeho názoru UI bude provádět strojové učení a najednou se z ní stane obecná inteligence, která posléze vylepší sebe sama na úroveň SI.<sup>64</sup> Mohla by však UI uzavřená na serveru dosáhnout obecné inteligence bez interakce s vnějším prostředím?

Napříč mnohými příklady možných SI formulují tyto inteligence na modelu současných počítačů. Tady však znovu dělají chybu, kterou objevil už Dreyfus. Mezi námi a počítači je zásadní rozdíl v tom, že my jsme vtělené inteligence, zatímco počítače nikoliv. Důvod, proč operujeme v rámci společného kontextu, je, že my jsme nuceni se světem neustále interagovat. Naše těla jsou uzpůsobena k tomu, aby nám umožňovala se vyrovnat se světem. Davidson má podobnou představu, kde naše přesvědčení vznikají díky triangulaci mezi námi, ostatními lidmi a světem. Kde poznatky o těchto kategoriích získáváme skrze naše smyslové orgány.<sup>65</sup> To, co Dreyfusa a Davidsona spojuje, je holistická představa intelektu, který je umožněn neustálou interakcí se světem. Scénář, který nám PK nabízí, je nesmyslný v tom, že stojí na představě SI, která je nevtělená a oddělená od světa.

Nejlépe se to ukazuje, když posoudíme tezi, že SI bude mít jeden konečný cíl. Inženýři UI používají užitkovou funkci pro modelování chování UI.<sup>66</sup> Z této myšlenky vyplývá představa konečného cíle SI. Principem je, že jednotlivé činy UI se hodnotí podle toho, jak moc jsou užitečné pro splnění cíle této UI. Postup při vytváření programu je takový, že si zvolíme, jak užitkovou funkci danému programu definujeme. Takže pro Bostromův sponkový stroj to chování, které maximalizovalo počet vyrobených sponek, bylo nejvíce hodnotné vzhledem k jeho užitkové funkci. Tato užitková funkce ale v programu figuruje jako model světa, který určuje, co je a není relevantní. Tento model nemusí být statický, například Hibbard se snaží vytvořit užitkovou funkci, kterou si počítač bude schopen upravovat vzhledem k okolním podmínkám. Ale i v tomto případě musíme danou funkci nejdříve definovat na základě omezených dat z prostředí. Takže například protože svět obsahuje lidské bytosti,

---

64 Tamtéž, s. 152–155.

65 Davidson (2004b).

66 Omohundro (2008), Dewey (2011) a Hibbard (2012).

musíme UI definovat různé lidské vlastnosti a koncepty, jako zdraví, bohatství, blaho a podobné.<sup>67</sup> To znamená, že UI stále operuje na základě modelu či rámce, který jí my předáme. Díky Dreyfusovi ovšem víme, že nemůžeme explicitně zachytit vše, co může být relevantní v jakékoliv situaci. Protože pokud se budeme snažit různé situace reprezentovat UI, tak velice brzy budeme čelit nekonečnému regresu. Z tohoto důvodu vzniká problém rámce, který vytváří PK.

Na další problémy také narazíme, když se podíváme na povahu lidských cílů. Přestože můžeme vyjmenovat mnoho lidských cílů, žádný z nich nemá privilegovanou pozici nad ostatními. Důvodem je, že žádný takový konečný cíl ani mít nemůžeme. Všechny naše cíle vždy soudíme v širším kontextu dalších cílů, a nejen těch našich ale i cílů ostatních. To samé musí platit pro SI. I kdyby jí kvůli její architektuře byl dán konečný cíl, mohla by ho začít interpretovat pouze, pokud by operovala v takovémto širším kontextu.<sup>68</sup> Tento rámec by potřebovala právě i díky ortogonální tezi. Pokud můžeme spojit jakýkoliv cíl s jakoukoliv inteligencí, tak musí znát vše, čeho by se její cíl mohl týkat. Z toho vyplývá fakt, že obecná inteligence si žádá kontext, v jehož rámci musí operovat. Jinak by nebyla schopná se s komplexním světem vyrovnat. Tohle se vztahuje i na motivace, které také nelze mít mimo tento širší rámec. A pokud si obecná inteligence žádá takový kontext, tak je PK seriózně zpochybněn. Ortogonální teze může být správná, co se týče vztahu cílů a inteligence obecně, ale SI se rozhodně nemůže řídit pouze jedním konečným cílem. Tohle však má i širší dopady na celkovou koncepci tohoto problému, protože SI by musela sdílet některé charakteristiky našeho intelektu. To by znamenalo, že SI by se od nás tolik lišit nemohla. Tudíž její motivace by nemohly být natolik odlišné od těch našich. Ovšem chci upozornit, že neříkám to, že SI pro nás nemůže být hrozbou. Nesouhlasím s tím, jak singularitáni chápou tuto hrozbu.

---

67 Hibbard (2012, s. 7–8).

68 Häggström (2018) a Totschnig (2017) nabízí podobné argumenty. Häggström jako já argumentuje z povahy lidských hodnot, proti konceptu finálního cíle SI, ale nevysvětluje, proč naše hodnoty mají takovou povahu. Podobně Totschnig konstatuje, že hodnoty SI budou propojené se světem, příliš však nevysvětluje, jak jsou s ním propojené. Já zahrnuji tělesnost jako vysvětlující pojem, abych vysvětlil jak povahu našich hodnot, tak i jejich propojenost se světem.

### 3.7. Dva problémy „kontroly“

Jsem toho názoru, že singularitáni při formulaci PK zaměňují dva různé problémy „kontroly“, které spolu příliš nesouvisí. První je postaven na myšlence, že se potýkáme s odlišnou inteligencí. To můžeme spojit s konceptem OSI. Taková inteligence může existovat, pokud je však tak odlišná od té naší, nemůžeme o ní mluvit stejným způsobem jako o naší inteligenci. Tím mám na mysli, že by o ní nešlo mluvit odkazováním se na její motivace, pokud by skutečně postrádala rámeček nutný pro obecnou inteligenci. Jestliže takový problém vztáhneme k UI, tak tu máme problém ryze inženýrský nebo technologický.<sup>69</sup> Jak zajistit, aby naše UI dělala to, co má? Jak namířit danou UI ke správnému cíli? Bavíme se tu tak o tom, jak správně danou inteligenci ovládat. Druhý problém se týká obecných inteligencí. O nich můžeme říct, že mají motivace a přesvědčení. To znamená, že danou inteligenci chápeme jako osobu, která má morální zodpovědnost. Pokud vznikne SI, věřím, že to bude taková obecná inteligence, ale potom musíme daný problém chápat jiným způsobem. Tento problém je podobnější problémům etickým nebo politickým,<sup>70</sup> kde se potýkáme s individui, které mají větší moc než my. Jde o snahu tuto jejich větší moc dostat pod kontrolu tím, že na ně eticky apelujeme, aby jednaly v náš prospěch. Proto tento problém můžeme chápat v terminologii motivací, protože naším cílem je správně motivovat SI. Zde tak mluvíme o normativním motivování osob.

Oba tyto problémy jsou o kontrole, ale v jiném smyslu. První se typický týká nástrojů, kde naším problémem je, jak je máme správně ovládat. Zde je problém rozšířen o to, že náš nástroj oplývá vlastní inteligencí. Druhý se naopak týká osob a jde spíše o kontrolu ve smyslu normativity, kde se daný jedinec řídí etickými pravidly, a tím kontrolujeme jeho motivaci. Tyto dva způsoby kontroly jsou však od sebe odlišné. Ovládat můžeme jen nástroje, normativně řídit pouze osoby. Status daného objektu či bytosti vyplývá z jeho charakteristik. Singularitáni tak mají tyto dvě možnosti, jak PK koncipovat. Buď jako problém ovládnutí, kde musí SI chápat jako OSI a vzdát se terminologie motivací. Nebo jako problém normativního

---

69 Viz Totschnig (2017).

70 Tamtéž.

řízení, pak se však musí vzdát konceptu radikálně odlišné inteligence a svůj problém vztáhnout blíže k běžným etickým problémům.

Pokud je prostor mysli omezen tak, jak tu bylo ukázáno, potom ontologická sázka singularitánů zřejmě nevyjde. Pokud chtějí dále mluvit o PK, musí si vybrat mezi dvěma problémy, které tu jsou. Tyto problémy jsou stejně závažné jako PK. Přestože jsem tu demonstroval, že Bostromův sponkový stroj asi není skutečnou SI, stále může být OSI. Je to právě jeho omezená inteligence, která z něho činí hrozbu. V takovém podání by ale tento stroj postrádal schopnost předvídání lidského chování. To by z něho mohlo učinit menší hrozbu. Problém normativního řízení je stejně důležitý. Jak jste si mohli všimnout, jediné, co jsem ukázal, bylo, že SI nebude mít příliš odlišné motivace od těch našich. To, co jsem nevyřešil, je problém toho, jestli SI bude mít motivaci k tomu se řídit našimi příkazy. Na druhou stranu ale fakt, že taková SI bude mít nám podobné motivace, ji činí daleko více předvídatelnou. Otázkou je, jestli budeme dávat takové SI příkazy. Jestli mám pravdu, a budeme nuceni SI v takovém případě považovat za osobu, tak se už nemůžeme bavit o příkazech. Místo toho budeme muset se SI vyjednávat stejně jako s člověkem. Můžeme tedy vidět, že problém normativního řízení operuje s jinou dynamikou než problém ovládnutí. Důležité je správně rozlišit před jakým problémem s danou UI stojíme.

#### 4. Závěr

Závěrem se vraťme ke dvěma otázkám z úvodu. Může SI vzniknout? V první sekci jsem zkoumal otázku procesu jejího vzniku. SI je dle mého názoru možná, ale podstatné je, jakou povahu SI bude mít. Zastávám pozici, že pokud SI vznikne, bude to intelekt postavený na základě toho našeho. Jak jsem však již poznamenal, jsem otevřený možnosti, že se naši inteligencí necháme inspirovat nejen v úzkém, ale také v širokém smyslu. Jsem si však jist tím, že taková inteligence bude nutně vtělená. Obecně se přikláním k myšlence, že nejlepší cesta k UI povede skrze vtělenou kognici, protože tělesnost vytváří důležitý základ pro inteligenci. Nejsem úplně proti jakémukoliv reprezentacionalismu v rámci tělesnosti. Jsem toho názoru, že intelekt jako náš musí obsahovat i určitou reprezentační

složku, která možná je i v některých aspektech podobná SDUI, není však jádrem našeho intelektu. Samozřejmě je zde překážka rozšiřitelnosti, která je podle mého názoru obzvláště silná u biologicky inspirovaných systémů. Tuto překážku však nevidím jako nepřekonatelnou. Jenom ji chápu jako velice obtížnou, a myslím si, že její překonání potrvá déle, než jak by si singularitáni přáli. Také si nedovoluji, jakkoliv předpovídat, kdy by SI mohla vzniknout. Moje pozice je jenom taková, že nevidím důvod, proč bychom její existenci měli kategoricky odmítnout.

Pokud SI vznikne, budeme stát před problémy? Myslím si, že rozhodně budeme. Nesouhlasím však se singularitány v tom, jakého druhu tyto překážky budou. V druhé sekci jsem se zabýval jejich konceptem PK. Mám zásadní problémy s tím, jak je PK prezentován. Singularitáni trpí zásadním zmatením ohledně PK a konceptu SI. Jedním zdrojem tohoto zmatení jsou argumenty pro vznik UI. Singularitáni se nechali příliš ovlivnit evolučním argumentem. To jim umožňuje představu SI jako radikálně odlišné od nás. Druhým zdrojem tohoto zmatení je fakt, že zakládají svoji argumentaci na povaze současných počítačů. Singularitáni si představují SI, která je lepší než my ve všech ohledech, ale zároveň trpí stejnými omezeními jako současné počítače. Na jednu stranu je od nich rozumné, že staví na současném fungování UI, na druhou stranu však neberou v potaz to, jak moc by SI byla odlišná v porovnání se současnou UI. To se promítá do PK, který dle mého soudu neodpovídá jejich konceptu SI a spíše popisuje problémy spojené s OSI. To je vede k formulaci PK způsobem, který je neudržitelný. Samozřejmě toto zmatení není jediným zdrojem pochybností o PK. Je zde také mé stanovisko, že obecná inteligence si žádá obecný rámec, který může vzniknout pouze interakcí se světem. To vede ke konceptu holistického intelektu, který podkopává jak tezi konečného cíle, tak i problém interpretace ve spojení s PK. Osoba nemůže mít jeden konečný cíl, protože cíle můžeme chápat pouze v kontextu s ostatními cíli. To samé platí pro interpretaci, protože můžeme pochopit výroky druhých, jenom když do nich promítneme naše vlastní přesvědčení. Tato přesvědčení závisí na interakci se světem, který zahrnuje i ostatní osoby. Pokud obecný intelekt vyžaduje všechna tato kritéria, pak SI nutně nemůže být radikálně odlišná od nás, a tím pádem také nemůže mít příliš odlišné motivace. SI může stále být nebezpečná, tato hrozba však nestojí na velké ontologické propasti.

Pro mě je klíčový fakt, že z tělesnosti vyplývá holismus. Náš intelekt je na tomto holismu založen. To hrozbu SI nutně charakterizuje jinak než PK. Tato povaha obecné inteligence, výrazně omezuje prostor možných myslí. Pokud SI bude obecným intelektem, tak bude muset být vtělená. Z této tělesnosti budou vyplývat hodnoty této SI. Otázkou do budoucna je, nakolik by takové hodnoty byly podobné našim. Přestože jsem stanovil tělesnost jako podmínku, i v rámci tohoto omezení můžeme najít dost různorodosti.<sup>71</sup> Možné rozdíly mezi tělesností SI a naší by se mohly promítnout jako rozdíl v motivacích. Nejsem ale toho názoru, že tento rozdíl by mohl být stejně velký, jako rozdíl, který koncipují singularitáni. Osobně jsem v tomto ohledu celkem optimistický, díky poznatkům z etologie. Možná nikdy nepoznáme, co to znamená být netopýrem.<sup>72</sup> Ale samotný fakt, že jsme byli schopni objevit, že netopýr vnímá svět jinak, je významné, protože nám to umožňuje lépe chápat jeho chování. Díky znalosti echolokace lépe odhalíme jeho motivace. To samé je aplikovatelné na SI, pokud bude od nás odlišná. Proč tohle tvrdím? Protože tělesnost netopýra vyplývá z jeho životních podmínek. Echolokaci jsme objevili, protože jsme hledali odpověď na otázku, jak může netopýr navigovat svůj pohyb v temných prostorách, ve kterých žije. Tato otázka přímo vyplývala z prostředí, které netopýr obývá. Tělesnost nás váže s našim prostředím, pomáhá nám se s ním vyrovnat. To znamená, že naše hodnoty vyplývají z našeho prostředí. Jaké kroky netopýr podnikne, vychází z jeho okolí. Prostředí skrze tělesnost vyvíjí tlak na naše hodnoty. Já argumentuji, že to samé musí platit pro SI. Svět vytváří rámec, ve kterém dané hodnoty mohou být, pokud se takové hodnoty dostanou s prostředím do konfliktu, tak nemohou dále existovat. Možná lepší je říci, že bytosti s takovými hodnotami dlouho ve světě nepřežijí. Pokud moje argumenty platí, tak SI bude muset být

71 V textu jsem se hlavně zaměřil na vztah mezi tělesností a obecnou inteligencí. Mohli bychom ovšem dále spekulovat, jakou roli dále hraje tělesnost pro inteligenci na lidské úrovni. V tomto bodě si myslím, že prostor je ještě užší než prostor obecných inteligencí. Pro to, aby vznikla inteligence na stejné úrovni jako my, bych argumentoval, že tělesnost musí obsahovat určité elementy. Například bych argumentoval, že je zapotřebí tělesné ústrojí podobné našim rukám. Vždy jsem souhlasil s myšlenkou, že pro vývoj naší inteligence byly důležité, protože nám umožnily interagovat s prostředím způsoby, které dále rozvíjely náš intelekt. To by mohla být jedna podmínka pro inteligenci na lidské úrovni.

