

# Umělé myšlení a kauzální paradox emergentních systémů

Jan Romportl  
Katedra kybernetiky  
Západočeská univerzita v Plzni

## 4.1 Úvod

Diskuse o možnostech a mezích vytvoření umělého vědomí je zcela samozřejmě předcházena nejen úvahami o podstatě vědomí přirozeného, ale také o vymezení a uchopení toho, co činí přirozené přirozeným a umělé umělým. Kognitivistické (resp. logicko-symbolické) paradigma klasické umělé inteligence zatím nedokázalo vysvětlit podstatu vědomí a vědomého prožívání, byť koncepce myslí jako výpočetních procedur operujících nad souborem mentálních reprezentací dokáže poměrně dobře popsat některé dílčí kognitivní funkce (Thagard, 2001). Teoretické úvahy nad původem a podstatou vědomí se tak stále častěji dostávají do emergentistického rámce uvažujícího vzájemné působení a ovlivňování velkého množství prvků, z jejichž interakce vyvstává jev, který lze jako vědomí chápat.

V této kapitole se nebudeme primárně zabývat otázkami vědomí jako takového – tj. samotným jeho ontologickým zakotvením či možnostmi jeho poznání a vědeckého zkoumání. Záměrně opomíjíme též důležitost rozlišování „vnitřního“ (v první osobě) chápání vědomého mentálního života a „vnějšího“ (ve třetí osobě), jakož i jeho fenomenální a performační složky.<sup>1</sup> Situaci si zjednodušíme tím, že dále budeme hovořit spíše jen o myšlení (umělém či přirozeném) nebo mentálních stavech a procesech, ať již jejich chápání bude zčásti intuitivní či nikoli. Naším hlavním cílem bude snaha alternativním způsobem ukázat některé zajímavé důsledky, jež se týkají emergence jevů, a tudíž i úvah o přirozeném myšlení či vytváření myšlení umělého. Výklad se přitom hojně opírá o první kapitolu třetího dílu knihy Umělá inteligence (Havel, 2001).

## 4.2 Emergence a myšlení

Termín *emergence* se v našem pojetí týká kolektivních systémů, jež se na jisté úrovni skládají z velkého množství vzájemně interagujících prvků. Na úrovni „vyšší“<sup>2</sup> pak toto bohaté strukturní uspořádání dává vzniknout novým, kvalitativně odlišným jevům, jež nelze redukovat na pouhé chování prvků podkladových či je z něj odvodit.

---

<sup>1</sup> To, co je na daném mentálním procesu objektivně, „zvnějšku“ popsateľné, je jeho performační komponenta; to, co přitom subjekt „vnitřně“ prožívá, je fenomenální komponenta. (Havel, 2001)

<sup>2</sup> Uvozovky jsou použity, aby podtrhly metaforický význam rozlišování „vyšších“ a „nižších“ úrovní, neboť – jak bude ukázáno později – se nechceme omezovat pouze na lineárně uspořádanou hierarchii, která by odpovídala stupnici měřítek různých fyzikálních veličin.

Význam pojmu emergence definuje I. M. Havel (Havel 1993, 2001) několika způsoby. Jedním z nich je vysvětlení jevů v ontologickém smyslu, kdy se jevy přirozeně vyskytují na určité úrovni, avšak něco brání jejich redukci na odpovídající nižší úroveň. Druhým je vysvětlení jevů v gnoseologickém smyslu – jistá vlastnost specifikovaná teorií  $T_1$  je emergentní, jestliže má skutečné výskyty, které jsou doprovázeny výskyty některé jiné vlastnosti definované redukující teorií  $T_2$ , avšak nelze ji na žádnou vlastnost teorie  $T_2$  redukovat.

Ve vztahu emergence k přirozenému myšlení můžeme přijmout tzv. *emergentistickou tezi* (Havel, 2001, str. 41):

*Mentální stavy a procesy lze pojmtout jako emergentní jevy na některé vyšší úrovni dostatečně složitého dynamického systému.*

Záměrně jsme zdůraznili, že jde o mentální stavy a procesy myšlení přirozeného. Myšlení umělé (ve smyslu kritérií „umělosti“, tj. zejména existence záměru člověka vytvořit daný artefakt a zároveň realizace intencionálního procesu vedoucího od záměru k vytvoření) může být nejspíše v rámci emergentistické teze také zahrnuto, avšak zde již musíme být mnohem opatrnější: zřejmě jsou epistemologické překážky, bránící poznání, zda to, co v umělém systému probíhá, je opravdu myšlení (čili nesdělitelnost fenomenální složky mysli); zároveň je nutno si uvědomit, že lidské myšlení (resp. vědomí) je situováno do specifického tělesného, fyzikálního a socio-kulturního prostředí – a právě takto je jako myšlení identifikujeme.

Co bychom měli nejprve hledat u umělého systému (např. počítače), je přítomnost emergentních jevů – tzn. jevů ve smyslu ontologické či gnoseologické definice emergence, lhostejno však, dají-li se považovat za projevy umělého myšlení jako duplikátu myšlení přirozeného. Takovou přítomnost emergentních jevů nazvěme *emergentní efekt* systému. Vztah mezi emergentním efektem a přirozeným myšlením (alespoň po stránce performační) může dále zkoumat kognitivní věda – o tom více (Pstružina, 1998).

Emergentním efektem systému umělé inteligence tak může být určitá jeho žádaná (či naopak spontánní) vlastnost, jako například různé predikční, rozhodovací a optimalizační schopnosti, či modely některých kognitivních funkcí (obrazové a zvukové rozpoznávání, apod.). Návrh systémů v rámci tzv. „Nové umělé inteligence“ (Havel, 2001; Pstružina, 1998; Brooks, 1999) může pak být vnímán jako snaha o vytvoření takového umělého „média“, na jehož základě díky emergentnímu efektu vyvstanou požadované vlastnosti či schopnosti systému jako celku.

Naskýtají se však důležité otázky: jaká je podstata emergentního efektu? Proč nelze emergentní chování redukovat na chování konstituujících prvků? Proč při dosažení určitého množství prvků a jejich vzájemných vztahů dojde k emergenci? Jaké musí být toto množství? Co dalšího musí emergentní systém splňovat? Co vůbec emergenci *způsobuje*?

Hovoříme-li o emergentních systémech, můžeme tak v zásadě (a velice zjednodušeně) činit ze dvou úhlů pohledu: z *analytického* a z *fenomenologického*. Analytický pohled považuje reálné emergentní systémy za realizace složitých abstraktních nelineárních systémů (tj. abstraktních systémů popsaných netriviálními nelineárními diferenciálními či diferenčními rovnicemi), čili emergence je brána toliko za epifenomén, zdání v člověku

vyvolané na základě neznalosti či neschopnosti zjistit nebo zajistit řešení příslušných rovnic a z toho plynoucí nemožnosti činit o systému a jeho chování jasné a ostré předpovědi a závěry. Emergentní jevy vznikají díky neúplnému popisu jisté části reality – jde tak v zásadě pouze o „jazykový problém“.

Jinými slovy: emergence v tomto pojetí je způsobena neznalostí pozorovatele či jeho technickou neschopností nalézt analytické řešení složitých rovnic popisujících chování systému (či nalézt přesný tvar těchto rovnic). Bude-li se pozorovatel dostatečně dlouho snažit, podaří se mu rozšířit obzor svého poznání, čímž rozptýlí zdání daného emergentního efektu. Jediné, co nadále bude způsobovat nepredikovatelnost a „chaotičnost“ chování systému, bude šum znemožňující přesné stanovení počátečních podmínek, což díky bifurkacím nelineárních rovnic povede k velkým rozdílům v možných průbězích stavové trajektorie. Ovšem přítomnost tohoto šumu může sama být vysvětlena dvěma způsoby: šum *emerguje* ze složitosti vnějšího prostředí, do něhož je systém umístěn; případně šum je způsoben metafyzickým principem čisté náhody. První vysvětlení zakládá počátek nekonečného regresu, který může být přerušen pouze redukováním vysvětlovaného jevu či přijetím druhého řešení – principu čisté náhody. Reduktivní vysvětlení emergentního efektu však rozhodně nepřispěje k pochopení jeho podstaty a přijetím čisté náhody získáme maximálně určitý explanační princip, nikoli vyvrátitelnou hypotézu.

Naproti tomu fenomenologický přístup zastává názor, že emergentní efekt je důsledkem *aktů vědomí*. Na půdě vědomí se setkáváme s jevy vnějšího světa a emergence může být jedním z prostředků vědomí, který při uchopování jevů dovoluje zachovat jejich soudržnost: například tvar předmětu *emerguje* z jeho vnitřní struktury a vědomí při daných rozlišovacích schopnostech uchopuje tento tvar, nikoli samotnou vnitřní strukturu. Díky bezprostřední danosti vědomých mentálních stavů (Havel, 2001) se s tvarem předmětu setkáváme jako s něčím, co *je*, nikoli jako s něčím *zdánlivým* (byť by to de facto zdánlivé i bylo – jako například *fata morgana*), stejně jako se na půdě vědomí setkáváme s vlastní bolestí (či jinými tělesnými pocity), o níž dozajista nikdy nelze mluvit jako o zdánlivé (tvrzení „zdánlivě mě bolí hlava“ je samozřejmě nesmyslné).

Chceme-li se tedy dopátrat vnitřní struktury předmětu, s jehož tvarem se setkáváme, zvolíme nejspíše analytický přístup. Avšak k vysvětlení, proč a jak se setkáváme právě s *emergujícím* tvarem a nikoli jednoduše rovnou s vnitřní strukturou, nám reduktivní pojetí emergence ve smyslu analytickém nepomůže. Podobně ačkoli *víme*, že červená barva je projevem elektromagnetického pole určitých parametrů, vždy se na půdě vědomí setkáváme s prožitkem červenosti, nikoli s prožitkem elektromagnetického pole a k nereduktivnímu kvalitativnímu popsání tohoto prožitku nám teorie elektromagnetického pole nepomůže.

Fenomenologický přístup je z hlediska exaktně-vědeckého poznání taktéž problematický, neboť naráží na překážky subjektivismu vědomého prožívání. Na druhou stranu alespoň nepopírá existenci jevu, jehož podstatu nyní chceme objasnit (tj. jevu emergence). V této kapitole se tedy pokusíme k podstatě emergentního efektu přistoupit způsobem, který se snaží do jisté míry propojit obě výše zmíněná pojetí, a to pomocí fenomenologicky motivovaného matematického aparátu, který do svého rámce včleňuje i pozorující vědomý subjekt.

### 4.3 Kauzální domény

Kauzální výklad fungování světa je ve vědeckém diskursu velmi důležitým prvkem. *Kauzální vztahy* – tj. vztahy umožňující výklad pozorovaného jevu jako způsobeného jevem jiným – jsou implicitně přítomny i v teleologickém výkladu (tj. nastal-li daný jev proto, *aby* bylo dosaženo jevu jiného, znamená to, že první jev má kauzální potenci, v níž je obsažen i jev druhý).

