
Rozhovor

O futuristické technice, vědomí a metafyzice¹

Susan Schneider – Tomáš Hříbek —

Susan Schneiderová je profesorkou astrobiologie při NASA, řídí skupinu AI, Mind and Society (AIMS) na Connecticutské univerzitě, je členkou Ethics and Technology Group v Yaleově Interdisciplinárním centru pro bioetiku a hostující členkou princetonského Institutu pro pokročilá studia. Působila jako poradkyně v otázkách financování vědy v americkém Kongresu. Proslavily ji publikace věnované filosofickým aspektům umělé inteligence, robotiky a dalších nových technologií. Jelikož se těmto otázkám filosofové budou patrně muset v následujících letech stále více věnovat, představují uznávané práce Susan Schneiderové v tomto ohledu jistý předvoj.

O těchto tématech publikovala i množství popularizujících příspěvků v prestižních médiích, např. v The New York Times, Wired Magazine, Smithsonian a Popular Mechanics; vystupovala také v televizi. Její kniha Artificial You: AI and the Future of Your Mind, určená pro široké čtenářské vrstvy, vyšla na podzim r. 2019. Několik svých odborných statí přetiskla Susan Schneiderová rovněž v úspěšné antologii Science Fiction and Philosophy,² která atraktivním způsobem kombinuje klasické sci-fi povídky s příspěvky současných filosofů tak, že objektivně ukazuje, že některé myšlenkové experimenty a úvahy týkající se povahy vědomí, osobní identity nebo času se poprvé – byť v beletristické formě – objevily právě v dílech spisovatelů.

Rozhovor mezi Susan Schneiderovou a Tomášem Hříbkem, pracovníkem Filosofického ústavu AV ČR, proběhl v červnu 2019.

T.H.

1 Přeloženo z angl. originálu. Transl. © Tomáš Hříbek. Pozn. red.

2 Schneider, S. (ed.), *Science Fiction and Philosophy: From Time Travel to Superintelligence*. 2nd Ed. Oxford, Wiley-Blackwell 2016.

Tomáš Hříbek (TH): *Susan, jak ses dostala k filosofii techniky, zvláště k filosofii umělé inteligence a robotiky?*

Susan Schneider (SS): Už od vysokoškolských studií jsem byla nadšenou čtenářkou science fiction. Takže si myslím, že to bylo díky četbě podivuhodných příběhů Isaaca Asimova a Arthura C. Clarka. Myšlenkové experimenty, které popisují, mě přiměly zajímat se o filosofický význam toho všeho, o vztah mezi technikou a filosofií, o možnosti a hranice poznání. Od té doby jsem se vždy zajímala o otázky spojené s budoucím rozvojem technologií: o cestování časem, umělou inteligenci, dystopické scénáře...

TH: *Máš ještě čas číst sci-fi?*

SS: Bohužel jsem teď tak zaneprázdněná, že je těžké najít čas na četbu. Ale užívám si alespoň filmy a v současnosti vzniká mnoho vynikajících sci-fi filmů a televizních pořadů. Pomáhají mi přemýšlet, přinášejí mi nové podněty...

TH: *...konečně ty filmy dělají pořádně a věrohodně, nemám pravdu?*

SS: Přesně tak! Před chvílí jsme hovořili o tom, jak se povedl film *Interstellar* – a já se také vždycky už moc těším na novou řadu seriálu *Black Mirror*. Jsem úplně nadšená z přístupu jeho autorů, z toho, jak dokáží zkoumat hranice současných technologií a dystopické scénáře. Tenhle seriál je tak inteligentní, až mám někdy strach ho sledovat. V dílu o mobilních robotických psech tvůrci použili jako modely skutečné roboty z laboratoře Boston Dynamics. Takže tyhle dystopie se stále víc přibližují realitě... Mým nejoblíbenějším typem sci-fi jsou taková zpodobení budoucí společnosti, v nichž se daří super-inteligentním bytostem obdařeným vědomím. Bohužel se právě tohle těžko zobrazuje (*smích*).

TH: *Mou nejoblíbenější epizodou v seriálu Black Mirror je příběh dvou starých dam, jež pomalu umírají. Jejich vědomí je ovšem uloženo do supervýkonného počítače, takže si posléze obě (a tisíce dalších zemřelých lidí) mohou užívat posmrtnou simulovanou existenci jako dvě šťastné mladé dívky v přímořském městečku. Na tomto příběhu mě oslovuje nejen ta úžasná futuristická technologie, ale také jeho lidskost...*

SS: *San Junipero!* Tahle epizoda je velmi podnětná. Je to jedna z mých vůbec nejoblíbenějších. Nejlepší na ní je, že překonává ty obvyklé dystopické obrazy umělé inteligence. Myslím, že Ray Kurzweil má pravdu, když tvrdí, že jen co

dokážeme umělou inteligenci aplikovat do lidského mozku a vyvineme čipy umožňující řešení různých handicapů, tak první, kdo bude mít z této technologie prospěch, budou lidé, kteří jsou dnes nepohybliví – mám na mysli ty, kteří trpí *locked-in* syndromem. To bude úžasná věc.