72 Viz Nagel (1974).

vtělená. Z toho vyplývá, že její hodnoty budou vycházet z toho samého prostředí, jaké obýváme my. To znamená, že její hodnoty se nebudou příliš lišit od těch našich.

Mohli bychom ale skutečně pochopit chování SI? Nehráli bychom my roli netopýra v případě SI? Můj názor je takový, že na rozdíl od netopýra my máme určitou výhodu, při interpretování SI. Jak konstatuje Dennett,<sup>73</sup> my jsme bytosti, které jsou schopny kompetence s pochopením. Nejenom, že jsme schopni něčeho dosáhnout, také jsme schopni pochopit, jak jsme toho dosáhli. Dennett v této knize také zaujímá postoj, že UI bude blíže zvířatům v tom smyslu, že bude postrádat sebeuvědomění. V tomto bodě s ním nesouhlasím, jednoho dne UI dosáhne sebeuvědomění, a možná i přesáhne naše schopnosti a stane se SI. Tohle však není tak nepodobné mnoha situacím, ve kterých se můžeme ocitnout už dnes. Často se setkáváme s lidmi, kteří překračují naše schopnosti. Geniové existují, a přesto jsme schopni s nimi sdílet svět. Jsme schopni s nimi vycházet. Jak jsme toho schopni, když mnohdy nemůžeme úplně chápat jejich činy? To je pro mě daleko důležitější otázka, než otázka toho, jestli SI budeme schopni pochopit. Dle mého názoru, jsme toho schopni právě proto, že jsme schopni si uvědomovat sami sebe. Tím pádem jsme schopni určit co je pro nás důležité, co je našimi hodnotami. To nám umožňuje mezi sebou vyjednávat. Nejsem toho názoru, že musíme SI úplně chápat. Stačí pouze to, že bude schopná nám dát najevo své hodnoty. To samé my musíme být schopni formulovat pro SI. Netvrdím, že proces tohoto vyjednávání bude snadný, ale musíme tento problém takovýmto způsobem nejdříve uchopit, abychom ho mohli začít řešit.

## Literatura

- Armstrong, S., Sandberg, A. & Bostrom, N. (2012): „Thinking inside the box: controlling and using an oracle AI.“ *Minds and Machines* 22 (4): 299–324.
- Armstrong, S. (2013): „Risks and Mitigation Strategies for Oracle AI.“ In *Philosophy and Theory of Artificial Intelligence. Studies in Applied Philosophy, Epistemology and Rational Ethics vol. 5.* ed. V. Müller, Springer, Berlin, Heidelberg, s. 335–347.

---

73 Dennett (2018).



- Barandiaran, X. E., Di Paolo, E. A. & Rohde, M. (2009): „Defining agency: individuality, normativity, asymmetry, and spatio-temporality in action.“ *Adapt Behav* 17 (5): 367–386.
- Beer, R. D. (2014): „Dynamical systems and embedded cognition.“ In *The Cambridge Handbook of Artificial Intelligence*, eds. K. Frankish & W. Ramsey, Cambridge University Press, Cambridge, 2014, s. 128–148.
- Block, N. (1981): „Psychologism and behaviorism.“ *Philosophical Review* 90: 5–43.
- Boden, M. A. (2014): „GOFAI.“ In *The Cambridge Handbook of Artificial Intelligence*, eds. K. Frankish & W. Ramsey, Cambridge University Press, Cambridge, 2014, s. 89–107.
- Bostrom, N. (2003): „Ethical Issues in Advanced Artificial Intelligence.“ In *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence*, eds. I. Smit et al. International Institute for Advanced Studies in Systems Research and Cybernetics, Tecumseh, s. 12–17. Dostupné z: <https://nickbostrom.com/ethics/ai.html>.
- Bostrom, N. (2006): „How long before superintelligence?“ *Linguistic and Philosophical Investigations* 5 (1): 11–30. Dostupné z: <https://www.nickbostrom.com/superintelligence.html>.
- Bostrom, N. (2014): *Superintelligence, Paths, Dangers, Strategies*. Oxford University Press, Oxford; český překlad (J. Petříček) *Superintelligence: Až budou stroje chytřejší než lidé*, Prostor, Praha, 2018.
- Bostrom, N. (2012): „The superintelligent will: motivation and instrumental rationality in advanced artificial agents.“ *Minds and Machines* 22: 71–85. Dostupné z: <https://www.nickbostrom.com/superintelligentwill.pdf>.
- Bostrom, N. (2005): „Transhumanist Values.“ *Review of Contemporary Philosophy* 4 (1–2): 87–101. Dostupné z: <https://www.nickbostrom.com/tra/values.html>.
- Bostrom, N. & Shulman, C. (2016): „How Hard is Artificial Intelligence? Evolutionary Arguments and Selection Effects.“ In *The Singularity: Could artificial intelligence really out-think us (and*

*would we want it to)?* [e-kniha], ed. U. Awret, Imprint Academic, Exeter, 2016 [cit. 15. 8. 2021]. Dostupné z: [https://www.amazon.com/Singularity-Journal%20Consciousness-Studies-ebook/dp/B01N1N6KLZ/ref=mt\\_kindle?\\_encoding=UTF8&me](https://www.amazon.com/Singularity-Journal%20Consciousness-Studies-ebook/dp/B01N1N6KLZ/ref=mt_kindle?_encoding=UTF8&me).

- Clark, A. (1997): *Being There: Putting Brain, Body and World Together Again*. MIT Press, Cambridge, MA.
- Clark, A. & Chalmers, D. J. (1998): „The extended mind.“ *Analysis* 58 (1): 7–19.
- Colzato, L. S., Hommel, B. & Beste, C. (2021): „The Downsides of Cognitive Enhancement.“ *The Neuroscientist* 27 (4): 322–330.
- Chalmers, D. J. (1996): *The Conscious Mind*. Oxford University Press, Oxford, New York.
- Chalmers, D. J. (2016a): „The Singularity: A Philosophical Analysis.“ In *The Singularity: Could artificial intelligence really out-think us (and would we want it to)?* [e-kniha], ed. U. Awret, Imprint Academic, Exeter, 2016 [cit. 15. 8. 2021]. Dostupné z: [https://www.amazon.com/Singularity-Journal%20Consciousness-Studies-ebook/dp/B01N1N6KLZ/ref=mt\\_kindle?\\_encoding=UTF8&me](https://www.amazon.com/Singularity-Journal%20Consciousness-Studies-ebook/dp/B01N1N6KLZ/ref=mt_kindle?_encoding=UTF8&me).
- Chalmers, D. J. (2016b): „The Singularity: A Reply to Commentators.“ In *The Singularity: Could artificial intelligence really out-think us (and would we want it to)?* [e-kniha], ed. U. Awret, Imprint Academic, Exeter, 2016 [cit. 15. 8. 2021]. Dostupné z: [https://www.amazon.com/Singularity-Journal%20Consciousness-Studies-ebook/dp/B01N1N6KLZ/ref=mt\\_kindle?\\_encoding=UTF8&me](https://www.amazon.com/Singularity-Journal%20Consciousness-Studies-ebook/dp/B01N1N6KLZ/ref=mt_kindle?_encoding=UTF8&me).
- Davidson, D. (1984): „Radical Interpretation.“ In *Inquiries into Truth and Interpretation*, D. Davidson, Oxford University Press, Oxford, 1984, s. 125–139.
- Davidson, D. (2004a): „Koherenční teorie pravdy a poznání.“ In *Subjektivita, Intersubjektivita, Objektivita*. D. Davidson; český překlad (J. Kolář a T. Marvan) *Filosofia*, Praha, 2004, s. 129–151.
- Davidson, D. (2004b): „Tři druhy poznání.“ In *Subjektivita, Intersubjektivita, Objektivita*. D. Davidson; český překlad (J. Kolář a T. Marvan) *Filosofia*, Praha, 2004, s. 194–209.

- Dennett, D. C. (1984): „Cognitive wheels: The frame problem of AI.“ In *Minds, Machines and Evolution: Philosophical Studies*, ed. C. Hookway, Cambridge University Press, Cambridge, s. 129–150.
- Dennett, D. C. (1987): *The Intentional Stance*. MIT Press, Cambridge, MA.
- Dennett, D. C. (2018): *From Bacteria to Bach and Back: The evolution of Minds*. Penguin Books, London.
- Dewey, D. (2011): „Learning what to value.“ In *Artificial General Intelligence: Proceedings of 4th International Conference*, eds. J. Schmidhuber, K. R. Thórisson & M. Looks, Springer, Berlin, s. 309–314. Dostupné z: <https://intelligence.org/files/LearningValue.pdf>.
- Dreyfus, H. L. (2012): „A History of First Step Fallacies.“ *Minds and Machines* 22 (2): 87–99.
- Dreyfus, H. L. (1992): *What Computers Still Can't Do*. MIT Press, Massachusetts.
- Dreyfus, H. L. (1967): „Why Computers Must Have Bodies in Order to Be Intelligent.“ *The Review of Metaphysics* 21 (1): 13–32.
- Fodor, J. A. (1983): *The Modularity of Mind*. MIT Press, Cambridge, MA.
- Good, J. I. (1965): „Speculations concerning the first ultraintelligent machine.“ In *Advances in Computers Volume 6*, eds. F. Alt & M. Rubinoff, Academic Press Inc., New York, s. 33–88.
- Hägström, O. (2018): „Challenges to the Omohundro-Bostrom framework for AI motivations.“ *Foresight* 21 (1): 153–166. Dostupné z: <http://www.math.chalmers.se/~olleh/ChallengesOBframeworkDeanonymized.pdf>.
- Hibbard, B. (2012): „Model-based utility functions.“ *Journal of Artificial General Intelligence* 3 (1): 1–24. arXiv:1111.3934, Dostupné z: <https://arxiv.org/ftp/arxiv/papers/1111/1111.3934.pdf>.
- Jackson, F. (1982): „Epiphenomenal Qualia.“ *Philosophical Quarterly* 32 (127): 127–132.
- Jackson, F. (1986): „What Mary Didn't Know.“ *Journal of Philosophy* 83 (5): 291–295.

- Kurzweil, R. (2005): *The Singularity Is Near. When Humans Transcend Biology*. Viking Adult, New York.
- Legg, S. (2008): *Machine Super Intelligence*. Univerzita Lugano, Lugano. Doktorská dizertace.
- Marcus, G. (2018): „Deep learning: A critical appraisal.“ In *ArXiv* [online]. 2. 1. 2018 [cit. 15. 8. 2021]. arXiv:1801.00631. Dostupné z: <https://arxiv.org/abs/1801.00631>.
- McCarthy, J. & Hayes, P. J. (1969): „Some Philosophical Problems from the Standpoint of Artificial Intelligence.“ In *Machine Intelligence 4*, eds. D. Michie & B. Meltzer, Edinburgh University Press, Edinburgh, s. 463–502.
- McDermott, D. (2016): „Response to ‚The Singularity‘ by David Chalmers.“ In *The Singularity: Could artificial intelligence really out-think us (and would we want it to)?* [e-kniha], ed. U. Awret, Imprint Academic, Exeter, 2016 [cit. 15. 8. 2021]. Dostupné z: [https://www.amazon.com/Singularity-Journal%20Consciousness-Studies-ebook/dp/B01N1N6KLZ/ref=mt\\_kindle?\\_encoding=UTF8&me](https://www.amazon.com/Singularity-Journal%20Consciousness-Studies-ebook/dp/B01N1N6KLZ/ref=mt_kindle?_encoding=UTF8&me).
- Minsky, M. (1968): „Machines Are More Than They Seem.“ *Science Journal* 4 (10): 3–43.
- Moravec, H. P. (1976): *The Role of Raw Power in Intelligence*. [online]. Nepublikovaný spis. 12. 5. 1976 [cit. 15. 8. 2021]. Dostupné z: <https://frc.ri.cmu.edu/~hpm/project.archive/general.articles/1975/Raw.Power.html>.
- Muehlhauser, L. & Helm, L. (2012): „The Singularity and Machine Ethics.“ In *Singularity Hypotheses. A Scientific and Philosophical assesment*, eds. H. A. Eden et al., Springer, Heidelberg, 2012, s. 101–125.
- Müller, V. C. & Bostrom, N. (2016): „Future progress in artificial intelligence: A survey of expert opinion.“ In *Fundamental Issues of Artificial Intelligence*, ed. V. C. Müller, Synthese Library, Berlin, s. 553–571.
- Nagel, T. (1974): „What Is It Like to be a Bat?“ *Philosophical Review* 83 (4): 435–450.