Jak ukazuje K. R. Popper (Popper, 1997), kauzalita je do daného fragmentu světa vnesena až (vědeckou) teorií (čili jazykem), která tento fragment popisuje (či vysvětluje) – čili není v něm přítomna a priori. Vědecká teorie je z formálního hlediska systém znaků nebo symbolů a pravidel manipulace s nimi. Je-li pak teorie nahlížena v empirickém kontextu, jde vlastně o universální tvrzení. Teprve prostřednictvím teorií se nám objekty vnějšího světa a vztahy mezi nimi zjevují.<sup>3</sup>

Kauzální vztahy (tj. vztahy příčinné souvislosti) spojují takové jevy (či události), u nichž je možné podat *kauzální vysvětlení*. „Podat *kauzální vysvětlení* nějaké události znamená dedukovat tvrzení, které ji popisuje, z jednoho či více *universálních zákonů* spolu s jistými singulárními tvrzeními, *počátečními podmínkami*.“ (Popper, 1997, str. 43) Ilustrace tohoto pojetí kauzality je zřejmá z jednoduchého Popperova příkladu: můžeme říci, že zavěšení dvoukilogramového závaží na vlákno o nosnosti 1 kg *způsobilo* přetržení vlákna, neboť naše teorie disponuje universálním tvrzením (resp. hypotézou), že je-li na vlákno o dané nosnosti zavěšeno závaží o větší hmotnosti, dojde k přetržení vlákna, a máme též dvě singulární tvrzení: „nosnost našeho vlákna je 1 kg“ a „hmotnost našeho závaží je 2 kg“. Mezi jevy zavěšení dvoukilogramového závaží a přetržení vlákna tudíž existuje kauzální vztah.

Hovoříme-li o světě, můžeme to v obecnějším pojetí činit tak, že je svět fragmentován do množství tzv. *domén diskursu*. Budeme-li se na tyto domény dívat prostřednictvím kauzálních vztahů, jež do nich při našem porozumění vneseme a které nás také budou nejvíce zajímat, můžeme domény diskursu omezit na *kauzální domény*, jež I. M. Havel zavádí takto (Havel, 2000, 2001):

*Kauzální doména je jakákoli oblast (výsek, fragment, komponenta) skutečnosti, v jejímž rámci se nám kauzální vztahy jeví jako zjevné, srozumitelné, a vzájemně koherentní – přinejmenším jsou zjevnější, srozumitelnější a vzájemně koherentnější, než vztahy mezi různými doménami.*

Hranice kauzální domény jsou často neostré a vágně vymezené, tudíž o dvou oblastech skutečnosti můžeme říci, že jsou různými kauzálními doménami až tehdy, jsou-li od sebe v nějakém smyslu dostatečně daleko. Zatímco interakce mezi jevy v rámci jedné domény jsou zřetelné a kauzální, vztahy mezi jevy napříč doménami se tomuto přímému popisu většinou vymykají a obvykle o nich předpokládáme, že mohou být též emergentní povahy. Při popisu světa tak můžeme myšlenkově přecházet z jedné kauzální domény do jiné, avšak jedním pohledem velice obtížně dokážeme pojmut více domén současně. Je nutné se též tázat, zda vztahy mezi doménami vůbec mohou být kauzální povahy, a pokud ano, čím se tyto vztahy liší od vztahů uvnitř domén.

<sup>3</sup> „Teorie jsou síť, házené, aby chytly to, co nazýváme ‚svět‘.“ (Popper, 1997, str. 42)

Zavedením kauzálních domén dostáváme možnost vyhnout se představě, že k dosažení emergentního efektu je zapotřebí hierarchický řetězec úrovní, u nichž lze vždy říci, která z nich leží „výše“ a která „níže“. Toto platí obzvláště v případě problému mentálních procesů: Lze vůbec tvrdit, že úroveň mentální leží nad úrovní neuronovou? A kde se například nachází úroveň evoluční či sociální? Nad, nebo pod těmito dvěma?

Na základě úvah týkajících se kauzálních domén formuluje I. M. Havel *zobecněnou emergentistickou tezi* (Havel, 2001, str. 48):

*Mentální stavy a procesy lze pojmut jako emergentní jevy nad rozsáhlou množinou vzájemně vázaných kauzálních domén.*

Tuto tezi lze ilustrovat mnoha příklady, z nichž některé I. M. Havel v jednom ze svých textů (Havel, 2000) analyzuje. Holistická provázanost velkého množství kauzálních domén je tam patrná zejména na příkladu motolice, jejíž rafinované (z pohledu člověka) chování se dá nejlépe vysvětlit teleologicky, zatímco kauzální vysvětlení nejčastěji inklinuje ke „snadné“ redukci opírající se o evoluční princip (tj. ze všech možných variant motolic se nejvíce dokázala množit ta, která náhodou „způsobovala“, že se její hostitel mravenec ve strnulé poloze na stéblu trávy nechal pozřít kolem procházející ovci či jiným přežvýkavcem).

Chceme-li se však dobrat konkrétnějších důsledků zobecněné emergentistické teze, jež by něco určitého vypovídaly o povaze emergentního efektu jako souboru jevů nedílně provázejících holistický pohled na síť vzájemně interagujících kauzálních domén, musíme přistoupit k formalizaci pojmu kauzálního vztahu a kauzální domény samotné. Nejzjevnějším a formálně nejuchopitelnějším rysem emergentního systému je jeho složitost chápaná jako *množství* vztahů mezi jeho prvky, proto se zaměříme zejména na jevy podložené právě množstvím, které odpovídá souborům *abstraktních objektů*, jež jistým způsobem zastupují objekty přirozeného světa. Tyto abstraktní objekty pak v námi vytvářené formalizaci vstupují do vzájemných vztahů místo skutečných prvků systému, přičemž narozdíl od těchto prvků nesou pouze vlastnosti vyplývající z množství, nikoli žádné další vlastnosti (kvality), jež u prvků samotných můžeme pozorovat. Jednou z nejvhodnějších formalizací těchto vztahů je vytvoření jejich množinového modelu, přičemž vztahy mezi abstraktními objekty v modelu obsaženými budou považovány za odpovídající vztahům, jež platí v modelované sféře světa přirozeného.

Klasická teorie množin však nenabízí prostředky k modelování kauzálních domén, neboť podstatou kauzální domény je právě její niterná svázanost s pozorujícím subjektem, zatímco klasická teorie množin ze své podstaty vylučuje jakéhokoli pozorovatele a její objekty musejí být nahlíženy zcela ostře, atemporálně a „odnikud“. Vytvořením modelu kauzální domény v klasické teorii množin bychom tak přišli o její klíčový fenomenologický rys, který se projevuje i ve výše zmiňované její neostrosti.

Z těchto důvodů na několika následujících stránkách provedeme velice stručný exkurs do alternativní teorie množin a zjednodušeně představíme některé její základní principy, na nichž dále vystavíme množinový model kauzální domény a mezidoménové interakce, jenž nám pomůže odpovědět na dílčí otázky vyplývající z výše nastíněných témat týkajících se povahy emergentního efektu.

#### 4.4 Dynamická teorie množství

*Dynamická teorie množství* (DTM) je matematickou teorií dále rozpracovávající myšlenky *alternativní teorie množin* P. Vopěnky (Vopěnka, 1989, 2001), jíž provází snaha, která by se dala stručně shrnout jako hledání odpovědi na otázku: „Jak je možné, že některé poznatky klasické infinitní matematiky jsou použitelné v přirozeném reálném světě a jiné nejsou použitelné dokonce ani při výkladech světa reálného<sup>4</sup>“ (Vopěnka, 2007).

Důležitou vlastností DTM je snaha o zachycení *neostroti* z fenomenologického hlediska, tj. snaha matematizovat to, jak se v přirozeném reálném světě přítomná neostrost dává vědomému subjektu, jenž si pro jevy tohoto světa zjednáva porozumění. Toto je zároveň základní požadavek, který klademe na formální aparát, v němž bychom měli být schopni zachytit model kauzálních domén.

Jev neostroti je velice těsně spjat s jevem *nekonečna*, jak ukazuje (Vopěnka, 2000) a (Vopěnka, 2001). Uchopení neostroti tak jde ruku v ruce s uchopením nekonečna, což se v následujících odstavcích pokusíme nastínit. Náš výklad se opírá zejména o knihu (Vopěnka, 2001), kde se však stále hovoří spíše o alternativní teorii množin (z této knihy při popisu DTM doslovně přebíráme i některé formulace, neboť se domníváme, že nejlépe vystihují podstatu dané věci). Termín „Dynamická teorie množství“ zavádí P. Vopěnka v připravované knize (Vopěnka, 2007), kde je zároveň z matematického hlediska tato teorie formulována důsledněji. Alternativní teorie množin samotná je z ryze množinově-teoretického hlediska plně popsána v (Vopěnka, 1989).

##### 4.4.1 Pohled, obzor a přirozené nekonečno

Každý *pohled* (nejedná se jen o pohled vizuální, ale spíše obecně o pohled jako akt lidského vnímání<sup>5</sup>), ať vedený v jakémkoli směru, je omezen. Buď mu stojí v cestě nějaká hranice, která jej ostře přerušuje, nebo je omezen *obzorem*, směrem k němuž se zřetelnost vytrácí a ostrost otupuje. Je-li pozorován nějaký předmět (nejen fyzický, ale obecně jakýkoli, i abstraktní objekt), z perspektivy pozorovatele je rozdělen na dvě části – *osvětlenou* (pohledem zachycenou) a *neosvětlenou* (pohledem nezachycenou). Mezi těmito dvěma částmi prochází hranice, jež je od sebe odděluje, a tak je přesně definována její podstata, nicméně neostře a vágně vymezen její rozsah.

Čili osvětlená a neosvětlená část předmětu dohromady pokrývají celý tento předmět, přičemž se v žádném místě neprotínají. Obzor nepatří ani do jedné z uvedených částí a na rozdíl od nich není ani částí pozorovaného předmětu – je přisouzen pohledu samotnému. Ježto je tedy obzor sám o sobě ostrý a přesně definovaný jev, na němž se z jeho podstaty zobrazuje neostrost, poslouží právě on k zachycení a popsání vágnosti a neostroti.

Část předmětu, která je nám známa, nazvěme *krajina známosti* (terra cognita) – přesahuje osvětlenou část předmětu, neboť i za obzorem pozorovaný předmět dále

<sup>4</sup> *Přirozeným reálným světem* rozumíme v podstatě to prostředí, jehož jevy nazíráme. *Reálným světem* pak jakési vědou předpokládané rozšíření přirozeného reálného světa (Vopěnka, 2007). Hlubší diskusi nad těmito pojmy lze nalézt v knize (Vopěnka, 2001).

<sup>5</sup> „Pohled je hledáním toho, s čím se lze shledat (to je toho, co lze zahlédnout), a hlídáním, hledáním si a ohledáváním toho, co jsem zahlédl.“ (Vopěnka, 2001, str. 11)

plynule pokračuje. *Přesahem krajiny známosti* rozumíme to, čím krajina známosti přesahuje osvětlenou část. Dále pak již je *terra incognita* („neznámá krajina“).

