Determinismus a svobodná vůle – postoje Rogera Penrose a Stephena Wolframa

Toto pojednání se zabývá otázkou determinismu – tzn. předurčenosti výsledku veškerých procesů. Představím zde několik názorů na tuto problematiku a zmíním hlavní argumenty jejich zastánců. Podrobně rozeberu dílo Stephena Wolframa a Rogera Penrose. Ve druhé části práce se pak budu zabývat tím, jaké důsledky by přijetí determinismu mělo v jiných oblastech, zejména v souvislosti s otázkou volního rozhodování a vědomí.

Úvod

Determinismus je filozofický názor, který tvrdí, že pro každou událost existuje příčina, která ji způsobila a tato příčina nemohla ani způsobit nic jiného. Každá událost je tedy jednoznačně určená (determinovaná) předchozími událostmi.

Fyzikální determinismus je postavený na dvou základních pilířích:

1. *Redukcionismus*: veškeré makroskopické jevy ve fyzikálním světě je možné v principu popsat pomocí fyzikálních zákonů mikrosvěta
2. *Determinismus mikrosvěta*: fyzikální zákony mikrosvěta jsou deterministické, tedy výsledný stav každého mikroskopického procesu je určen jednoznačně svým počátečním stavem.

Představa redukcionismu je vnitřně blízká mnoha běžným lidem i vědcům[1]. Zvykli jsme si, že pohyb letadla ve vzduchu můžeme vysvětlit pomocí proudění částic, komplexní sloučeniny vznikají z jednodušších pomocí chemických vazeb založených na sdílení elektronů a podobně. Představa, že fyzika elementárních částic může vysvětlit makroskopické jevy se tak může zdát realistická. Zastánci redukcionismu, mezi něž patří například známý fyzik Stephen Hawking, věří, že biologické jevy lze redukovat na jevy chemické a jevy chemické dále na jevy fyzikální. Tento postup lze aplikovat i na člověka a veškeré lidské chování i ostatní projevy tak redukovat na fyzikální zákony.

Stejně tak myšlenka determinismu fyzikálních zákonů mikrosvěta je pro většinu lidí intuitivně přijatelná. Naše běžná zkušenost říká, že pokud nějaký proces zopakujeme s naprosto totožnými počátečními podmínkami, pak výsledek bude stejný a to nejen na mikroskopické úrovni. Nikdo se nepozastavuje například nad scénami ze sci-fi filmů, kde se hrdina vrátí zpátky v čase a všechno se odehrává stejně jako poprvé.

Determinismus má však několik následků, které – pokud ho přijmeme – musíme také akceptovat. Mezi nejčastěji zmiňované patří absence svobodné vůle a ztráta morální odpovědnosti za své činy.

Pojem determinismu je širší než jak jsem ho zde popsal a lze ho aplikovat i v mnoha jiných odvětvích než ve fyzice.[2] Je možné s ním pracovat i bez vztahu k redukcionismu a fyzikalismu[3] vůbec (v tom případě bychom místo „fyzikální zákony mikrosvěta" říkali pouze „přírodní zákony"). Zastánci těchto názorů však

[1] například fyzikové sní o vytvoření tzv. sjednocené teorie („teorie všeho"), která by dokonale popisovala všechny fyzikálně měřitelné jevy ve vesmíru.
[2] Teologický determinismus (Dvořák, 2009), Fatalismus, Technologický determinismus a jiné.
[3] Fyzikalismus tvrdí, že existují pouze entity obsažené ve fyzikálních teoriích, a že vědecký rozvoj nakonec umožní vysvětlit mysl v pojmech této teorie. Popírá tedy existenci nefyzických entit, například spirituálních

ve své argumentaci redukcionistické postupy většinou aspoň částečně používají, proto jsem determinismus zavedl v této podobě.

Svobodná vůle

Svobodná vůle je pojem dosti zvláštní, skoro až mystický. Většina lidí věří, že ji má, ale je těžké ji formálně definovat. Stejně tak je problematické tvrdit, že nějaký jiný subjekt – ať už člověk, zvíře nebo věc – svobodnou vůli má, či nemá. Při podrobnějším zkoumání této otázky narazíme na pojmy jako *vnímání, vědomí, konání* a také *čas* (zejména *jednosměrnost času*).

Že to se svobodnou vůlí není tak snadné, můžeme ilustrovat na těchto příkladech:

- Host v restauraci si prohlédl jídelní lístek a objednal si pizzu. *Bylo to jeho svobodné rozhodnutí?*
- Řidič auta zastavil na semaforu, kde svítila červená. *Bylo to jeho svobodné rozhodnutí?*
- Buňka vstoupila do rozmnožovací fáze a vytvořila svou kopii. *Bylo to její svobodné rozhodnutí?*
- Strom na podzim shodil listy. *Bylo to jeho svobodné rozhodnutí?*
- Oheň spálil dřevěný špalek. *Bylo to jeho svobodné rozhodnutí?*
- Elektron oběhl kolem jádra atomu. *Bylo to jeho svobodné rozhodnutí?*

Vědci, kteří s pojmem svobodné vůle pracují, ji definují několika způsoby. Například:

„Svobodná vůle je zdánlivá kapacita agentů činit rozhodnutí svobodná od jakýchkoliv omezení.“

Nebo:

„Svobodná vůle je specifický druh kapacity racionálního agenta vybírat směřování svých akcí z množství různých alternativ.“

Tímto tématem se filozofie i jiné vědy zabývají už od pradávna. Dříve to bylo zejména pomocí myšlenkových experimentů, kde se řešily otázky etiky a morálky, dnes už můžeme vědomí a svobodnou vůli zkoumat i experimentálně pomocí nástrojů psychologie a neurověd[4]. K tématu svobodné vůle se vyjadřovaly a vyjadřují mnohé disciplíny: od filozofie, psychologie, biologie, kognitivní vědy, přes fyziku až po teologii. Tato zkoumání přinesla mnoho poznatků, názorů a souvislostí, jejichž přehled by dalece přesahoval rámec této práce. Zmíním zde pouze názory Rogera Penrose a Stephena Wolframa a pokusím se je zasadit do širšího kontextu.

Různé postoje k předurčenosti

K problematice předurčenosti můžeme zaujímat několik různých postojů a determinismus je jedním z nich. Klasické rozdělení názorů v souvislosti se svobodnou vůlí vypadá následovně:

	Svobodná vůle neexistuje	Svobodná vůle existuje
Svět funguje deterministicky	Tvrdý determinismus	Kompatibilismus
Svět nefunguje deterministicky	Tvrdý indeterminismus	Libertarianismus

[4] Viz např. známý experiment Benjamina Libeta

Tvrdý determinismus předpokládá deterministické fungování světa a nepřipouští existenci svobodné vůle. Bývá kritizován za to, že v podstatě odmítá etiku a morální odpovědnost jedince.

