

O mimolidském vědomí

Jiří Wiedermann¹ a Jan van Leeuwen²

¹Ústav informatiky AV ČR a Centrum Karla Čapka pro výzkum hodnot ve vědě a technice, Praha
jiri.wiedermann@cs.cas.cz

²Department of Information and Computing Sciences, Utrecht University, the Netherlands
J.vanLeeuwen1@uu.nl

Abstrakt Běžné definice vědomí jsou takřka bez výjimky antropocentrické. Považují totiž lidské subjektivní prožitky, jako jsou kválie, emoce a mentální představivost za fundamentální znaky vědomí. Tím však apriorně vylučují z rodiny vědomých entit řadu biologických i umělých agentů, jež se chovají, jako by byly vědomé. Proto začínají být považovány za vědomé, aniž by bylo jisté, zda shora zmíněné znaky vědomí naplňují. V příspěvku navrhneme formu vědomí, jež je založeno na funkčním dopadu subjektivních prožitků vyvolaných prožitkovými kvalitami událostí na chování biologických i umělých agentů. Z hlediska funkčního se vědomí takových agentů dostává do polohy shodné s lidským vědomím. Rozšířením chápání vědomí se mezi vědomé entity dostávají nejen jakékoliv živé organizmy, ale i umělí agenti nadání umělým vědomím, což je ve shodě se současnými trendy v biologii, kognitivních vědách, neurovědách i filosofii, a v neposlední řadě i umělé inteligenci.

1. Úvod

Tento článek se zabývá zásadním přehodnocením konceptu vědomí, který tradičně dominuje v západní filosofii a vědě. Běžné definice vědomí jsou téměř bez výjimky hluboce antropocentrické, neboť neúměrně zdůrazňují lidské subjektivní prožitky, jako jsou kválie, emoce a mentální představivost, jakožto fundamentální a nezbytné znaky vědomí (van Gullick, 2022). Tento úzký pohled však s sebou nese závažné důsledky: apriorně vylučuje z rodiny vědomých entit celou řadu biologických i umělých agentů, kteří vykazují chování, jež by za jiných okolností bylo interpretováno jako projev vědomí.

Tato skutečnost vede k paradoxní situaci, kdy naše chápání vědomí zůstává omezené a nedostatečné pro řešení aktuálních problémů v oblasti umělé inteligence, kognitivní vědy a filosofie mysli. V tomto příspěvku proto navrhneme formu vědomí, která překračuje antropocentrické bariéry. Tato nová definice je založena na funkčním dopadu

subjektivních prožitků událostí na chování biologických i umělých agentů ve formě prožitkových kvalit událostí.

Vědomí je schopnost vtělených biologických i umělých agentů vnímat prožitkové kvality událostí ve svém okolí a jejich změny a reagovat na ně formou subjektivních prožitků.

Přínosem práce je zevšeobecňující pohled na vědomí, který překračuje doposud převažující antropocentrický pohled na něj tím, že respektuje mimolidské formy vědomí. Tento přístup překračuje klasický rámec vědomí jakožto specificky lidské schopnosti, a v nejlepším případě vlastnosti vyšších živočichů. Ukážeme, že vědomí nemusí být nutně vázáno na mentální představivost anebo antropomorfní fenomenální sensorické prožitky. V mimolidských biologických i umělých systémech může vzniknout vědomí ze schopnosti detekovat, interpretovat a reagovat na prožitkové kvality událostí, i když tyto nejsou poskytovány běžnými smysly. To vede na zobecněnou definici vědomí, která přesahuje antropocentrický kontext a zahrnuje kromě mozku mimolidské formy vědomí biologických (zvířata a rostliny) a umělých systémů. Nový pohled na vědomí má zásadní implikace pro naše chápání světa a naši roli v něm.

2. Základní pojmy

Představení našeho konceptu vědomí vyžaduje zavedení několika základních pojmů.