TH: *Někteří z těch, co píší o podobných futuristických technologiích, jsou technooptimisté, jako Tebou zmíněný Ray Kurzweil, jiní jsou technopesimisté, jako Nick Bostrom.³ V jaké části tohoto spektra se vidíš Ty?*

SS: Osobně zastávám „vyčkávací postoj“. Možná bych si měla pro svou pozici vymyslet nějaké exaktnější označení? Když se to pokusím shrnout, tak si prostě myslím, že je ještě příliš brzy na to, abychom mohli odhadnout, jak věci dopadnou. Ve hře je spousta neznámých, takže v budoucnosti se stane mnoho nepředvídatelného jen v důsledku nahodilostí, jež v současnosti nedokážeme předvídat. Takže nevíme, zda technologie umělé inteligence, které dnes vyvíjíme, povedou k dystopiím, které líčí spisovatelé žánru *cyberpunk*. Nevíme tedy, zda nám budou do mozku implantovány čipy kontrolující naše myšlenky; nebo zda se nám může stát, že nebudeme mít peníze na to, abychom si mohli koupit licenci umožňující oživení našich vzpomínek; nebo zda se pokusíme přenést své vědomí na *cloud* a ukáže se, že v důsledku toho přestaneme být vnímatelny bytostmi [*sentient beings*], případně přijdeme o svoji osobnost... Zkrátka, těžko říci. Nevíme, zda nějaké stroje mohou získat vědomou zkušenost [*conscious experience*]. Nevíme, zda firmy vyvíjející umělou inteligenci budou podléhat sociální regulaci, která zajistí, aby tyto technologie skutečně prospívaly lidem. Nevíme ani, jak bude vypadat strategie vývoje superinteligentní umělé inteligence. V současnosti toho hodně visí ve vzduchu, a právě proto je důležité, abychom se o všechny tyto věci zajímali a pokoušeli se je ovlivňovat. Věřím, že se nacházíme v bodu zvratu: pokud uděláme, co budeme moci, abychom se domluvili na povaze vyvíjených technologií, všechno dopadne lépe, než by mohlo.

3 Futuristický vizionář a vedoucí vývojář společnosti Google Ray Kurzweil je znám svými spekulacemi o tzv. singularitě čili zlomovém bodu, v němž dojde k explozivnímu nárůstu schopností umělé inteligence, jež posléze ve všech ohledech překoná tu lidskou. Kurzweil vyložil své vize v knihách *The Age of Spiritual Machines: When Computers Exceed Human Intelligence* (New York, Penguin 1999) a *The Singularity is Near: When Humans Transcend Biology* (New York, Penguin 2005), které nebyly přeloženy do češtiny. V češtině je dostupný pouze jeho návod na prodloužení vlastního života do doby, než bude dostupná nesmrtelnost: *Fantastická cesta. Jak radikálně prodloužit svůj život* (společně s Terry Grossmanem. Přel. B. Grygová. Praha, Anag 2007). Nick Bostrom je švédský filosof působící na Oxfordské univerzitě, propagátor posthumanismu. V českém překladu je dostupná jeho nejznámější kniha *Superinteligence* (Přel. J. Petříček. Praha, Prostor 2017).

Vždycky jsem si přála žít v budoucnosti, snažím se jít jí naproti. Teď například jezdím jedním z těch kvazi autonomních aut *Tesla*. Jeho interiér vypadá jako kosmická loď. Sleduji všechny novinky. Myslím si, že budoucnost se všemi těmi novými technologiemi by mohla být skutečně vzrušující. Zároveň je zde ovšem i hodně důvodů k obavám. Například vůbec nevěřím, že by lidé hájící korporátní zájmy brali významný ohled na zájmy a dobro ostatních. Aféra Cambridge Analytica byla v tomto směru velmi poučná. Bude hodně záležet na penězích, právě ty do značné míry určí, jak věci dopadnou. Což je dost smutné. Už dnes lze na umělé inteligenci tolik vydělat a tolik se do ní investuje, že lidé, kteří se na tom podílejí, mohou bohužel ztratit ze zřetele budoucnost lidstva.