Indeterminismus naopak předpokládá, že vývoj světa je daný pouze pomocí pravděpodobností a není determinovaný, ale člověk ani jiné organismy nemůžou tento běh svou vůlí nijak ovlivňovat.

Kompatibilismus uznává deterministické fungování přírody, ale současně vidí i prostor pro svobodnou vůli jedinců. V tomto pojetí je to však spíše „iluze svobodné vůle“, protože výsledek je určený jednoznačně. Odpůrci toho směru často poukazují právě na nevhodnou definici svobodné vůle s tím že „iluze svobodné vůle není svobodná vůle“.

Libertarianisté svobodnou vůli uznávají a tvrdí, že je neslučitelná s determinismem, který proto odmítají. Fungování světa vysvětlují buď jako nefyzikalistické – tzn. odmítají materialismus s tím, že mysl přesahuje fyzikální realitu, nebo fungování pravděpodobnostní, kde jedinci obdaření svobodnou vůlí můžou tyto pravděpodobnosti ovlivňovat.

Autoři Wolfram a Penrose mají svými názory nejblíže ke kompatibilismu.

Postoje Stephena Wolframa

Stephen Wolfram je matematik, fyzik a informatik, zakladatel firmy Wolfram Research, která vyvíjí software Mathematica a tzv. „odpovídací stroj“ Alpha[5]. Své názory na otázky týkající se fungování přírody sepsal v rozsáhlé knize „A New Kind of Science“ (Wolfram, 2002).

Wolfram je zastáncem determinismu. Věří, že se přírodní zákony na nejnižší úrovni dají popsat algoritmicky pomocí programů. Pro popis těchto programů využívá Wolfram tzv. celulární automaty, které zde v krátkosti představím.

Celulární automaty

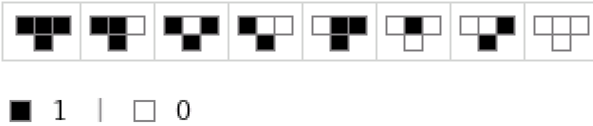
Celulární automat sestává ze struktury „buněk“, které se mohou nacházet v různých stavech a přechodového pravidla, které jednoznačně určuje, jak se budou stavy těchto buněk měnit. Nový stav buňky je určený současným stavem této buňky a také stavy okolních buněk.

Uvažme následující příklad: automat je tvořený jednou řadou buněk, které můžou být ve dvou stavech – 0 a 1. Stav 0 budeme nazývat stav pasivní a 1 stav aktivní. Buňky budeme znázorňovat jako čtvercová políčka a stav buněk zobrazíme pomocí barev – buňka ve stavu 0 bude znázorněná bílou barvou a ve stavu 1 černou barvou.

Vývoj automatu probíhá po krocích následovně: buňka se „podívá“ na svůj současný stav, na stav sousední buňky nalevo a sousední buňky napravo. Na základě těchto tří informací určí svůj stav v následujícím kroku pomocí přechodového pravidla. Tuto operaci provedou všechny buňky současně a tak vznikne nový stav automatu. Celá procedura se opakuje a vzniklá posloupnost stavů ukazuje vývoj automatu v čase.

[5] http://www.wolframalpha.com/

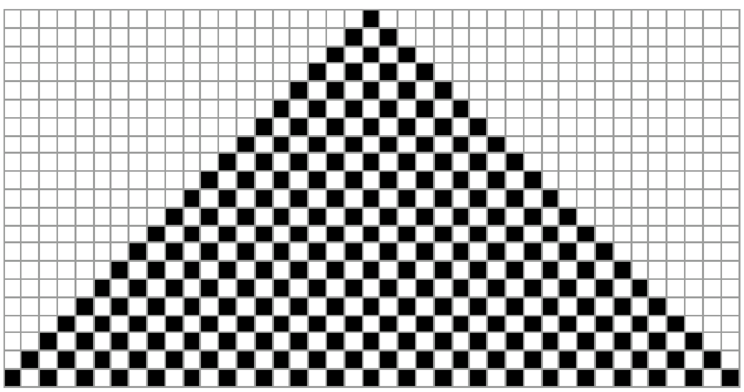
Přechodové pravidlo automatu lze zobrazit v následující podobě:



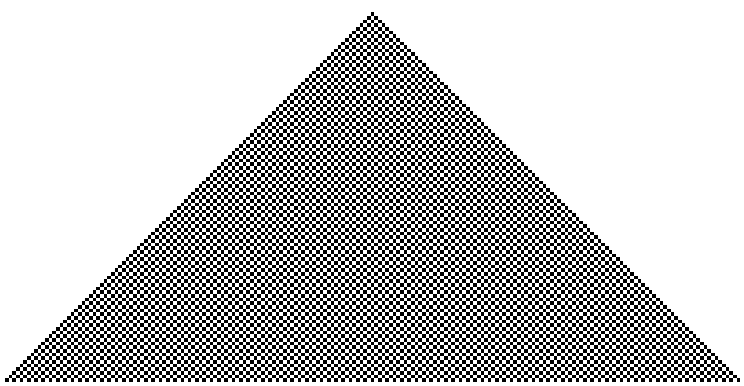
Horní tři políčka ukazují stav buňky nalevo, uprostřed a napravo (všech 8 možných konfigurací). Spodní políčko pak ukazuje nový stav centrální buňky pro každou z těchto konfigurací. Přechodové pravidlo je pro všechny buňky stejné. Toto konkrétní přechodové pravidlo se dá slovně popsat jako „buňka bude v dalším kroku aktivní právě tehdy, když aspoň jedna sousední buňka je aktivní“. Všech možných přechodových pravidel je 256 a můžeme je jednoznačně očíslovat. Toto konkrétní pravidlo budu nazývat „pravidlo 250“.

Vývoj automatu pak lze přehledně zobrazit na čtvercové mřížce, kde na prvním řádku bude zadaný počáteční stav (tedy stav v kroku 1), a na dalších řádcích postupně následující stavy automatu po aplikacích přechodového pravidla. Na i-tém řádku bude stav automatu v i-tém kroku.