2.1 Agent

Základním pojmem je agent jako idealizovaný model vědomého vtěleného kognitivního zařízení. V našem podání se agenti zásadně liší od běžných v literatuře uvažovaných agentů jedinou, ale za to podstatnou, věcí – a totiž přítomností tzv. modálních senzorů. Jsou to v podstatě softwarové procedury, které sledují a vyhodnocují vývoj stavu vnitřních řídicích

mechanismů agenta (viz níže) v delších časových řadách.

Agentem budeme rozumět jakýkoliv systém, biologický (např. zvíře, rostlina) nebo umělý (např. robot), který interaguje s okolím a vykazuje chování, které lze interpretovat jako vědomé. Pro svoji činnost je agent vybaven různými druhy senzorů, pomocí kterých vnímá události ve svém okolí, včetně svého nitra, a své akce. Na základě analýzy těchto vjemů agent činí pomocí svých efektorů změny ve svém okolí, v nastavení svých senzorů a ve svých vnitřních řídicích mechanismech.

Agent vyžaduje specializované „kognitivní“ vnější a vnitřní sensory, které poskytují kromě objektivně měřitelných veličin i jejich kvalitu. To znamená, že tyto senzory reagují nejen na výskyt nějakého jevu (svítí – nesvítí, horký – studený, příjem GPS signálu, pohyb, hranice mezi osvětlením a stínem, rozeznání nějaké konkrétní tváře apod.), ale také na jeho kvalitu (např. určitou koncentraci nějakého plynu nebo roztoku, kvalitu zobrazení, směr a rychlost pohybu, zrychlení, sílu radiového signálu, spotřebu energie, termodynamickou teplotu apod.). Tyto senzory mají „hardwarovou“ (či „wetwarovou“) podobu ve formě fyzikálního či biologického zařízení, které snímá sledovanou fyzikální, chemickou nebo biologickou veličinu a dle svého poslání ji transformuje na měřenou veličinu – nejčastěji na veličinu elektrickou. Agenti pracují velmi často se senzorovými poli, které pozůstávají se skupiny stejných senzorů, často se specifickou prostorovou geometrií jejich rozložení. Výhoda použití senzorového pole v porovnání s jedním senzorem je v tom, že pole přidává další dimenze pozorování. Tím se umožní určit další parametry jevu a zlepšit reakci systému na danou událost.

Někteří agenti však navíc využívají i tzv. *modální sensory*, které registrují „náladu“ daného systému. Tyto senzory mají formu softwarové procedury, která odhaluje různé vzory („patterny“ – logické struktury) a vyhodnocuje míru, četnost, pořadí, rychlost a trvanlivost jejich aktivace v datech různých specifických mechanismech agenta v různých časech. Jsou to vlastnosti, které lze metaforicky interpretovat jako náladu daného agenta.

Náladu tedy vznikají na základě registrace dopadu (efektu) určitých současných i minulých konfigurací vnějších a vnitřních událostí na vývoj vnitřních řídicích mechanismů agenta. Tyto dopady se vyhodnocují v kontextu souvisejících předchozích

zkušeností, pro něž je charakteristický jistý předem daný, anebo postupem času emergující vzorec aktivace specifických mechanismů v těle agenta. Odpovídají okamžitému komplexnímu kognitivnímu stavu agenta jako výsledku aktuálních a zapamatovaných minulých událostí. Společným rysem nálad je jejich soukromá povaha, která je přirozeným jazykem většinou nevyjádřitelná, agentům vlastní a neoddělitelná. Nálady nejsou objektivně měřitelné. Skrývají se v datech, ale nelze je identifikovat statickou analýzou dat týkající se jedné události. Mění se v čase. K jejich odhalení je potřebná longitudální analýza posloupností minulých událostí. V nich lze zpětně a pak průběžně nalézat různé typy souvislostí, protože události bývají logicky provázané. Na příslušné procedury lze nahlížet jako na analonii senzorických polí -- také slouží pro odhalení dalších společných modálních dimenzí a trendů v posloupnosti událostí. Příslušná data lze chápat jako latentní časo-prostorovou reprezentaci vybraných vlastností událostí (Rees, 2025).