TH: *Jedna z těch budoucích technologií, jež komentuješ ve své nové knize Artificial You,⁴ se týká možného splnutí lidské a umělé inteligence. Jde tu o dva filosofické problémy: osobní identitu a vědomí. Tvrdíš, že by bylo možné empiricky testovat, zda je či není možný transfer vědomí do nějakého jiného substrátu. Píšeš, že si máme představit pacienta, jehož organický mozek je postupně nahrazován křemíkovými čipy, přičemž tento pacient během operační procedury hlásí doktorovi, zda vše probíhá v pořádku. A najednou řekne: „Doktore, právě jsem ztratil část svého zrakového pole.“ Právě to by byl důkaz, že transfer vědomí z organického do anorganického substrátu není možný. Čili bychom mohli takto empiricky zjistit, zda je možný, nebo ne. – Naopak osobní identita je patrně čistě metafyzický problém, takže neexistuje žádný prostředek, kromě filosofických argumentů, jež do diskurzu uvedl Derek Parfit, jak zjistit, zda replikovaná osoba je stále táž, nebo už je někým jiným.⁵ Na rozdíl od transferu vědomí neexistuje žádný test, kterým by to bylo možné zjistit.*

SS: Ano, souhlasím s Tebou. Sama jsem tyhle dva přístupy neporovnávala, ale jak jsi je teď popsal, tak se mi zdá, že máš pravdu. Je totiž možné, že se mýlíme ohledně strojového vědomí, protože je tu ještě další problém, který s ním souvisí, totiž problém druhých myslí [*other minds*]. Já mám vědomí a myslím si, že i Ty máš vědomí, ale je možné, že ve skutečnosti všichni kromě mne postrádají vědomí. O strojích tohle říci můžeme. Avšak v případech lidských bytostí se můžeme spolehnout jen na úsudek vedoucí k nejlepšímu vysvětlení [*inference to the best explanation*]. To znamená, že mohu říci: „Máš podobnou

4 Schneider, S., *Artificial You: AI and the Future of Your Mind*. Princeton, NJ, Princeton University Press 2019. Test transferu vědomí, o němž je řeč, tzv. „čip test“ [*the chip test*], je představen ve 4. kapitole této knihy.

5 Viz Parfit, D., *Reasons and Persons*. Oxford, Oxford University Press 1984, odd. III. Srov. též týž, *Divided Minds and the Nature of Persons*. In: Schneider, S. (ed.), *Science Fiction and Philosophy: From Time Travel to Superintelligence*, c.d., kap. 8.

nervovou soustavu jako já, chováš se podobně jako já, takže máš vědomí jako já.“ Podobně se můžeme domnívat, že lze testovat i přítomnost strojového vědomí, čili že nejlepší vysvětlení chování stroje, který pozorujeme, je to, že má vědomí. Na tom testu, v němž jsou neurony postupně nahrazovány čipy, se mi líbí, že vychází z perspektivy první osoby. Jediné, co je potřeba k jeho úspěšnému provedení, je testovaný člověk, jenž pravdivě hlásí, zda si je toho či onoho vědom. Tímto testem se vlastně pokoušíme zjistit, zda mikročip nahrazuje ty části lidského mozku, které jsou základem vědomé zkušenosti. A pokud v laboratoři nebo prostřednictvím hlášení testovaného subjektu neobjevíme žádný deficit, tak jsme zjistili alespoň to, že právě tento druhu čipu principiálně může být základem vědomí. Takže v takovém případě mohu věřit, že máš vědomí. Avšak v případě stroje musím navíc zjistit, zda jeho vnitřní architektura, která sestává z čipů, vůbec umožňuje, aby nějakým vědomí disponoval.

Sice je tu třeba řešit hodně složitých věcí, ale pořád mi to připadá jako zvládnutelný úkol. Není to tak obtížně řešitelné jako osobní identita. Právě ta je skutečně jedním z nejtěžších problémů, které před námi stojí. V tom máš úplnou pravdu. Připadá mi, že na psychologické kontinuitě je cosi velmi speciálního, ale nevím, zda výhradní lpění na ní je absolutně nutnou podmínkou dalšího postupu. Osobní identita by mohla být prostě fundamentální. Zatím to sice nedokážu analyticky zdůvodnit, ale říkám si, zda by řešení těžkého problému vědomí⁶ [*the hard problem of consciousness*] nemohlo být zároveň řešením i povahy jáství [*the self*]. Takže by možná bylo ontologicky rozumné postulovat cosi jako kontinuitu vědomí ve vesmíru. Vyřešili bychom tím dva problémy naráz.

TH: *Je toto onen idealismus, který v poslední době hájíš?*

SS: Ano, napsala jsem článek, který jsem na počest Jaegwona Kima, z jehož práce jsem měla vždy velké intelektuální potěšení a prospěch, nazvala „Idealism, or Something Near Enough“.⁷ A v tomto článku i v dalších statích jsem se zabývala tím, co je podle mého názoru metafyzicky fundamentální, a s čím tedy musí souviset i řešení psychofyzického problému. Domnívám se, že chceme-li přijít na kloub podstatě psychofyzického problému, musíme vel-

6 Jde o otázku, proč a jak stavy vědomí nabývají fenomenálního charakteru. Ve své knize *The Conscious Mind* (Oxford, Oxford University Press 1996) tento problém zpopularizoval australský filosof David Chalmers.