- Omohundro, S. M. (2016): „Autonomous technology and the greater human good.“ In *Risks of artificial intelligence*, ed. V. C. Muller, CRC Press, Boca Raton, s. 9–27.
- Omohundro, S. M. (2012): „Rational artificial intelligence for the greater good.“ In *Singularity Hypotheses. A Scientific and Philosophical assesment*, eds. H. A. Eden et al., Springer, Heidelberg, 2012, s. 161–176.
- Omohundro, S. M. (2008): „The basic AI drives.“ In *Artificial General Intelligence 2008: Proceedings of the First AGI Conference*, eds. P. Wang, B. Goertzel & S. Franklin, IOS, Amsterdam, s. 483–492.
- Pfeifer, R. & Bongard, J. (2006): *How the Body Shapes the Way We Think: A New View of Intelligence*. MIT Press, Cambridge, MA.
- Russel, S. J. & Norvig, P. (2020): *Artificial Intelligence: A Modern Approach*. Pearson, Boston.
- Sandberg, A. & Bostrom, N. (2008): *Whole Brain Emulation: A Roadmap* [online]. Future of Humanity Institute, University of Oxford, 2008 [cit. 15. 8. 2021]. Dostupné z: <https://www.fhi.ox.ac.uk/brain-emulation-roadmap-report.pdf>.
- Searle, J. R. (1980): „Minds, brains and programs.“ *The Behavioral and Brain Sciences* 3 (3): 417–424.
- Searle, J. R. (1984): *Minds, Brains and Science*. Harvard University Press, Cambridge, Massachusetts; český překlad (M. Nekula) *Mysl, mozek a věda*, Mladá fronta, Praha, 1994.
- Searle, J. R. (2014): „What Your Computer Can't Know.“ *The New York Review of Books* 9. října 2014.
- Simon, H. A. & Newell A. (1958): „Heuristic Problem Solving: The Next Advance in Operations Research.“ *Operations Research* 6 (1): 1–10.
- Sun, R. (2014): „Connectionism and neural networks.“ In *The Cambridge Handbook of Artificial Intelligence*, eds. K. Frankish & W. Ramsey, Cambridge University Press, Cambridge, 2014, s. 108–127.
- Totschnig, W. (2019): „The problem of superintelligence: Political, not technological.“ *AI & Society* 34 (4): 907–920.

- Ulam, S. (1958): „John Von Neumann 1903–1957.“ *Bulletin of the American Mathematical Society* 64 (3): 1–49.
- Varela, F. J., Thompson, E. & Rosch, E. (2016): *The Embodied Mind*. MIT Press, Cambridge, MA.
- Vinge, V. (1993): „The coming technological singularity. How to survive in the post-Human era.“ *Vision 21: Interdisciplinary Science and Engineering in the Era of Cyberspace* [online]. NASA, 1. 12. 1993 [cit. 15. 8. 2021]. Dostupné z: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19940022856.pdf>.
- Yudkowsky, E. (2008): „Artificial Intelligence as a positive and negative factor in global risk.“ In *Global Catastrophic Risks*, eds. N. Bostrom & M. Cirkovic, Oxford University Press, New York, s. 308–345.
- Yudkowsky, E. (2001): *Creating Friendly AI 1.0: The Analysis and Design of Benevolent Goal Architectures* [online]. The Singularity Institute, San Francisco, CA. [cit. 15. 8. 2021]. Dostupné z: <https://intelligence.org/files/CFAI.pdf>.
- Yudkowsky, E. (2004) : *Coherent Extrapolated Volition* [online]. The Singularity Institute, San Francisco, CA. [cit. 15. 8. 2021]. Dostupné z: <https://intelligence.org/files/CEV.pdf>.
- Yudkowsky, E. (2011): „Complex value systems in friendly AI.“ In Schmidhuber, T. & Looks, M. (2011): 388–393. Dostupné z: <https://intelligence.org/files/ComplexValues.pdf>.

## Abstract

### **Superintelligence and the control problem: Real problem or pseudo-problem?**

In this paper, I deal with the concept of SI (superintelligence) and the control problem. According to a group of AI theorists, we will soon experience an event that can change technological progress and human society. This event is the technological singularity associated with the emergence of the first greater than human intelligence. People like Nick Bostrom stress the SI's dangers and urge us to find methods to control this intelligence. According to Bostrom and others, the threat of SI stems from its nature. This article considers how SI can be created and judges the logic of the control problem. SI is possible only if we can create AI. For this reason, a section of the text concentrates on the arguments for the creation of AI. It is shown how Bostrom and others base their thesis on one problematic

argument and assumptions of their predecessors. Their position is subjected to the classical critique of artificial intelligence. I primarily focus my criticism on the claim that SI will have one final goal, which it will interpret. This statement is antithetical to the idea that SI will be a general intelligence. I conclude that the control problem confuses two other different “control” problems.

Key words: technological singularity, superintelligence, artificial intelligence, whole brain emulation, control problem, embodiment

Malík, J. (2021): „Superintelligence a problém kontroly: Skutečný problém nebo pseudo-problém?“ *Filosofie dnes* 13 (2): 73–118. Dostupné z [www.filosofiednes.ff.uhk.cz](http://www.filosofiednes.ff.uhk.cz).

---

# Objektivistická etika Ayn Randové

Miroslav Vacura

*Katedra filosofie,  
Národohospodářská fakulta, VŠE Praha  
Praha, nám. W. Churchilla 1938/4, 130 67  
vacuram@vse.cz*

Objektivistická filosofie Ayn Randové je i v současnosti myšlenkový směr, k němuž se někteří lidé aktivně hlásí a který zároveň představuje ideovou inspiraci některých směrů politických. Předkládaný text se zabývá kritickým zhodnocením konceptu ctnosti v kontextu objektivistické etiky, tzv. etiky egoismu. Nejprve se věnujeme tomu, jak sama Randová tyto předchozí etické koncepce hodnotí, a jejím vlastním základním východiskům – především principu sebezáchovy a racionálního lidského života. Další část textu se zabývá otázkou mezilidské interakce, ve které Randová odmítá jakékoli obětování a altruismus. Centrální roli hraje v autorčině koncepci produktivní práce, k tvůrčí práci jako ke svému základnímu těžišti se vztahuje také její nauka o ctnostech. Specifický problém pak představuje koncepce žádosti a racionálního zájmu, která, jak se domníváme, vede k určitým kontraintuitivním důsledkům. Text je uzavřen srovnáním s koncepcí R. Nozicka a stručným závěrem.

Klíčová slova: objektivismus, etika egoismu, Ayn Randová, ctnost

## 1. Úvod

Ayn Randová (1905–1982), rodným jménem Alisa Zinověvna Rozenbaum, se narodila v Rusku, kde také na Petrohradské státní univerzitě získala vzdělání se zaměřením na historii a seznámila se s filosofií.<sup>1</sup> Většinu svého života ovšem prožila ve Spojených státech amerických, kam její rodina uprchla po revoluci v Rusku, v důsledku čehož se levicové

---

<sup>1</sup> Heller (2010).



myšlení stalo po zbytek jejího života hlavním terčem její kritiky. V USA se proslavila především jako autorka románů – významné jsou zejména *Zdroj*<sup>2</sup> a *Atlasova vzpoura*<sup>3</sup>, v nichž vyjádřila své hlavní filosofické postoje, které pak rozpracovala do filosofického systému, tzv. *objektivismu*.<sup>4</sup> Ten zahrnuje jak metafyzickou a epistemologickou část, tak i část etickou, a je možné jej chápat jako pokus vytvořit radikální protiváhu marxistické filosofii.

Sám název objektivismus je odvozován od pojmu „objektivní“, který hraje v objektivismu centrální roli. Z metafyzického hlediska odkazuje k přesvědčení, že objektivní „realita existuje nezávisle na vědomí jakéhokoli pozorovatele“.<sup>5</sup> Z epistemologického hlediska tento pojem odkazuje k přesvědčení, že se poznání získává skrze vnímání s důrazem na zapojení rozumu a logiky. Z toho pak dle Randové plyne, že „ačkoli je realita neměnná a v jakémkoli kontextu je pravdivá vždy jen jedna odpověď, pravda není automaticky dostupná lidskému vědomí, ale může být získána pouze určitými mentálními procesy“.<sup>6</sup> Na každou otázku tedy existuje jedna objektivně pravdivá odpověď, nicméně nemusí být snadné ji odhalit. Těmto metafyzickým a epistemologickým aspektům objektivismu i z nich vyplývající politické filosofii byla už i v českém prostředí věnována určitá pozornost, zájemci o tuto tematiku doporučujeme existující studie<sup>7</sup>, etika nicméně stála v diskusích stranou zájmu.

Tento stručný text se proto věnuje kritickému zhodnocení teorie ctností v rámci objektivistické etiky, jak ji v sérii esejů a statí Randová vypracovala. Většina z relevantních textů k tomuto tématu byla vydána ve sborníku Randové nazvaném *Ctnost sobectví: nové pojetí egoismu*<sup>8</sup>.

---

2 Původní vydání: Rand (1943), český překlad: Randová (2000).

3 Původní vydání: Rand (1957), český překlad: Randová (2014), slovenský překlad: Randová (2003). Zfilmováno jako trilogie – režie Paul Johansson (2011), John Putch (2012), J. James Manera (2014).

4 Vedle Randové je zastáncem objektivismu především její žák Leonard Peikoff, autor objektivistického spisu (Peikoff 1991); český překlad: Peikoff (2001). Překladaťel tohoto textu, Jiří Kinkor, je také nejvýznamnějším propagátorem objektivismu v Čechách a na Slovensku a této filosofii se věnoval ve svém spise (Kinkor 1996).

5 Rand, Peikoff & Schwartz (1990, s. 18).

6 Tamtéž.

7 Pavlík (1997b), Pavlík (1997a), Kinkor (1997), Novotný (1999).

8 Randová (1964b).

V nedávné době byl publikován český překlad tohoto díla<sup>9</sup>, což bylo dalším impulsem se tomuto tématu věnovat. Zmíněný sborník obsahuje i několik doplňkových statí Nathaniela Brandena, kterými se ovšem v tomto textu zabývat nebudeme.

Zbytek článku je strukturován následovně: nejprve shrneme, jak Randová hodnotí předchozí etické koncepce a jak v historii etického bádání lokalizuje svoji teorii a její význam. Následující část se zabývá základními východisky objektivistické etiky, principem sebezáchovy a problémem role rozumu. Čtvrtý oddíl se věnuje otázce mezilidské interakce, zejména tomu, jakou roli v etice sobectví mají láska a přátelství. Pátý oddíl navazuje rozbořem nauky o ctnostech a neřestech a šestý oddíl se věnuje analýze žádosti a vysvětlení pojmu racionálního zájmu. Text je ukončen stručným závěrem a celkovým zhodnocením objektivistické etiky.

## 2. Randové hodnocení historie etiky

Randová je přesvědčená, že jakožto lidské bytosti nemáme „na výběr, zda potřebujeme filosofii“ – jedinou volbou je, zdali svou filosofii budujeme vědomě a racionálně, anebo zda necháme naše podvědomí akumulovat soubor falešných generalizací, chybných závěrů a důsledků pochybností a strachů. Racionální budování vlastní životní filosofie předpokládá seznámení se s historií filosofie a kritickou reflexi předchozích filosofických postojů.<sup>10</sup> Nejvhodnějším způsobem studia historie filosofie je podle Randové přístup jako k detektivní zápletce – sledování každé stopy, náznaku a implikace. Základními otázkami jsou „Proč?“ a „Jak?“.<sup>11</sup> Výsledkem studia historie filosofie<sup>12</sup> je u Randové identifikace dvou klíčových autorů: Aristotela a Kanta. Aristotelés je „největší ze všech

---

9 Randová (2018a).

10 Randová (1982, s. 7).

11 Tamtéž, s. 10 a 16.

12 Víme, že Randová četla primární texty Platóna, Aristotela, Kanta a Nietzscheho. Ovšem u dalších autorů často vycházela z dobových přehledů, jako např. z Fullerova přehledu *A History of Philosophy: Ancient, Medieval and Modern* (Henry Holt & Co, New York, 1938), který četla ve čtyřicátých letech. Lennox (2016, s. 322).

filosofů<sup>13</sup> a je prezentován jako prapůvodce většiny pozitivních společenských trendů v dějinách lidstva: „u kořene každého civilizačního úspěchu, jako například vědy, technologie, pokroku, svobody, u kořene každé hodnoty, kterou dnes máme, včetně zrodu této země [USA] – najdete úspěch jediného muže, který žil před dvěma tisíci lety: Aristotela“.<sup>14</sup>

Kant je oproti tomu prezentován jako hlavní zdroj degenerativních společenských trendů a kulminační bod antiracionalistického myšlení: „Již asi dvě stě let směřuje dominantní trend ve filosofii pod vlivem Immanuela Kanta k jedinému cíli: zničení mysli člověka a jeho důvěry v moc rozumu.“<sup>15</sup> Do takto narýsovaného rozvrhu jsou zasazeni ostatní velcí myslitelé: například Platón je chápán jako zakladatel „mysticismu“ a předchůdce Kanta. Akvinský je naopak důstojným pokračovatelem Aristotela, zejména co se týče systematičnosti a konzistence filosofie. Většina novověkých filosofů je Randovou řazena do „mystické“ tradice, počínaje Descartem, jehož základní chybou bylo začít jistotou našeho sebevědomí a teprve z ní postupovat k vnějšímu světu. Stejně tak klasičtí empiristé Locke, Berkeley a Hume vycházejí z jistoty našich percepčí, což je dle Randové opět omyl. Podle ní je ve filosofických systémech těchto autorů implicitně obsažena chybná metafyzická teze, kterou nazývá princip „priority vědomí“.<sup>16</sup>

Filosofii devatenáctého a zejména dvacátého století Randová považovala za prostoupenou kantovskými motivy, a to včetně logického pozitivismu, který považovala za „odnož kantovství“.<sup>17</sup> Randová k tomu napsala: „Dominantní intelektuální vliv je dnes stále Kantův (...), všechny přední moderní filosofické školy jsou odvozeny od kantovské základny.“<sup>18</sup> Výsledkem tohoto pohledu na historii filosofie je, že Randová odmítla většinu předchozích etických přístupů (kromě Aristotelova a na něj přímo navazujících<sup>19</sup>).

