TEORIE VĚDY / THEORY OF SCIENCE / XLV / 2023 / 1
https://doi.org/10.46938/tv.2022.556

////// studie / article //

VON NEUMANN, TURING A GÖDEL: O MYSLI A STROJÍCH

Abstrakt: *Stať pojednává o některých nedostatečně prozkoumaných vazbách mezi pojmovými systémy logiky u Kurta Gödela, teorie automatů u Alana Turinga a teorie sebe-reprodukujících se automatů u Johna von Neumanna. Stranou jsou ponechány tradiční polemiky (především opozice Gödela a Turinga v pojetí mysli) a pozornost je soustředěna na podobnosti mezi všemi třemi autory. V jednotlivých kapitolách se text věnuje postupně: podobě odlišení syntaxe a sémantiky formálního systému u Gödela, Turinga a von Neumanna; von Neumannově variantě Gödelova důkazu a von Neumannově a Gödelově pojetí Turingova stroje; a konečně stejnému základu pojetí vztahu mezi myslí a strojem u všech tří autorů.*

Klíčová slova: *syntax a sémantika formálního systému; Gödelovy věty; Turingův stroj; von Neumannova sonda*

Von Neumann, Turing, and Gödel: On Mind and Machines

Abstract: *The paper discusses some of the poorly explored links between the conceptual systems of logic in Kurt Gödel, the theory of automata in Alan Turing, and the theory of self-reproducing automata in John von Neumann. Traditional controversies are left aside (especially the opposition of Gödel and Turing in the view of mind) and attention is focused on the similarities between all three authors. In individual chapters, the text deals with: the form of differentiation of syntax and semantics in formal system in Gödel, Turing and von Neumann; von Neumann's variant of Gödel's theorem and von Neumann's and Gödel's conception of Turing machine; and finally the same basis of the view of the relation between mind and automaton in all three authors.*

Keywords: *syntax and semantics in formal system; Gödel's theorem; Turing machine; von Neumann's probe*

BARBORA JURKOVÁ
Katedra obecné lingvistiky
Filozofická fakulta UP
Křížkovského 14, 779 00 Olomouc
email / barbora.jurkova01@upol.cz
iD 0000-0002-7816-3710

LUKÁŠ H. ZÁMEČNÍK
Katedra obecné lingvistiky
Filozofická fakulta UP
Křížkovského 14, 779 00 Olomouc
email / lukas.zamecnik@upol.cz
iD 0000-0002-0965-4583

Příspěvek vznikl za podpory MŠMT ČR udělené UP v Olomouci (IGA_FF_2021_048 – „Možnosti formalizace sémantických a pragmatických aspektů zdrojového materiálu ve vstupní fázi kvantitativně-lingvistického výzkumu“).

> I don't want to give a definition of thinking, but if I had to I should probably be unable to say anything more about it than that it was a sort of buzzing that went on inside my head. But I don't really see that we need to agree on a definition at all. The important thing is to try to draw a line between the properties of a brain, or of a man, that we want to discuss, and those that we don't. (Turing, BBC, 1952)[1]

1. Úvod

Souvislost mezi objevy Kurta Gödela v oblasti metamatematiky a Alana Turinga v oblasti teorie automatů byla vícekrát konstatována,[2] ale pokud víme, nebyla podrobena pečlivější filosofické reflexi. Vztah obou objevů k von Neumannově koncepci sebe-reprodukujícího se automatu je filosofy reflektován ještě méně.[3] V publikacích věnujících se informatice[4] se sice nachází konstatování o dvou původních architekturách počítače (Turingově a von Neumannově), ale jen zřídka[5] se setkáme se zjištěním, že von Neumann činí v teorii automatů zřejmým požadavek distinkce syntaxe a sémantiky formálního systému, který přinesl Gödelův objev.[6]

V článku se nesnažíme navázat na linii známých textů, které zkoumají implikace Gödelova objevu ve filosofii matematiky,[7] filosofii mysli a filosofii umělé inteligence.[8] Do určité míry totiž platí, že tento známý přístup zastřel

[1] Přepis tohoto rozhovoru byl převzat z Jack B. Copeland, ed., *The Essential Turing* (Oxford: Clarendon Press, 2004), 494.