Při zkoumání předmětu zrak (opět v abstraktním smyslu) postupuje v jednotlivých krocích směrem k obzoru. Takto vykonávanou cestu je možno rozlišit na *cestu na obzor* a *cestu směřující k obzoru*. V obou případech jde o cestu, která postupuje osvětlenou částí předmětu a míří k nějakému bodu na obzoru, jenž vyznačuje její směr. Cesta na obzor vede přímo do tohoto bodu a ani o kousek dále. Zde se ztrácí a její případné pokračování by na ni z tohoto bodu plynule navazovalo a šlo dále do přesahu krajiny známosti. Cesta směřující k obzoru končí někde v osvětlené části předmětu a její případné pokračování by začínalo až někde dále za obzorem (v závislosti na velikosti kroku překračujícího obzor), a to v místě zakrytém oním bodem, jenž udává směr. Je účelné krokům na těchto cestách přiřazovat přirozená čísla a tyto cesty – seskupení kroků – vykládat jako abstraktní třídy přirozených čísel.<sup>6</sup>

Cestu směrem k obzoru je zřejmě možné popsat ostře vymezenou konečnou množinou kroků. Poněvadž tato končí v oblasti před obzorem, je určitě kratší než cesta na obzor. Avšak cestu na obzor není možné popsat konečnou třídou kroků – při cestě na obzor se k němu stále „blížíme“, nicméně nebude existovat poslední krok, o němž bychom mohli prohlásit, že vede přesně na obzor. Jako příklad je možné uvést situaci, kdy zmenšujeme kousek papíru odtrháváním jeho částí – v jistém kroku již nejsme schopni zachytit ani velikost papíru, která se propadne do neosvětlené části světa poznávaného tělesnými smysly, zanechávajíc po sobě na obzoru pouze stopu v podobě geometrického bodu. I kdybychom zvyšovali smyslové schopnosti nade všechny meze, nepodaří se nám vytvořit z papíru geometrický bod. Cesta na obzor je tedy nekonečná.

Ačkoli tomu tak je, můžeme stále vykonat cestu, která je jejím prodloužením do přesahu krajiny známosti – za obzor. Třída všech kroků na této cestě je množinou, a to konečnou. Přitom třída těch z nich, které leží pouze na cestě na obzor, je zřejmě její neostře vymezenou podtřídou. Neostře vymezené (nekonečné) podtřídy ostře vymezených konečných množin nazýváme *polomnožinami*.

Polomnožiny ilustruje kupříkladu notoricky známý paradox hromady – polomnožina je třída těch přirozených čísel, pro něž platí, že odebereme-li z dané hromady písku takovýto počet zrníček, stále ještě zůstane hromada písku. Odebereme-li totiž kdykoli z hromady písku jedno zrníčko, vždy zůstane hromada, avšak původní hromada jistě obsahovala konečný počet zrníček (zde je zřejmá polomnožinová povaha jazykových pojmů). Jiným příkladem polomnožiny je váha nákladu, kterou unese osel (abychom zmínili opět tradičně uváděné zvíře): přidáme-li kdykoli oslovi k nějakému nákladu jeden gram, víme, že osel takto zvětšenou zátěž unese. Víme však též, že celou tunu dozajista neunese, proto množství jednogramových přídavek nákladu, které osel unese, odpovídá polomnožině.

Další známý příklad polomnožiny je třída všech těch přirozených čísel, která lze v českém jazyce jednoznačně popsat v několika větách neobsahujících dohromady více

---

<sup>6</sup> Zde zatím chápeme pojem „třída“ jako něco velmi podobného pojmu „množina“. Rozdíl mezi třídou a množinou je vysvětlen o několik odstavců níže. *Abstraktní* třída navíc v tomto kontextu znamená třídu, která obsahuje pouze abstraktní objekty (tj. objekty, které primárně nenesou žádnou jinou vlastnost než „být objektem“ a v množinovém modelu zastupují objekty přirozeného světa, případně objekty, které obsahují pouze jiné abstraktní objekty), nikoli například objekty přirozeného světa.

než tisíc písmen. Všechny nejvýše tisícčlenných posloupností české abecedy je jen konečně mnoho, a tedy všechna čísla zachytitelná uvedeným způsobem jsou menší než nějaké hodně vysoké přirozené číslo, takže shora uvedená třída je podtřídou nějaké množiny. Kdyby však tato třída sama byla množinou, bylo by nejmenší přirozené číslo, které do ní již nenáleží, jednoznačně popsáno, a sice těmi větami, které jsme právě pronesli. Jelikož však tyto věty mají dohromady méně jak tisíc písmen, náleželo by i toto číslo do uvedené množiny, což je spor.

Klasický výklad nekonečna umísťuje nekonečné množství až za všechna množství klasicky vykládaná jako konečná. Toto je natolik zažitý zvyk, že je velice obtížné jej překonat, přitom ze samotného pojmu „nekonečno“ není nikde patrné, proč by tak tomu mělo být – nekonečno netvrdí, že musí obsahovat více objektů než konečno; nekonečno znamená, že pouze není jeho konec. To, co pozorujeme na polomnožinách, nazvěme *přirozené nekonečno*. Vztah přirozeného nekonečna k jiným nekonečným a jeho úlohu v lidském poznávání hlouběji analyzuje (Vopěnka, 2000).

#### 4.4.2 Množinový popis přirozeného nekonečna

To, co se jeví na kauzálních doménách, je právě přirozené nekonečno, které pomocí přirozených čísel uchopíme na oboru abstraktních tříd a jehož popsání nám umožní přiblížit se k podstatě emergentního efektu. K tomu však musíme použít určitého formálního vyjadřovacího aparátu.<sup>7</sup> Jelikož potřebujeme nějak jednoduše nazývat dohromady množiny a polomnožiny, začneme užívat termínu „třída“ jako souhrnného názvu pro množiny a polomnožiny. Každá množina je tak zároveň třídou (a to ostře a konečně vymezenou), avšak nikoli každá třída je též množinou – může to být i polomnožina.

Nyní stručně uvedme přehled základních operací, které nám umožní přirozené nekonečno kauzálních domén uchopit (vesměs jde o klasické a běžně známé množinové operace). Užijme následujícího značení:  $A \times B$  označuje kartézský součin tříd  $A$ ,  $B$ , tj. třídu všech uspořádaných dvojic  $\langle x, y \rangle$  takových, že  $(x \in A) \wedge (y \in B)$ . Relací (binární) rozumíme takovou třídu, jejíž každý prvek je uspořádanou dvojicí.<sup>8</sup>  $A|B$  označuje parcializaci třídy  $A$  na třídu  $B$ , tj. třídu všech uspořádaných dvojic  $\langle y, x \rangle$ , kde  $(\langle y, x \rangle \in A) \wedge (x \in B)$ . Definiční obor relace  $R$  značíme  $dom(R)$  a je to třída všech  $x$  takových, že  $\langle y, x \rangle \in R$ , obdobně obor hodnot relace  $R$  značíme  $rng(R)$  a je to třída všech  $y$  takových, že  $\langle y, x \rangle \in R$ . Je-li  $A$  třída, jejíž prvky jsou též třídy, pak  $\cup A$  označuje sjednocení všech tříd, jež jsou prvky třídy  $A$ . Analogicky je definován i průnik  $\cap A$ .<sup>9</sup>

Relace  $F$  je funkcí, když ke každému  $x \in dom(F)$  existuje právě jedno  $y$  takové, že  $\langle y, x \rangle \in F$ ; toto  $y$  značíme  $F(x)$ . Relace  $R$  má konjugovanou funkci, jestliže pro každé

<sup>7</sup> Zde uvedeme jen pro nás nejnütnější základy. Podrobnosti necht' čtenář hledá v (Vopěnka, 2001, 2007).

<sup>8</sup> Množinové relace modelují vztahy mezi objekty – je-li mezi objekty  $x$  a  $y$  daný vztah, pak je uspořádaná dvojice  $\langle x, y \rangle$  prvkem třídy, která tento vztah reprezentuje. Vztahy mohou být například „být těžší než“, „být sourozencem někoho“, ale i „být příčinou něčeho“ (tj. kauzální vztah mezi dvěma jevy).

<sup>9</sup> Je-li například  $A = \{\{1,2,3\}, \{1,3,4\}, \{1,3,5\}\}$ , pak  $\cup A = \{1,2,3,4,5\}$  a  $\cap A = \{1,3\}$ .



$x \in \text{dom}(R)$  je třída  $y$  takových, že  $\langle y, x \rangle \in R$ , množinou. Konjugovanou funkcí té relace je funkce  $F$  taková, že  $\text{dom}(F) = \text{dom}(R)$  a pro každé  $x \in \text{dom}(R)$  je  $F(x)$  množina všech  $y$  takových, že  $\langle y, x \rangle \in R$ . Konjugovanou relací funkce  $F$  rozumíme třídu všech  $\langle y, x \rangle$  takových, že  $y \in F(x)$ .

Třídu všech objektů spadajících do oboru abstraktních tříd označíme  $V$ . Třídu všech abstraktních objektů náležejících do  $V$ , jež nejsou množinami, označíme  $V_0$ . Malými písmeny latinské abecedy budeme označovat pouze prvky třídy  $V$ . Podtřídy třídy  $V$  budeme označovat velkými latinskými písmeny. Třídu přirozených čísel (náležejících do  $V$ ) označíme  $N$ . Přirozená čísla samotná budeme označovat malými řeckými písmeny.

Řezem na přirozených číslech rozumíme takovou třídu přirozených čísel, která spolu s každým v sobě obsaženým číslem obsahuje i všechna přirozená čísla menší. Je-li  $\alpha \in N$ , pak třída všech přirozených čísel menších než  $\alpha$  je řezem na přirozených číslech, je to množina a značíme ji  $[\alpha]$ . Tento řez nazýváme řezem hlavním a číslo  $\alpha$  je jeho (vnější) hlavou.

Je-li  $H$  řez na přirozených číslech, říkáme, že posloupnost nějakých objektů na řezu  $H$  je na něm stabilní, jestliže pro každé přirozené číslo  $\alpha \in H$  je úsek této posloupnosti až do čísla  $\alpha$  (včetně) ostře vymezen.

Klíčovým pojmem k uchopení prve zmiňované podoby přirozeného nekonečna (a s tím i kauzálních domén) je *obzorný řez*. Obzorným řezem rozumíme takový řez  $H$  na přirozených číslech, pro který platí:

a) Řez  $H$  nemá poslední (tj. největší) prvek a třída  $H$  je polomnožinou.

b) Každou posloupnost na řezu  $H$ , která je na tomto řezu stabilní, lze prodloužit do posloupnosti, která je stabilní na nějakém hlavním řezu na přirozených číslech, jehož podtřídou je třída  $H$ .

Abychom mohli obzorný řez formalizovat, zavedeme ještě následující termíny: řekneme, že funkce  $F$  je stabilní na řezu  $X$ , je-li  $\text{dom}(F) = X$  a pro každé  $\gamma \in X$  je  $F | [\gamma+1]$  množina náležející do třídy  $V$ . Řekneme, že funkce  $f$  je tvrdým prodloužením funkce  $F$  definované na řezu  $X$ , jestliže existuje  $\alpha$  takové, že  $\text{dom}(f) = [\alpha]$ , a  $F = f | X$ .

Na základě toho řekneme, že neprázdna třída  $H \subseteq N$  je obzorným řezem, jestliže platí:

a)  $H$  je řez, který v uspořádání přirozených čísel nemá největší prvek.

b) Je-li  $F$  stabilní funkce na řezu  $H$ , pak existuje její tvrdé prodloužení  $f$ .