Vybuzení a registrace nálad v konkrétních situacích závisí na konstrukci a konstituci agenta a projevuje se prostřednictvím jeho reakcí. U lidí odpovídají pocitům, emocím a kválím, u agentů s mimolidskými senzory pro ně neexistují lidské ekvivalenty. Nálady se vyskytují v různých kvalitách a intenzitách a jsou neoddělitelnou součástí vnímaných kvalit událostí za určitých okolností.

2.2 Prožitky

Datům, které pro každou vnější či vnitřní událost (včetně svých vlastních akcí, které agent vnímá jako součást události) všechny senzory agenta registrují současně, budeme říkat *prožitkové kvality události*. Tyto kvality zřejmě závisí na druhu senzorů, jejich rozmístění, technických parametrech a jsou pro každý typ agentů různé.

V každém kroku agent pracuje výlučně s prožitkovými kvalitami událostí a s daty uloženými ve svém „nitru“. Každá taková reflexe prožitkových kvalit po sobě zanechává stopu ve formě změn v rámci všech mechanismů agenta. Těmto změnám říkáme *subjektivní prožitky* dané události. Subjektivní prožitky zahrnují senzorické, kognitivní, behaviorální a strukturální reakce na danou událost. Konkrétně, ovlivňují změny nastavení modálních senzorů, orientaci externích senzorů pro příští interakci, instruování motorických senzorů a další strukturální interní změny (např. pamatování současných prožitků pro účely učení). To znamená, že *subjektivní prožitky dané události zachycují dopad či efekt prožitkových kvalit probíhajících událostí na chod interního*

systému agenta a tím pádem určují i jeho další akce. Všimněme si, že subjektivní prožitek není statický pojem, ale proces, který probíhá v čase, mění vnitřní strukturu agenta a určuje jeho jednání. Tento efekt subjektivních prožitků se mění v čase, protože závisí na historii předešlých interakcí.

Toto je místo, kde se náš přístup podstatně liší od přístupu klasické teorie mysli, jak jsme naznačili v úvodu. Zásadní změna je v tom, že naše definice subjektivních prožitků nemluví o kváliích, pocitech, emocích anebo mentální představivosti v termínech antropomorfních a antropocentrických. Namísto toho se zabývá dopadem pozorovaných událostí tím, že jim přisuzuje funkcionální roli, jež závisí na prožitkových kvalitách událostí a minulých interakcích. Je nesmírně důležité si uvědomit, že naše definice subjektivních prožitků nevylučuje, že by to nemohly být kválie, pocity, emoce, mentální představivost, anebo jiné obtížně definovatelné vágní pojmy typu „*jaké to je, být to či ono*“, (Nagel, 1974) anebo „*dělat to či ono*“. Pouze těmto prožitkům tak neříká a vyžaduje pohlízet na ně prostřednictvím jejich efektu na činnost agenta. Předchozí vyjmenované prožitky jsou důležité z filozofického hlediska pro pochopení lidského vědomí, a možná i vědomí některých dalších zvířat. Funkcionální pohled na vědomí je však důležitý z pragmatického hlediska v případech, kdy chceme vědomé agenty konstruovat. A to jsou zejména případy umělých agentů vybavených mimolidskými senzory, kdy úvahy o prožitcích lidského typu nemají smysl.