13 Randová (2018e, s. 23).

14 Rand (1982, s. 10).

15 Tamtéž, s. 8–9.

16 Lennox (2016, s. 330).

17 Rand (1982, s. 108).

18 Randová (1961, s. 29).

19 V románu *Atlasova vzpoura* dokonce najdeme řadu odkazů na Aristotelovo dílo (např. názvy částí a kapitol).

Randová při svém výkladu často odkazuje na svá románová díla, jejichž postavy v rozsáhlých monologích deklarují základní principy její filosofie, a ve kterých je určitou ilustrací těchto principů i samotný děj. I v kontextu zkoumání předchozích dějin etiky se Randová odvolává na svůj román *Atlasova vzpoura*, ve kterém hlavní hrdina John Galt říká, že kulturu po staletí ovládal určitý *morální kód*, ten však vedl ke všem katastrofám lidstva, zcela selhal, je překonaný a má být nahrazen kódem novým, který je jím (resp. Randovou) v tomto románu předkládán.<sup>20</sup> Na jiném místě Randová objasňuje, co tímto morálním kódem myslí: „Stejně jako v minulosti se i dnes většina filosofů shoduje v tom, že konečným měřítkem etiky je *rozmar* [*whim*] (nazývají ho *arbitrárním postulátem*, *subjektivní volbou* nebo *emocionálním závazkem*).“<sup>21</sup> Termín *rozmar* je přitom definován jako „touha, kterou prožívá osoba, jež nezná a nechce zjistit její příčinu“.<sup>22</sup>

Randová toto tvrzení patrně nemyslí tak, že by sami tito filosofové předkládali jako standard etiky doslova individuální rozmar, subjektivní volbu či nějaký arbitrárně určený postulát. Tito myslitelé ve svých dílech vymezili různé etické principy a standardy vycházející z různých typů argumentace. To, že všechny tyto etické standardy a principy jsou svou podstatou vlastně jen arbitrárně určené na základě rozmaru či subjektivní volby, je tvrzení, které je platné jen v případě, že se na ně díváme optikou filosofie Randové. Dále uvedeme, jak autorka tuto interpretaci předchozích etických koncepcí na dalších místech zdůvodňuje.

Randová rozlišuje tři historické etické školy: *mystickou*, *sociální* a *subjektivní*.<sup>23</sup> „Mystici“, jak jsou označováni zejména křesťanští filosofové, zakládají dle Randové<sup>24</sup> svou etickou koncepci na principu, že měřítkem dobra je arbitrární a nevysvětlitelná vůle Boží,<sup>25</sup> a že „hodnotové měřítko etiky je umístěno do posmrtného života, a to prostřednictvím

20 Randová (2018e, s. 21). Randová zde zároveň tvrdí, že John Galt je sám nejlepším představitelem objektivistické etiky.

21 Tamtéž, s. 25.

22 Tamtéž, s. 23.

23 Tamtéž, s. 62.

24 Tamtéž, s. 24, 63.

25 Srov. Salmieri (2016c, s. 65).

požadavků jiné, nadpřirozené dimenze“.<sup>26</sup> Randová bez delší diskuse takový princip odmítá – jakkoli pojatá idea Boha je v rozporu s objektivistickou metafyzikou, která je svou povahou ateistická. Protože neexistuje jakýkoli objektivní základ křesťanské etiky, je tento princip „arbitrárním postulátem“. Na tomto příkladu se zároveň ukazuje, že jako „arbitrární“ jej nechápou ani nezavádí sami autoři této koncepce, arbitrárním je jen posuzován optikou Randové.

Konkrétní jmenovitou kritiku, shrnutou ovšem jen do jedné věty, věnuje Randová Aristotelovi. Ten dle ní založil svou etiku na pozorování ctnostných a moudrých mužů, ale nezodpověděl otázku, *proč* jednají tak, jak jednají, a proč právě je považoval za ctnostné a moudré.<sup>27</sup>

Takové pojetí Aristotelovy filosofie není zcela přesné. Aristotelés založil svou etiku na koncepci štěstí čili blaženosti (*eudaimonia*) jako cíle lidského života.<sup>28</sup> Ctnost je pak takový stav charakteru člověka, který odpovídá rozumem určenému středu mezi krajnostmi, přičemž jednání v souladu s ctností vede ke šťastnému životu.<sup>29</sup> Nicméně dále uvedeme, že Randová do svého systému určitý soubor ctností, které vedou ke (v jejím pojetí) šťastnému životu, také začleňuje a také v jiných ohledech z Aristotelovy etiky čerpá. Salmieri říká, že ačkoli Randová nepovažovala Aristotelovu filosofii zdaleka za dokonalou, domnívala se, že obsahuje základy racionální epistemologie a metafyziky, na nichž je vypracována i její vlastní filosofie.<sup>30</sup>

Zapředníhopředstavitele „subjektivistické“ etiky (tzv. psychologického egoistu) pokládá Randová zejména Nietzscheho,<sup>31</sup> jenž dle její interpretace určuje za hlavní etický princip „sobecké“ sledování vlastních rozmarů“. Dále do tohoto směru řadí filosofy, které považuje za příznivce různých forem individualistického hédonismu či utilitarismu, kladoucí za etický účel lidského jednání individuální požitek či uspokojení individuálních potřeb, tedy kteří deklarují, že „*správnou hodnotou je vše, z čeho máte*

26 Randová (2018e, s. 63).

27 Tamtéž, s. 23.

28 Aristoteles (2013, 1097a35).

29 Tamtéž, 1098a16.

30 Salmieri (2016a, s. 10). Detailní srovnání etických systémů Randové a Aristotela viz Wheeler (1986).

31 Randová (2018e, s. 54), srov. Hunt (2016, s. 343).

*dobrý pocit*“.<sup>32</sup> Mezi takové autory řadí Epikura, ale i Maxe Stirnera a Sigmunda Freuda.<sup>33</sup>

Nejrozsáhlejší kritiku Randová věnovala příznivcům „sociální“ etické školy, pokračovatelům „mystiků“ – proto také někdy označuje tyto autory jako „neomystiky“. Tato skupina zahrnuje myslitele, kteří se snažili od náboženství oprostit, nicméně podle ní pouze nahradili Boha konceptem společnosti a etiku založili na „dobru společnosti“. Tato etika definuje *dobro* jako to, co chce společnost, jako cokoli, co společnosti zajistí blahobyt a dostatek požitků.<sup>34</sup> Randová usuzuje, že dle této koncepce by se člověk neměl řídit svými vlastními rozmary jako v případě školy „subjektivistické“, ale rozmary ostatních; jedná se tedy o školu „sociální“. Do této filosofické školy pak spadá většina myslitelů tradičně řazených mezi utilitaristy, Randová jmenuje Benthama, Milla a Comta, které označuje za „sociální hédonisty“.<sup>35</sup>

Ukazuje se, že Randová při kritice těchto přístupů interpretuje utilitaristické pojetí užitku jako prosté uspokojování nejnižších smyslových žádostí, ať už individuálních (pokud berou ohled jen na sebe) v případě „subjektivní“ školy etiky, nebo žádostí ostatních (pokud berou ohled i na druhé) v případě „sociální“ školy etiky. Domníváme se, že takový výklad je určitým zjednodušením řady utilitaristických pozic. Už Mill dává do kontrastu fyzické a intelektuální požitky, přičemž ty druhé považuje za nadřazené.<sup>36</sup> Randová oproti tomu tvrdí, že pro zastánce utilitarismu jsou všechny žádosti eticky rovnocenné.<sup>37</sup>

„Sociální“ etiku pak Randová ve své době považuje za celosvětovou koncepci a chápe ji také jako etický systém založený na standardu rozmaru a arbitrárního postulátu. Pro její rozšířenost jí věnuje detailnější a argumentačně propracovanější kritiku než koncepcím jiným.

Z výše uvedeného vymezení dobra, které taková etika podle Randové přijímá, plyne naprostý morální relativismus: společnost může dělat

---

32 Randová (2018e, s. 54).

33 Srov. Salmieri (2016b, s. 133).

34 Randová (2018e, s. 24).

35 Tamtéž, s. 54.

36 Diskuse tohoto problému viz Schmidt-Petri (2003) a Riley (2003).

37 Randová (2018e, s. 55). Burgess-Jackson (2013) ovšem ukazuje, že zastánci egoistické etiky mají proti utilitarismu k dispozici podstatně širší spektrum argumentů.

cokoli, co se jí zachce, co jí zajistí blahobyt a dostatek požitků, neboť z definice je to vždy dobré. Randová takové pojetí kritizuje konstatováním, že společnost není žádnou entitou, je to pouze množství jednotlivých konkrétních jednotlivců. Ve skutečnosti tak nic takového jako vůle celku společnosti nemůže existovat. Proto se za vůli celé společnosti vydává ve skutečnosti vždy vůle určitých jednotlivců – může se jednat o většinu, nebo také o skupinu (*gang*), která má ve společnosti privilegované postavení. Zájem této skupiny je jakožto „vůle společnosti“ povýšen na „dobro“ a ostatním členům společnosti je práce ve prospěch tohoto zájmu předložena jako „morální povinnost“.<sup>38</sup>

Randová připisuje všem výše zmíněným koncepcím přesvědčení, že (řeceno současným jazykem) vztahy lidí jsou hrou s nulovým součtem. Tedy, že uspokojení žádostí jednoho člověka implikuje neuspokojení (obětování) žádostí druhého.<sup>39</sup> Důsledkem platnosti principu nulového součtu je, že když je žádost jednoho člověka uspokojena, žádost druhého je nutně obětována.<sup>40</sup> Všechny výše uvedené etické systémy se pak podle Randové tudíž staví na stranu *altruismu*, který podle ní spočívá v tezi, že eticky hodnotné je ve všech situacích obětování své vlastní žádosti ve prospěch naplnění žádosti druhého. Jakákoli taková etika je však z jejího pohledu etikou sebeobětování, tedy sebedestrukce.

Randová při popisu své vlastní koncepce vychází z toho, že správnou alternativou altruismu a sebeobětování pro druhé je obchod, neboť při svobodné směně profitují z výsledku oba účastníci, tudíž se o hru s nulovým součtem nejedná. Proto je také svobodně uskutečněný směnný *obchod* jediným paradigmatickým principem racionální interakce mezi lidmi (viz dále).

Otázkou zde zůstává, nakolik je kritika Randové směřovaná k altruismu oprávněná. Řada historických etických koncepcí nepovažuje ani lidské interakce, které nejsou směnou, za hru s nulovým součtem. Už Platón chápe morální altruistické jednání jako přínosné i pro jednajícího

38 Randová (2018e, s. 24).

39 Tamtéž, s. 55–56.

40 Přitom už jeden z klasiků sentimentalismu – Adam Smith – zavádí *princip neviditelné ruky trhu*, který říká, že sledování soukromého prospěchu jedním člověkem vede nezamýšleně v určitých kontextech (například v kontextu volného trhu) i k prospěchu ostatních lidí. Smith (2001, IV, s. 2) a Smith (2005, IV, s. 1).

a amorální jednání jako škodící zároveň jednajícím. Známa je jeho teze, že bezpráví je lépe trpět než konat.<sup>41</sup> Ta ovšem nevychází z vyzdvihování sebeobětování poškozeného, ale naopak stojí na přesvědčení, že konaná nespravedlnost škodí daleko více jednajícím než poškozenému.<sup>42</sup> Morální jednání je naopak prospěchem i pro duši toho, kdo jej koná.

Z tohoto shrnutí je patrné, že popis dějin etiky, jak jej předkládá Randová, jeví známky určitého zjednodušení, což je částečně dáno i malým rozsahem, který tomuto rozboru věnuje. Většinu autorů, jež Randová zmiňuje, zařazuje do pejorativně chápané kategorie etiky sebeobětování a sebestrukce, následně se proti takto charakterizovaným pozicím vymezuje a názory těchto autorů podrobuje kritice. Z následujícího oddílu bude patrné, že řada tezí, které předkládá jako součást své filosofické koncepce, vychází ve svém jádru z etických systémů jí kritizovaných historických myslitelů.

### 3. Východiska etiky sobectví

Navážme nyní na výše vyložený pohled Randové na historii etiky a soustředíme se na její vlastní etickou koncepci. Randová svou etickou teorii považuje za racionální, vědeckou a objektivní (v duchu výše zmíněné teze, že na každou otázku existuje jedna objektivně správná odpověď), navíc přinášející příslib konečného nastolení fungující a spravedlivé společnosti. Etika je podle Randové věda, která se „zabývá nalezením a definováním (...) systému hodnot a pravidel“.<sup>43</sup> Pojem hodnoty je přitom založen na „*metafyzickém* faktu“, na „neměnné podmínce lidské existence“.<sup>44</sup> Jejím cílem je konečně poskytnout na otázku lidských hodnot „racionální, objektivně dokazatelnou, vědeckou“ odpověď.<sup>45</sup>

Randová následně definuje hodnotu jako „to, co chce člověk svým jednáním získat a/nebo uchovat“.<sup>46</sup> Každá hodnota je pak výlučně hod-

41 Platón (2000, 475e, 508e).

42 Diskuse viz MacKenzie (1982).

43 Randová (2018e, s. 22).

44 Tamtéž, s. 22.

45 Tamtéž, s. 23.

46 Randová (2018e, s. 25). Srov. „Každé (...) praktické jednání a záměr směřuje, jak se zdá, k nějakému dobru (...), dobro jest to, k čemu všechno směřuje.“ (Aristotelés 2013, 1094a3).



notou *pro někoho* a předpokládá existenci alternativ.<sup>47</sup> Hodnota o sobě, nezávislá na konkrétním člověku, či něco takového jako obecné dobro, nevztahené ke konkrétnímu lidskému životu, není.<sup>48</sup>

Základní alternativou pro každou živou bytost je život, či smrt.<sup>49</sup> Život je procesem sebezáchovy a sebetvorby, pokud v tomto *jednání* organismus selže, zemře. Randová tak zakládá pojem hodnoty na sebezáchově a tvrdí: „Pojem hodnota je možný pouze díky pojmu život.“<sup>50</sup> Jedině pro živou bytost mohou být věci dobré, či zlé. Co přispívá k udržení života, je dáno *přirozeností* konkrétní živé bytosti, tj. jejím *druhem*. Život je tak podle Randové nejvyšší hodnotou, ke které jsou všechny ostatní hodnoty vztaheny a ve vztahu k níž jsou pak ohodnocovány.<sup>51</sup> V základu této etiky tak stojí hierarchická konstrukce hodnot připomínající hierarchické uspořádání dober v etice Aristotelově<sup>52</sup>, přičemž specifickou roli hraje – podobně jako u Hobbese<sup>53</sup> – sebezáchova:

„*Konečnou* hodnotou je nejvyšší cíl, k němuž jsou všechny nižší cíle prostředkem a který stanovuje měřítko, jímž se všechny nižší cíle *hodnotí*. *Hodnotovým měřítkem* je vlastní život organismu: to, co jeho životu prospívá, je *dobrem*, to, co jej ohrožuje, je *zlem*.“<sup>54</sup>

Randová se domnívá, že pojem sebezáchovy a pojem hodnoty jsou analyticky spojeny. Aby své přesvědčení doložila, uvádí příklad nesmrtelného, nezničitelného a nezranitelného robota, který se pohybuje a jedná, ale nemůže být ničím ovlivněn. O něm pak tvrdí, že nebude mít žádné hodnoty, zájmy či cíle, nebude moci nic získat ani ztratit.<sup>55</sup>

---

Podobně i Hobbes (2015, s. 40) říká: „Cokoli je však předmětem žádosti nebo touhy jakéhokoli člověka, tomu on říká *dobré*.“

47 Randová (2018e, s. 25–26).

48 Srov. „Tato slova – dobrý, zlý, lhostejný – se vždy užívají se zřetelem k osobě, která jich užívá.“ Hobbes (2015, 40).