Řekneme, že třída  $A$  je přehledná na obzorném řezu  $H$ , jestliže existuje stabilní funkce  $F$  na řezu  $H$  taková, že  $A = \text{rng}(F)$ . Řekneme, že  $X$  je  $\sigma$ -třída (vzhledem k řezu  $H$ ), jestliže existuje přehledná třída množin  $A$  (tj. třída  $A$ , jejíž prvky jsou množiny) taková, že  $X = \cup A$ . Řekneme, že  $X$  je  $\pi$ -třída (vzhledem k řezu  $H$ ), jestliže existuje přehledná třída množin  $A$  taková, že  $X = \cap A$ .

V rámci DTM mají mezi všemi třídami obzvláštní význam právě  $\sigma$ - a  $\pi$ -třídy, jejichž vzájemné vztahy ukazují na souvislosti, jež klasická teorie množin vytěsnila z oblasti svého studia. Není naším cílem uvádět a odvozovat zde všechna tvrzení týkající se  $\sigma$ - a

$\pi$ -tříd – za tímto účelem čtenáře odkážeme na knihy (Vopěnka, 2001, 2007), odkud zde vybíráme dvě základní tvrzení, jež mají pro nás v danou chvíli zásadní význam (důkazy těchto tvrzení lze samozřejmě nalézt tamtéž).

*Tvrzení 1.*

Bud'  $X \subseteq w$ , kde  $w$  je množina. Necht'  $X$  je  $\sigma$ -třída (resp.  $\pi$ -třída). Potom  $w - X$  je  $\pi$ -třída (resp.  $\sigma$ -třída).

*Tvrzení 2.*

Třída  $X$  je současně  $\sigma$ -třída i  $\pi$ -třída právě tehdy, když  $X$  je množina.

Je důležité povšimnout si rozdílů mezi vnitřní strukturou  $\sigma$ -třídy a  $\pi$ -třídy (tento rozdíl později využijeme při modelování interakce kauzálních domén), který je dán použitím odlišného operátoru při jejich konstrukci (operace sjednocení proti vůči ní v jistém smyslu komplementární operaci průniku). Ve výsledné podobě  $\sigma$ -třídy se totiž mnohem lépe odráží charakter funkce  $F$  definované na daném obzorném řezu, jež stojí za vznikem oné  $\sigma$ -třídy, neboť tato  $\sigma$ -třída obsahuje všechny prvky, na něž funkce  $F$  poukazuje prostřednictvím svých funkčních hodnot (díky jejich sjednocování). Naproti tomu  $\pi$ -třída – byť vztažená ke stejnému obzornému řezu – tento charakter stále více zakrývá tím, jak se s postupným přibližováním pohledu k obzoru zmenšuje a své prvky „ztrácí“ (po operaci průniku vždy zůstanou jen ty objekty, jež byly prvky současně všech funkčních hodnot funkce  $F$  až do daného kroku).

Dále zavedme pojem *prvoevidovatelný jev*. Prvoevidovatelným jevem nazvěme takový jev, který jsme schopni evidovat, tj. vidět ho a vědět, že ho vidíme, jakmile ho evidovat můžeme. Například jev červenosti je prvoevidovatelný, zatímco jev nečervenosti prvoevidovatelný není, neboť k jeho seznání potřebujeme mít již uchopen jev červenosti. Obdobně jev rozlišitelnosti je prvoevidovatelný, jev nerozlišitelnosti prvoevidovatelný není. Obzvláště důležité je uchopení rozdílnosti ve strukturách jevů prvoevidovatelných a jejich doplňcích (negacích).

Výše nastíněné matematické vztahy je vhodné interpretovat tak, že v oboru abstraktních tříd odpovídají prvoevidovatelným jevům  $\sigma$ -třídy. Znamená to tedy, že za účelem studia neostrosti prvoevidovatelných jevů je možné tyto jevy podkládat různými  $\sigma$ -třídami a z vlastností takových  $\sigma$ -tříd pak zpětně nacházet vlastnosti jimi podložených prvoevidovatelných jevů.

Z výše uvedeného Tvrzení 1 pak vyplývá, že doplňkem určité  $\sigma$ -třídy do její nadmnožiny je  $\pi$ -třída, čili negace prvoevidovatelných jevů mají jinou strukturu než prvoevidovatelné jevy samotné. Negace prvoevidovatelných jevů je tedy možné popsat  $\pi$ -třídami. Vyjádříme-li se více v duchu analytické filosofie, abychom neužili fenomenologicky laděného termínu „jev“, pak můžeme říci, že např. extenzí pojmu „červený“ je  $\sigma$ -třída všech objektů, které jsou červené, a extenzí pojmu „nečervený“ (jako negace „červeného“) je  $\pi$ -třída všech objektů, které nejsou červené. Oproti klasickému množinovému vyjadřování tak získáváme účinnou formalizaci důležitého kvalitativního rozdílu spočívajícího v tom, že červené věci jsou červenými *jinak*, než nečervené věci nečervenými. Ze znalosti toho, jaké to je *být nečervené*, nelze díky

charakteru  $\pi$ -třídy nijak odvodit, jaké to je *být červené*, což dobře odpovídá naší běžné zkušenosti.

Podle Tvzení 2 dále můžeme usuzovat, že jev, který je prvoevidovatelný stejně jako jeho negace, je jevem ostře vymezeným, jelikož doplněk  $\sigma$ -třídy  $X$  je též  $\sigma$ -třídou pouze tehdy, je-li  $X$  množinou. Má-li naopak doplněk třídy  $X$  být pouze  $\pi$ -třídou, musí být  $X$  pouze  $\sigma$ -třídou, a tudíž neostře vymezená.

## 4.5 Mezidoménová interakce a emergentní efekt

### 4.5.1 Model kauzální domény

V tuto chvíli by bylo vhodné poněkud zúžit a precizovat, co v dané situaci míníme, užijeme-li slova „jev“. Z hlediska fenomenologického můžeme jako jev chápat všechno, „...co lze pohledem zachytit a odlišit od všeho ostatního, na co lze takto ukázat, [pro co] si mohou zjednat první porozumění spočívající v tom, že to nějak ‚je‘ a že to je ‚to‘.“ (Vopěnka, 2001, str. 11) Nás však spíše zajímají takové jevy, které s sebou nesou určitou kauzální potenci, tzn. mohou něco způsobit. Fenomenologickému vymezení jevu by tak odpovídala i objektová individua – např. nějaká kočka. Nelze však říci, že by kočka sama o sobě jako jsoucno něco způsobovala. Jazykové zkratky sice dovolují říci větu: „Kočka způsobila, že pes začal štěkat.“, avšak ve skutečnosti štěkot psa byl způsoben tím jevem, že se kočka objevila na zahradě. V tomto případě výrazy „kočka“ a „pes“ máme na mysli jednu konkrétní kočku a jednoho konkrétního psa, např. Mindu a Žeryka. Označuje-li výraz jednotlivý předmět, nazvěme jej *vlastní jméno* (Fiala, 2006). Za jevy v zúženém smyslu budeme tedy nadále pro naše potřeby považovat propozice,<sup>10</sup> v nichž figurují pouze vlastní jména.

Omezením se na vlastní jména sice možná ztrácí takto chápaný jev na své obecnosti, avšak naší snahou je vyhnout se co možná nejvíce rozličným logicko-jazykovým či jiným problémům, které by vyvstaly například při zahrnutí propozic, jež by mimo vlastní jména obsahovaly i pojmy. K modelování mezidoménové interakce a emergentního efektu zjednodušeně chápané jevy bohatě postačí – dokonce i v případě mentálních stavů lze učinit dohodu, že každá mysl bude v našem modelu disponovat svými vlastními „exempláři“ rozličných mentálních stavů – takže například deprese či migréna jednoho člověka bude jiným individuem než u člověka druhého. Za jevy tedy budeme, zhruba řečeno, považovat jednotlivé události, které se ve světě mohou stát, nikoli universální tvrzení (universální tvrzení – jak bylo diskutováno dříve – nám umožňují kauzální výklad jevů).

Předpokládejme, že můžeme uvažovat dostatečně rozsáhlé universum jevů přirozeného reálného světa – natolik rozsáhlé, aby v sobě obsahovalo alespoň dvě kauzální domény. Každému objektu z tohoto universa jednoznačně přiřadíme jeden abstraktní objekt (resp. objekt z oboru abstraktních tříd). Takto vzniklou třídu abstraktních objektů označme  $U$  (jde vlastně o množinový model universa jevů).

Zabývejme se nyní kauzální doménou, již označíme symbolem  $i$ . Tato doména nechť vypovídá o dostatečně bohatém fragmentu přirozeného reálného světa (bohatém v tom

<sup>10</sup> Propozicí chápeme větu, u níž lze hovořit o její pravdivostní hodnotě.

smyslu, aby například neobsahoval pouze chování nějakého jednoduchého logického automatu), který je zároveň součástí universa  $U$ . Třídou všech jevů domény  $i$  označme  $H_i$ . Zřejmě  $H_i \subset U$ .

Kráčíme-li od jádra<sup>11</sup> domény  $i$  směrem vytyčeným její hranicí, můžeme se v každém kroku u každého jevu zeptat, zda stále patří do této domény (tj. zda o něm konzistentně a přirozeně můžeme uvažovat ve vztahu k ní). Odpovíme-li si kladně, pak zároveň víme, že u bezprostředně následujícího jevu (tj. jevu v epistemickém smyslu „nejbližšímu“) odpovíme též kladně. Zároveň však víme, že pokud bychom naráz přeskočili dostatečně velké množství jevů, našli bychom takový jev, který už do kauzální domény  $i$  bezpečně nepatří. V jedné kauzální doméně tak budou například jevy týkající se fungování neurotransmiterů; jsou-li tam tyto jevy, pak tam nejspíše budou i jevy týkající se mitochondrií<sup>12</sup>, jsou-li v doméně zahrnuty mitochondrie a jejich kauzální vztahy, pak tam zřejmě mohou být kauzální vztahy, v nichž vystupují synapse či celé neurony; obsáhne-li tato kauzální doména i neurony, pak může popisovat též dynamiku celých neuronálních struktur a je zřejmě možné, aby v ní byly konzistentně zahrnuty i dílčí kognitivní funkce. Takto můžeme kauzální doménu rozšiřovat o další a další jevy (v daném případě nejspíše s rozvojem příslušného vědního oboru) a vždy, když do ní určitý jev vložíme, nemůžeme do ní nezahrnout i jev jemu nejbližší. Víme však s určitostí, že do ní nepatří kupříkladu aktuální počasí, dopravní situace v sousedním městě, výkyvy na burze či erupce na Slunci.

Výše ukázané bychom mohli charakterizovat jako projev principu jakési „přirozené indukce“, který na universu jevů vytváří projevy přirozeného nekonečna. Z toho budeme usuzovat, že  $H_i$  je obzorný řez.

To, jak se věci mezi sebou v kauzální doméně  $i$  mají, nazíráme prostřednictvím jisté teorie  $T_i$  – v nejobecnějším případě jazykem, který nám umožňuje vztahy mezi jevy strukturovat a zachytit. Tato teorie vlastně říká, které kauzální vztahy jsou zjevné, srozumitelné a vzájemně koherentní, tudíž je v epistemologickém smyslu vlastně konstituujícím elementem kauzální domény. Naproti tomu má-li jakákoli teorie o kauzální doméně vůbec vzniknout, musí tato kauzální doména již nějakým způsobem existovat (v ontologickém smyslu). Teorie je tak se svojí kauzální doménou spjata v hermeneutickém kruhu, který úzce souvisí se zjednáváním světa jakoukoli vnímající bytostí (Brooks, 1999).