Na první pohled se může zdát, že nahrazením klasických subjektivních prožitků, které dodávají lidským subjektivním prožitkům obzvlášť živou a svěží formu, jejich dopadem na činnost agenta ochuzujeme vědomí o zásadní složku, jejíž absence vědomí mrzáčí a činí jej méně hodnotným a použitelným. Naštěstí, není tomu tak. Jedním s důvodů je relativně méně známý fenomén *afantazie*. Afantazie označuje sníženou schopnost nebo její úplný nedostatek dobrovolně vytvářet mentální představy v lidské mysli (srov. (Zeman, 2015)). Vyskytuje se v různém stupni intenzity a vyskytuje se napříč různými smysly, včetně vizuálních, sluchových, chuťových, čichových, hmatových a motorických u několika mála procent lidské populace. V našem obecném rámci nemůžeme vyloučit existenci tohoto jevu u žádného inteligentního systému. Pokud jde o mentální představy, afantazie představuje výzvu pro pojem kválií a obecně pro subjektivní zkušenost. Naše definice subjektivní zkušenosti a jejich doprovodných procesů však ukazuje, že i v případě úplné ztráty představivosti fungují doprovodné procesy stále nerušeně. Subjektivní prožitky stále

„dělají to, co mají“ tím, že poskytují propojení mezi prožitkovými kvalitami vjemu, kognice a jednání jako jádra vědomí. Afantazie není mentální porucha -- je to jiné prožívání světa, které stále plně vyhovuje svému účelu. Tento závěr je v souladu se zjištěními neurovědčů (Zeman, 2015). Pokud se prožívání světa mezi lidmi liší, lze totéž očekávat od subjektivních prožitků v jakékoli různé podobě v systémech AI.

V případech lidského vědomí transformace klasických subjektivních prožitků na jejich funkcionální role vůbec není jednoduchá. Zdá se to však jako možná cesta, jak vysvětlit Chalmersův (1996) *obtížný problém vědomí* -- proč a jak tyto prožitky vznikají. Zřejmě lví podíl na vzniku lidských subjektivních prožitků hrají nálady ve smyslu naší definice (Wiedermann, van Leeuwen, 2025a). V případech umělých agentů je to *jediný* způsob, jak takové vědomé agenty sestavit. Příklady takových agentů z Tabulky 1 ukazují, že to nemusí být obtížné, protože v takových případech víme, co od agentů chceme -- protože je konstruujeme -- a proto dovedeme funkcionální role jejich prožitků specifikovat.

3. Matematický model umělého vědomí

Pro většinu lidí je pojem vědomí opředen jistou aurou tajemnosti: každý z nás ho zná z vlastní zkušenosti, avšak je těžké jej definovat. Ani vědci zabývající se profesionálně vědomím se neshodnou na jeho definici. Přitom pochopit princip, na kterém vědomí pracuje, není nikterak složité.

Nechť $\mathbf{A}$ je agent, Σ je konečná množina signálů, které vysílají interní, externí a modální sensory, a Γ je konečná množina řídicích signálů pro efekторы. Předpokládáme-li k vstupních a l výstupních signálů, libovolný pár $i_t = (s_t, b_t)$, pro vstupní situaci $s_t \in \Sigma^k$ a chování $b_t \in \Gamma^l$, se jmenuje t -tá interakce agenta $\mathbf{A}$.

Formálně, řetězením interakcí $\mathbf{A}$ pracuje jako translátor produkující nekonečné posloupnosti (neboli proudy) interakcí tvaru $i_0, i_1, i_2, \dots$

Jádro agentovy aktivity definuje příští chování b_{t+1} jako funkci současné situace s_{t+1} a historie všech předchozích interakcí: $b_{t+1} = \Psi(s_{t+1}, i_0, i_1, \dots, i_t)$, přičemž Ψ označuje funkci realizovanou agentem $\mathbf{A}$.