Vývoj automatu, kde v počátečním stavu je právě jedna buňka aktivní a přechodové pravidlo je 250 vypadá následovně:

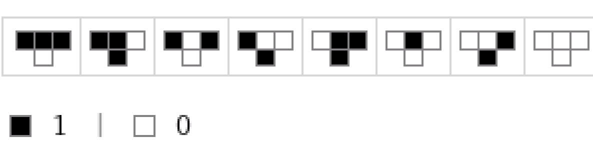


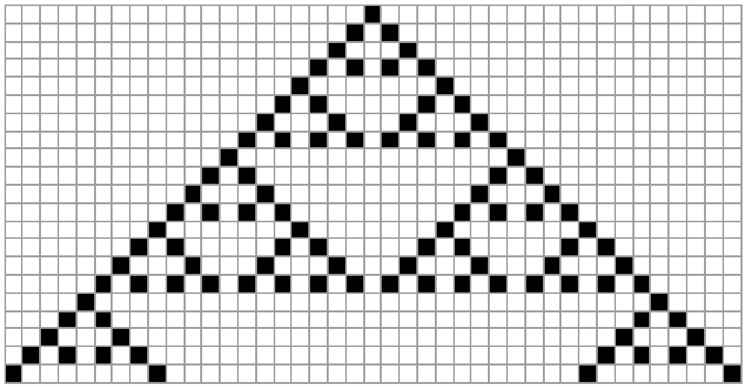
20 steps



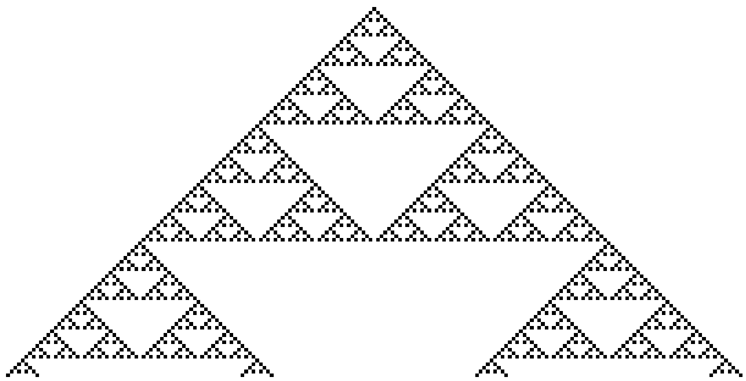
100 steps

Vývoj automatu silně závisí na použitém přechodovém pravidle. Kupříkladu automat s pravidlem 90 se chová následovně:



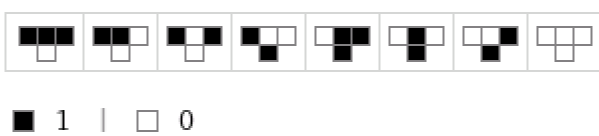


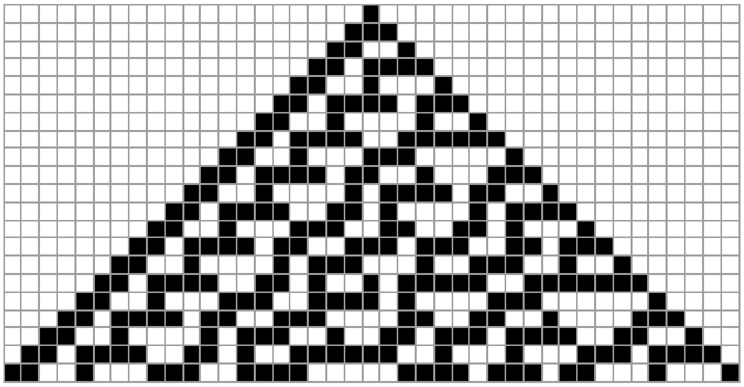
20 steps



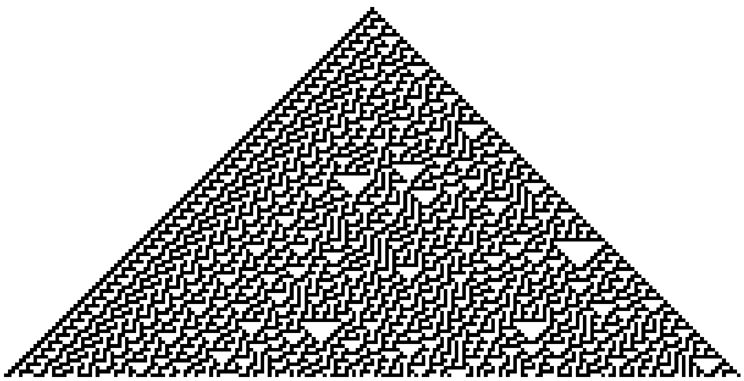
100 steps

A s pravidlem 30 takto:





20 steps



100 steps

Chování každého takového automatu je plně deterministické – pokud známe počáteční stav (nebo kterýkoliv stav v průběhu) a přechodové pravidlo automatu, můžeme přesně určit celý následující vývoj automatu.

Výpočetní síla celulárních automatů

Jedním z důvodů, které přivedly S. Wolframa k myšlence použití celulárních automatů pro vysvětlení chování přírody bylo zjištění, že tyto automaty jsou schopné provádět univerzální výpočetní operace. Bylo dokázané, že automat s pravidlem 30 dokáže simulovat Turingův stroj, což znamená, že dokáže „spočítat“ všechno, co lze spočítat na jakémkoli počítači.

Pokud například do stavu automatu vhodným způsobem zapíšeme[6] přirozené číslo k a necháme automat běžet po dostatečně dlouhou dobu, pak jistá buňka automatu se stane aktivní pokud číslo k je prvočíslo a bude neaktivní, pokud číslo k prvočíslem není. Automat tak dokáže řešit problém testování prvočíselnosti.

Pokud zakódujeme (jiným způsobem než v předchozím případě) do počátečního stavu automatu dvě čísla x a y, pak stav automatu po dostatečně velkém počtu kroků bude kódovat číslo x+y. Automat je tedy schopný provádět součty.

Takto bychom mohli pokračovat dál. Zmíním pro představu ještě jeden příklad. Existuje způsob, jak do počátečního stavu automatu zakódovat šachovou pozici tak, že po určitém počtu kroků bude nový stav automatu kódovat tah, který je v této pozici nejlepší možný.

[6] Přesněji řečeno „zakódujeme“. Kromě čísla bychom samozřejmě museli zapsat i algoritmus, který má automat používat. Veškeré konečné objekty lze vždy kódovat pomocí posloupnosti nul a jedniček například tak, že objekt popíšeme na počítači v textovém souboru a tento soubor uložíme. Na disku je informace zapsaná právě pomocí posloupnosti nul a jedniček.