[2] Poprvé se o souvislosti zmiňuje Alan Turing, srov. Alan Turing, „On Computable Numbers, with an Application to the Entscheidungsproblem," *Proceedings of the London Mathematical Society* s2-42, no. 1 (1937): 230–65. V současné době je běžné tyto dva logické problémy uvádět společně. Zajímavé zpracování případu, kdy by se univerzální automat musel přímo vypořádat s Gödelovým důkazem lze nalézt v knize Torkel Franzén, *Gödel's Theorem: An Incomplete Guide to its Use and Abuse* (Wellesley: A. K. Peters, 2005), 59–76.

[3] Výjimkou může být Abrir U. Igamberdiev and Joseph E. Brenner, „Mathematics in Biological Reality: The Emergence of Natural Computation in Living Systems," *BioSystems* 204 (2021): 104395.

[4] Martin Davis, *The Universal Computer. The Road from Leibniz to Turing* (New York: W. W. Norton & Company, 2000), 139–75.

[5] Lukáš H. Zámečník and Jaroslav Krbec, „Describing Life: Towards the Conception of Howard Pattee," *Linguistic Frontiers* 2, no. 1 (2019): 1–9.

[6] Názorně představeno v knize David Papineau, *Philosophical Devices* (Oxford: Oxford University Press, 2012), 137–77.

[7] Ucelený pohled na význam Gödelových vět pro filosofii matematiky nalezne čtenář v knize Pavol Zlatoš, *Ani matematika si nemôže byť istá sama sebou: Úvahy o množinách, nekonečne, paradoxoch a Gödelových vetách* (Bratislava: IRIS, 1995).

[8] Debata o Gödelově důkazu, jakožto argumentu proti umělé inteligenci, je velice rozsáhlá. Současné debata se ale odvíjí především od článku John R. Lucas, „Minds, Machines and

některé jiné zajímavé aspekty (nejen) Gödelova objevu a vedl u některých autorů ke skepsi vůči časté argumentaci opřené o Gödelovy věty.[9] Protože úvahy v těchto případech vedou příliš často k závěrům o principiálních nemožnostech, které se vztahují ke schopnostem myslí strojů[10] a kterým lidská mysl, vybavena určitou extra-kvalitou, nějak dokáže uniknout.[11]

Pravděpodobně nepřekonatelné interdisciplinární zhodnocení Gödelova objevu je obsaženo v díle Douglase Hofstadtera. Přesto v něm nacházíme pouze stručné pojednání o Turingově stroji[12] a zcela v něm absentuje odkaz na von Neumanna a jeho úvahy o sebe-reprodukujících se automatech. Tato skutečnost, spolu s významem, který připisuje Hofstadter (respektive Nagel a Newman)[13] postupu Gödelova číslování, nás vedla k myšlence prozkoumat vztah všech tří autorů, u kterých nacházíme odlišování syntaxe a sémantiky formálního systému a s ním i centrální koncept kódu, tj. možnosti reprezentovat tvrzení o struktuře systému objekty systému samotného.

V článku se snažíme ukázat, jak byl objev distinkce syntaxe a sémantiky formálního systému paralelně vyjádřen a do jaké míry vedl každého ze tří uvedených autorů k odlišným pojetím výpočetního stroje a k odlišným konceptualizacím mysli. Tradičně uváděnou odlišnost pojetí mysli u Gödela a u Turinga (a obdobně u von Neumanna) ukazujeme jako sekundární záležitost, která se u všech tří autorů váže na soulad s ohledem na uváděnou distinkci. Právě toto odhalení společného jádra jejich teorií, které je přitom

Gödel," *Philosophy* 36, no. 137 (1961): 112–27, kde je Gödelův důkaz používán k popření mechanistického modelu mysli. Podobně též viz Roger Penrose, *Shadows of the Mind* (Oxford: Oxford University Press, 1994), 64–116.

9 Například v textech Vladimír Havlík, „Kurt Gödel a AI," in *Meze formalizace, analytičnosti a prostoročasu*, eds. Tomáš Čana a Vladimír Havlík (Praha: Filosofia, 2007), 161–77; Filip Tvrdý, *Turingův test: Filozofické aspekty umělé inteligence* (Praha: Togga, 2014).