V oboru abstraktních tříd můžeme vytvořit jednoduchý model kauzálních vztahů v rámci kauzální domény pomocí relace, do níž budou vstupovat právě ty objekty, které jsou ve vzájemném kauzálním vztahu. Necht' tedy existuje relace  $R_i$  taková, že  $dom(R_i) = H_i$ ,  $rng(R_i) \in H_i$  a dva objekty (jevy)  $x, y$  jsou v relaci  $R_i$  (tj.  $\langle y, x \rangle \in R_i$ ) právě tehdy, když  $x$  přímo způsobuje  $y$  (resp.  $y$  je přímým důsledkem  $x$ ). Například

<sup>11</sup> „Jádrem“ domény nazvěme tu podtřídou jejích jevů, u nichž shledáváme nejvyšší „zřejmost“ kauzálních vztahů a „přirozenost“ nacházet se v této doméně.

<sup>12</sup> Neurotransmitery (nervové přenašeče) jsou objekty, které se významným způsobem podílejí na synaptickém signálovém přenosu. Mitochondrie je částice nacházející se v tzv. presynaptické části chemické synapse neuronových buněk.

specifická hladina fluoxetinu (jednoho z inhibitorů zpětného vychytávání serotoninu<sup>13</sup>) způsobí změnu zpětného vychytávání serotoninu, což způsobí změnu hladiny serotoninu.

Nechť dále  $F_i$  je konjugovaná funkce k relaci  $R_i$ . Abychom mohli zajistit její existenci, budeme po teorii  $T_i$  požadovat, aby třída všech přímých důsledků každého jevu  $x$  byla ostře vymezena, a tudíž byla množinou. Teorie ve smyslu souborů formálních tvrzení budou nejspíše tento požadavek splňovat, stejně tak i teorie v obecnějším smyslu teorií vědeckých. Teorie ve smyslu vágně vymezených a formulovaných oblastí lidského poznání, vyjádřené přirozeným jazykem, tento požadavek samozřejmě splňovat vždy nemusí, podobně si lze i představit speciálně vytvořenou formální teorii, která jej splňovat nebude záměrně. Tyto případy z našeho zkoumání vyloučíme z technických důvodů, abychom si usnadnili práci v oblasti operací s abstraktními třídami.

Pro každé  $x \in H_i$  je  $F_i(x)$  množinou všech důsledků jevu  $x$  v rámci kauzální domény  $i$ . Funkce  $F_i$  je tak vlastně modelem její vnitřní struktury. Relace  $R_i$  je stabilní na řezu  $H_i$ , proto je na něm stabilní i funkce  $F_i$ , a tedy třída  $A_i = \text{rng}(F_i)$  je na obzorném řezu  $H_i$  přehledná. Z toho vyplývá, že třída  $D_i = \cup A_i = \cup \text{rng}(F_i)$  je  $\sigma$ -třída.  $D_i$  zjevně obsahuje všechny jevy z domény  $i$ , jež jsou důsledky jiných jevů ze stejné domény (čili  $D_i$  neobsahuje „počáteční podmínky“ – tj. jevy nemající v dané doméně žádnou příčinu). Třídou  $D_i$  nazvěme *kostra kauzální domény  $i$* .

Interpretace kostry kauzální domény jako  $\sigma$ -třídy odpovídá přirozené představě kauzální domény jako jevu prvoevidovatelného. Člověk obklopený spletí jevů, jíž si zjednáva porozumění, jaksi automaticky poznává, co jsou kauzální domény (opírajíc se o jejich kostry v podobě prvoevidovatelných seskupení objektů, z nichž domény sestávají), avšak z této schopnosti či znalosti nikterak neurčí, co kauzální domény nejsou (čili co není v nějaké kauzální doméně, pro to si nezjedná porozumění). Ve vztahu k teorii  $T_i$  je kostra kauzální domény  $i$  vlastně třídou všech jevů, které jsou v této teorii kauzálně vysvětlovány.

#### 4.5.2 Model mezidoménové interakce

Jak již bylo dříve naznačeno, máme-li se dopátrat zajímavějších výsledků, musí být universum  $U$  dostatečně bohaté, aby v sobě obsahovalo přinejmenším dvě kauzální domény – jednu výchozí a druhou, z níž budeme na tu výchozí pohlížet. O jaké konkrétní domény půjde, je vcelku lhostejné, avšak v souladu s naším prvotním tématem týkajícím se emergentního efektu v systémech s umělým či přirozeným myšlením můžeme model mezidoménové interakce ukázat na příkladu, v němž výchozí kauzální doménu  $i$  nazvěme třeba doménou nervových procesů<sup>14</sup> a pokusíme se ukázat, jak se tato vztahuje k doméně jiné, která složité nervové procesy obvykle doprovází, a to doméně mentálních stavů. Dále v textu budeme tyto domény zkráceně nazývat „nervová“ a „mentální“.

<sup>13</sup> Serotonin je jeden z řady neurotransmiterů. Zpětné vychytávání (či znovuvstřebávání, angl. reuptake) je jedna z fází procesu tohoto přenosu, která reguluje počet uvolněných neurotransmiterů. Více např. (Novák et al., 1998).

<sup>14</sup> „Nervové procesy“ jak ve smyslu původním (biologickém), tak i ve smyslu přeneseném – „in silicio“ (tj. „v počítači“).

Nechť tedy všechny jevy domény mentální jsou zastoupeny třídou  $C_m$ , přičemž platí, že  $C_m \subset U$  a zároveň  $H_i \cap C_m = \emptyset$  (tj. doména nervová nemá žádné společné jevy s doménou mentální – jinak by to nebyly dvě různé kauzální domény).

Zkoumáme-li mezidoménovou interakci, chceme vlastně vědět, co způsobují jevy z výchozí kauzální domény v kauzální doméně cílové – jinými slovy se můžeme ptát, jaké jevy z domény nervové způsobily pozorovaný jev v doméně mentální. Zajímají nás tedy kauzální vztahy, do nichž vstupují jevy z nervové domény na jedné straně a jevy z domény mentální na straně druhé. Pakliže bychom se z vědeckého hlediska opravdu zajímali o tyto dvě konkrétní domény, chtěli bychom též vytvořit teorii, která by tyto mezidoménové kauzální vztahy shrnovala. Model této teorie bychom v oboru abstraktních tříd mohli zachytit funkcí podobnou funkci  $F_i$ . V rámci našeho problému nás samozřejmě nezajímá konkrétní podoba této teorie či její modelující funkce, ale spíše epistemologické hranice, jež možnosti jejího nalezení vymezují.

Náš pohled na výchozí (nervovou) doménu  $i$  rozšíříme nyní tak, aby dosáhl až k doméně mentální a pokryl alespoň ty jevy z ní, které nás zajímají. Vytvoříme tedy množinu  $w$  takovou, že  $D_i \subset w$  a současně  $w \cap C_m \neq \emptyset$ . Množina  $w$  obsahuje všechny jevy kostry kauzální domény  $i$  a zároveň k nim přidává třídu dalších jevů, mezi nimiž jsou ty z mentální domény, jež nás zajímají.

Nyní se pohledem přesuneme z nervové domény směrem k doméně mentální a z této perspektivy se pokusíme nahlédnout, jak jsou jevy ze třídy  $X_i = w - D_i$  kauzálně závislé na jevech z  $D_i$ . Analogicky k funkci  $F_i$  se pokusíme najít stabilní funkci  $F_{X_i}$ , která bude též definována na řezu  $H_i$  (tj. stejné třídě příčin jako má  $F_i$ ), avšak obor hodnot rozšíří o množiny důsledků, jež nejsou v  $D_i$ , tj. patří do mentální domény (či do „temného prostoru“ mezi oběma doménami, který nás však nyní nezajímá).

A zde se dostáváme k velice důležitému bodu: podle *Tvrzení 1* je třída  $X_i$   $\pi$ -třída a tudíž  $X_i = \text{rng}(F_{X_i})$ , a jelikož  $D_i$  a tedy ani  $X_i$  nejsou množiny, tak podle *Tvrzení 2* ani neexistuje takové  $F_{X_i}'$ , aby platilo  $X_i = \text{Urng}(F_{X_i}')$ . Z toho vyplývá klíčový poznatek, že všechny prvky třídy  $X_i$  nutně musejí být rovněž obsaženy ve *všech* prvcích třídy  $\text{rng}(F_{X_i})$  (které samy jsou množinami).<sup>15</sup> Vezmeme-li v úvahu interpretaci funkce  $F_{X_i}$ , pak to znamená, že *každý* jev z mentální domény musí být v nově vznikající teorii principiálně zahrnut mezi důsledky *všech* jevů z domény nervové!

Jinými slovy řečeno: pokud se zeptáme z hlediska mentální domény, které jevy nervové domény *způsobují* daný jev v doméně mentální, a budeme trvat na kauzálním vysvětlení, odpovědí bude, že všechny jevy obsažené v nervové doméně jsou jeho příčinou. Znamená to tedy, že při doménové fragmentaci světa můžeme vykládat vztahy mezi jevy napříč doménami kauzálně, avšak v tom případě musíme připustit, že jde o kauzalitu poněkud zvláštního rázu. Může se totiž zdát, že tento výsledek není v souladu se „zdravým rozumem“ a základními pozorováními jevů reálného světa: je přece „jasné“, že inhibujeme-li v nervové kauzální doméně zpětné vychytávání serotoninu, v mentální doméně člověka trpícího depresí to *způsobí*, že deprese poleví či zmizí. Podle námi

<sup>15</sup> Toto vyplývá z logiky operace průniku; tak např. máme-li funkci  $G$  definovanou tak, že  $G(1) = \{1,2,3,4\}$ ,  $G(2) = \{1,2,3\}$ ,  $G(3) = \{2,3\}$ , pak  $\text{rng}(G) = \{\{1,2,3,4\}, \{1,2,3\}, \{2,3\}\}$  a  $\text{rng}(G) = \{2,3\}$ .

prezentovaných výsledků však je vymizení deprese *způsobeno* i například jevem zcela opačným, tj. posílením zpětného vychytávání serotoninu, byť by tento jev nebyl v danou chvíli pozorován. Z těchto důvodů nazvěme náš poznatek o mezidoménové kauzalitě jako *kauzální paradox vícedoménových systémů* (v užším případě též kauzální paradox emergentních systémů).

Zcela záměrně se nebudeme zabývat jakýmkoli klinickým faktem, jako například že někdy inhibice zpětného vychytávání serotoninu nejen k vyléčení deprese v danou chvíli nepomůže, ale depresivní stavy se dále prohloubí<sup>16</sup> (můžeme pak říci, že tato inhibice *způsobila* zhoršení deprese?) – tímto bychom totiž odváděli pozornost od „zásadnějšího“ problému (zásadnějšího pouze v rámci našeho textu – pochopitelně nikoli pro člověka trpícího depresí): je vůbec možné tvrdit (ve smyslu hlubším, než je pouhá jazyková konvence), že nějaká změna stavu věci v jedné kauzální doméně *způsobí* nějaký jev v jiné kauzální doméně? Dle našeho názoru nikoli, neboť pak se nelze vyhnout kauzálnímu paradoxu. Dozajista existuje nějaký vztah mezi stavem serotoninu na synapsích a depresí, avšak tento vztah nelze konzistentně chápat jako kauzální (tj. příčinný).