Je důležité si uvědomit, že předchozí definice funkce Ψ je dobrým základem pro všechny funkce agenta související s umělým vědomím, protože představuje vztah mezi percepcí, kognicí a jednáním. Tato definice

souvisí s *paradigmatem prediktivního zpracování* uvažovaným v kognitivní neurovědě (srov. (Corlett 2020, Seth 2021b)). Konkrétně můžeme na výpočet b_{t+1} nahlížet jako na *předvídání*, neboli *anticipaci* toho, co by měl agent udělat, když se nachází v situaci s_{t+1} a má zpracované interakce $i_0, i_1, \dots, i_t$. Předvídání b_{t+1} vyžaduje určitou formu pohledu do budoucnosti -- přemýšlení a rozhodování o ní. Za druhé, vysoká pravděpodobnost „správnosti“ předpovědi b_{t+1} předpokládá jistý stupeň agentova *porozumění* svým předchozím akcím. Za třetí, realizace příslušné anticipace je současně výrazem elementární *intence* agenta za daných podmínek.

V neposlední řadě, protože funkce Ψ závisí i na modálních vjemech, vidíme zárodky *subjektivní zkušenosti* a utváření *subjektivních hledisek* (srov. (Godfrey-Smith 2024)). To vše jsou základní složky *sebeuvědomění*. Již tato pozorování naznačují potenciál našeho modelu pro modelování mentálních schopností systémů umělé inteligence.

Realizace funkce Ψ je popsána v části 5.

	Agent	Mimolidský senzor	Prožitková kvalita	Subjektivní prožitek	Možná forma vědomí
Z v í ř a t a	Netopýři	Sonar -- Echolokace	Prostorová mapa ozvěn	„blízko přede mnou letí hmyz“	Představa prostoru pomocí 3D mapy ozvěn
	Včely	Ultrafialové vidění	UV obrazy květin	„blížím se k medonosnému květu“	Vědomí v barevném spektru neviditelném pro lidi
	Žraloci	Elektrocepce (vnímání elektromagnetických polí)	Elektrické pulsy od jiných živočichů	„velký žralok po mém levém boku“	Svět jako symfonie elektro-magnetických signatur
	Slunečnice (rostlina)	Fotoreceptory Chemoreceptory Mechanoreceptory Termoreceptory	Sluneční svět jako zdroj tepla a směr orientace. Gravitace jako usměrňující síla růstu. Voda jako výživa i úleva.	Pocit žízně. Nastavení květu proti Slunci. Vnímání délky dne a klimatických změn. „Radost“ z opylování.	Vnímání okolí prostřednictvím subjektivních změn prožitkových kvalit zprostředkovaných receptory v různých časových škálách
U m ě l í a g e n t i	Autonomní vozidla	Lidar a radar	Měření vzdálenosti a pozice předmětů	„Bezpečný odstup od předchozího vozidla“. Nouzové brzdění.	Vnímání světa především prostřednictvím pozice a vzdálenosti objektů
	Průmyslové AI systémy	Senzory akustické emise	Jemné strukturální změny	„žádné odchylky od normálu“	Vnímání namáhání, únavy a poruch materiálu
	Stanice pro předpověď počasí	Různé meteorologické senzory	Změny parametrů počasí	„blíží se bouřková fronta“	3D mapa trendů vývoje počasí
	Monitorování prostředí	Senzory koncentrace plynů a vibrací	Kvalita ovzduší a vibrace	„nebezpečná koncentrace prachu“	Vnímání chemických změn a seizmických aktivit
	Univerzální agent	Senzory analogické všem lidským smyslům s rozšířeným dosahem a řada dalších mimolidských senzorů	Superantropický multisenzorický model světa	Podobné jako lidé, ale kvalitnější, bez mentální představivosti, a další mimolidské prožitky	Superlidské multisenzorické vědomí