7 Schneider, S., *Idealism, or Something Near Enough*. In: Goldschmidt, T. – Pearce, K. L. (eds.), *Idealism: New Essays in Metaphysics*. Oxford, Oxford University Press 2007, s. 275–290. Titul článku je narážkou na název monografie korejsko-amerického filosofa Jaegwona Kima *Physicalism, or Something Near Enough* (Princeton, NJ, Princeton University Press 2005).

mi poctivě přemýšlet o základní povaze skutečnosti. Ve svých úvahách při tom vycházím z předpokladu, že tato povaha je kvazi duševní, třebaže odmítám panpsychismus, protože podle mého zachází až příliš daleko. Nicméně jsem přesvědčena, že fundamentální ingredience reality nejsou fyzikální v tradičním smyslu slova, protože prostě z řady důvodů odmítám označení „fyzikální“.

TH: V současnosti je však právě fyzikalismus tou nejakceptovanější metafyzickou koncepcí. Převážná část badatelů dnes tvrdí, že skutečnou pravdu o světě odhalujeme prostřednictvím fyziky a jejích postupů. Je snad právě toto názor, který odmítáš?

SS: Měli bychom si přiznat, že vlastně ani přesně nevíme, co to znamená, když řekneme, že je něco fyzikální. Snad je to možné říci o něčem, co je součástí úplné fyzikální teorie – a k tomu je patrně možné doplnit i nějaké další podmínky, třeba že „je to nementální“. Ovšem při bližším pohledu na současnou fyzikální teorii vyvstávají hlavně problémy, které plynou z rozporu mezi teorií relativity a kvantovou mechanikou. A tady se dostáváme na skutečně nejistý terén, na jehož bázi se celý systém fyzikalismu jeví jako poněkud podivný. Máme tu teorii strun, teorii smyčkové kvantové gravitace a další teorie, jež se pokoušejí sjednotit relativitu s kvantovou mechanikou. Zastánci těchto přístupů naznačují, že časoprostor vlastně není fundamentální veličinou a že základní úroveň reality – alespoň podle nich – sestává pouze ze strukturních vztahů, jež fungují v jisté časoprostorové struktuře, kterou bychom mohli označit jako protoprostorovou [*proto-spatial*] nebo protočasovou [*proto-temporal*].

Jestliže by však z definice fyzikalismu měl vypadnout předpoklad, že fyzično je časoprostorové, tak mi to celé začíná připadat až příliš abstraktní. A vypadá to, že podle těchto teorií je realita opravdu abstraktní. Je to sice pořád ještě jistá forma fyzikalismu, ale pokud věci ve skutečnosti nemají časoprostorový charakter, myslím si, že by někdo měl konceptuálně vyjasnit, co je tedy podle fyzikalistického světonázoru realita. Na to sice můžeš namítnout: „Tak co, tuhle práci je třeba udělat – to je v pořádku.“ Ale nemohu si pomoci: když slyším, že je řeč o něčem, co je protoprostorové a protočasové, tak mi na mysli vytane panprotopsychismus [*panprotopsychism*]. Představuji si pak realitu jako specifický základ figurující v těchto teoriích kvantové gravitace, které evokují nikoli fyzické komponenty, ale protoprostorové a proto-mentální ingredience.

Mimochodem, já jsem monistka, takže všechny tyto věci jsou podle mého názoru téhož druhu. Nicméně dneska se člověk ve fyzice setkává s téměř pythagorejskými teoriemi – a podobné přístupy jsou k vidění i ve filosofii: lidé

tvrdí, že vše je informace. Pro mě však reálná otázka zní: „Co vdechne život rovnicím?“⁸ Je to otázka, kterou si položil i Steven Hawking – a já si myslím, že odpovědí na ni je, že fundamentální ingredience jsou buď protofyziké, nebo protomentální, ale že existují věci jen jednoho druhu. Víím, že to, co právě říkám, v tuto chvíli ještě není propracovaný argument, je to zatím pouze předpoklad, ale nedávno jsem měla příležitost o tom blogovat.

TH: „Vdechnout život rovnicím“ – není to snad jen metafora?

SS: Ano, je pravda, že i některé významné osobnosti, například známí filosofové mohou říkat absurdní věci. Vezmi si třeba Davida Lewise a ta jeho „possibilia“... Moje kamarádka jednou vešla do módního butiků, a co tam nevidí: citáty z Quina na stěnách. Smály jsme se tomu: na ramínkách tam visely šaty a za nimi si nakupující mohli číst o „síti přesvědčení“ atd.; bylo to velmi legrační. Takže ano, máš pravdu, to, co jsem řekla předtím, opravdu nebyl solidní argument, který by dostatečně osvědčoval můj výchozí předpoklad. Ale už jsem se jím vážně zabývala a podrobněji ho zformulovala. Můžeš ho najít v mém článku, který zahajoval nedávné číslo *Journal of Consciousness Studies*.⁹ Týkal se matematické povahy fyziky a rozvedla jsem v něm svou tezi, že skutečným problémem fyzikalismu dnes je, že není jasné, co by měl být pravdivý [truthmaker] výroků fundamentální fyziky. Obecně vycházím z předpokladu, že fyzika má velmi matematickou povahu.