10 Srov. Kurt Gödel, „Some Basic Theorems on the Foundation of Mathematics and Their Philosophical Implications," in *Kurt Gödel: Unpublished Philosophical Essays*, ed. Francisco A. Rodriguez-Consuegra (Basel: Birkhäuser, 1995), 129–70; dále také Hubert L. Dreyfus, *Alchemy and Artificial Intelligence* (Santa Monica, CA: Rand Corporation, 1965), 46–64; nebo Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (Hoboken: Pearson, 2021), 1032–37.

11 Především viz Hao Wang, *A Logical Journey: From Gödel to Philosophy* (Cambridge, MA: MIT Press, 1996), 183–208; ale také Solomon Feferman, „Are There Absolutely Unsolvable Problems? Gödel's Dichotomy," *Philosophia Mathematica* 14, no. 2 (2006): 134–52.

12 Hofstadter se ve svém díle omezuje na základní modely Turingova stroje s některými běžně přijímanými rozšířeními, srov. Douglas R. Hofstadter, *Gödel, Escher, Bach: existenciální gordická balada: metaforická fuga o mysli a strojích v duchu Lewise Carrolla* (Praha: Argo/Dokořán, 2012), 46, 409–10, 446–51.

13 Ernest Nagel and James Roy Newman, *Gödel's Proof*, ed. Douglas R. Hofstadter (New York: New York University Press, 2002), 68–80.

následováno odlišnými filosofiemi mysli, může vrhnout nové světlo na některé problémy analytické filosofie mysli.[14]

Oporou v našem snažení jsou nám primární zdroje, v prvé řadě vzájemná korespondence,[15] kterou spolu autoři udržovali zejména ve 30. letech 20. století, než se Gödel a von Neumann setkali v Institute for Advanced Study v Princetonu. Spouštěčem korespondence byla Gödelova přednáška v Královci v roce 1930, na které Gödel upozornil na důkaz neúplnosti formálních systémů. Jak uvidíme níže, von Neumann v té době pravděpodobně již pracoval na podobném důkazu, s tím rozdílem, že jej stavěl na intuicionistických základech. Von Neumann dosáhl podobného závěru s ohledem na nedokazatelnost konzistence formálních systémů, rozhodl se ale na toto téma nepublikovat.

Opíráme se také o další texty, o Turingovu dizertaci[16] a o von Neumannovy přednášky,[17] ve kterých se věnoval zejména tématu „*high speed computing machines*", rozvíjející se kybernetice a umělým neuronovým sítím. Přednášky ukazují, jak dobře byl von Neumann s celou problematikou Turingova stroje seznámen, což vrhá nové světlo na některá tradiční pojetí vztahu Turinga a von Neumanna, především v otázce prvenství v designování moderního počítače.

Článek jsme rozčlenili na tři hlavní části věnované postupně: srovnání syntakticko-sémantické distinkce ve formálních systémech u všech tří au-

[14] Máme na mysli standardní argumentaci proti Turingovu testu, opřenou o závěry plynoucí z Gödelova důkazu. Srov. např. Roger Penrose, *The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics* (Oxford: Oxford University Press, 1990), 129–92.

[15] Existuje pětisvazkový soubor Gödelových textů s komentáři od předních odborníků, který zahrnuje i jeho nepublikované stati. Poslední dva svazky jsou věnované Gödelově osobní korespondenci. Výjimečně jsou zde uvedeny obě strany korespondence, což v případě von Neumannovy korespondence neplatí. Srov. Kurt Gödel, *Collected Works, Volume I (1986, Publications 1929–1936) Volume II (1989, Publications 1938–1974), Volume III (1995, Unpublished Essays and Lectures), Volume IV (2003, Selected Correspondence, AG), Volume V (2003, Selected Correspondence, HZ)*, eds. Solomon Feferman et al. (New York: Oxford University Press, 1986–2003). Další důležité výbory textů představuje Rodriguez-Consuegra, *Kurt Gödel*; výbor von Neumannovy korespondence nalezneme v Miklós Rédei, *John von Neumann: Selected Letters* (Providence, RI: American Mathematical Society, 2005).