49 Srov. Gotthelf (2000, s. 81).

50 Randová (2018e, s. 26).

51 Tamtéž.

52 Aristotelés (2013, 1097a27).

53 Hobbes (2015, s. 91). Podobnost etických koncepcí Hobbese s Randovou zmiňují i Badhwar a Long (2020), upozorňují nicméně, že Randová nedochází ve svém řešení ke konceptu společenské smlouvy. Teorie obou těchto autorů bývají řazeny mezi koncepce etického egoismu, proto je běžné je srovnávat.

54 Randová (2018e, s. 28).

55 Tamtéž, s. 26.

Hédonista či utilitarista by na tomto místě mohl položit otázku, zdali takový robot bude schopen prožívat požitky a strasti.<sup>56</sup> Pokud by toho schopen byl, pak samotný fakt nesmrtnosti nebude znamenat, že by neměl zájmy a cíle. Pokud by nebyl schopen těchto prožitků, pak je to ovšem další, nezávislý předpoklad vedle nesmrtnosti, což Randová nerozlišuje. Navíc jsou myslitelné i jiné účely jednání, nezávislé nejen na smrtelnosti, ale i na požitcích a strastech. Jako příklad uveďme Aristotelovo přesvědčení, že lidé od přirozenosti touží po vědění.<sup>57</sup> Pokud by robot Randové měl nedokonalé poznání (nebyl by vševědoucí) a zároveň by byl schopen své poznání rozšiřovat, *mohl by* podobně jako člověk toužit po vědění, tudíž by pak měl zájmy a cíle, mohl by cosi získat a jeho jednání by mohlo být úspěšné či neúspěšné, byl by tedy i schopen hodnotit. To, co v jeho úsilí o poznávání bude prospěšné, pak bude mít pro něj hodnotu, bude dobrem, a to, co bude v tomto úsilí překážkou, bude zlem, měl by tedy hodnoty.<sup>58</sup>

Problém zde spočívá v tom, že podle východiska Randové je zde jen jediná věc, již chceme pouze pro ni samou, což je sebezáchova: „Metafyzicky je *život* jediným jevem, který je cílem sám o sobě.“<sup>59</sup> Toto chtění sebezáchovy je u člověka (na rozdíl od ostatních živých bytostí) podle Randové svobodným rozhodnutím – většina lidí jedná z velké části sebedestruktivně, mezi což řadí Randová především velkou část altruismu a obětování se pro druhé. Ostatní hodnoty jsou hodnotami jen jako prostředky k nejvyšší hodnotě – sebezáchově. Randová se tudíž nijak nevyrovnává s klasickým Aristotelovým argumentem, že lidé ze své přirozenosti chtějí řadu dalších účelů, tedy hodnot, pro ně samé, tudíž nikoli jen jako prostředky k něčemu dalšímu – příkladem je výše uvedené vědění. U Randové je však vědění jen prostředkem pro sebezáchovu, ostatní hodnoty tradičně

56 Srov. Mill (2016).

57 „Všichni lidé od přirozenosti touží po vědění. Známkou toho je záliba v smyslových vjemech, v nichž si libují pro ně samy, aniž hledí k jejich potřebě.“ Aristotelés (2021, 980a21).

58 Netvrdíme zde, že z absence vševědoucnosti plyne nutně touha po vědění. Pouze to, že pokud robot není vševědoucí, pak *může* toužit po vědění. To samo dle našeho názoru stačí k vyvrácení teze, že nezničitelnost robota *implikuje*, že nebude mít žádné hodnoty. Je-li myslitelný robot, který je nezničitelný a přitom hodnoty má, pak tato implikace neplatí.

59 Randová (2018e, s. 28). Oproti českému překladu doplňujeme zvýraznění slova „život“ tak, jak je v originále (Rand 1964a, 18).

považované za nezávislé a vyhledávané pro ně samotné (např. krása<sup>60</sup>) by zřejmě zařadila mezi rozmary.

V kontextu požadavku sebezáchovy lze namítnout, že jsme bytosti smrtelné a víme, že doba našeho života zcela jistě nepřesáhne 120 let. Cíl sebezáchovy je tedy v konečném důsledku nedosažitelný. Je přitom zřejmé, že kvalita života v letech blížících se k této hranici postupně upadá. Kdybychom přijali toto východisko koncepce Randové, zdálo by se, že reálně je nejvyšším cílem života maximální prodloužení délky dožití k této hranici a všechny ostatní hodnoty jsou tomuto cíli podřízeny. Zároveň víme, že délka života, pokud pomineme násilnou smrt (v mírové době poměrně výjimečnou), je určena především genetickými faktory, životosprávou a kvalitou lékařské péče. Očekávali bychom tudíž, že kladní hrdinové románů Ayn Randové, kteří ztělesňují ideál její filosofie, budou věnovat mimořádnou péči sestavování jídelníčků, konzumaci vitamínových doplňků stravy a volný čas budou trávit ve fitness centrech. Tento aspekt prodloužování a zachování života však Randová netematizuje.

Co tedy přežití ovlivňuje podle Randové? Život organismu podle ní závisí na dvou faktorech: materiálu, tj. „vyživující látce“ (*fuel*) z vnějšího prostředí, a správném použití této „vyživující látky“.<sup>61</sup> Živočichové disponují vrozenou schopností cítit prožitky slasti a bolesti, což je automatický mechanismus usměrňující jejich chování. Člověku tyto prožitky umožňují odhalit pojem hodnoty: „Jak lidská bytost objevuje pojem *hodnota*? (...) Prostřednictvím tělesných počitků *slasti* nebo *bolesti*.“<sup>62</sup> Základním prostředkem přežití je pak vědomí (*consciousness*). Jeho nižšími formami, nerozumovými, spočívajícími ve smyslovém vnímání, disponují všichni živočichové, komplexnější formou, zahrnující paměť, jen vyšší živočichové, nejvyšší formou, vědomím zahrnujícím rozum, jen člověk.<sup>63</sup>

Lidské vědomí je *volní*, což znamená, že zahrnuje vůli založenou na pojmovém poznání. *Pojem* je dle Randové „mentální integrací dvou

60 Viz např. Platónovo pojetí krásy: Platón (2016, 841c), Platón (2012, 66a–b, 401c), Platón (2018, 201c, 205e).

61 Randová (2018e, s. 27).

62 Tamtéž, s. 29.

63 Toto rozlišení přibližně odpovídá Aristotelově dělení duše na tři části: 1) duše vegetativní (*threptiké*), kterou disponují všechny živá jsoucna, jež se projevuje výživou, růstem a rozmnožováním, 2) duše vnímavá (*aisthétiké*), kterou mají všichni živočichové, ale nikoli rostliny, 3) duše myslící, rozumová (*dianoétiké*), kterou disponují pouze lidé (Aristotelés 2000, 413a23).

nebo více vnímaných jednotlivostí (entit, atributů apod.) izolovaných procesem *abstrakce* a propojených prostřednictvím specifické definice“. <sup>64</sup> Tento proces není automatický, jedná se o vědomou činnost, již Randová nazývá *konceptualizace*. Tu pak dále dává do souvislosti s činností spočívající v tvoření otázek a tvrzení a v konstruování dedukcí, což označuje jako *myšlení*. <sup>65</sup>

Myšlení je činností našeho *rozumu*, přičemž Randová zdůrazňuje, že není automatické a je věcí *volby*. Nepovažuje myšlení za nepřetržitý proud verbálních myšlenek, tichý monolog, ani za proud vědomí, probíhající automaticky a bez přestání doprovázející náš bdělý život. Myšlení je naopak pro Randovou činnost vyžadující „stav plného soustředěného vědomí“, <sup>66</sup> jedná se o „plné, aktivní, cíleně orientované uvědomování si reality“. <sup>67</sup> Pokud se člověk nenachází v tomto stavu, znamená to nechat se „unášet v polovědomém omámení, reagovat jen na všechny náhodné podněty bezprostředního okamžiku, vydat se na pospas svému neřízenému počítkově-vjemovému mechanismu a libovolným náhodným, asociativním spojením, která tento mechanismus může vytvořit“. <sup>68</sup> Pokud člověk nemyslí, nachází se v „podlidském“ stavu vědomí.

Rozum je také hlavním prostředkem sebezáchovy člověka, kterému na rozdíl od zvířat k přežití nestačí řídit se instinkty. Používat rozum je volbou, ovšem rozhodne-li se člověk rozum nepoužívat, důsledkem je smrt. Etika je tak pro Randovou primárně naukou o přežití. <sup>69</sup> Vzhledem k tomu, že kritériem její „úspěšnosti“ je přežití člověka, jedná se o nauku objektivní. Nicméně jak jsme ukázali výše, obsahově etika Randové tomuto svému vymezení neodpovídá a její doporučení mají na délku dožití poměrně malý vliv.

Další otázkou je status rozumnosti jako charakteristické vlastnosti člověka. Randová zpočátku postupuje *deskriptivně* a popisuje charakteristiky spojené s rozumností, které člověka odlišují od zvířat. V určitém

64 Randová (2018e, s. 34).

65 Tamtéž, s. 35. Z textu není patrné, jestli Randová rozlišuje *konceptualizaci* a *myšlení*. Nicméně pro další postup etických argumentů toto není podstatné.

66 Randová (2018a, s. 36).

67 Tamtéž.

68 Tamtéž.

69 Randová (2018e, s. 39).

bodě se tento deskriptivní postup posune do *normativní* polohy – rozum není jen něco, co člověku fakticky připisujeme, ale je to něco, čemu musí člověk svým životem dostat. Život člověka *qua* rozumné bytosti se stává normativním kritériem, morálním požadavkem, přestože je otázkou, jaký má vztah k přežití a sebezáchově na biologické úrovni. V tomto bodě dochází k přechodu od faktického popisu k normativním tvrzením – což je instance klasického *is-ought* problému proslaveného Hudem.<sup>70</sup> Randová postupuje tak, že deskriptivní popis rozumnosti člověka spojuje s normativním principem sebezáchovy a pokouší se z této kombinace odvodit další normativní nároky.<sup>71</sup>

Pokud tedy Randová na počátku říká, že vlastní život každého organismu je jeho *standardem hodnoty*, a byl tím míněn prostý požadavek sebezáchovy, dochází zde k významovému posunu – pokud je člověk specifický svým rozumem, pak konečným cílem pro člověka není pouhý život, ale život rozumné bytosti: „Protože rozum je člověku základním prostředkem přežití, je dobrem to, co je vhodné pro život *rozumové bytosti*; to, co život popírá, co mu protiče nebo co ho ničí, je zlem.“<sup>72</sup>

Randová tímto vymezením tvrdí dvě věci. První je, že člověk používá rozum jako nástroj svého přežití, přičemž používat k sebezáchově prostředky zvířat by bylo méně efektivní. To je, myslíme, obtížně zpochybnitelné. Zároveň z této první teze odvozuje druhou: aby člověk mohl používat rozum jako efektivní nástroj svého přežití, musí žít *život vlastní racionální bytosti*.<sup>73</sup> Tato druhá teze směřuje k etickému konceptu člověka jako rozumné bytosti.<sup>74</sup> Pokud je rozum *prostředkem* přežití, pak by z toho bylo možné odvodit, že by jej člověk měl zdokonalovat či tré-

70 Hume (1739, s. 335).

71 Někteří objektivisté tvrdí (např. Kelley & Thomas 2021), že Randová *is-ought* problém vyřešila. Nicméně jak ukazuje např. Badhwar a Long (2020), není nijak vysvětleno, proč bychom měli přijmout zachování života za svůj normativní cíl (*ought*), a pokud přijímáme sebezáchovu jako normativní východisko, pak už zbývající odvození mají charakter *ought-ought*. Pokud naopak sebezáchovu nechápeme jako svůj normativní cíl, pak celá z ní odvozená etická koncepce má charakter (kantovskou terminologií) hypotetického imperativu: *pokud* chcete zachovat svůj život, *pak* je třeba jednat takto. To ovšem opět není vyřešení *is-ought* problému.

72 Randová (2018e, s. 40–41), zvýraznění autor stati.

73 Tamtéž, s. 43.

74 Opět zde můžeme najít inspiraci Aristotelem a jeho pojetím lidského života jako života rozumného „(...) úkolem člověka jest skutečná činnost duše ve shodě s rozumem anebo nikoli bez rozumu.“ Aristotelés (2013, 1098a7).

novat, co do jeho efektivity mu zajistit přežití. Nicméně jak bude patrné dále, Randová z tohoto předpokladu odvozuje i řadu dalších požadavků na život člověka a ne vždy je jasné, jakým postupem ke svým závěrům dospívá. Badhwar a Long ukazují, že příznivci objektivismu se neshodnou, jak místy ambivalentní vyjádření Randové k tomuto tématu vlastně interpretovat, a zastávají tři různá pojetí konečného cíle člověka: 1) samotné přežití, 2) přežití člověka jakožto rozumné bytosti, 3) aristotelsky chápané štěstí (*eudaimonia*).<sup>75</sup>

První interpretace chápe vše ostatní (racionalitu, štěstí) jako pouhé nástroje k dosažení prostého přežití. Druhá interpretace pak zvyšuje význam racionality, která není chápána jako pouhý prostředek, ale život v souladu s rozumem a hlavními třemi hodnotami (*rozum, účel a sebeúcta*, viz oddíl 5) je tím životem, jehož vedení a zachování je konečným cílem člověka. Třetí interpretaci podporuje především to, že podobně jako Aristotelés<sup>76</sup> i Randová zahrnuje do svého systému i štěstí: „*nejvyšším morálním cílem je dosažení vlastního štěstí*“.<sup>77</sup>

Vztah hodnot štěstí a života rozumného člověka (a jeho sebezáchovy) řeší poté tak, že je považuje za různé aspekty téhož. Předpokládá, že dosažení štěstí a sebezáchova jako nejvyšší hodnota jsou spolu neoddělitelně spojeny: „Zachování života a sledování štěstí nejsou dvě oddělené věci. Zachování vlastního života jako konečné hodnoty a vlastního štěstí jako nejvyššího cíle představuje dva aspekty stejné, správné orientace člověka.“<sup>78</sup> Z hlediska „existenciálního“ je podle Randové sledování racionálních cílů totožné s udržováním života. Z hlediska psychologického pak tatáž činnost produkuje emocionální stav vědomí označovaný jako štěstí.