Automat je tedy jen „jeden řádek barevných políček“, který se mění podle triviálních pravidel, ale přesto dokáže provádět libovolné matematické operace, porazit mistra světa v šachu a vůbec provést jakýkoliv výpočet, který člověka napadne[7]. Mohl by tedy v principu vytvářet i naši velmi složitou fyzikální realitu.

Svět jako celulární automat?

Zmíněné příklady automatů pracovaly pro jednoduchost pouze s jednou řadou buněk, byly tedy jednorozměrné. Je však možné si představit automat pracující na dvourozměrné čtvercové mřížce, kde nový stav buňky je opět určen stavem okolních buněk, kterých je v tomto případě víc než 2[8]. Stejně tak lze zavést trojrozměrný automat, kde buňky budou odpovídat krychličkám.

V trojrozměrném automatu pak můžeme jednotlivé buňky přirovnat k miniaturním oblastem našeho prostoru a aktivní buňka by označovala, že v tomto prostoru probíhá jistá elementární „aktivita“[9]. Větší skupiny buněk by pak odpovídaly větším oblastem prostoru a celkový stav automatu by reprezentoval stav celého vesmíru. Změny stavů automatu v jednotlivých krocích by představovaly běh času.

Tento výpočetní pohled na fungování světa je některým fyzikům velmi blízký (např. James P. Crutchfield) a existují výzkumné skupiny, které se snaží takový automat skutečně navrhnout a ukázat, že dobře simuluje fyzikální realitu a jeho predikce jsou v souladu s výsledky experimentů[10]. Takové teorie spadají do oblasti zvané Digital physics – digitální fyzikální svět. Wolfram tento přístup zpracovává spíše z teoretického hlediska. Navrhuje mnoho typů automatů i jiných výpočetních systémů, které by v principu mohly vytvářet naši realitu a zabývá se i otázkami rychlosti běhu času a v neposlední řadě problémem svobodné vůle v tomto modelu.

---

[7] Samozřejmě bychom měli dodat, že výsledek dostaneme až po provedení dostatečného počtu kroků automatu, kde slovo „dostatečný“ znamená miliony, miliardy a nebo ještě mnohem víc. Důležitý je spíše fakt, že automat vůbec dokáže takové výpočty provádět – tzn. vždy dojde ke správnému výsledku, nikdy se nesplete, funguje (např. v případě sčítání) pro čísla libovolné velikost a podobně. Počet kroků, který bude automat potřebovat k provedení výpočtu lze předem zjistit z velikosti problému, který má automat vyřešit.

[8] viz známý systém Johna Conwaye „Game of Life“

[9] Pro zastánce tzv. M-teorie může být touto „aktivitou“ vibrace struny, pro jiné to může být nějaká „základní“ částice (ať už hmotná či energetická), ze které jsou všechny větší částice tvořené. Pokud bychom v našem modelu potřebovali, aby těchto „základních“ částic bylo více než jedna, můžeme použít automat s více stavy – tj. buňky budou mít více „barev“ než jen dvě

[10] Zatím se podařilo použít celulární automaty pouze k simulacím specifických jevů (například vlnění vodní hladiny, šíření nákazy v populaci), nikoli celé fyzikální reality.

Shrnutí

Názory Stephena Wolframa lze shrnout v těchto bodech:

- fyzikalismus + redukcionismus: vše lze redukovat na fyzikální zákony elementárních prvků (například částic).
- interakce elementárních prvků jsou velmi jednoduché a lokální. Složité makroskopické projevy vznikají pouze pomocí emergence[11]
- celulární automaty (nebo jiné masivně paralelní uniformní výpočetní systémy[12]) mohou v principu popisovat naši fyzikální realitu
- determinismus: fungování světa lze plně popsat (deterministickým) celulárním automatem nebo podobným výpočetním systémem

Postoje Rogera Penrose

Sir Roger Penrose je matematik, teoretický fyzik a filozof, který se zabývá mnoha tématy, od kosmologie a fungování vesmíru až po problematiku lidské mysli. Své názory publikoval v knihách „Císařova nová mysl“, „Stíny mysli“ a „Makrosvět, mikrosvět a lidská mysl“.

Penrose věří, že svět funguje v principu deterministicky. Vidí však velký rozdíl mezi zákony, které řídí chod mikrosvěta a makrosvěta.

Klasická a kvantová fyzika

Penrose uznává, že chod světa lze plně popsat fyzikálními zákony. Chování mikrosvěta se řídí zákony kvantové fyziky (konkrétně kvantové mechaniky a kvantové teorie pole) a chod makrosvěta je určen zákony klasické fyziky, kam spadá zejména speciální a obecná teorie relativity. Přechod od kvantové úrovně ke klasické, který Penrose nazývá objektivní redukce, je podle něj jev nesimulovatelný a nepředpověditelný, a právě pomocí tohoto jevu se snaží vysvětlit vznik vědomí a fungování lidské mysli. Pojďme se na klasickou a kvantovou fyziku podívat podrobněji.

Klasická fyzika:

Klasická makroskopická úroveň je nám blízká, jelikož se v ní běžně pohybujeme. Má tyto přirozené vlastnosti:

- existuje prostor a čas, v tomto prostoročase se vyskytují hmotné objekty
- každý hmotný objekt má dobře definované hranice, polohu a rychlost
- vlastnosti objektů jsou lokální a měřitelné

Kvantová fyzika:

Kvantová úroveň je poměrně odlišná a v některých aspektech z našeho pohledu neintuitivní. Tyto zvláštní jevy lze shrnout následovně:

---

[11] Kvalita systému, která vzniká interakcemi jeho částí, ale nelze ji přisoudit žádné konkrétní části
[12] Masivně paralelní = obsahující obrovské množství nezávislých jednotek. Uniformní = všechny jednotky se řídí stejnými pravidly

- *Částicový a vlnový dualismus*

  Na elementární objekty můžeme pohlížet buď jako na částice – tzn. „malé kuličky“, nebo jako na vlnění – elektromagnetické působení. Částice při svém chování využívají obě tyto formy a představa, že prostorem poletují malé míčky nemusí být v některých situacích vůbec vhodná
- *Nelokální jevy*

  Vlastnosti elementárních objektů nelze omezit na lokální prostor kolem nich. Existují částice v tzv. *provázaném stavu*, jejichž vlastnosti se navzájem ovlivňují, přestože jsou od sebe velmi vzdálené
- *Superpozice a nulová měření*

  Elementární částice nemusí v prostoru zaujímat konkrétní pozici, ale může se – podle tzv. Kodaňské interpretace – vyskytovat ve více pozicích současně. Pokud se pokusíme polohu změřit, částice si jednu z těchto pozic „vybere“ a tam ji naměříme. Na které z pozic se částice skutečně vyskytne, nelze podle současné teorie určit, lze však přesně spočítat, s jakou *pravděpodobností* to nastane.