Trochu to souvisí s tématem, o němž jsme před chvílí hovořili, ale přece jen je to poněkud odlišné. Otázka, o níž tady jde, zní takto: Co plyne z faktu, že podíváme-li se na úplně fundamentální fyziku, tak jsou to samé rovnice? Podle mě jde o to, abychom pochopili, že máme-li individuovat, resp. identifikovat vlastnosti a substance, které vyplývají z fyzikální teorie, musíme porozumět povaze čísel. A to zasahuje do problematiky filosofie matematiky, jež je mi velmi blízká. O povaze čísla v matematice se dnes vedou skutečně pozoruhodné diskuse a mně připadá, že jejich sledování by mělo být vrcholně zajímavé právě pro fyzikalisty. Účastníci těchto diskusí mají za sebou už hodně práce, o níž by se fyzikalisté mohli přirozeně opřít a dobrat se k rozumnému pojetí povahy čísla. Jenže je tu jeden velký problém: závěry, které z těchto diskusí plynou, s dnešním fyzikalismem ve skutečnosti vůbec neladí. Takže osobně si myslím, že prostředky fyzikalismu jsou mnohem omezenější, než

8 Narážka na pasáž ze „Závěru“ populární knihy britského fyzika Stephena Hawkinga *Stručná historie času* (Přel. V. Karas. Praha, Mladá fronta 1991, s. 185): „I v případě, že pouze jediná teorie bude s to popsat přírodu, je to stále jen soubor pravidel a rovnic. Co jim však vdechne život a stvoří vesmír, o kterém hovoří?“

9 Schneider, S., Does the Mathematical Nature of Physics Undermine Physicalism? *Journal of Consciousness Studies*, 24, 2017, No. 9–10, s. 1–33.

se obecně předpokládá. Fenomén vědomí mě navíc utvrzuje v přesvědčení, že se tu děje něco zvláštního.

TH: *Znamená to, že bereš fenomenální vědomí jako danost?*

SS: Určitě. Domnívám se, že eliminativisté zacházejí příliš daleko – a navíc, že stejně implicitně předpokládají vědomí. Divím se, že vůbec potřebujeme vést diskusi o eliminativistické pozici (*smích*). Vědomí je přímo dostupné prostřednictvím introspekce a zdá se mi, že je velmi odlišné od čehokoli jiného, co existuje ve světě kolem nás. Vědomí je skutečně hluboká záhada a já si nejsem jistá, zda jsme schopni ji vyřešit. Můj vlastní pokus o řešení tohoto problému spočívá v předpokladu, že na dně leží fundamentální fenomény, které identifikuje fyzika. Jejich základní povaha ale nemůže být časoprostorová; pouze dávají vzniknout časoprostoru a z definice jsou tím, co generuje jak časoprostor, tak mentalitu – to je ve zkratce definice toho, čím tyto entity jsou podle panprotopsychemu. Jestliže postulujeme tyto základní ingredience, pak tím dokážeme vysvětlit mnohé z fenoménů, jež existují kolem nás.

TH: *V této souvislosti považuji za velmi případné zmínit názor jiného amerického filosofa, s nímž jsem hovořil: Alexe Rosenberga. Na rozdíl od Tebe je přesvědčeným fyzikalistou, ale při tom si je také vědom problémů, které pro pozici, již obhájuje, představuje matematika. Nicméně tvrdí, že matematika je problém pro každého, takže fyzikalismus na tom není o nic hůř než jiné metafyzické pozice. A co se týče vědomí, tak svou pozici nedávno jasně zformuloval v replice na Searlův „Argument čínského pokoje“.¹⁰ Rosenbergova pozice je reductio ad absurdum tohoto argumentu. Jak známo, Searle tvrdí, že v počítači je jen syntax, a žádná sémantika, která je nutnou podmínkou fungování mysli. Rosenberg to akceptuje, a v podstatě říká: „Tím hůř pro sémantiku, a je-li sémantika nutná pro mysl, tím hůř pro mysl.“ Ty akceptuješ Systémovou odpověď na „Argument čínského pokoje“, čili vnímáš sémantiku na této úrovni. Moje otázka zní: Chápeš sémantiku jako danost, jako fenomén, který má být vysvětlen, nikoli eliminován, podobně jako chápeš jako dané vědomí?*

SS: Tohle je tak zajímavé! Je to velmi podnětná otázka, protože když jsem uvažovala o Systémové odpovědi, tak jsem měla na mysli rozumění. Rozu-

10 Argument čínského pokoje viz v: Searle, J., *Mysl, mozek a věda*. Přel. M. Nekula. Praha, Mladá fronta 1994. Rosenbergovu repliku viz v: Rosenberg, A., *How History Gets Things Wrong: The Neuroscience of Our Addiction to Stories*. Cambridge, Mass., The MIT Press 2018, kap. 9. Rozhovor s Rosenbergem vyjde pod názvem „O ateismu, scientismu a „manifestním obrazu““ v: Hříbek, T., *Jak být dobrý i bez pána baha a další eseje o náboženství, moralce a vědě ve svobodné společnosti*. Praha, Academia 2021.