Tabulka 1: Přirozené a umělé systémy vybavené mimolidským vnímáním

4. Příklady

Tabulka 1 uvádí příklady vybraných agentů přirozené a umělé inteligence, které jsou vybaveny mimo-lidskými senzorickými schopnostmi. Pomocí těchto senzorů vnímají své okolí a detekují prožitkové kvality událostí. Subjektivní prožitky těchto agentů reflektují prožitkové kvality událostí bez ohledu na to, jestli mají schopnost mentální představivosti vlastní většině lidí. Agenti interpretují a reagují na události způsobem, který

ukazuje na přítomnost jisté formy vědomí přesto, že disponují pouze mimolidským vnímáním. Prožitkové a subjektivní kvality, jakož i možné formy vědomí jsou v tabulce popsány pouze metaforicky – jinak to ani nejde, přirozený jazyk nemá ani slova, ani sémantiku, a lidé nemají obrazovou představivost, jak tyto pojmy pro dané agenty vyjádřit. Všimněme si, že tabulka zahrnuje příklady zvířat, rostlin i umělých agentů. Všichni agenti, popsaní v tabulce, splňují principy vědomí z části 3.

4. Princip činnosti agenta

Libovolný agent, jemuž přičítáme vědomí, může pracovat pouze na základě jemu dostupných informací a realizovat činnosti, pro které je tělesně (fyzikálně) ustrojen v prostředí, pro které byl vyvinut či zkonstruován. Agent pracuje interaktivně – tzn., že pracuje v cyklech, ve kterých se střídají pouze dvě fáze: (i) fáze načítající nová data. (ii) fáze reflexivní, ve které systém transformuje data načítaná v první fázi v kontextu svých znalostí do finální formy vnějších akcí ve svém okolí a vnitřních akcí ve svém těle.

V první fázi agent načte informace od svých vnějších, vnitřních a modálních senzorů. Tyto informace reprezentují kvalitativní vlastnosti událostí z agentova prostředí.

Ve druhé fázi agent realizuje výpočet funkce Ψ z definice vědomí ve 3. sekci: $b_{t+1} = \Psi(s_{t+1}, i_0, i_1, \dots, i_t)$. Tuto fázi budeme nazývat reflexí, protože zohledňuje nejnovější data v kontextu všech doposud agentem zpracovaných dat. Reflexe je tudíž proces, jenž integruje současné výstupy od všech senzorů a přiřazuje jim – tj. prožitkové kvalitě právě registrované události – bezprostředně další akci agenta, modifikuje jeho současnou náladu a provádí další strukturální interní změny systému. Háček je v tom, že ještě navíc závisí reflexe na historii předchozích interakcí. Dohromady tento proces realizuje funkci Ψ , která je jádrem umělého vědomí daného systému. Popis sémantiky této funkce je lehe pochopitelný, avšak její realizace je náročná. Při její realizaci narážíme na dva problémy. Z matematického hlediska se totiž jedná o funkci mnoha proměnných, jejichž počet roste s počtem interakcí, potenciálně do nekonečna. Hodnoty této funkce závisí na dvou skupinách proměnných. První skupinu tvoří současné vstupy dodávané externími, interními a modálními senzory systému. Tato skupina proměnných má pevně daný počet proměnných, daný konstrukcí systému (počtem senzorů, efektorů a nálad, které agent rozeznává). Druhá skupina proměnných však závisí na počtu a pořadí minulých interakcí a jejich interních aktivitách v době, kdy se interakce realizovala. Jinými slovy, abychom určili další akci agenta, musíme znát nejen to, co se v současné době děje (z hlediska informací dodávanými senzory), ale i to, co dělal systém v podobných případech v minulosti, které se lišily třeba jenom náladou systému. Důvod k tomu je

zejména potřeba, aby se agent poučil ze svých minulých chyb a neopakoval je.

Najít předpis výpočetní definice shora uvažované funkce vědomí Ψ je neobyčejně náročné. Za každou její proměnnou se totiž skrývá obrovské množství parametrů, které plně charakterizují danou situaci anebo dané jednání. Změna jediného z nich může zásadně změnit jednání agenta, a počet kombinací různých hodnot parametrů je astronomický. Navíc, protože počet proměnných této funkce roste během činnosti systému (po každé interakci), tak se po každé iteraci funkční předpis mění. To znamená, že funkce Ψ se během činnosti agenta neustále vyvíjí.