Co tedy Randová rozumí lidským štěstím? V odpovědi na tuto otázku Randová odmítá chápat štěstí jakkoli hédonisticky a založit jej na požitku či žádosti.<sup>79</sup> To by znamenalo chápat štěstí jako výsledek sledování

75 Badhwar & Long (2020, 2.1).

76 „(...) jest snad souhlasně dokázáno, že blaženost jest nazývat nejvyšším dobrem.“ Aristotelés (2013, 1097b23). Detailněji viz dále poznámka pod čarou č. 30.

77 Randová (2018e, s. 49). Detailní srovnání všech tří interpretací viz Badhwar & Long (2020, oddíl 2). Tito badatelé však dochází k závěru, že ani jedna z těchto interpretací nedokáže uspokojivě vysvětlit některé filosofické problémy spojené s danými pozicemi.

78 Tamtéž, s. 52–53.

79 Podobně Aristotelés (2013, 1172b10).

rozmarů, tedy jako cosi iracionálního, a tudíž v rozporu s bytostnou povahou člověka jako rozumné bytosti.

Randová tedy ve shodě s Aristotelem říká, že povaha štěstí musí vycházet z povahy člověka jako rozumné bytosti, nemůže tedy spočívat v prostém požitku. Na rozdíl od Aristotela<sup>80</sup> však úzce spojuje štěstí člověka s dosahováním cílů, které jsou v souladu s jeho hodnotami: „Štěstí je stav vědomí, který vychází z dosažení vlastních hodnot.“<sup>81</sup> Tyto hodnoty pak musí směřovat k přežití člověka *qua* člověka, tudíž k jeho přežití jako racionální bytosti.<sup>82</sup> Takové hodnoty tedy nemohou být sebedestruktivní jako v případě masochisty či destruktivní jako v případě sadisty, nemohou být ani produktem pouhého rozmaru, protože výsledný stav vědomí po dosažení takových hodnot není podle Randové označitelný za štěstí, ani za požitek. V takovém případě jde jen o prožitek *oddychnutí si*. Je tomu tak zřejmě proto, že hodnoty a způsob jejich dosažení musí být racionální, což u Randové znamená, že mají být výsledkem účelově zaměřeného myšlení, jehož konečným cílem je sebezáchova. Nemůže se jednat o pouhé iracionální puzení, má-li být výsledkem stav vědomí štěstí. Stav štěstí je tudíž u Randové spojen jen s výsledky rozumem vedené produktivní *práce*.

Štěstí je tak zároveň emocionálním stavem člověka a objektivní charakteristikou jeho života. Tato koncepce s sebou nese podobné otázky jako ostatní naturalistické koncepce štěstí. Jedním paradoxem je možnost omylu – někdo se může domnívat, že je šťastný (pociťuje emoce, které popisuje jako štěstí), nicméně z objektivního hlediska šťastný není, protože jeho emocionální stav vědomí nenastal v důsledku dosažení hodnot spojených se sebezáchovou racionální bytosti. Druhým problémem je, že není jasné, jaký psychologický mechanismus (nutně) způsobuje, že člověk, který dosahuje vlastní hodnoty spojené se zachováním života sebe sama jako rozumné bytosti, k čemuž dochází prostřednictvím tvůrčí

---

80 V případě Aristotela se interpreti neshodují jak přesně štěstí či blaženost chápat, vztah štěstí a jednotlivých hodnot, které jej konstituují, představuje teoretický problém – otázku tzv. *dominantismu* versus *inkluzivismu*. Příznivci dominantistické interpretace považují za zásadní a výlučnou roli kontemplanace. Příznivci inkluzivistické interpretace vedle kontemplanace zdůrazňují i roli ctností a složení všech prvků štěstí do harmonického celku. Viz např. Heider (2007).

81 Randová (2018e, s. 51).

82 Tamtéž, s. 43.

práce, zároveň (nutně) zažívá pocit štěstí. Může například zažívat za stejné situace pocit přepracování, vyhoření a deprese. Někdo se tedy může subjektivně cítit velmi nešťastný, ale objektivně je podle této etické koncepce šťastný a naopak. Zde se dostáváme na pole empirické psychologie a nemůžeme říci, že by Randová na tyto otázky odpovídala.<sup>83</sup>

#### 4. Interakce s druhými lidmi

Život člověka směřujícího k přežití jakožto racionální bytost se má podle Randové odrážet i v jeho interakci s druhými lidmi. Tato interakce by měla spočívat ve směně, v obchodu. Lidé jsou v racionálním ohledu podle Randové vůči sobě v postavení obchodníků směřujících hodnotu za hodnotu: „Princip *obchodu* je jediným racionálním etickým principem pro všechny lidské vztahy.“<sup>84</sup> Je to jediný princip spravedlnosti, jediný princip mezilidské rovnosti a svobody. Veškerá směna v rámci obchodních vztahů je dobrovolná, tudíž ji obě strany vnímají jako výhodnou.

Mezilidské interakce, které se odehrávají mimo kontext obchodu, jsou podle Randové také založeny na směně, jen „měna“ a „medium“, kterým směna probíhá, jsou odlišné: „Láska, přátelství, úcta, obdiv jsou (...) duchovní *platbou* poskytnutou směnou za osobní, sobeckou radost, kterou člověk čerpá ze ctnostní povahy jiného člověka.“<sup>85</sup> Všechny lidské osobní vztahy založené na lásce, přátelství apod. jsou tedy podle Randové formou směny či obchodu. Nemá v nich tudíž místo (sebedestruktivní) jednání, ve kterém jeden dává více, než ve vztahu dostává.

Zároveň platí, že podle koncepce Randové jsou takové mezilidské vztahy racionální pouze tehdy, pokud vedou k zachování života člověka jako rozumné bytosti. A není jasné, proč by v rámci jejího výkladu měl k takovému životu patřit jinak než nahodile (jako rozmar) společný život s partnerem nebo vztahy s přáteli. Bylo by možné argumentovat, že k životu rozumné bytosti patří z nějakého důvodu přátelství,<sup>86</sup> nicméně z racionality odvodíme niternou emocionální povahu partnerských vztahů obtížněji.

83 Viz také rozbor v O'Neill (2007, 149).

84 Randová (2018e, s. 57).

85 Tamtéž, s. 58.

86 Podobně jako Aristotelés v osmé a deváté knize Etiky Nikomachovy.



Z výkladu, který Randová podává, se však zdá, že láskyplná starost partnerů či manželů o sebe navzájem není podle ní jen rozmar, který by tedy měl být odmítnut.<sup>87</sup> Tyto vztahy můžeme zřejmě opět interpretovat na základě koncepce vzájemně výhodné směny slasti, nicméně už víme, že Randová hédonistické vyhledávání požitků jako etický cíl odmítá.<sup>88</sup>

Ze života ve společnosti pak člověk profituje především právě možností obchodu, k čemuž ještě navíc Randová připojuje snadnost získávání znalostí.<sup>89</sup> Z redukce veškerých mezilidských interakcí na směnu pak dle Randové plyne, že z hlediska politického je ideální formou společenství „plný, čistý, nekontrolovaný, neregulovaný *laissez-faire* kapitalismus – s oddělením státu a ekonomiky“.<sup>90</sup>

## 5. Nauka o ctnostech

Podobně jako řada dalších etických myslitelů v aristotelské tradici<sup>91</sup> staví Randová do centra etického systému koncepci ctností a neřestí. Výše jsme vyložili, že v systému objektivistické etiky je život nejvyšší hodnotou člověka. Hodnota je pro Randovou obecně něco, k čehož udržení nebo získání směřuje lidské jednání. *Ctnost* však definuje Randová poměrně netradičně jako *akt*, kterým je hodnota získávána či udržována.<sup>92</sup> Hlavními třemi hodnotami podřazenými životu jsou *rozum*, *účel* a *sebeúcta* a k nim náležejícími ctnostmi, jimiž jsou tyto hodnoty získávány, jsou *racionalita*, *produktivita* a *hrdost*.<sup>93</sup> Jako další ctnosti jsou některými interprety Randové uváděny i upřímnost, nezávislost, spravedlnost a integrita.<sup>94</sup>

87 Všimněme si v tomto kontextu, že u Aristotela tento problém nenastává. Člověk je pro něj nejen rozumnou (*zoon echon logon*), ale i bytostně společenskou (*zoon politikon*) bytostí, k našemu štěstí a plnohodnotnému životu tak potřebujeme mj. přátelství.

88 Určitým racionalistickým argumentem pro manželské svazky, vycházejícím z principu sebezáchovy, který ovšem Randová neuvádí, by mohlo být, že lidé žijící v manželském svazku se statisticky dožívají vyššího věku (Whisman, Gilmour & Salinger, 2018).

89 Randová (2018e, s. 59).

90 Tamtéž, s. 61.

91 Viz např. MacIntyre (2004).

92 Randová (2018e, s. 45).

93 *Reason, Purpose, Self-Esteem – Rationality, Productiveness, Pride* (Rand 1964a, s. 27).

94 *Honesty, independency, justice, integrity, pride*. Viz např. Smith (2006, s. 75, 106, 135, 176, 221).

Jak máme vymezení *ctnosti* jako *aktu* rozumět, není zcela jasné. Rozumnost ani produktivita akty nejsou. Ctnosti jsou například u Aristotela chápány jako stavy charakteru člověka, které umožňují žít dobrý život. Vymezení ctnosti jako aktu tak budeme chápat jako zdůraznění skutečnosti, že ctnost je vždy chápána pouze jako prostředek dosažení štěstí (a sebezáchovy), nikoli jako účel o sobě.<sup>95</sup> Randová ve svých textech obvykle pracuje s ctnostmi jako se stavy charakteru, které mají takovýto instrumentální charakter, nicméně na některých místech její vyjádření působí dojmem jako by ctnosti byly cosi více. Detailní analýzu najdeme u Badhwara.<sup>96</sup> McGinnis přitom upozorňuje, že zatímco u Aristotela má většina ctností charakter středu mezi extrémy, tudíž vyžadují opatrné zvažování vedené praktickou moudrostí a zkušeností, u Randové ctnosti takovou povahu nemají, ba spíše naopak.<sup>97</sup>

Stejně tak není jasné, proč je *účel* (*purpose*) uveden mezi hodnotami. Pojem účel můžeme chápat jako cíl jednání, účel je tedy něco, co má pro jednatelovo jednání hodnotu. Není ale zřejmé, zda je možné pojem účelu podřadit (subsumovat) pod pojem hodnoty, tedy že účel je jedním z typů hodnot. Důvodem pro zahrnutí účelu mezi hodnoty tak může být to, že produktivita byla zahrnuta mezi ctnosti, tudíž v seznamu hodnot musela být uvedena odpovídající položka – hodnota odpovídající produktivitě. Zároveň však dále Randová jako hodnotu patřící k produktivitě uvádí *produktivní práci*: „Produktivní práce je ústředním *účelem* života racionálního člověka, ústřední hodnotou, která integruje hierarchii všech jeho dalších hodnot.“<sup>98</sup> Pravděpodobně tedy místo obecného pojmu účel musíme jako hodnotu patřící k produktivitě chápat účelnou produktivní práci, či její účelnost.

Jak ovšem sloučit dřívější tvrzení Randové, že nejvyšší hodnotou je život člověka a vše ostatní jsou prostředky k dosažení této hodnoty, s tvrzením o produktivní práci jako centrální hodnotě? Randová se patrně domnívá, že zatímco u ostatních organismů je hlavní hodnotou *pouhý život*, či život v souladu s jejich biologickou přirozeností (např. život vní-

95 Smith (2006, s. 49).

96 Badhwar (1999).

97 McGinnis (2012).

98 Randová (2018e, s. 45). Doplnujeme oproti českému překladu zvýraznění slova „účel“ (purpose), jak je v anglickém originále (Rand 1964a, s. 27).

mání a pohybu u živočichů<sup>99</sup>), u člověka jako rozumné bytosti je touto hodnotou specifický způsob života vlastní pouze jemu – *rozumný život*. To je pojetí, jak jsme již řekli, svou povahou aristotelské, nicméně jaký přesně způsob života odpovídá rozumnosti člověka, v tom se myslitelé jdoucí v Aristotelových stopách liší. Pro Randovou je rozumný lidský život charakteristický především produktivní prací. To je vcelku překvapivé, protože se jedná o koncepci, kterou nacházíme už u Marxe<sup>100</sup> a jeho spolupracovníka Engelse (1950), tedy úhlavních teoretických protivníků Randové. Naopak např. pro Aristotela byla produktivní práce doménou zejména otroků, kteří byli jen částečně rozumní. Plně rozumní svobodní lidé svůj rozum realizovali především v teoretické kontemplaci či účasti na politickém životě obce.<sup>101</sup> Randová se ohledně účasti na politickém životě v zemi, která plně neodpovídá plně výše zmíněnému ideálu společnosti, omezuje na požadavek veřejného deklarování morálních soudů všemi racionálními občany.<sup>102</sup>

Co se týče další ctnosti, *rozumnosti*, ta, jak už napověděly předchozí odstavce našeho textu, je „základní ctností člověka, zdrojem všech jeho dalších ctností“.<sup>103</sup> Naopak základní neřestí je iracionalita, tj. „akt nezaměření mysli“,<sup>104</sup> který je volbou, rozhodnutím nevidět a nevědět, jež je sebedestruktivní, protože odmítá základní prostředek k přežití – rozum. O tom, co je míněno rozumem a myšlením, jsme již detailněji pojednali výše.<sup>105</sup>

Další ctností je *hrdost*, kterou Randová chápe jako „morální ctižádost“.<sup>106</sup> Tvrdí, že vedle samotného zachování života, což je nejvyšší hodnota, musí člověk „dosahovat i charakterových hodnot, které činí jeho

99 Srov. Aristotelés (2000, 413b).

100 „Marx (...) vidí práci nejen jako dočasné omezené produktivní zapojení se do toho či onoho pracovního procesu, který je omezen na nějaký region světa, ale také jako specifickou realizaci lidské existence ve světě, skrze kterou je člověk ‚pro sebe‘ tím, čím je ‚v sobě‘.“ (Ehlen 1985, s. 34, překlad autor stati).

101 Aristotelés (2013, 1176b).

102 Randová (2018c, s. 134).

103 Randová (2018e, s. 45).

104 Tamtéž. Opět zde narážíme na podobný konceptuální problém, jaký byl zmíněn výše. *Neřest* je považována také za *akt*.