### 4.5.3 Interpretace kauzálního paradoxu

Kauzální paradox a jeho důsledky se zde pokusíme interpretovat ze dvou základních hledisek: z hlediska epistemologického – tj. co nám kauzální paradox vypovídá o možnostech, způsobech a hranicích poznání přirozeného reálného světa, speciálně pak jaké z něj můžeme učinit závěry v rámci logiky vědeckého bádání (ve smyslu „teorie teorií“, viz (Popper, 1997)); a z hlediska umělé inteligence, resp. kognitivní vědy – tj. jak se kauzální paradox promítne do prve nastíněné problematiky umělého a přirozeného myšlení.

Z logiky množinové formulace kauzálního paradoxu vidíme, že důležitým aspektem je fakt, že třída  $X_i$  je  $\pi$ -třídou vzhledem k obzornému řezu  $H_i$ , stejně jako je vzhledem k němu třída  $D_i$   $\sigma$ -třídou. Z toho plyne, že výklad příčinné souvislosti v mezidoménové interakci nutně požaduje, aby kauzální vztahy byly mezi všemi jevy výchozí kauzální domény (tj. všemi prvky řezu  $H_i$ ) a všemi zkoumanými jevy mimo tuto doménu. I kdyby se nám však podařilo klást na mezidoménovou „kauzální“ funkci  $F_{X_i}$  takové požadavky, aby nebyla definována na celém obzorném řezu  $H_i$  (tj. na celé kauzální doméně  $i$ ), ale pouze na některé jeho části – čímž by se zúžila třída myslitelných příčin (například kdybychom do celého formálního popisu vnesli časovost a nějak bychom zužovali třídu jevů z domény  $i$ , jež mohou být myslitelnými příčinami, toliko na ty jevy, které jsou v daném časovém okamžiku pozorované, resp. „aktuální“) – pořád bychom se nezbavili těch vlastností třídy  $X_i$ , které z ní činí  $\pi$ -třidu, a kauzální paradox by se nikterak neredukoval. Stále by totiž všechny jevy třídy  $X_i$  byly důsledky všech jevů ze zúžené třídy příčin (tj. z definičního oboru funkce  $F_{X_i}$ ).

Z třídy  $X_i$  činí  $\pi$ -třidu právě přirozeně nekonečné narůstání  $\sigma$ -třídy  $D_i$ , jemuž lze dle *Tvrzení 2* „zabránit“ pouze ostrým množinovým vymezením. Kauzální paradox tak můžeme obejít jedině tím, že vybereme nějakou ostrou podmnožinu jevů ze třídy  $H_i$ , s níž budeme nadále pracovat. Tímto „seříznutím“ původní kauzální domény se však všechna další tvrzení budou týkat nikoli původní kauzální domény, jak ji empiricky

<sup>16</sup> K činění podobných tvrzení se samozřejmě ani zdaleka necítíme být kompetentní.

seznáváme, ale jejího více či méně zkráceného ostře vymezeného množinového modelu. K výsledné redukované (či zkomolené<sup>17</sup>) doméně lze připojit podmnožinu jevů z třídy  $C_m$  a nad tímto můžeme dále hledat kauzální funkci již bez kauzálního paradoxu. Dostáváme však kauzální doménu novou, zcela odlišnou od obou původních kauzálních domén, jejichž vztah nás nejdříve zajímal. Naše závěry činěné nad touto novou doménou potom však hovoří (být naprosto konzistentně a kauzálně uspokojivě) o něčem zcela jiném, než o co nám zprvu šlo.<sup>18</sup>

Budeme-li se naopak snažit o nereduktivní výklad daného fragmentu světa, kauzálnímu paradoxu se nevyhneme. Jelikož na základě empirických kritérií nemůžeme jednoduše přijmout představu, v níž nelze každý jev jedné domény kauzálně vysvětlit jinak než jako důsledek všech jevů druhé domény, musíme nalézt jiné východisko. To spočívá v zamítnutí předpokladu kauzální interakce mezi jevy z různých kauzálních domén. Nemůžeme tedy tvrdit (rozumí se jinak, než v rámci prosté jazykové konvence), že jev z jedné kauzální domény (například již výše zmiňované zpětné vychytávání serotoninu) *způsobuje* jev z jiné kauzální domény (opět například fenomenologicky prožívané psychické stavy, jako deprese, apod.). Nemůžeme ani tvrdit, že by tento vztah byl kupříkladu charakteru pravděpodobnostního – do pravděpodobnostních vztahů mohou vstupovat pouze jevy, které jsou prvky pravděpodobnostní algebry, a tudíž jde v důsledku o jevy jiné (zástupné), než jsou původně zkoumané jevy z kauzálních domén. Pravděpodobnostní výklad je opět zkomolením (být nesmírně účinným) výchozí vícedoménové situace, neboť opět nahrazuje dvě kauzální domény doménou jedinou, v rámci níž jsou pravděpodobnostní jevy pak již kauzálně soudržné a bezesporné.

Podobně lze v rámci snahy o porozumění vědomí charakterizovat Chalmersův tzv. „*snadný problém*“ (Chalmers, 1996, cit. podle Havel, 2001). Problém, který lze takto pojmut, je v jistém (metodologickém) smyslu „*snadný*“, neboť jeho vymezení odpovídá redukcí mezidoménové interakce na kauzální vztahy v rámci jedné kauzální domény, vytvořené například prostřednictvím teorie pravděpodobnosti. Příkladem může být například snaha odpovídat jazykem neurologie na otázky kladené jazykem (objektivní) psychologie. Naproti tomu „*těžký problém*“ formulujeme takto (Havel, 2001, str. 69):

*Jak vůbec může fyzické (chemické, biologické) dění v mozku vést k mému subjektivnímu prožitku?*

Jinými slovy řečeno: co *způsobuje*, že projevy naší nervové soustavy nevnímáme jako projevy naší nervové soustavy, ale jako subjektivně prožívané mentální stavy, jež disponují svojí fenomenální komponentou? Kupříkladu proč excitaci jistých skupin neuronů nezpracováváme jednoduše jako excitaci jistých skupin neuronů, ale jako pohled na rozkvetlou louku? Ve formulaci „*těžkého problému*“ je z naší perspektivy jasně patrná snaha o nereduktivní zachování mezidoménové interakce.

Jakého charakteru však mohou být vztahy, které působí mezi jevy napříč kauzálními doménami, nejsou-li to vztahy kauzální? Chápeme-li kostry kauzálních domén jako třídy

<sup>17</sup> Termínu „*zkomolené jevy*“ užívá P. Vopěnka pro ostré „*aproximace*“ jevů neostrých – neostrý přírodní jev se stane ostrým pouze tehdy, vytvoří-li se seznam všech jeho výskytů; takový seznam však původní jev komolí a jeho zkomolenina je přísně vzato jevem novým. (Vopěnka, 2007)

<sup>18</sup> Vhodným zamaskováním tohoto aspektu pak dochází k dnes leckdy vídaným sebevědomým prohlášením typu „*podstata vědomí vysvětlena*“, apod.



jevů, které mohou být teoriemi kauzálně vysvětleny, pak je naše interpretace kauzálního paradoxu v souladu s představou, že kauzální vysvětlení nemůže překračovat hranice vymezené těmito teoriemi. I v případě, že bychom ve vztahu ke kauzalitě zaujali radikální postoj po vzoru D. Huma – že totiž všechny vztahy mezi jevy, které považujeme za kauzální, nejsou ve skutečnosti ničím jiným než pouhou koincencí (tj. souvíváním) daných jevů, a jen induktivní přístup spolu s pravidelným opakováním této koincidence nám umožňuje *věřit*, že vztahy jsou svázány kauzálně – nepodaří se nám uspokojivě překonat problematičnost mezidoménových vztahů. Po formální stránce se totiž v našem modelu nic nezmění – pouze relaci  $R_i$  budeme vykládat tak, že do ní vstupují dva jevy tehdy, jestliže jejich výskyty spolu pravidelně koincidují. Výsledkem pak bude paradox týkající se pravidelné koincidence jevů, zcela analogický ke kauzálnímu paradoxu.

Činění soudů o povaze vztahů či zákonů mezi doménami je dle našeho názoru záležitostí metafysiky, neboť jakékoli tvrzení o nich, jež by neredukovalo podstatu disjunktnosti obou neostře vymezených domén, by bylo nefalsifikovatelné.<sup>19</sup> Totiž i samotný princip kauzality stojí mimo vědu jako takovou:

*Víra v kauzalitu je metafysická. Není to nic než typická metafysická hypotéza dobře zdůvodněného metodologického pravidla – vědcova rozhodnutí nikdy se nevzdat hledání zákonů.* (Popper, 1997, str. 266)

A právě jedním z projevů tohoto metodologického pravidla je snaha interpretovat mezidoménové vztahy jako vztahy kauzální, a to výše zmíněným komolením vzájemné interakce dvou či více kauzálních domén do kauzální domény jediné, pokud možno ještě ostře vymezené.

Tolik bychom mohli poznamenat k epistemologickým závěrům plynoucím z kauzálního paradoxu. Emergentní efekt, o němž byla řeč v první části této kapitoly, lze též vykládat jako projev kauzálního paradoxu. Ať je totiž v daném časovém okamžiku vstup emergentního systému jakýkoli, systém se z vnějšího pohledu jaksi „rozhodne“, jaký výstup bude indukovat (dalo by se říci, že projevuje určitou „svobodnou vůli“). Jak toto „rozhodování“ probíhá, je otázka v zásadě shodná s otázkou po povaze mezidoménových vztahů. Bezpochybnou, avšak ryze metafysickou odpovědí může být například tvrzení, že v daném časovém okamžiku výběr aktuálně nastávajících jevů ze všech ekvivalentně příhodných důsledků provede princip čisté náhody, Bůh či nějaký strukturně nadřazený systém.

Emergentní efekt jako projev kauzálního paradoxu je (s přihlédnutím ke vztahu kauzálních domén a teorií) tak vlastně důsledkem nahlížení světa prostřednictvím dvou či více vzájemně nekompatibilních teorií (v širším slova smyslu). To by mohlo na první pohled znamenat, že emergence je opravdu pouhý epifenomén, jak jsme jej diskutovali z analytického hlediska dříve v této kapitole. Tento zjednodušující názor ale předpokládá, že neznalost<sup>20</sup> pozorovatele, která dle něj zapříčiňuje emergentní efekt,

<sup>19</sup> Kritérium falsifikovatelnosti navrhuje K. R. Popper jako konvenci k rozlišování mezi vědecko-empirickými a metafysickými systémy (mezi „metafysické“ systémy by se zde řadila i matematika a logika): empirické systémy lze z principu falsifikovat (tj. ukázat nějakou empirickou evidenci, která je v rozporu s universálním tvrzením daného systému), zatímco metafysické nikoli.