mět totiž znamená znát pravdivostní podmínky, a k tomu, abychom se k nim dostali, myslím stačí pouze syntaktický stav mozku. Takže jsem neuvažovala o mapujícím vztahu mezi myslí a realitou. V době, kdy jsem psala knihu na toto téma, *The Language of Thought*,¹¹ jsem byla externalistkou. Externalismus byla teorie o povaze mentálního obsahu a o tom, jak lze vyvodit obsah z kauzálních a nomologických vztahů ke světu. Byla jsem žákyňou Jerryho Fodora a převzala tuto pozici od něj. A pořád k ní mám blízko, i když bych ani na vteřinu nechtěla naznačovat, že význam je jednoznačný pojem. Domnívám se totiž, že Putnam byl ve svém slavném eseji „The Meaning of „Meaning“¹² až příliš optimistický, protože u pojmu „význam“ lze pravděpodobně rozlišit tolik různých typů...

TH: ... nejen intenzi a extenzi.

SS: Ano, přesně tak. V každém případě mě Systémová odpověď zajímala hlavně proto, že jsem se domnívala, že mozek je složen z různých komponent, z nichž ovšem žádná není vědomá, nebo přinejmenším není vědomá na té úrovni, na jaké jsou si samy sebe vědomy osoby. Takže lze postavit vědomý systém z nevědomých komponent, a to podle mne stačí k tomu, abychom vrátili Searlovo tvrzení, že není možná umělá inteligence, která buď dokáže rozumět, nebo má vědomí.

TH: *Dobrá, rád bych se teď zeptal na Tvůj názor na vztah mezi filosofií a vědou. Hovořil jsem o něm i s fyzikem Lawrenceem Kraussem, jenž zastává svého druhu pozitivistický postoj a domnívá se, že pokrok vědy vytlačuje filosofii. Připouští sice, že možná ještě ve studiu vědomí existuje prostor, kde se může uplatnit filosofie, ale je přesvědčen, že pokud jde o fyziku, byla zcela ovládnuta vědou. Ve svém bestselleru Vesmír z ničeho¹³ v podstatě tvrdí, že vědeckými prostředky rozlousknul starou filosofickou hádanku, otázku, jak něco může vzniknout z ničeho.*

SS: To není uspokojivá odpověď. Domnívám se, že dokonce i fyzika musí postulovat nějaké primitivní danosti. Je mi nepříjemné, když se vědci tváří, jako

11 Schneider, S., *The Language of Thought: A New Philosophical Direction*. Cambridge, Mass., The MIT Press 2011.

12 Česky viz Putnam, H., Význam slova „význam“. Přel. O. Tomala. In: Peregrin, J. (ed.), *Logika ve filosofii, filosofie v logice*. Praha, *Filosofia* 2006, s. 259–323.

13 Krauss, L. M., *Vesmír z ničeho. Proč existuje něco, místo aby neexistovalo nic*. Přel. J. Langer. Praha, Knižní klub, 2013. Viz též rozhovor T. Hříbka s L. Kraussem: *Vesmír smysl nemá, naše existence však ano*. *Lidové noviny*, 26. 4. 2015, s. 21–22 (vyjde rovněž znovu v: Hříbek, T., *Jak být dobrý i bez pánaoha a další eseje o náboženství, morálce a vědě ve svobodné společnosti*, c.d.).

by filosofické otázky byly směšné a jako by byli schopni odpovědět na ně lépe. Považuji to za vrchol arogance. Na druhé straně si kontakt s vědci nesmírně užívám. Hovořím nyní pravděpodobně o filosofii více s vědci než s ostatními filozofy. Při svých pokusech fušovat do astrofyziky a astrobiologie, jsem potkala mnoho zvědavých duchů, které těší rozhovor s filosofkou a kteří si uvědomují, že hovoří-li o těchto otázkách, tak hovoří o filosofii.

TH: Vlastně jsem Ti tuhle otázku položil proto, že na mě udělal dojem postřeh o vztahu mezi vědou – zvláště fundamentální fyzikou – a filosofií, který jsi zmínila během naší dřívější konverzace. Naznačila jsi totiž jistý přístup, který je úplným opakem Kraussova pojetí. Krauss si myslí, že vědci nahradili filozofy, kdežto Ty jsi řekla, že ve fundamentálním výzkumu, který se odehrává například v kosmologii, je tomu právě naopak. Vědci sice mají k dispozici technický jazyk matematiky, ten ale nestačí k tomu, aby si s otázkami na této úrovni dokázal poradit.