  Tento jev se interpretuje tak, že částice se nachází v *„prostoru možností“*. Pokud je k tomu donucena měřením, jednu z těchto možností si vybere náhodně. S částicemi však lze pracovat i bez měření – tzn. v superpozicích. Můžeme tak některé možnosti částicím omezit a jiné naopak učinit pravděpodobnějšími. Při interakci částic v superpozicích může dokonce dojít k situaci, kdy jev, který **nenastal**, ovlivňuje realitu[13].
- *Problém měření*

  Při měření převádíme kvantový stav na klasický, tedy přecházíme z mikroskopické úrovně na makroskopickou. Tento přechod je zatím pro vědu velkou neznámou – lze sice spočítat pravděpodobnost jednotlivých výsledků, ale není možné výsledek přímo určit.

  Můžeme se také ptát, co přesně je „měření“ – je to situace, kdy kvantový jev ovlivňuje jakýkoli dostatečně *velký počet jiných částic*, nebo kdy se výsledek dozví člověk (tedy když se lidské tělo/mozek stane součástí tohoto *„dostatečně velkého počtu jiných částic“*)[14]

Rovnice pro fungování na klasické i kvantové úrovni jsou plně deterministické. Nejasný je však způsob přecházení z kvantové úrovně na klasickou. Penrose zastává názor, že i tento proces je deterministický, ale nevýpočetní – tedy, že výsledek je v principu určený jednoznačně, ale proces, který k výsledku vede, je natolik složitý, že jej nelze spočítat. Odvolává se při tom na několik matematických tvrzení, která zde ve stručnosti přiblížím.

Meze výpočetních prostředků

Dá se spočítat všechno? Může počítač v principu najít řešení jakéhokoliv matematického problému, pokud mu dáme dostatek času? Dá se jakákoli matematická věta jednoznačně dokázat nebo vyvrátit? Jak ukázali Alan Turing a Kurt Gödel, odpovědi na tyto otázky jsou NE, NE, NE.

Problém zastavení:

Problém je definovaný následovně: na vstupu je zdrojový kód programu P a data D. Otázka zní, jestli se program P zastaví, pokud bude pracovat s daty D. Chtěli bychom tento problém umět rozhodovat automaticky, tedy mít k dispozici program X, který pro jakékoliv zadané P a D rozhodne, zda se P(D) zastaví, nebo ne. Je však dokázané, že takový program nemůže existovat. Problém zastavení je tzv. *algoritmicky nerozhodnutelný* a problémů s touto vlastností existuje více.

[13] Viz např. myšlenkový experiment „Kvantové testování bomby“

[14] Nebo snad stačí, když se do oblasti, kterou kvantový jev ovlivňuje dostane kočka? Viz známý paradox Erwina Schrödingera

Věty o neúplnosti

Úplnost se zabývá otázkou dokazatelnosti matematických vět z axiomů. Matematická teorie je postavená na konečné množině axiomů – výchozích tvrzení (například *„Existuje číslo 0“, „Ke každému číslu existuje číslo o jedna větší“* a další) a odvozovacích pravidel (například *„Pokud platí tvrzení A a také platí tvrzení A -> B, pak odvoď platnost tvrzení B“*). *Důkaz* (nebo také *Odvození*) věty *V* je pak definován jako *„Posloupnost tvrzení, kde každé tvrzení je buď axiom nebo bylo odvozené z předchozích podle odvozovacích pravidel, a posledním tvrzením této posloupnosti je věta V“*.

Snadno se dá navrhnout systém, který je tzv. *korektní*, což znamená, že všechny dokazatelné věty jsou skutečně platné a není možné dokázat větu, která neplatí. Toto je stěžejní vlastnost, bez níž je systém v praxi nepoužitelný. Cílem matematiků bylo ukázat, že systém aritmetiky je také *úplný*, což znamená, že každá věta se dá buď dokázat, nebo vyvrátit (dokázat její negace). Kurt Gödel však ukázal, že každý formální matematický systém je buď nekorektní, nebo neúplný. Za předpokladu korektnosti tedy říká, že vždy existuje pravdivé tvrzení, které se nedá dokázat. Takových tvrzení navíc existuje nekonečně mnoho.

Nevýpočetní objektivní redukce

Objektivní redukci kvantového stavu na klasický vidí Penrose takto: (pro jednoduchost jen jedna částice a dva možné stavy)

1. kvantová částice se nachází v superpozici stavů *x* a *y*
2. proběhne měření
3. v tuto chvíli se příroda rozhoduje, do jakého *skutečného* stavu částice přejde
4. tento skutečný stav se vypočítá jako řešení nějakého *algoritmicky nerozhodnutelného problému* – například stav *x* se interpretuje jako zdrojový kód programu a stav *y* jako data. Pak částice přejde do stavu *x,* pokud se program *x* na datech *y* zastaví, a přejde do stavu *y*, pokud se program *x* na datech *y* nezastaví. Řešení tohoto problému je sice určené jednoznačně (každý program se buď zastaví, nebo ne), ale není možné ho zjistit žádnou výpočetní procedurou. Vývoj světa je tedy *deterministický*, ale nelze ho předvídat, vypočítat, ani simulovat (jevy s touto vlastností označuje jako *nevýpočetní*).

Lidská mysl

Penrose věří, že lidská mysl taktéž funguje nevýpočetním způsobem, což znamená, že ji nelze simulovat v počítači. Ve svých dílech představil argumenty, proč si myslí, že je *nutné*, aby takto fungovala a navrhl i způsob, jak toto fungování vysvětlit právě pomocí redukce kvantových stavů.

Vysvětlení je založené na strukturách zvaných *mikrotubuly*, které se nacházejí v mozku a podle Penrose umožňují existenci částic v superponovaném stavu. Při mozkové činnosti některé z těchto částic kolabují do klasických stavů a ovlivňují tak mikrotubuly, které následně ovlivňují příslušné neurony a tím i lidské chování. Kolapsy těchto částic jsou však řízené nevýpočetní procedurou a v důsledku toho je i veškeré lidské chování v principu nevýpočetní.