<sup>20</sup> „Neznalost“ v tom smyslu, jak byla formulována v diskusi o analytickém pojetí emergence.

může být poznáváním eliminována (čímž je eliminováno i zdání emergence), přičemž bariéry tohoto poznávání nejsou principiální, ale pouze technické. Za tímto předpokladem se však skrývá představa, jež je v jistém smyslu neoprávněná. Za neznalost je zde považována situace, kdy na jistý úsek reality pohlížíme prostřednictvím dvou či více vzájemně neslučitelných (či velmi těžko slučitelných) teorií, jimž odpovídají různé kauzální domény. „Poznávání“ v tomto kontextu pak znamená vytváření nové teorie, která „spojí“ původní teorie neslučitelné. Tím vzniká též nová kauzální doména, jež se rozepne přes původní disjunktní domény, čímž je sloučí a umožní koherentní kauzální výklad všech jevů v nich obsažených, vyhýbajíc se tak kauzálnímu paradoxu a emergentnímu efektu. Jak jsme však ukázali výše, dojde tím ke zkomolení původního uchopení stavu věcí, čímž nakonec poznáme něco jiného, než jsme nejdříve zamýšleli.

Doménová fragmentace světa je nedílnou součástí lidského vnímání. Emergentní efekt je pak důsledkem této fragmentace. Lidské poznávání v podstatě přeskupuje strukturu rozvrstvení domén, nové domény vytváří či staré opouští, takže emergentní efekt postupně mění svoje nositele, na nichž se ukazuje. Spojení jedné či více teorií<sup>21</sup> nevysvětlí dříve přítomný emergentní efekt, pouze jej opustí a ponechá zcela mimo dosah svého zájmu (spolu s celou řadou jevů, které „epistemicky zaniknou“ tímto zkomolením původních kauzálních domén). Navíc tím vznikne nová kauzální doména, díky jejímuž obzoru se objeví nový emergentní efekt. Dalo by se říci, že cílem poznání je tedy zachycení celého světa v rámci jedné kauzální domény. To by však mělo velmi závažné důsledky v dalších úvahách o emergujícím myšlení.

#### 4.5.4 Vztah kauzálního paradoxu a emergujícího myšlení

Vraťme se nyní k zobecněné emergentistické tezi. Dle ní lze mentální stavy a procesy chápat jako emergentní jevy nad třídou kauzálních domén. Již jsme poznali, že kauzální domény (i jejich modely v rámci DTM) jsou úzce spjaty s pohledem pozorujícího subjektu. Emergující myšlení tedy vyvstává nad kauzálními doménami, avšak aby tomu tak mohlo být, musí již nějaký pozorující subjekt předem existovat, aby se k němu mohly tyto domény vztahovat. Pozorující subjekt však sám (jsouc subjektem vědomým) musí vycházet z emergentního efektu. Dostáváme se tímto opět k hermeneutickému kruhu. Můžeme si představit, že jeden z těchto dvou objektů (třída kauzálních domén či pozorující subjekt) byl dán a priori (například Bohem), nebo žádný z těchto objektů nebyl dán a priori, přičemž druhá možnost by mohla být vysvětlena „evolučně“ (resp. „vývojově“): vědomá mysl a kauzální domény vznikaly souběžně krok za krokem ve vzájemném zjednávání až do takového komplexního stavu, v jakém o nich mluvíme dnes.

Nicméně i druhé vysvětlení je problematické, neboť aby bylo toto postupné vzájemné zjednávání možné, musí již ve světě být přítomná neostrost, která by umožnila realizaci principu „přirozené indukce“<sup>22</sup> namísto principu klasické neslábnuocí indukce. Není-li totiž nějaký systém v daném časovém okamžiku pozorujícím subjektem, pak jím nebude ani v bezprostředně následujícím časovém okamžiku. Neslábnuocí indukce by tudíž narozdíl od „přirozené indukce“ nedovolila pozorujícímu subjektu vzniknout, neboť by

<sup>21</sup> „Spojení teorií“ ve smyslu vytvoření nové teorie, která vysvětluje jak všechny jevy původních neslučitelných teorií, tak i vztahy mezi jevy napříč těmito teoriemi.

<sup>22</sup> „Přirozená indukce“ je zde opět chápána ve smyslu komplementárního principu k přirozenému nekonečnu – viz odstavce o DTM.

nikdy nemohlo dojít k takovému „fázovému přechodu“. Neostrost umožňující tuto „přirozenou indukci“ je v oboru abstraktních tříd také podložena obzorným řezem, takže dochází k jakési emergenci „druhého řádu“, z níž emergence „prvního řádu“ emerguje. Hermeneutický kruh je tímto rozšířen do „hermeneutické sféry“. Můžeme nyní dále analogicky pokračovat a předpokládat emergence stále vyšších řádů (spolu s „hermeneutickými hyper-sférami“), a to až do nekonečna (zřejmě též přirozeného). Všechny tyto kroky mají společnou jednu základní věc, jíž je přítomnost neostrosti. Má-li tedy mysl (resp. mentální stavy a procesy) být emergentním jevem, musíme přijmout alespoň jeden metafyzický princip, jímž je právě apriorní přítomnost neostrosti ve světě.

Takto chápaná mysl přináší ale zároveň dostatek prostoru pro *svobodnou vůli*, neboť vymaňuje lidské (a možná i umělé) myšlení ze zajetí přímé kauzální determinace dané slepým během procesů v biologických a fyzikálních doménách, z nichž mysl emerguje. Jak jsme již totiž ukázali, emergentní efekt žádnou takovou kauzální determinaci nemá. To, co „rozhoduje“ o výsledných projevech emergentních systémů, a tedy i myslí (čili to, co „způsobuje“ svobodnou vůli), je skryto v neprostupnosti  $\pi$ -třídy, z níž na podkladové kauzální domény hledíme, a při snaze o zachycení prostřednictvím jiné kauzální domény (tedy  $\sigma$ -třídy) to nenávratně mizí (zde bychom mohli poukázat na analogii s Heisenbergovým principem neurčitosti ve fyzice).

Úvahy o možnostech zkonstruování umělého myšlení se také nevyhnou konfrontaci se závěry plynoucími z kauzálního paradoxu. Vytvořit umělé myšlení znamená na základě konstrukčního plánu sestavit systém, který by duplikoval myšlení přirozené, tedy lidské. V rámci emergentistického paradigmatu by šlo o vytvoření takového prostředí *in silicio*, v němž by díky vzájemné interakci jeho prvků došlo ke vzniku (při pohledu z vnějšku) emergentního efektu, který by odpovídal přirozenému myšlení. Konstrukční plán by tedy měl charakterizovat všechny relevantní kauzální vztahy mezi prvky podkladového prostředí a požadovanými projevy systému jako celku. Jak je však z kauzálního paradoxu zřejmé, přítomnost emergentního efektu vylučuje smysluplné charakterizování takovýchto mezidoménových kauzálních vztahů. V kontextu tohoto nejsilnějšího pojetí umělého myšlení se tudíž domníváme, že umělé myšlení nelze v tomto smyslu zkonstruovat.

Jiným případem však je vytvoření takového umělého emergentního systému, jehož primárním cílem je simulace určitého typu evoluce. Sekundárně se pak v tomto systému mohou objevit emergentní jevy, které by určitému typu myšlení odpovídaly, avšak vztah konstruktéra k těmto jevům je již zcela jiného typu než u prve zmiňovaného případu. Další variantou může být situace, kdy konstruktér začne iterativně vytvářet takovéto emergentní systémy, přičemž bude experimentálně (čili metodou pokus-omyl) měnit jejich parametry (což odpovídá jakési „meta-evoluci“ systému) do té doby, než se po nesmírném množství iterací v systému objeví nějaký emergentní efekt, jenž by odpovídal přirozenému myšlení. Tehdy by však bylo na pováženu mluvit o systému *umělém*, když většina jeho zajímavých vlastností mu byla dána odjinud než konstruktérem, pro něhož by byla podstata emergentního efektu stejně nepřístupná, jako u přirozeného myšlení samotného. Stejná situace by nastala i tehdy, kdyby někdo dokázal buňku po buňce sestavit repliku skutečného člověka. Šlo by pak vůbec o člověka umělého?

Obecně lze říci, že podaří-li se uměle vytvořit emergentní systém, o němž se prohlásí, že explicitně simuluje přirozenou kognitivní funkci či mentální proces a zároveň se

svému konstruktérovi zcela nevymkne kontrole, pak přesto dochází k redukci požadovaného jevu (tj. mentálního či kognitivního procesu) na některé jevy podkladové kauzální domény – tj. o požadovaných jevech se řekne, že jsou ekvivalentní s některými jevy domény, v rámci níž je systém zkonstruován. Interpretace výsledku je však zcela v rukou pozorovatele systému – systém sám není interpretantem. Tento fakt však není v souladu s požadavkem přítomným v sémiotické formulaci umělého myšlení, která vychází z pojetí sémiotiky (nauky o symbolech) C. S. Peirce, jenž symboly charakterizuje triadickým popisem, skládajícím se z těchto aspektů: 1) znak, neboli representamen; 2) bezprostřední objekt; 3) interpretant. Základní problém umělé inteligence se dá v tomto kontextu formulovat tak, že je třeba určit, jaké kroky se mají provést, aby se počítač octnul v roli interpretanta. (Markoš, Kelemen, 2004) Lze to ilustrovat na velice jednoduchém příkladu umělé neuronové sítě, jejímž cílem je rozpoznávat ručně psaná písmena. Rozpoznané písmeno je indikováno aktivací daného neuronu ve výstupní vrstvě sítě, avšak interpretovat, že tento jev je rozpoznáním písmene, rozhodně není v kompetenci sítě, ale jejího uživatele. Stejně tak v potenciálním umělém modelu lidské mysli nemůže přicházet v úvahu, že by aktivace neuronu s názvem „strach“ jakkoli odpovídala skutečnému mentálnímu stavu „mít strach“. Z těchto skutečností je zřejmé, že sémiotický aspekt hraje v úvahách o přirozeném a umělém myšlení důležitou roli.

Také doménová fragmentace světa je svázána se způsobem, jímž lidé provádějí interpretaci symbolů, které je obklopují. Možný výklad sémiotického přístupu chápe symbol takovým způsobem, že jeho prostřednictvím interpretant přímo vyděluje bezprostřední objekt z chaosu vjemů okolního světa. V časovém běhu však nelze celou symbolovou triádu chápat staticky – znak, bezprostřední objekt i interpretant jsou v neustálé přeměně, celá triáda se může stát jedním z prvků triády jiného symbolu (tj. celý symbol může z perspektivy jiného symbolu figurovat například jako bezprostřední objekt) a dochází k neustálému a nekonečnému interpretačnímu kroužení. Na základě aktivní snahy, již interpretant projevuje při tomto *zjednávání*, se mu začnou vyjevovat zákonitosti panující mezi jevy okolního světa. Uvnitř těchto zákonitostí vyvstávají z bezprostředních objektů objekty dynamické, které narozdíl od objektů bezprostředních nejsou ve světě jaksi „pasivně“ přítomny, jednoduše svázány poukazem k nějakému symbolu, ale aktivně jsou interpretantem *zjednávány* v rámci celého sémiotického procesu, aby mu výklad okolního světa „dával smysl“. Zákonitosti panující ve světě potom lze v tomto kontextu považovat za *zvyklosti* zavedené dlouhým interpretačním procesem (tradicí).