*SS: Ano. Jako příklad je možné uvést třeba určitý typ publikací o teorii strun, v nichž se analyzuje emergence časoprostoru. Kdyby se jejich autoři-vědci více zabývali filosofickými problémy vztahujícími se k této oblasti, například obsahem pojmu emergence či otázkou, jak dostat časoprostorovou strukturu z nečasoprostorových entit, mohlo by to být jen ku prospěchu věci. Znam hodně vědců, hodně fyziků, kteří si to také myslí, ale obávám se na druhé straně, že se současná filosofie stala příliš nesrozumitelnou, a to nefilozofy odrazuje. Nedávno jsem o tom hovořila s Freemanem Dysonem, a ten mi tento dojem také potvrdil. Říkal, že takoví filozofové, jako byl například Descartes, byli důležití nejen pro filozofii, ale i pro rozvoj matematiky (konkrétně diferenciální počet), anatomii atd. V dnešní době prohlubující se specializace ovšem není realistické očekávat, že by filozofové tohle všechno zvládli. Já si ale myslím, že skutečný problém spočívá v tom, že lidé v rámci našeho oboru hovoří jen mezi sebou navzájem a noří se do terminologie, jíž rozumí jen oni sami. Proto nedokáží zaujmout vědce, i když ti to velmi potřebují a mnozí by to i uvítali. Je nesmírně těžké pochopit rozdíl mezi vědeckým popisem a hodnocením a mezi vědeckým a filosofickým tvrzením – a přitom to ve fyzice či informatice nikdy nebylo důležitější než dnes. V současnosti také slyšíme mnohé autority v oboru informatiky hovořit o tom, čeho všeho bychom mohli dosáhnout po splynutí s umělou inteligencí, po *uploadu* mysli – ale to jsou přece odvěké filosofické otázky. A jsem přesvědčená, že jen díky statusu, slávě a bohatství těchto vědeckých autorit mají lidé tendenci neochvějně přijímat jejich tvrzení jako pravdivá. Lidé mají sklon tato tvrzení přijímat, protože si neuvědomují rozdíl mezi vědou a filosofií. My filozofové se proto musíme snažit, aby naše práce byla dostatečně interdisciplinární a zaujala veřejnost. Odpovídáme za to, aby se tyhle věci lidem ujasnily. Pevně věřím, že*

nás filosofů je zapotřebí. Takoví autoři, jako je Lawrence Krauss, mi připadají úplně mimo. Když formuluje ta svá tvrzení, tak ve skutečnosti zaujímá určitou metafyzickou pozici, jejímž jádrem je sdělení: „Toto je vše, co existuje.“ Prostě to nevyplývá, že by dělal čistou vědu.

TH: *Jiní současní myslitelé, s nimiž jsem hovořil, například Daniel Dennett nebo Alex Rosenberg, byli velmi výmluvní ohledně tématu náboženství. Otázku vztahující se k náboženství mám i pro Tebe. Do antologie Science Fiction and Philosophy: From Time Travel to Superintelligence,¹⁴ kterou jsi uspořádala, jsi zařadila i slavný esej Nicka Bostroma, v němž uvažuje o tom, že jsme možná součástí počítačové simulace navržené nějakými superinteligentními bytostmi. Bostrom mimo jiné naznačuje, že pokud takový scénář akceptujeme, pak bychom si z něj mohli vzít ponaučení, jak vést svůj vlastní život. Konkrétně uvažuje o možnosti, že stvořitelem naší simulace je vzdálený potomek křesťanských fundamentalistů. V takovém případě bychom měli vážně uvažovat o možnosti, že skončíme v pekle apod. A to mě přivádí k otázce: Myslí to podle Tebe Bostrom vážně? Už dnes žijeme ve světě, v němž je náboženství buď „na odchodu“, nebo tomu tak je přinejmenším v té nejrozvinutější části světa. Nebude pro superpokročilé designéry celých vesmírů náboženství dávná prehistorie?*

SS: Nejsm si jistá... Tvůrci simulovaných vesmírů by mohli mít své vlastní náboženství. Myslím si, že když jde o Boží existenci, tak toho nedokážeme ani moc dokázat, ani moc vyvrátit. Máme tu ohromné dějiny pokusů o důkaz Boží existence, a na druhé straně je tu ten strašný problém zla, který všechny trápí. Možná že existuje nějaký nevysvětlitelný element skutečnosti, Bůh, a my ho nechápeme jednoduše proto, že je superinteligentní.

TH: *Dobrá, chápu, že nějaký druh spirituality by mohl přetrvat i u velmi pokročilých bytostí. Ale domnívám se, že by šlo o něco velmi odlišného od primitivního křesťanského fundamentalismu, který si vymyslel Stvořitele, jehož zajímá náš sexuální život a který nám hrozí trestem, pokud porušíme pravidla jeho sexuální morálky. Takže je mi zatěžko uvěřit, že nějaká o miliardy let pokročilejší civilizace, schopná simulovat celé vesmíry, by ještě inklinovala k podobné víře. To se mi zdá přehnané.*

SS: Aha, to je pravda, s tím souhlasím.