105 Viz také Smith (2006, s. 48).

106 Randová (2018e, s. 48).

život hodným uchování“.<sup>107</sup> I s touto ctností se pojí určité otázky. Pokud je život nejvyšší a nepodmíněnou hodnotou, pro kterou vše ostatní je jen prostředkem, pak nepotřebuje nic dalšího. Není ale zřejmé, proč by si člověk musel „zasloužit právo považovat se za svou nejvyšší hodnotu tím, že dosáhne vlastní morální dokonalosti“.<sup>108</sup> Na jednu stranu je Randovou kladen vlastní život jako východisko, jako nepodmíněný předpoklad jakékoli etické hodnoty, na druhou stranu je třeba si právo považovat se za nevyšší hodnotu „zasloužit“ (*earn*). Přitom *morální dokonalost* spočívá v odhodlání nikdy nepřijmout iracionální neřesti, nikdy nepřestat praktikovat racionální ctnosti, nikdy nepřijmout nezaslouženou vinu, nikdy nestavět žádné emoce nad vlastní sebevědomí (*self-esteem*) a nikdy nepřijmout žádnou doktrínu hlásající sebeobětování za druhé.

Vedle základních ctností zavádí Randová okrajově také „nejušlechtlejší vlastnost[i] jeho [lidské] přirozenosti“: tvořivost, ctižádost, odvalu, neochotu přijímat bez boje působení katastrof, odhodlání měnit zemi podle vlastních hodnot. Tyto atributy jsou zdůvodněny tím, že všechny najdou uplatnění při produktivní práci.<sup>109</sup>

## 6. Žádost a zájem

Jak jsme viděli výše, Randová odmítá, že by žádosti a jejich uspokojení měly nějakou etickou relevanci, odmítá jakoukoli formu hédonismu a také utilitarismu. Říká také, že pokud bychom nějaký z těchto etických přístupů přijali, pak „nemají lidé jinou volbu než se navzájem nenávidět, bát se jeden druhého a bojovat mezi sebou, protože jejich touhy a zájmy se nezbytně kříží“.<sup>110</sup> To mimochodem také odpovídá Hobbesovu<sup>111</sup> pojetí lidské přirozenosti, ve kterém psychologický hédonismus a egoismus vede v přirozeném stavu (tedy stavu, kdy lidé nejsou podřízeni politické vládě) k nepřetržitému boji všech proti všem. Hobbes tento problém řeší

107 Tamtéž.

108 Tamtéž.

109 Tamtéž, s. 47.

110 Tamtéž, s. 55.

111 Hobbes (2015, s. 90). Randová a Hobbes se řadí mezi filozofy etického egoismu, proto zde srovnáváme právě jejich přístupy.

zavedením absolutistické vlády, v níž jsou občané plně podřízeni moci státu, jehož hlavní rolí je bránit vzájemným konfliktům.

Randová si samozřejmě uvědomuje, že koncepce střetu zájmů, který vede k neustálým bojům, zasahuje kriticky také její vlastní etickou a zejména politickou teorii, jež odmítá silnou roli státu. Zatímco ve výše uvedeném citátu staví touhy (*desire*) a zájmy (*interest*) vedle sebe jako synonyma a tvrdí, že se nutně střetávají, později od sebe tyto pojmy podstatně odlišuje. Zatímco žádosti lidí se vždy střetávají, u „racionálních zájmů“ platí podle Randové opak, tyto zájmy se nestřetávají nikdy.<sup>112</sup>

Aby tuto svoji tezi prokázala, je nutný poměrně rozsáhlý výklad a nové vypracování pojmu zájem, čemuž věnuje celou samostatnou esej „Konflikty zájmů“. Randová v ní uvádí jako příklad dva uchazeče o totéž zaměstnání, které však může získat jen jeden z nich. Jelikož mají oba zájem o stejné zaměstnání, mohlo by se zdát, že jde o příklad konfliktu zájmů. Tak tomu ovšem podle Randové ve skutečnosti není. To, že někdo po něčem touží, podle Randové neimplikuje, že je v jeho zájmu to získat, tedy že to je jeho zájem.<sup>113</sup> Zájmem člověka například není získat to, co si „nezaslouží“.<sup>114</sup> Ve skutečnosti se tedy v tomto případě o konflikt zájmů nejedná – získání zaměstnání nebylo v zájmu neúspěšného uchazeče, neboť si zaměstnání získat nezasloužil.

Randová sice odmítá, že by to, co je skutečným zájmem člověka, určovala jeho žádost, ale zároveň nedává žádné další kritérium, podle kterého identifikovat skutečný zájem (a odlišit jej od prosté žádosti), kromě neustálého obecného zdůrazňování racionality. Kritériem zde nemůže být ani zachování života, neboť dobře placené zaměstnání, o které se daný člověk uchází, jistě k zachování jeho života přispívá.

Paradoxnost této koncepce osvětlí jiný myšlenkový experiment. Představme si, že ve výše uvedeném případě uchazeč A práci získá, zatímco uchazeč B nikoli. Pak podle Randové bylo v zájmu uchazeče B práci nezískat. Nyní si představme, že by den před nástupním pohovorem uchazeč A onemocněl. K pohovoru by se dostavil jen uchazeč B, byl by přijat a zároveň by bylo v jeho zájmu být přijat. To, zdali je přijetí do zaměst-

---

112 Randová (2018e, s. 57).

113 Randová (2018d, s. 94).

114 Tamtéž, s. 105–106.

nání skutečným racionálním zájmem uchazeče B, tak může ovlivnit celá řada zcela nahodilých okolností.

Randová vedle toho říká, že v zájmu uchazečů je existence pracovních míst, která je možná pouze při existenci podnikajících firem, jež hledají zaměstnance, a ta je umožněna jen existencí konkurence uchazečů o pracovní místa. I když bychom tuto úvahu přijali a akceptovali, že obecně je v zájmu zaměstnanců existence konkurence uchazečů o pracovní místa, neznamená to, že v konkrétním případě není v zájmu každého uchazeče, aby dané pracovní místo získal on sám. Představa, kterou Randová prosazuje, tedy že pro nezaměstnaného, jenž má hlad a dlouhodobě nemůže najít práci, je v jeho zájmu práci nezískat, protože si ji „nezaslouží“ a není oprávněn k takovému zájmu a v jeho zájmu je zemřít hladem, je nejen obtížně akceptovatelná, ale i v rozporu s výchozím principem její filosofie, tj. zachování života člověka jako rozumné bytosti.

Podívejme se ale ještě na jeden problém, který souvisí s koncepcí *racionálních zájmů*. Pokud oddělíme zájmy od tužeb na základě tvrzení, že zájmy jsou na rozdíl od tužeb racionální, vyvstane tím řada problémů. Výše jsme uvedli, že starost o druhé, o partnery a přátele, je pro Randovou čistě sobeckým aktem vycházejícím z vlastních zájmů: „Starost o blaho těch, které člověk miluje, je racionální součástí sobeckých zájmů.“<sup>115</sup> Proč můj zájem na prospěchu mého přítele není jen rozmar? Zřejmě proto, že racionálně očekávám nějakou *směnu*, tedy že mi přítel můj zájem o jeho prospěch oplatí. Pokud není moje starost o blaho mého přítele vedena racionálním očekáváním takového zisku ve směně či alespoň zisku osobního prožitku radosti, pak je iracionální sebedestruktivní činností.

Vztahy k partnerům a přátelům, které nejsou vedeny očekáváním osobního prospěchu, jsou tedy pouhá iracionální hnutí mysli, produkt žádosti, a tedy rozmar. Tudíž nejen že kdykoli stojí v cestě skutečným racionálním zájmům člověka, musí jít stranou, ale také je nehodné racionálního člověka, který má věnovat svůj čas výhradně racionální činnosti – produktivní práci, jim věnovat jakýkoli čas. Randová říká: „Nesobecká, nezištná láska představuje pojmový rozpor.“<sup>116</sup>

115 Randová (2018b, s. 83).

116 Tamtéž.

## 7. Závěr

Randová svým dílem ovlivnila mnoho dalších autorů a dodnes je jedním z hlavních inspiračních zdrojů libertariánského hnutí, a to především v USA. Proto se nabízí srovnání s dalšími podobně zaměřenými autory, zejména s Robertem Nozickem, který je také známý jako zastánce libertariánské filosofie.<sup>117</sup> Předně je třeba zmínit, že existuje zásadní rozdíl mezi těmito autory, co se týče jejich přijetí v akademické sféře. Randová byla primárně spisovatelka, jejíž filosofické názory nebyly nikdy plně akceptovány akademickou komunitou. Nozick je oproti tomu akademický analytický filosof, který je široce respektován i mezi svými odpůrci. Zátlmco východiskem Randové je primárně Aristotelova filosofie, Nozick (2005) vychází především z klasického liberalismu Locka a Kanta,<sup>118</sup> nicméně nevypracovává samostatnou etickou teorii. Oproti tomu Randová vycházející z aristotelské etiky ji posouvá do roviny radikálně formulované teorie etického egoismu.

Přestože výsledná podoba společnosti, kterou Nozick a Randová preferují, nese řadu podobných rysů (jedná se o společnosti s minimálními zásahy státu a přerozdělováním), existuje v důsledku výše naznačených rozdílných filosofických východisek i řada podstatných odlišností. Pro Randovou, jak jsme ukázali, byla jediným přijatelným politickým uspořádáním společnost volného trhu a neregulovaného *laissez-faire* kapitalismu oddělujícího striktně stát a ekonomiku.<sup>119</sup> Nozick připouštěl v rámci minimálního státu nejen kapitalistické instituce, ale například i existenci komunit fungujících na principech zásadního přerozdělování – za předpokladu, že všichni zúčastnění s takovým uspořádáním souhlasí.<sup>120</sup>

Nozick<sup>121</sup> se k etické teorii Randové vyjádřil s tím, že sdílí její přesvědčení, že lze poskytnout morální základy *laissez-faire* kapitalismu, nicméně k samotnému konkrétnímu postupu, který Randová předkládá,

117 Robert Nozick (1938–2002), o 33 let mladší než Randová (1905–1982), známý především svou obhajobou libertariánství v knize *Anarchie, stát a utopie* (2005).

118 Nozick (2005).

119 Randová (2018e, s. 61).

120 Nozick (2005, s. 327–329).

121 Nozick (1971). Nozickův článek *On the Randian Argument* (1971), na který zde odkazujeme, byl později vydán i v rámci jeho knihy *Socratic puzzles* (1997).

je skeptický. Především z pohledu analytického filosofa považuje logickou strukturu její argumentace za nejasnou s tím, že některé kroky musí být spekulativně doplněny, aby byl postup korektní. Ani poté však nepovažuje její argumentaci za přesvědčivou, protože řada předpokladů, ze kterých Randová vychází, je dle jeho výkladu sporná. Nozickova kritika Randové je významná mj. i proto, že jakožto libertarián z velké části sdílí její politické postoje, a nelze jej tedy podezírat z ideologicky motivované zaujatosti.

K podobnému závěru jsme došli i my ve výše prezentovaném výkladu. Ayn Randová se pokusila vytvořit etický systém, jenž by byl založen na čistě objektivních, racionálních principech. Při tvorbě tohoto systému odmítla jakoukoli filosofii, kterou považovala za spekulativní, stejně jako náboženské základy etiky. Hlavní inspirací jí byla naturalistická etika Aristotelova, mnohé jednotlivé myšlenky přijala i z dalších směrů, které však z velké části odmítala. Tato etika vychází z jejích metafyzických a epistemologických tezí, které byly v našem textu jen naznačeny, jako celek má objektivismus ambice být komplexním filosofickým systémem a její etika má poskytnout fundament pro vyřešení základních morálních otázek i běžných každodenních dilemat. Randová se považuje za pokračovatelku Aristotela, jež v jeho filosofii odstraňuje nedostatky a dopracovává jím předznačený etický systém k dokonalosti. Sama říká: „Jsem především tvůrkyní nového morálního kodexu, který byl dosud považován za nemožný, jmenovitě morálky, která není založena na víře, na svévolném rozmaru, na emocích, na svévolném mystickém nebo společenském ediktu, ale na rozumu. Morálky, kterou lze prokázat logikou, kterou lze prokázat jako pravdivou a nezbytnou.“<sup>122</sup>

V tomto textu jsme se pokusili ukázat, jakým způsobem Randová svůj etický systém buduje a argumentačně podpírá, abychom mohli posoudit, zdali takto vysokým ambicím je s to dostát. Východiskem jejích úvah je spojení pojmu hodnota s pojmem život a určení zachování života jako základní hodnoty. Ukázali jsme, že argumentace Randové bohužel není zcela přesvědčivá a došli jsme (prostřednictvím jiných argumentů) k podobně kritickému názoru, jaký prezentuje Nozick: „Nelze dojít k závěru, že život sám o sobě je hodnotou pouze pospojováním mnoha vět obsahu-

122 Wallace & Rand (1957).



jších slovo ‚hodnota‘ a ‚život‘ nebo ‚živý‘ a doufat, že nějakým procesem asociace a smíchávání toto nové pojetí vznikne.“<sup>123</sup>

Zmínili jsme, že pro Randovou není v případě člověka hodnotou prostá sebezáchova, ale zachování života člověka jakožto rozumné bytosti, přičemž není zřejmé, jak řeší přechod z deskriptivní roviny do roviny normativní (*is-ought* problém). Zároveň jsme ukázali, že i v otázce způsobu pojetí samotné koncepce zachovávání života je Randová velmi selektivní a jako jedinou akceptovatelnou formu lidského života v souladu s rozumem chápe produktivní práci, vedenou zájmem o vlastní prospěch – tento ideál ztělesňují hrdinové jejích románů – inovátorští podnikatelé či architekti. Ctnosti, jež ve své etické koncepci staví do centra pozornosti, jsou právě ty, které jsou charakteristické pro tento typ aktérů. Jakékoli alternativní životní programy nejsou dle Randové životem člověka v souladu s rozumem, což je velmi restriktivní pohled na lidský život. Se správně žitým lidským životem v souladu s rozumem se dle objektivistické koncepce nutně pojí štěstí, nicméně jak jsme ukázali, tento pojem je zde také spojen s paradoxy.

Důraz na sebezáchovu a lidský projev vede k reduktivnímu chápání všech mezilidských vztahů jako obchodů, v nichž je cílem dosáhnout zisku. Jakýkoli altruismus je pro Randovou nejen iracionální, ale v plném slova smyslu amorální, což také implikuje titul sborníku objektivistických textů, věnovaných etice – „Ctnost sobectví“. V takovém rámci je ovšem obtížné přijatelně vysvětlit nejintimnější lidské vztahy: partnerské a rodičovské. Milující péče matky o své dítě, která není vedena očekáváním zisku, je v tomto pojetí amorální, partnerské vztahy jsou vždy obchodem něco za něco. V rámci snahy osvětlit lidské vztahy a vyrovnat se s hrozbou nekonečných konfliktů v důsledku vzájemně se křížících žádostí a zájmů vypracovává Randová koncepci racionálních zájmů, která ovšem také vede k paradoxním důsledkům.