Na svět tedy Peircova sémiotika nahlíží jako na nikdy nekončící výkladový proces (sémiosis). K sémiotické triádě navíc přistupuje další faktor – zvyk – díky němuž můžeme celou sémiosi *jakoby* pozastavit a v rámci tohoto vymezení času a prostoru definovat významy *přesně*. Proces sémiose však ve skutečnosti běží dál, proto jsou tyto definice přesné pouze v daném zafixovaném výřezu, nikoli ve *zjednávaném* světě jako takovém. Při jeho poznávání se však chováme, jako by definované byly přesně, a to tak dlouho, dokud trvá zvyk. Poznání vědecké pravdy tak jde ruku v ruce s kulturou, neboť závisí na dlouhodobém úzu celého společenstva.<sup>23</sup>

---

<sup>23</sup> Výklad o procesu sémiose zde byl volně převzat z (Markoš, Kelemen, 2004).

Aktuální rozložení kauzálních domén tedy zrcadlí průběh sémiose a odpovídá osobním, kulturním i evolučním „zvyklostem“. Zachycení tohoto rozložení prostřednictvím polomnožin v rámci DTM v sobě obsahuje i dynamiku sémiotického procesu. Prve zmiňované komolení jevů zastoupených polomnožinami pak odpovídá fixování statických „výřezů“ tohoto dynamického procesu. Emergentní efekt je tedy zejména projevem dynamiky nikdy nekončícího výkladového procesu nahlížení světa, jeho různé podoby a přítomnosti se však mění, a to s postupnými změnami zvyklostí daných proměnami společenského úzu, evolučními změnami či osobními postoji.

Netvrdíme, že by z principu nebylo možné sestrojiti umělou bytost myšlením stejnou jako člověk. Záměrně se však vyhýbáme nazvat tuto bytost „umělou bytostí s vědomím“. Kauzálnímu paradoxu, který by bránil vytvoření „umělé bytosti s vědomím“, se totiž „jednoduše“ můžeme vyhnout redukcionistickým vyprázdněním mentální domény s tím, že mentální stavy a procesy (strach, láska, bolest, vědomí, apod.) nejsou *nic jiného než* jejich koreláty z domény nervové. V důsledku tohoto redukcionismu dojde k postupnému odstranění neostrosti, čímž zanikne zde přítomný emergentní efekt a bude již možné kauzálně popsat vztahy mezi nervovými jevy a redukovanými „mentálními“ jevy.

Kupříkladu je-li v našem porozumění chování mravence přítomná neostrost a mezidoménová interakce, je toto chování provázeno emergentním efektem, a mravence tak vnímáme jako intencionální bytost, která projevuje jisté aspekty svobodné vůle (např. „mravenec se *chce* dostat tam a tam“, či „mravenec cítí bolest, a proto běží pryč“, apod.). Nicméně podaří-li se nám (nyní alespoň v rovině myšlenkového experimentu) redukovat celé mravenčí chování a život do domény jeho nervového fungování, mravenec přestává být intencionálním tvorem a stává se prostým reaktivním agentem, u něhož nemá příliš smysl hovořit o tom, že „žije“.<sup>24</sup>

Podobně většina živých bytostí je „sestavena“ z miliard reaktivních agentů a pouze přítomnost neostrosti umožňuje prostřednictvím mezidoménové „paradoxní kauzality“ vnímat je jako živé bytosti, nikoli jako množiny miliard reaktivních agentů. Jelikož však toto vnímání vychází z aktuálního rozvržení doménové fragmentace světa, můžeme předpokládat, že se s postupem času bude v rámci nekonečného procesu sémiose měnit.

Z výše uvedených úvah je patrné, že konstitutivním prvkem vědomí<sup>25</sup> (ať přirozeného, nebo umělého) je „stav věcí“, který bychom mohli nazvat „*fundamentální agnosticismus*“, a to nikoli jako pouhý záměrný epistemologický přístup nějaké osoby, ale ústrojnou součástí (zjednávaného) vědomě vnímaného světa. Můžeme říci, že vědomé mysli zjednávají sebe samé tím, že „nevidí všechno“. Každé vědomí může z principu rozšířit svůj obzor a vysvětlit jevy, které vnímalo předtím, jako něco jiného (tj. prostřednictvím jevů nových, dříve neviděných). Nicméně to, co bylo vnímáno předtím, je pak navždy ztraceno a nemůže být znovu vyvoláno. Pro jednoduchou ilustraci nechť si čtenář představí libovolný dům (tj. ve své mysli vyvolá obraz domu). Má-li již představu domu hotovou, nechť nyní odpoví na otázku, jakou barvu mají okapy tohoto domu. V naprosté většině případů si čtenář bude muset přimyslet k domu ještě okapy, čímž

<sup>24</sup> Snad s výjimkou vymezení „živého systému“ jako systému autopoietického, tj. systému se specifickým případem rovnováhy sítě procesů, která průběžně regeneruje sama sebe (resp. svoji vlastní strukturu) a zároveň konstituuje systém jako konkrétní individuum v prostoru. (Maturana, Varela, 1980)

<sup>25</sup> Od tohoto místa již záměrně užíváme termínu „vědomí“ místo původního „myšlení“, byť tím klademe větší požadavky na jeho intuitivní uchopení.

začne pracovat s jinou představou, než původního domu, o němž nebylo možné říci, že by okapy neměl, pouze že jejich (ne)přítomnost byla skrytá za obzorem. Nyní však již čtenář tento experiment zodpovědně zopakovat nemůže, neboť se nelze vrátit k původní představě stejného domu s okapy za obzorem – dokážeme si představit již pouze dům s okapy či dům bez okapů, případně představu o tom, jaké to je představit si tento stejný dům s okapy za obzorem. Jiným příkladem je představa Boha, která v naší společnosti (s přihlédnutím ke stavu poznání v rámci sémieose, jež v naší společnosti probíhá) již nejspíše nikdy nebude taková, jako před tisíci lety. Můžeme se pouze snažit představit si, jak si tehdejší člověk Boha (či bohy) představoval.

Otázkou zůstává, zda takovéto neustálé sémiotické rozšiřování obzoru může být nekonečné, nebo se jednou zastaví (resp. zkolabuje). V případě, že by bylo přirozeně nekonečné (neboť to je právě ten typ nekonečna, který se v přirozeném reálném světě nachází), pak by nejspíše mohlo dojít k jeho kolapsu, a to v situaci, kdy by nastalo úplné vyprázdnění mentální domény a mysl by byla definitivně vysvětlena pouze v pojmech výše nastíněného redukcionismu. Chtělo by se říci, že tato změna by de facto neměla vliv na naše vědomí, které je nám (jako druhu) jednou dané, bezprostředně je cítíme, a tudíž bychom pouze poznali jeho podstatu, aniž by ji to ovlivnilo. Musíme si však uvědomit, že naše vědomí je emergentním efektem nad nesmírně složitým souborem vzájemně interagujících kauzálních domén, jejichž rozvržení je dáno právě sémiotickým procesem, který sám je zpětně ovlivňován činností našeho vědomí (resp. našich vědomí – v plurálu). Vliv poznávání (ve smyslu reduktivním) vědomí v tomto procesu vzájemného ovlivňování a tvarování vede, přísně vzato, k mizení vědomí samotného, tj. zjednávaní světa bez vědomí jednotlivců, jak je známe dnes. Samozřejmě jde o časově velice dlouhodobý proces – nemůže se stát, že by se dotkl života jednotlivců – avšak máme za to, že již dnes člověk vnímá své vlastní vědomí jinak, než například před dvaceti generacemi. Důkazem může být třeba pocit jakési nepatřičnosti, máme-li se v dnešní době vyjadřovat o vědomí v rámci vědeckém diskursu – ve vyostřeném případě dokonce předstíráme, že žádné vědomí ani nemáme.

Společnost (a tudíž i její kulturní úzus) je v procesu sémieose ovlivňována a směřována svými vlastními předchozími interpretačními procesy. Díky otupujícímu vlivu zjednodušujícího a zavádějícího způsobu přenosu informací v dnešním světě (zejména prostřednictvím současných sdělovacích médií, jejichž styl lze parafrázovat prohlášením typu: „Američtí a japonské vědci *objevili* robota, který dokáže mít rád svého majitele.“) začíná již množství lidí žít tak, že kupříkladu láska pro ně není *nic než* „vzbouření hormonů“ či navyklé vzorce chování, nebo že za činy jednotlivce nemůže on sám, nýbrž pouze jeho psychika (či dokonce mozek). Všechny tyto nekonečně malé (opět v přirozeném smyslu) kroky se však z vývojového hlediska podílejí na postupných kvalitativních změnách vědomí jako takového.

Tudíž pokud by se nám (jako lidstvu) nakonec přece jen podařilo vytvořit *umělé vědomí*, nebylo by to pro nás již stejně žádné *vědomí*, neboť by už nebylo žádné „my“ a ani žádné přirozené vědomí, které by dokázalo rozhodnout, zda to, co jsme vytvořili, je opravdu to správné *Umělé Vědomí*. Nebyli bychom totiž *nic jiného než* reaktivní agenti, neboť poslední kauzální paradox by byl odstraněn a s ním i poslední emergence, přinášející nám náš život, jak jej známe nyní.

**Literatura**

- Brooks R.: *Cambrian Intelligence: The Early History of the New AI*. MIT Press, Cambridge, MA, 1999.
- Chalmers D. J.: *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press, Oxford, 1996.
- Fiala J.: *Analytická filosofie – úvod*. OPS, Plzeň – Nymburk, 2006.
- Havel I. M.: *Artificial thought and emergent mind*. In: Proceedings IJCAI'93, Morgan Kaufman Profesional Book Center, Denver, CO, USA, 1993, s. 758-766.
- Havel I. M.: *Causal domain and emergent rationality*. In: B. Smith, B. Broguard (editoři): *Rationality and Irrationality*. Proceedings of the 23rd International Wittgenstein Symposium, Kirchberg, Austria, 2000.
- Havel I. M.: *Přirozené a umělé myšlení jako filozofický problém*. In: Mařík V. a kol.: *Umělá inteligence 3*. Academia, Praha, 2001.
- Markoš A., Kelemen J.: *Berušky, andělé a stroje*. Dokořán, Praha, 2004.
- Maturana H. R., Varela F. J.: *Autopoiesis and Cognition*. D. Reidel, Dordrecht, Holland, 1980.
- Popper K. R.: *Logika vědeckého bádání*. Oikoymenh, Praha, 1997.
- Pstružina K.: *Svět poznávání (K filozofickým základům kognitivní vědy)*. Nakladatelství Olomouc, Olomouc, 1998.
- Thagard P.: *Úvod do kognitivní vědy (Mysl a myšlení)*. Portál, Praha, 2001.
- Vopěnka P.: *Úvod do matematiky v alternativnej teórii množín*. Alfa, Bratislava, 1989.
- Vopěnka P.: *Úhelný kámen Evropské vzdělanosti a moci*. Práh, Praha, 2000.
- Vopěnka P.: *Meditace o základech vědy*. Práh, Praha, 2001.
- Vopěnka P.: *Dynamická teorie množství (připravováno)*. OPS, Plzeň – Nymburk, vyjde 2007.