14 Schneider, S. (ed.), *Science Fiction and Philosophy: From Time Travel to Superintelligence*, c.d. Viz tamtéž zvláště esej Nicka Bostroma „Are You a Computer Simulation?“ (s. 22–25).

TH: *V této souvislosti bych se Tě také rád zeptal, jak nahlížíš na vztah mezi transhumanismem a náboženstvím? Některé proudy v transhumanismu jsou náboženské, že?*

SS: Ano. Sama jsem transhumanistka a vzpomínám si, jaké to bylo, když jsem poprvé četla Raye Kurzweila... Ptám se sama sebe, zda v době, kdy jsem byla mladší, mě transhumanismus nepřitahoval proto, že mi připadal skoro jako náhrada za Boha – v tom smyslu, že sliboval technologickou transcenci a vyšší formu vědomí. Lidé, kteří působí v různých humanitních a sociálních vědách, poukazují na křesťanské a platónské podtóny transhumanistického myšlení. Dokonce i Kurzweil ve svých knihách věnovaných singularitě místy hovoří o Novém Nebi. Zkrátka: kdo by nechtěl žít navěky? Je to lákavá vidina – *upload* vědomí nebo kryogenické zmražení. Idea, že bychom mohli být nějak technologicky znovuoživeni. Touha žít déle, ne-li navždy, je přirozená. Víím, že otázka, zda *skutečně* chceme žít navěky, je předmětem diskuse. Osobně bych ráda žila alespoň dlouho.

TH: *Nebo bys ráda měla možnost rozhodnout se, kdy chceš přestat žít?*

SS: Ano, a ráda bych měla věcírek na rozloučenou!

TH: *Poslední téma. Když jsem četl Tvé odborné stati a Tvou knihu Artificial You, měl jsem dojem, že se soustřeďuješ na hrozby spojené s futuristickými technologiemi, které plynou z těchto technologií samých – např. na nebezpečí, jež hrozí lidem, kteří naivně věří, že prodlouží svůj život splynutím s AI, ačkoli ve skutečnosti spáchají technologickou sebevraždu. Ovšem řada jiných odborníků si myslí, že tyto technologie nejsou hrozbami ze své vlastní podstaty, nýbrž proto, že je v našem světě vlastní a užívají mocné soukromé společnosti. Víím, že tato otázka patří do oblasti sociální filosofie, jíž se primárně nevěnuješ. Byl bych nicméně rád, kdybys mi rekla, co si o tomto problému myslíš.*

SS: Těmto otázkám se velmi dobře věnují autoři, kteří píší v žánru *cyberpunk*, a já mám pocit, že nemám moc co dodat k tomu, co již bylo řečeno...

TH: *O těchto tématech snad ale mají co říci i někteří marxisté, resp. autoři z oblasti kritické teorie, kteří se zaměřují na vědu a techniku, nebo ne? A všiml jsem si, že i Ty jsi na začátku našeho rozhovoru zmínila obavu z jednání takových firm, jako je Cambridge Analytica...*

SS: Určitě! Ano, zmiňuji se o těchto věcech, zejména když hovořím s médii, a to kvůli své sociální odpovědnosti. Nepovažuji však jejich analýzu za svůj

originální příspěvek na poli vědeckého bádání. Pokud jde o umělou inteligenci, pokouším se vysvětlit, že rozkvět lidstva závisí na takových věcech, jako je mysl, vědomí, já, a že by ze strany vědců – případně ze strany zástupců médií, veřejnosti nebo politiků – bylo vrcholně arogantní, kdyby tvrdili, že lze vytvořit vědomou bytost pouze tím, že vložíte mikročip do mozku; nebo že když sestrojíte robota, tak tím sestrojíte vědomou bytost; případně že lidská bytost může přežít nahrazení mozku umělými komponentami. Takže si myslím, že máš pravdu, když říkáš, že na osobní identitě je cosi bytostně záhadného. Možná je to cosi, nač nikdy nenajdeme odpověď. Podobně asi nikdy nebudeme schopni najít důkaz, který by definitivně objasnil, jak se to ve skutečnosti má s mozkiem v kádi [*brain in a vat*].¹⁵ A myslím si, že mi stačí princip nejlepšího vysvětlení u problému druhých myslí, ačkoli se obvykle vyhýbám řešením jakýchkoli otázek, které se týkají skepticismu.

TH: *Skvělé, Susan, díky za Tvé odpovědi.*

15 Mozek v kádi je inovací klasické skeptické pochybnosti Reného Descarta týkající se našeho poznání vnějšího světa. Zatímco Descartes se snažil ukázat, že má představa o materiálním světě – včetně mého těla – by mohla být pouhou iluzí, kterou mi vnutil klamavý démon, v současné verzi jde o to, že táž představa by mohla být iluzí pouhého mozku v kádi, který je klamán výkonným počítačem. Viz Putnam, H., *Reason, Truth and History*. Cambridge, Cambridge University Press 1981, kap. 1.