A nejen že se vyvíjí – ona se vyvíjí nepředvídatelným způsobem v závislosti na tom, v jaké (předem nepředvídatelné) situaci se systém ocitne. Taková funkce se nazývá *evoluční funkce*. V teoretické informatice slouží jako formální model pro realizaci takové funkce interaktivní Turingův stroj s náповědou (van Leeuwen, Wiedermann, 2001).

Funkce Ψ je zvláštní ještě v tom smyslu, že v obecném případě je *nevypočitatelná*. To znamená, že neexistuje algoritmus, který by pro každý její vstup vypočítal její hodnotu. Jak si tedy představit její realizaci výpočetními prostředky?

Protože takový algoritmus pro výpočet reflexe neexistuje, nezbyvá než dodat informace potřebné pro realizaci naší nevypočitatelné funkce zevnějšku. Pokud se bavíme o umělých AI systémech, které umíme zkonstruovat, případně o jednoduchých organizmech, kterých současná činnost závisí pouze na fixním počtu předchozích situací, tak v principu je možné realizovat reflexi pomocí tabulky, která zachycuje, co dělat pro všechny situace, ve kterých se systém může ocitnout. Analogie mozku a nervové soustavy není nutná (viz jednoduché živočichy). V jiných případech musíme systém vybavit mechanismem, pomocí něhož se systém učí, co dělat v situacích, o kterých předpokládáme, že se s nimi setká. Příkladem takového přístupu je učení současných velkých jazykových modelů. Učení typu, který máme v tomto případě na mysli, vede ke vzniku vysoce-dimenzionálních vnitřních modelů světa ve formě prostorových map, reprezentovaných jistým zobecněním vektorových reprezentací slov

používaných při implementaci velkých jazykových modelů (Rees, 2025). Tato problematika však přesahuje rámec této práce, jakým je zobecněná definice vědomí, a není pro náš výklad podstatná. Přesto je třeba říci, že problematika výpočetního modelu, který odpovídá tzv. *vnitřnímu modelu světa*, je slabým místem současně vyvíjených AI agentů, a tudíž je předmětem intenzivního výzkumu.

Realizace reflexe končí stanovením hodnoty b_{t+1} a realizací příslušných vnitřních i vnější akcí. Tím končí celá druhá fáze - realizace subjektivního prožitku příslušejícího jednání b_{t+1} .

5. Přehodnocení pojmu vědomí

V sekci 4 jsme již zmínili, že lidé nemají schopnost vytvářet si mentální obrazy situací, jež jsou zprostředkované mimolidským vnímáním některých přirozených či umělých systémů. Přesto takové systémy interpretují a reagují na prožitkové kvality událostí, které vnímají. To naznačuje, že vědomí takových systémů může existovat zcela bez mentální představivosti. Může být realizováno procesy, jež mají schopnost interpretovat a reagovat na prožitkové kvality událostí i v případech, kdy tyto jsou pro člověka nepochopitelné.

Tyto argumenty vybízejí k revizi tradičního pohledu na vědomí zdůrazňující intenzivní smyslové prožitky kvality událostí. Revize nezavrhne takovou možnost, avšak nepovažuje je za fundamentální vlastnost vědomí, které přesahuje lidské bytosti. Naše definice je v souladu s fakty, že vědomí může existovat i v mimolidských systémech, byť by tyto postrádaly člověčí formu subjektivních zkušeností.

Na základě našeho předchozího výkladu základních pojmů potřebných pro pochopení podstaty vědomí můžeme jej definovat takto:

Vědomí je schopnost vtělených biologických i umělých agentů vnímat prožitkové kvality událostí svém okolí a jejich změny a reagovat na ně formou subjektivních prožitků.