[16] Turingova disertace z Princetonu z roku 1938: Alan Turing, „Systems of Logic Based on Ordinals," *Proceedings of the London Mathematical Society* s2-45, no. 1 (1939): 161–228.

[17] Přednášky jsou k dispozici v online archivu American Philosophical Society Library, *John Von Neumann – Folder 5*, https://diglib.amphilsoc.org/islandora/object/john-von-neumann-folder-5#page/1/mode/1up. Na tyto materiály upozorňuje především Thomas Haigh and Mark Priestley, „Von Neumann Thought Turing's Universal Machine Was ‚Simple and Neat.': But That Didn't Tell Him How to Design a Computer," *Communications of the ACM* 63, no. 1 (2019): 26–32.

torů (část 1), analýze nových zjištění obsažených v primárních zdrojích (část 2) a představení implikací obou dřívějších částí pro pojetí mysli u Gödela na jedné a Turinga (a von Neumanna) na druhé straně (část 3).

2. O třech způsobech odlišení syntaxe a sémantiky ve formálních systémech

Zohlednění sémantiky formálního systému je v kontextu filosofie matematiky a filosofie mysli spojováno pravidelně s důsledky Gödelova důkazu.[18] Rozpoznatelnost pravdivosti věty (sémantické hledisko) jako odlišné od dokazatelnosti věty (syntaktického hledisko) představovala pokrok v matematické logice a současně omezení pro některé formalistické výklady matematiky.[19] Naší ambicí je ukázat, že tato distinkce je stejně tak přítomna v Turingově odlišování akceptovatelnosti a rozhodnutelnosti množiny řetězců daným automatem (obecně Turingovým strojem) a také ve von Neumannově koncepci sebe-reprodukujícího se automatu (von Neumannovy sondy), v odlišování reprezentovatelnosti a konstruovatelnosti automatu.[20]

Důkazy zde nebudeme reprodukovat, všechny tři důkazy čtenář nalezne v originální podobě,[21] v případě Gödela a Turinga také v jejich podrobném výkladu v sekundární literatuře.[22] Von Neumannův důkaz nebyl dosud ke Gödelovu a Turingovu důkazu vztahován, reflektován byl v kontextu sémiotiky, v souvislosti s hledáním spodní hranice sémiotického systému.[23]

[18] Srov. Matt Carter, *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence* (Edinburgh: Edinburgh University Press, 2007), 52–69; Wang, *Logical Journey*, 247–85; Patrick Allo et al., *Three Views of Logic: Mathematics, Philosophy and Computer Science* (Princeton: Princeton University Press, 2005), 95–121.

[19] Modelově spojované především s dobovým úsilím Russella a Whiteheada: Alfred North Whitehead and Bertrand Russell, *Principia Mathematica. Volume I* (Cambridge: Cambridge University Press, 1963).

[20] Pojmy *reprezentovatelnost* a *konstruovatelnost* zde zavádíme, jejich anglické ekvivalenty *representability* a *constructability* jsou zavedeny a použity ve stati Zámečník and Krbec, *Describing Life*, 4–5.

[21] Turing, *On Computable Numbers*; Kurt Gödel, „Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I," *Monatshefte für Mathematik und Physik. Akademische Verlagsgesellschaft, Leipzig* 38 (1931): 173–98; John von Neumann, *The Theory of Self-Reproducing Automata* (Urbana: University of Illinois Press, 1966).

[22] Barbara Hall Partee, Alice ter Meulen, and Robert E. Wall, *Mathematical Methods in Linguistics* (Dordrecht: Kluwer Academic Publishers, 1990), 520–25; Nagel and Newman, *Gödel's Proof.*

[23] Howard Pattee and Kalevi Kull, „A Biosemiotics Conversation: Between Physics and Semiotics," *Sign System Studies* 37, no. 1/2 (2009): 311–31; Howard Pattee, „The Physics

Soustředíme se přímo na stručnou rekapitulaci jednotlivých důkazů a na jejich srovnání.