Přesvědčení, že lidská mysl funguje nevýpočetně, staví Penrose na pozorování, že lidé přistupují k řešení problémů principiálně odlišně než počítače. Tento fakt demonstruje na hraní šachů a hledání důkazů matematických vět, kde je potřeba jistého *vhledu* či *porozumění* a nestačí jen výpočetní síla.

Penrose předkládá naivní způsob dokazování vět, kde stroj zkouší najít protipříklad k větě systematickým zkoušením všech čísel. Pokud věta platí, pak takový protipříklad neexistuje a stroj nikdy neskončí, což lze považovat za projev jeho neinteligence ve srovnání s člověkem.

Tento svůj postoj pak zobecňuje využitím tvrzeních o mezích výpočetních prostředků, které jsem zmínil výše a činí závěr, že žádná výpočetní procedura nemůže být schopná dokazovat matematické věty tak, jak to dělá člověk.

Kritika

Za své teorie o povaze lidské mysli sklidil Penrose jednak obdiv za zajímavou a neotřelou teorii, ale také tvrdou kritiku od matematiků, informatiků, fyziků, biologů i filozofů. Terčem kritiky jsou některé unáhlené a nedostatečně podložené závěry na úrovni spekulací a také velmi svérázné používání vět o mezích výpočetních procedur. Je však třeba ocenit, že Penrose představil skutečně ucelenou teorii a pokusil se vysvětlit příčiny a domyslet do konce všechny důsledky, často i mezioborové. Vzal si tedy mnohem těžší úkol, než například Wolfram.

Necítím se být kompetentní vyjadřovat se ke všem částem Penroseovy teorie, můžu však popsat svůj pohled na problematiku výpočetních prostředků.

Předně Penrose tvrdí, že člověk postupuje při řešení těžkých problémů jinak něž počítač, neříká však jak. I když připustíme, že člověk používá jisté nevýpočetní prostředky, není jasné proč a jak by takové prostředky mohly pomoct při řešení šachových úloh nebo hledání matematických důkazů.

Pokud by takovými prostředky člověk skutečně disponoval, měl by být v principu schopný řešit mnoho úloh, které jsou dnes pro počítač těžké – například prolamování šifer. Problémy jsou na sebe převoditelné a můžeme tedy například dekódování zašifrované zprávy převést na hledání matematického důkazu. Kdyby toto člověk jednoduše zvládnul, většina dnes používaných zabezpečení by se zhroutila.

Velmi sporné je použití věty o neúplnosti: Věta říká, musejí existovat nedokazatelná tvrzení, což znamená, že důkaz skutečně **neexistuje**. Pokud důkaz věty neexistuje, pak ani počítač, ale ani člověk nemůže takový důkaz najít bez ohledu na to, jakými prostředky disponuje.

Co se týče problému zastavení: je pravda, že hledání důkazů vět lze převést na problém zastavení a ten je algoritmicky nerozhodnutelný. Z toho vyplývá, že neexistuje algoritmus, který by našel důkaz **jakékoli** dokazatelné věty. Ale to nedokáže ani člověk. I v současnosti existuje mnoho hypotéz, jejichž důkaz se teprve hledá. Navíc existuje několik vět, jejichž důkaz byl nalezen právě pomocí počítače. Férové tvrzení by tedy mělo znít, že existují věty, jejichž důkaz umí najít člověk a neumí to počítač, a také existují věty, jejichž důkaz umí najít počítač a neumí to člověk. To stejné platí i o pozicích v šachové hře. Argument, že člověka nelze s počítačem takto srovnávat, protože počítač má k dispozici obravskou výpočetní kapacitu není příliš přesvědčivý, protože i lidský mozek je schopný vykonávat obrovské množství výpočtů, pouze tuto svou kapacitu využívá jiným způsobem.

Při hledání důkazů počítačem Penrose zmiňuje jen velmi naivní způsob, a to sice hledání protipříkladu procházením všech možností. Existuje mnoho jiných mnohem sofistikovanějších metod pro automatické hledání důkazů (a obecně automatické řešení problémů). Lze zmínit například automatické dokazovače[15], Induktivní logické programování, Genetické programování, systémy pro splňování omezujicích podmínek a další.

A konečně využití zmíněných vět v reálném světě je vždy problematické. Problém zastavení i věty o neúplnosti popisují **matematické** objekty – tedy čísla, nikoli objekty fyzické. Obě tyto věty jsou silně závislé na faktu, že čísel existuje nekonečně mnoho. Například určitě **existuje** program, který rozhoduje o zastavení programů *P* s daty *D*, takových, že (velikost *P* + velikost *D*) je menší než např. 1 000 000. A obecně pokud množina vstupů a tedy i možných výsledků je pouze konečná, pak problém lze vždy řešit výpočetně.

I kdybychom připustili nekonečnost vesmíru, tak oblast, se kterou může člověk za života přijít do styku, je pouze konečná - obsahuje konečné množství částic, konečné množství různých stavů těchto částic, a tedy i konečné množství problémů, které je potřeba vyřešit. V principu tedy existuje výpočetní procedura, která bude všechny tyto problémy řešit úspěšně a stejně dobře, jako člověk. Taková procedura sice nebude skutečně fungovat pro **všechny** vstupy, ale z praktického hlediska bude od ní nerozlišitelná. Dokonce ani nevýpočetní charakter fungování světa na tom nic nezmění, protože počet možných výsledných stavů je stále konečný. V tomto případě by zmíněná procedura také existovala, ale neexistoval by žádný algoritmický způsob, jak ji najít.

Deterministická svobodná vůle

Oba zmínění autoři pracují se slabší definicí svobodné vůle. Jelikož uznávají determinismus, je zřejmé, že výsledky všech rozhodnutí jsou dané. Pokud nám však bude stačit, aby výsledky rozhodnutí jednotlivců byly *nepředvídatelné jinými jedinci*, pak je možné iluzi svobodné vůle ještě „zachránit“.

Stephen Wolfram ve spojení s tím zavádí tzv. *výpočetní neredukovatelnost*[16], kterou lze popsat takto: Existuje mnoho výpočtů, řešících daný problém a tyto výpočty trvají určitý čas. Některé výpočty jsou neefektivní a lze je nahradit jinými, které povedou ke stejnému výsledku v kratším čase. Některé výpočty však nelze zkrátit.