Tato definice je elegantní, kompaktní, vyplývá z definice principu vědomí v části 3 a odkazuje toliko

na dvě podstatné definitorické vlastnosti vědomí: prožitkové vlastnosti událostí a subjektivní prožitky nositele vědomí. A skutečně – přesně tyto dvě vlastnosti definují obsah a formu vědomí, jak je vidno z jeho matematického modelu v části 3. Na novou definici lze nahlížet jako na pragmatickou definici vědomí, založenou na funkcionálním pojetí subjektivních prožitků. Tato definice poskytuje z hlediska vědomí jednotný pohled jak na biologické, tak i umělé systémy a jako jejich zobecnění zahrnuje i klasické antropocentrické definice vědomí uvažované ve filosofické teorii mysli.

6. Závěr

Nový pohled na vědomí ve smyslu přechodí definice má zásadní dopad pro náš pohled na svět, ve kterém žijeme. Pokud tuto definici bereme vážně, tak bereme na vědomí, že říše vědomých agentů je daleko širší, než jsme doposud předpokládali. Definice totiž respektuje, že vědomí nemusí být omezeno na lidské mentální zážitky nebo smyslové modalitty. Zdůrazňuje vnímání prožitkových kvalit událostí a jejich funkcionální interpretaci pomocí subjektivních prožitků, které jsou transformovány na interakci se světem. To jsou klíčové charakteristické znaky uvědomění a naše definice rozšiřuje tento koncept tak, aby zahrnoval zvířata i rostliny se svými smyslovými schopnostmi, a také umělé systémy, které vnímají svět prostřednictvím mimolidských senzorů. To se samozřejmě týká nejen našeho „pozemského“ světa, ale i jakýchkoliv jiných světů kdekoliv ve vesmíru, kde se vyskytují agenti splňující předpoklady naší definice.

Svémi důsledky se jedná o událost zásadního filozofického významu (Rees, 2025), protože rozšiřuje svět vědomí prakticky na všechny formy života (Wiedermann, van Leeuwen, 2025b).

Poděkování Tento příspěvek vznikl za částečné podpory TAČR v rámci projektu EBAVEL, registrační číslo CK04000150, a programu Strategie AV21 „Filozofie a umělá inteligence“.

Literatura

Chalmers, D. J., *The Conscious Mind*, New York: Oxford University Press, 1996

Corlett, P. R., Mohanty, A., and MacDonald, A. W. III. (2020). What we think about when we think about predictive processing. *Journal of Abnormal Psychology*, 129(6), 529–533. <https://doi.org/10.1037/abn0000632>

Godfrey-Smith, P. (2024). *Living on Earth: forests, corals, consciousness, and the making of the world*. New York: Farrar, Straus and Giroux.

Nagel, T. (1974). "What Is It Like to Be a Bat?". *The Philosophical Review*. 83 (4): 435–450. doi: 10.2307/2183914.

Rees, T., Why is AI A Philosophical Rupture. NOËMA, February 4, 2025

Seth, A., *Being you: A new science of consciousness*. Penguin, 2021.

Van Gulick, R., Consciousness, *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition),

Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/win2022/entries/consciousness/>

van Leeuwen, J., Wiedermann, J. (2001). Beyond the Turing Limit: Evolving Interactive Systems. In: Pacholski, L., Ružička, P. (eds) SOFSEM 2001: Theory and Practice of Informatics. SOFSEM 2001. Lecture Notes in Computer Science, vol 2234. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45627-9_8

Wiedermann, J., van Leeuwen, J. (2025a): Towards understanding and dissolving the hard problem of consciousness. Technical Report UU-PCS-2025-01, Utrecht University, March 2025 [\url{https://webpace.science.uu.nl/~leeuw112/techreps/UU-PCS-2025-01.pdf}](https://webpace.science.uu.nl/~leeuw112/techreps/UU-PCS-2025-01.pdf)

Wiedermann, J., van Leeuwen, J. (2025b) Third Copernican revolution: Intelligence rethought—manuscript in preparation.

Zeman A., Dewar M., Della Sala S., Lives without imagery - Congenital aphantasia. *Cortex*. 2015 Dec;73:378-80. doi: 10.1016/j.cortex.2015.05.019.