Důležitým prvkem v konstrukci Gödelova důkazu je provedení nějaké podoby Gödelova číslování,[24] přiřazení vybraných přirozených čísel (např. po sobě jdoucích prvočísel)[25] k jednotlivým prvkům formálního systému,[26] jehož úplnost (a bezespornost) zkoumáme. Po očíslování jednotlivých prvků kalkulu je možné podle předepsaného algoritmu[27] vypočítávat Gödelova čísla pro jakoukoli větu, která je v kalkulu vyjádřitelná. A dokonce i pro věty, které referují o větách systému samotného – je tak možné provést aritmetizaci metamatematiky, každému metamatematickému tvrzení je možné také přiřadit Gödelovo číslo. Jak je známo, lze pak naleznout také Gödelovo číslo pro větu *G*, která tvrdí, že je *G* nedokazatelná. Tato sémanticky pravdivá věta je nedokazatelná ve formálním systému, který je bezesporný.

Gödel ukázal, že matematická pravda nemusí korespondovat s matematickým důkazem. Možnost konstrukce vět typu *G* vede k případům, kdy pravdivost dané věty přesně koresponduje s neexistencí důkazu této věty. Toto odlišení dokazatelnosti a pravdivosti odpovídá emancipaci sémantiky formálního systému (pravdivost) na syntaxi formálního systému (dokazatelnost). Podmínku konzistence systému je možné naplnit za předpokladu

and Metaphysics of Biosemiotics," *Journal of Biosemiotics* 1 (2005): 281–301; Dan Faltýnek, *Sémiotické primitivy v konstrukci gramatik* (Olomouc: Univerzita Palackého v Olomouci, 2011), 49–88.

[24] Používáme termín „číslování", který by bylo možné nahradit termínem „kódování", typicky používaným v soudobé informatice. Děkujeme anonymnímu recenzentovi za upozornění, že způsob, jakým jsou celá čísla rozkódována na textové řetězce, které simulují, poskytuje intuitivní porozumění Gödelovým větám. Gödelovy věty platí v „dostatečně silných systémech", tj. takových, ve kterých je dokazatelná vlastnost aritmetické operace dělení se zbytkem (tj. operace používaná k rozkódování celých čísel na textové řetězce).

[25] Prvočísla ovšem nejsou nezbytná, viz např. Raymond Smullyan, *Gödel's Incompleteness Theorems* (Oxford: Oxford University Press, 1992), 22–24.

[26] Gödel vychází z vymezení formálního systému ve Whitehead, Russell, *Principia Mathematica*. Obecně se uvádí, že Gödelovy závěry se vztahují na takové formální systémy, ve kterých jsou zavedena celá kladná čísla a lze aplikovat operace sčítání a násobení, srov. Nagel and Newman, *Gödel's Proof*, 68, pozn. 15.

[27] Gödel pro formální systém vytvořil dva slovníky, zahrnující proměnné a konstanty. Jednotlivým položkám ve slovnících jsou přiřazena čísla. Nejprve konstantám, což jsou především základní symboly logiky a následně proměnným, které jsou rozděleny do tří kategorií. Pro každou kategorii existují specifická pravidla, která k jejich proměnným přiřazují prvočísla. Musí se současně vždy jednat o prvočísla větší, než jsou použita v případě konstant. Pro každou matematickou formuli lze takto vypočítat Gödelovo číslo. Srov. Nagel and Newman, *Gödel's Proof*, 68–80.

existence pravdivých nedokazatelných vět. Zatímco dokazatelné věty musí být všechny pravdivé, pokud je zkoumaný systém aritmeticky korektní.[28]

I když v popředí komentářů ke Gödelovu důkazu bývá většinou rozbor implikací neúplnosti formálního systému, my bychom chtěli obrátit pozornost čtenářů především ke Gödelovu číslování. Klíčem k nalezení sémantické úrovně formálního systému je totiž možnost reprezentovat něco – prvky kalkulu formálního systému – něčím jiným – přirozenými čísly. Jestliže existuje příslušné Gödelovo číslo, existuje i příslušná pravdivá matematická věta, kterou toto číslo reprezentuje, ačkoliv k ní nemusí existovat důkaz. Ve formálním systému je tak sémantika (pravdivost věty) autonomní na syntaxi (dokazatelnost věty).