Příklad:

Dejme tomu, že chceme spočítat výraz 3^3. Je možné postupovat takto: $3^3 = (3+3+3) + (3+3+3) + (3+3+3)$. Takový výpočet vyžaduje 8 matematických operací. Je však možné počítat také $3^3 = 3*3*3$, kde použijeme pouze 2 operace. První výpočet je tedy možné zkrátit. Pokud bychom ale mohli takto zkracovat *jakýkoli* výpočet, pak lze cokoli spočítat například pomocí pouhé jedné elementární operace, což evidentně není možné. Princip výpočetní neredukovatelnosti v podstatě říká, že *existují výpočty, které nelze zkrátit*.

Pokud nyní chování člověka popíšeme celulárním automatem (jak Wolfram předpokládá), pak pro předvídání výsledku tohoto chování bychom potřebovali spočítat nějaký budoucí stav takového automatu. K tomuto našemu výpočtu ovšem používáme nějaké fyzické prostředky, například neurony v mozku, nebo elektrony v elektrickém počítači. Tyto fyzické prostředky jsou také součástí světa a jejich chování je řízené tímtéž celulárním automatem. Pouze jsme tak nahradili jeden výpočet automatu za jiný. Chování člověka označme jako výpočet A a předpověď výsledku tohoto chování jako výpočet B. Aby bylo možné chování předvídat, musel by druhý výpočet skončit dřív. To je ale možné pouze tehdy, když

[15] Viz např. úspěšný dokazovač Vampire - http://www.vprover.org/index.cgi
[16] V originále *computational irreducibility*

výpočet A bude redukovatelný. Jelikož existují neredukovatelné výpočty, pak existují procesy ve fyzickém světě, jejichž výsledek není možné předvídat dřív, než nastane. Nejlepší možný způsob, jak zjistit výsledek takového procesu je prostě počkat až proces fyzicky proběhne v reálném světe. Tímto způsobem lze do naprosto deterministického modelu dostat nepředvídatelnost a iluzi svobodné vůle.

Stejného efektu dosahuje i Roger Penrose, který však k tomu využívá mnohem silnější prostředky a sice nevýpočetní charakter procesů. Takové procesy pak nejen že nelze předvídat dřív než nastanou, ale nelze je předvídat *vůbec*.

Oba autory tak můžeme řadit ke kompatibilistům, protože svobodnou vůli v této slabší verzi připouštějí.

Závěrem

Osobně jsou mi teorie Penrose i Wolframa blízké, protože věřím v determinismus i v iluzi svobodné vůle. Autoři předkládají zajímavé způsoby jak tyto zdánlivě protichůdné postoje spojit a své teorie prezentují na pevném vědeckém základě. Chtěl bych zmínit ještě pár postřehů a asociací, které se k jejich názorům pojí.

Nevýpočetní charakter = zásah shůry?

Nevýpočetní charakter fyzikálních procesů je možné simulovat tzv. *orákulem* – tj. „vědmou“, která dává odpovědi na příslušné nevýpočetní otázky. Proces by tedy probíhal algoritmicky a v případě, že by pro pokračování bylo nutné znát výsledek nějakého nerozhodnutelného problému, by se algoritmus zeptal orákula na odpověď a pokračoval správnou cestou dál.

Problematika výpočtů s orákulem je dobře teoreticky prozkoumaná[17] a ví se například, že i když přidáme k algoritmu orákulum, které zná odpovědi na problém zastavení, tak ani tento posílený algoritmus nebude schopný vyřešit vše. Opět bude existovat nějaký těžší problém, na který algoritmus s orákulem nebude stačit. Tento proces lze opakovat: pokud k našemu algoritmu přidáme ještě další, silnější orákulum, které bude dávat odpovědi na tento težší problém, opět bude existovat ještě těžší problém, který algoritmus stále nebude schopný řešit, a tak dále.

To by se dalo interpretovat tak, že mohou existovat bytosti s ještě vyšší „úrovní vědomí“ než my, které se od nás liší asi tak, jako my se lišíme od počítačů a dokonce může existovat celá nekonečná hierarchie inteligencí takových bytostí.

Orákulová představa chodu světa případně lidské mysli umožňuje ještě jednu zajímavou interpretaci. Existují teorie, že vědomí nevzniká v mozku, ale že mozek je pouze jakýmsi „přijímačem“ a skutečný zdroj je jinde, mimo náš fyzický svět. Tato představa je blízká např. transpersonální psychologii[18]. Penroseova teorie poskytuje vědecky dobře definovanou „bránu“, kterou by případný mimofyzický zdroj mohl komunikovat s fyzikální realitou – byl by skrytý právě ve zmíněném orákulu.

Reálné nevýpočetní problémy

Představa, že v přírodě existují nevýpočetní procesy, nemusí být až tak neintuitivní. Známe několik příkladů reálných situací, které jsou dobře definované, ale neumíme je spolehlivě simulovat. Například:

- *Problém tří těles*

---

[17] Teorie se nazývá *relativizovaná složitost*

[18] Nutno dodat, že transpersonální psychologii velká část vědců neuznávaná a je považovaná spíše za pavědu

Jedná se o simulaci pohybu tří planet působících na sebe silou. Přestože definice problému je poměrně jednoduchá, současné metody pro simulaci nejsou příliš úspěšné

- *Turbulentní proudění kapalin*

  Kapaliny mohou proudit buď „poklidně“ (laminárně), nebo turbulentně. Zatímco pro laminární proudění máme přesné rovnice, turbulentní proudění je mnohem složitější simulovat

- *Útěk kořisti před predátorem*

  Když kořist kličkuje při útěku před predátorem, snaží se cíleně vytvořit nepredikovatelné chování. Kdyby dráhu jejího běhu bylo možné predikovat, predátor by tuto možnost v průběhu evoluce pravděpodobně už objevil a využil

Další příklady zejména v souvislosti s teorií chaosu lze nalézt v knize „Odsud až do nekonečna“ (Stewart, 2006).

Souvislost se self-referencí

V důkaze Gödelovy věty o neúplnosti se používá trik zvaný self-reference, kdy matematická věta odkazuje sama na sebe. Trik se dá přibližně popsat takto: můžeme očíslovat všechny možné matematické věty a psát „věta s číslem 1“, „věta s číslem 2“ atd. Stejně tak můžeme očíslovat všechny důkazy – tedy posloupnosti vět a opět můžeme psát „důkaz číslo 1“, „důkaz číslo 2“, atd. Lze pak pouze pomocí čísel, kvantifikátorů a běžných matematických operátorů sestavit matematickou větu, která říká zhruba toto:

„Pro každé číslo X platí, že důkaz číslo X není důkazem věty číslo Y“. Pokud tato konkrétní věta má v našem pořadí číslo Y, pak se jedná o pravdivou, ale nedokazatelnou větu19.