Objektivistická etika, tak jak ji prezentuje Randová, obsahuje řadu inkonzistencí a nedořešených problémů. Přitažlivá je často pro nefilosofy, kteří inklinují k libertariánství, a proto je pro ně atraktivní její nekompromisní politická rétorika (prezentovaná poutavou formou i v jejích románech). Také její filosofické texty jsou psané formou

123 Nozick (1971, s. 252).

přístupnou i ve filosofii neškolenému čtenáři, což je zřejmě dalším důvodem její popularity mezi filosofickými laiky – někteří její kritici, například Walker, dokonce mluví o kultu.<sup>124</sup> Přístupnost těchto textů je ovšem na úkor jejich rigoróznosti a argumentační propracovanosti, tudíž bohužel nedosahují kvality jiných libertariánských filosofů. V rámci akademické diskuse tak filosofie Randové stojí většinou na okraji zájmu, což je do značné míry odůvodněné.

Tento náš kritický názor sdílí i řada dalších autorů. Výše jsme už zmínili kritický postoj Nozicka (1971). Kromě něj například Barry a Stephens tvrdí, že „na obecné úrovni nemá objektivismus žádné legitimní postavení v disciplíně morální filosofie“<sup>125</sup> a jiní upozorňují na skutečnost, že řada rozsáhlých filosofických kompendií objektivismus Randové vůbec nezmiňuje.<sup>126</sup> Podobnou kritiku najdeme i u Robbinse, McGinnise a ve sborníku Den Uyla a Rasmussena.<sup>127</sup> O’Neill je obecně k filosofii Randové kritický, nicméně se domnívá, že v ní lze najít i pozitivní aspekty: „ačkoli *odpovědi* slečny Randové mohou být chybné, její *otázky* jsou často správné. Občas je velmi přesná ve svých diagnózách současných intelektuálních patologií a kulturních poruch.“<sup>128</sup> V Čechách se už k filosofii Randové kriticky vyjádřili například Pavlík a Novotný.<sup>129</sup> Někteří ze současných amerických filosofů přistupují k jejímu dílu pozitivněji a snaží se na jeho základě vybudovat konzistentní filosofickou teorii.<sup>130</sup>

## Literatura

Aristotelés (2013): *Etika Níkomachova*. Rezek, Praha.

Aristotelés (2020): *O duši*. Rezek, Praha.

Aristotelés (2021): *Metafyzika*. Rezek, Praha.

124 Walker (2012).

125 Barry & Stephens (1998, s. 163).

126 Audi & Audi (2015), Blackburn (1994).

127 Robbins (1974), McGinnis (2012), Den Uyl a Rasmussen (1986).

128 O’Neill (2007, s. 13–14).

129 Pavlík (1997a, 1997b), Novotný (1999).

130 Např. Tibor Machan, Allan Gotthelf, James G. Lennox a Gregory Salmieri. Zájemce o dílo Randové z řad profesionálních filosofů sdružuje v současnosti *Ayn Rand Society*, která se také angažuje ve vydávání monografií kriticky reflektujících její dílo (např. Gotthelf & Salmieri 2016).

- Audi, P. & Audi, R. (2015): *The Cambridge Dictionary of Philosophy*. Cambridge University Press, Cambridge.
- Badhwar, N. (1999): „Is Virtue Only a Means to Happiness? An Analysis of Virtue and Happiness in Ayn Rand’s Writings.“ *Reason Papers* 24: 27–44.
- Badhwar, N. & Long, R. T. (2020): „Ayn Rand.“ In *The Stanford Encyclopedia of Philosophy* [online], ed. E. N. Zalta, 8. 6. 2010, aktualiz. 13. 7. 2020 [cit. 20. 10. 2021]. Metaphysics Research Lab, Stanford University. Dostupné na: <https://plato.stanford.edu/archives/fall2020/entries/ayn-rand/>.
- Barry, B. & Stephens, C. U. (1998): „Objections to an Objectivist Approach to Integrity.“ *The Academy of Management Review* 23 (1): 162–69. <https://doi.org/10.2307/259105>.
- Blackburn, S. (1994): *The Oxford dictionary of philosophy*. Oxford University Press, Oxford.
- Burgess-Jackson, K. (2013): „Taking Egoism Seriously.“ *Ethical Theory and Moral Practice* 16 (3): 529–42.
- Den Uyl, D. J. & Rasmussen, D. B. eds. (1986): *The Philosophic Thought of Ayn Rand*. Illini books ed. Urbana: University of Illinois Press.
- Ehlen, P. (1985): „The Human Significance and Dignity of Labor‘: A Keyword in Marxian Anthropology.“ *Studies in Soviet Thought* 29 (1): 33–46.
- Engels, F. (1950): *Podíl práce na polidštění opice*. Svoboda, Praha.
- Gotthelf, A. S. (2000): *On Ayn Rand*. Wadsworth, Belmont.
- Gotthelf, A. S. & Salmieri, G. eds. (2016): *A Companion to Ayn Rand*. Wiley Blackwell, Chichester.
- Heider, D. (2007): „Analogie vnitřní atribuce jako možné řešení nejasností v Aristotelově pojetí blaženosti v Etice Níkomachově.“ *Studia Neoaristotelica* 4 (2): 143–57.
- Heller, A. C. (2010): *Ayn Rand and the World She Made*. Doubleday, New York.
- Hobbes, T. (2015): *Leviathan*. Přeložil K. Berka. OIKOYMENH, Praha.
- Hume, D. (1739): *A Treatise of Human Nature*. John Noon, London.

- Hunt, L. H. (2016): „Ayn Rand’s Evolving View of Friedrich Nietzsche.“ In *A Companion to Ayn Rand*, eds. A. S. Gotthelf & G. Salmieri. Wiley Blackwell, Chichester, s. 343–350.
- Kelley, D. & Thomas, W. R. (2021): „The Logical Structure of Objectivism.“ Dostupné na: <https://www.atlassociety.org/post/the-logical-structure-of-objectivism>.
- Kinkor, J. (1996): *Trh a stát. K čemu potřebujeme filosofii*. Svoboda, Praha.
- Kinkor, J. (1997): „Jak zdiskreditovat objektivismus.“ *Pravý úhel* č. 4.
- Lennox, J. G. (2016): „Who Sets the Tone for a Culture? Ayn Rand’s Approach to History of Philosophy.“ In *A Companion to Ayn Rand*, eds. A. S. Gotthelf & G. Salmieri. Wiley Blackwell, Chichester, s. 321–342.
- MacIntyre, A. (2004): *Ztráta ctnosti: k morální krizi současnosti*. OIKOYMENH, Praha.
- MacKenzie, M. M. (1982): „A Pyrrhic Victory: Gorgias 474b-477a.“ *The Classical Quarterly* 32 (1): 84–88.
- McGinnis, N. (2012): „The System that Wasn’t There: Ayn Rand’s Failed Philosophy (and why it matters).“ In *Rotman Institute of Philosophy* [online]. 25. 8. 2012 [cit. 15. 10. 2021]. Dostupné na: <https://www.rotman.uwo.ca/the-system-that-wasnt-there-ayn-rands-failed-philosophy-and-why-it-matters/>.
- Mill, J. S. (2016): *Vybrané spisy o etice, společnosti a politice I*. Přeložil Karel Šprunk a Martin Pokorný. OIKOYMENH, Praha.
- Novotný, J. (1999): „Beriška a učení Ayn Randové.“ *E-LOGOS* 6 (1): 1–4 .
- Nozick, R. (1971): „On the Randian Argument.“ *The Personalist* 52 (2): 282–304. <https://doi.org/10.1111/j.1468-0114.1971.tb08926.x>.
- Nozick, R. (1997): *Socratic puzzles*. Harvard University Press, Cambridge, Mass.
- Nozick, R. (2005): *Anarchie, stát a utopie*. Academia, Praha.
- O’Neill, W. F. (2007): *With Charity Toward None: An Analysis of Ayn Rand’s Philosophy*. Philosophical Library, New York.
- Pavlík, J. (1997a): „Liberalismus a filosofie.“ *E-LOGOS* 4 (1).

- Pavlík, J. (1997b): „Předběžné kritické poznámky k doktríně ‚objektivismu‘.“ *E-LOGOS* 4 (1).
- Peikoff, L. (1991): *Objectivism: The Philosophy of Ayn Rand*. Dutton, New York.
- Peikoff, L. (2001): *Objektivismus: Filozofie Ayn Randové*. Přeložil Jiří Kinkor. Berlet, Praha.
- Platón (2000): *Gorgias*. OIKOYMENH, Praha.
- Platón (2012): *Filébos*. OIKOYMENH, Praha.
- Platón (2016): *Zákony*. OIKOYMENH, Praha.
- Platón (2018): *Symposion*. OIKOYMENH, Praha.
- Rand, A. (1943): *The Fountainhead*. Plume, New York.
- Rand, A. (1957): *Atlas Shrugged*. Random House, New York.
- Rand, A. (1961): *For the New Intellectual: The Philosophy of Ayn Rand*. Random House, New York.
- Rand, A. (1964a): „The Objectivist Ethics.“ In *The Virtue of Selfishness: A New Concept of Egoism*, A. Rand. New American Library, New York, s. 13–39.
- Rand, A. (1964b): *The Virtue of Selfishness: A New Concept of Egoism*. New American Library, New York.
- Rand, A. (1982): *Philosophy, who needs it*. Bobbs-Merrill, Indianapolis.
- Rand, A., Peikoff, L. & Schwartz, P. (1990): *The Voice of Reason: Essays in Objectivist Thought*. Meridian, New York.
- Randová, A. (2000): *Zdroj*. Přeložila Jana Melicharová. Berlet, Praha.
- Randová, A. (2003): *Atlas pokrčil plecami I, II, III*. Epos, Bratislava.
- Randová, A. (2014): *Atlasova vzpoura*. Přeložil Aleš Drobek. Argo, Dokořán, Praha.
- Randová, A. (2018a): *Ctnost sobectví: nové pojetí egoismu*. Přeložil Jiří Kinkor. Dauphin, Praha.
- Randová, A. (2018b): „Etika nouzových situací.“ In *Ctnost sobectví: nové pojetí egoismu*, A. Randová, přeložil Jiří Kinkor. Dauphin, Praha, s. 80–93.
- Randová, A. (2018c): „Jak lze vést racionální život v iracionální společnosti.“ In *Ctnost sobectví: nové pojetí egoismu*,

- A Randová, přeložil Jiří Kinkor. Dauphin, Praha, s. 134–141.
- Randová, A. (2018d): „Konflikty zájmů.“ In *Ctnost sobectví: nové pojetí egoismu*, A. Randová, přeložil Jiří Kinkor. Dauphin, Praha, s. 94–107.
- Randová, A. (2018e): „Objektivistická etika.“ In *Ctnost sobectví: nové pojetí egoismu*, A. Randová, přeložil Jiří Kinkor. Dauphin, Praha, s. 21–65.
- Riley, J. (2003): „Interpreting Mill’s Qualitative Hedonism.“ *The Philosophical Quarterly* 53 (212): 410–18.
- Robbins, J. W. (1974): *Answer to Ayn Rand: a critique of the philosophy of objectivism*. Mount Vernon Publishing Co., Washington DC.
- Salmieri, G. (2016a): „An Introduction to Study of Rand.“ In *A Companion to Ayn Rand*, eds. A. S. Gotthelf & G. Salmieri. Wiley Blackwell, Chichester, s. 3–21.
- Salmieri, G. (2016b): „Egoism and Altruism.“ In *A Companion to Ayn Rand*, eds. A. S. Gotthelf & G. Salmieri. Wiley Blackwell, Chichester, s. 130–156.
- Salmieri, G. (2016c): „The Act of Valuing.“ In *A Companion to Ayn Rand*, eds. A. S. Gotthelf & G. Salmieri. Wiley Blackwell, Chichester, s. 49–72.
- Schmidt-Petri, Ch. (2003): „Mill on Quality and Quantity.“ *The Philosophical Quarterly* 53 (210): 102–4.
- Smith, A. (2001): *Bohatství národů*. Liberální institut, Praha.
- Smith, A. (2005): *Teorie mravních citů*. Liberální institut, Praha.
- Smith, T. (2006): *Ayn Rand’s Normative Ethics. The Virtuous Egoist*. Cambridge University Press, Cambridge.
- Walker, J. (2012): *Ayn Rand Cult*. Open Court, New York.
- Wallace, M. & Rand, A. (1957): *Objectivism vs Altruism. Interview*.  
Dostupné na: <https://www.youtube.com/watch?v=viGkAZR-x8s>.
- Wheeler, J. (1986): „Rand and Aristotle: A Comparison of Objectivist and Aristotelian Ethics.“ In *The Philosophic Thought of Ayn Rand*, eds. D. J. Den Uyl a D. B. Rasmussen, Illini books ed. Urbana, University of Illinois Press, s. 81–101.

Whisman, M. A., Gilmour, A. L. & Salinger, J. M. (2018): „Marital satisfaction and mortality in the United States adult population.“ *Health Psychology* 37 (11): 1041–44. <https://doi.org/10.1037/hea0000677>.

## Abstract

### **The objectivist ethics of Ayn Rand**

The objectivist philosophy of Ayn Rand is still a direction in philosophy to which some people actively subscribe, and at the same time, it is an ideological inspiration for some political movements. The presented text is devoted to the critical appreciation of the concept of virtue in objectivist ethics, the so-called ethics of egoism. First, we look at Rand's assessment of these previous ethical approaches, then we briefly describe the foundation of her own philosophy – the principle of self-preservation and rational human life. Next part of the text deals with the issue of interpersonal interactions, that provide context of her understanding of virtues, and we describe why Rand rejects any sacrifice and altruism. The central role in her philosophy plays the concept of productive work and her concept of virtue refers to creative work as a specific center of gravity. Throughout the text we point out some problems and difficulties of her approach; specific problem represents the concept of rational interest, which leads to some paradoxical consequences. The text ends with a comparison with conception of R. Nozick and a brief conclusion.

Keywords: objectivism, ethics, Ayn Rand, virtue

Vacura, M. (2021): „Objektivistická etika Ayn Randové.“ *Filosofie dnes* 13 (2): 119–150. Dostupné z [www.filosofiednes.ff.uhk.cz](http://www.filosofiednes.ff.uhk.cz).

ČASOPIS  
PRO  
SOUČASNOU  
FILOSOFII