Jiné příklady self-reference jsou například „Tento výrok není pravdivý.“ A nebo „Holič holí všechny ty, kteří se neholí sami.“ Může se tento holič sám oholit nebo neoholit?

Problematikou self-reference a různých jejích aplikací se zabývá zajímavá kniha „Gödel, Escher, Bach“ (Hofstadter, 2012), kde autor prosazuje názor, že právě schopnost self-reference je nutnou podmínkou a příčinou vzniku vědomí.

Je s podivem, že Penrose tuto knihu vůbec nezmiňuje, přestože se zabývá velmi podobným tématem a používá i podobné prostředky. Obzvláště uvědomíme-li si, že Gödelovy věty a tedy i argument self-reference - i když kupodivu také nevyslovený - používá Penrose právě jako argument existence nevýpočetního vědomí.

Vyšší vědomé struktury

Pokud přijmeme zde popsaný deterministický model světa a vědomí, které vzniká jako emergentní jev, můžeme se přirozeně ptát, zda vědomím oplývají i jiné struktury než lidé. Tyto otázky se často kladou v souvislosti se zvířaty, rostlinami, případně jednoduchými organismy či dokonce neživými objekty. Mnohem zajímavější mi přijde ptát se, zda vědomí můžou mít i struktury větší a složitější než člověk – například skupinka lidí, rodina, obec, město, rozvášněný dav, skupina na Facebooku, stát, kontinent atd.

Tyto struktury obsahují jednotlivé lidi jako své složky a měly by tedy vykazovat aspoň tak složité chování jako lidé. Stejně tak by měly být na vyšším stupni emergence a mít vyšší stupeň vědomí, než lidé. Otázka teda zní jak se toto jejich vědomí projevuje (pokud existuje), jakým způsobem tyto entity

[19] Nechť si čtenář sám rozmyslí, že všechny ostatní případy (tedy pokud by věta byla nepravdivá, nebo dokazatelná) vedou ke sporu.

vnímají okolí, rozhodují se a jednají. Můžeme s nimi nějak komunikovat, nebo je takový pokus odsouzený k neúspěchu podobně, jako kdyby bakterie chtěla komunikovat s člověkem?

Těmito otázkami se seriózně mnoho lidí nezabývá. Pohled na Zemi jako na živý organismus lze nalézt v hermetismu, ale ten se příliš nesnaží tyto jevy vědecky vysvětlovat a modelovat.

Jedno z možných vysvětlení, proč spolu bytosti na různých úrovních vědomí nemůžou dobře komunikovat, může být rozdílné vnímání času. Mikroskopické organismy (například buňky těla) mají poměrně krátký život a také jejich „vjemy" přicházejí v krátkých časových intervalech (například 5-krát za sekundu). Buňky jsou však schopné na tyto vjemy reagovat a vykonávat příslušné akce i v takto krátkých intervalech. Jejich časová rozlišovací schopnost je velká. (Jinak řečeno 10 minut je pro ně velmi dlouhý čas.)

U člověka je jeho život i interval příchozích vjemů mnohem delší. V běžných situacích je také mnohem nižší frekvence vykonávání vědomých akcí – například jednou za 5-20 sekund. Rozlišovací schopnost v čase je tedy nižší a například interval 10 minut je pro nás krátký čas. (Z hlediska dlouhodobých plánů a cílů je třeba i 10 hodin či 10 dní stále poměrně krátký čas. To už by pro buňku mohlo představovat dobu jejího celého života).

Pokud jako další stupeň vezmeme hvězdy (například Slunce), tak jejich doba existence je mnohonásobně větší a stejně tak jejich „akce" přicházejí v mnohem delších intervalech. V duchu předchozích úvah bychom mohli předpokládat, že jejich rozlišovací schopnost v čase bude mnohem hrubší a například období 100 let vůbec nezaznamenají jako relevantní.

Takovéto vnímání času se dá velmi dobře popsat a simulovat pomocí tzv. mobilních automatů, které zavedl ve své knize Wolfram. Mobilní automat je v mnoha směrech podobný klasickému celulárnímu automatu s tím rozdílem, že je zde jistý token, který cestuje mezi jednotlivými buňkami a pouze buňka, která právě „drží" token, může měnit svůj stav. Ostatní buňky zůstávají neměnné až do doby, než se k nim token dostane. (Jako by se pro ně v tu dobu zastavil čas.) Přechodové pravidlo automatu pak říká, jak se bude měnit stav buňky a také určuje, kam se token přesune (pouze lokálně do některé z vedlejších buněk).

Nejkratší interval, který je vědomá bytost schopná vnímat, by pak odpovídal jednomu oběhu tokenu po celém „těle" této bytosti. Tedy počtu kroků automatu, které uplynou od doby, kdy token odejde z jisté buňky, projde všechny buňky odpovídajicí tělu bytosti a opět se vrátí. Tento interval by očividně závisel na velikosti těla bytosti a to tak, že čím větší tělo, tím delší by tento interval byl.

Věřím, že takováto úvaha, pokud se rozvede a dotáhne do důsledků, by mohla vést k novým zajímavým pohledům na fungování našeho světa a na interakce mezi jeho součástmi.

Pro úplnost ještě doplním, že jiná zajímavá aplikace mobilních automatů, kterou zmiňuje i Wolfram, je „celulární" vysvětlení jevů známých z teorie relativity. Teorie říká, že pozorovatelé, kteří se pohybují v prostoru různou rychlostí, zažívají také různou rychlost plynutí času. Tento jev by v principu mohl být simulovaný mobilním automatem tak, že části prostoru pohybující se velkou rychlostí bude token navštěvovat méně často.

Zdroje

Dvořák, P. (2009). Teologický determinismus plynoucí z Boží příčinnosti a jeho predeterminacionistické řešení v raně novověké scholastice. *Aither: Časopis pro studium řecké a latinské filosofické tradice*, 63-77.

Hofstadter, D. R. (2012). *Gödel, Escher, Bach - Existenciální gordická balada. Metaforická fuga o mysli a strojích v duchu Lewise Carrolla.* Dokořán.

Penrose, R. (1989). *Císařova nová mysl.* Cambridge University Press.

Penrose, R. (1994). *Stíny mysli.* Cambridge University Press.

Penrose, R. (1999). *Makrosvět, mikrosvět a lidská mysl.* Cambridge University Press.

Stewart, I. (2006). *Odsud až do nekonečna.* Argo.

Wolfram, S. (2002). *A New Kind of Science.* Wolfram Media, Inc.