Postup, jaký je využit v případě Gödelova číslování, používá také Turing, když zavádí univerzální automat (univerzální Turingův stroj), tedy automat, který je schopen simulovat výpočet libovolného Turingova stroje.[29] Důležité je, že jednotlivé Turingovy stroje je možné reprezentovat řetězcem symbolů s konečnou abecedou (typicky s abecedou složenou z 0 a 1) a že Turingových strojů je spočetně nekonečné množství. Vzhledem k tomu, je možné jejich reprezentace uspořádat zvoleným jednoznačným způsobem.[30]

Tak jako existují Gödelova čísla, která reprezentují pravdivé matematické věty, které nejsou dokazatelné, také v tomto případě budou existovat množiny řetězců (tvořené z 1 a 0), které budou akceptovatelné Turingovým strojem, ale které nebudou Turingovým strojem rozhodnutelné. S určitou mírou zjednodušení lze konstatovat, že akceptovatelnost je pro Turingův stroj definována tak, že se po přečtení každého řetězce z dané množiny řetězců Turingův stroj zastaví, zatímco rozhodnutelnost požaduje navíc,

[28] Z neúplnosti systému vyplývá, že existují jeho bezesporná aritmeticky nekorektní rozšíření (např. systém, který obsahuje negaci tvrzení *G* jako nový axiom). Děkujeme anonymnímu recenzentovi za toto doplnění.

[29] Turingův stroj je nejobecnějším typem automatu, tj. abstraktního výpočetního prostředku (jednoduššími typy jsou především zásobníkový a konečný automat), který může číst, mazat a přepisovat jednotlivé symboly řetězce (v definované konečné abecedě), a to ve směru zprava doleva i zleva doprava. Řídí se při tom množinou pravidel, která uvádějí, jak se při čtení mění jednotlivý stav automatu (z konečné množiny stavů). Dále je pro každý automat definován počáteční stav a množina finálních stavů. V kontextu soudobé informatiky můžeme prohlásit, že univerzální Turingův stroj je plnohodnotným programovacím jazykem.

[30] Konkrétní postup implementace vynecháváme, viz např. Partee et al., *Mathematical Methods,* 520–21.

aby se Turingův stroj po přečtení každého řetězce z dané množiny řetězců zastavil v některém z finálních stavů.[31]

Opět zde nebudeme reprodukovat důkaz, který představil Turing (1937) a který je názorně reprodukován na řadě míst.[32] Jen připomeneme, že řešení tzv. úlohy zastavení (*halting problem*) Turingova stroje vede skrze důkaz sporem ke zjištění, že rozhodnutelnost určité akceptovatelné množiny řetězců by implikovala kontradikci.[33] Z Turingovy ideje univerzálního stroje vyplývá distinkce akceptovatelnosti a rozhodnutelnosti dané množiny řetězců. Stejně jako v případě Gödelovy dvojice pravdivost a dokazatelnost se v Turingově případě ukazuje, že akceptovatelnost nemusí korespondovat s rozhodnutelností, ačkoliv všechny rozhodnutelné množiny řetězců jsou současně akceptovatelnými množinami řetězců. Bude tedy existovat nekonečné množství akceptovatelných, ale nerozhodnutelných množin řetězců.

Neformálně vyjádřeno, se snahou přiblížit čtenáři podobnost mezi prací Gödela a Turinga, můžeme prohlásit, že jsme-li schopni ve formálním systému syntakticky odvodit teorém (provést důkaz) u Gödela, pak tomu u Turinga odpovídá schopnost identifikovat konečnou množinu finálních stavů pro daný Turingův stroj. Neexistence důkazu je pak analogická neexistenci postupu, který by umožňoval automaticky vybírat množiny finálních stavů pro libovolný Turingův stroj. Opět nám ve formálním systému sémantická úroveň akceptovatelnosti množiny řetězců získává autonomní postavení nad syntaktickou úrovní rozhodnutelnosti množiny řetězců.

Jako již výše, chceme se pozastavit opět nad tím, že klíčem k Turingovu důkazu je možnost reprezentovat něco – libovolný Turingův stroj – něčím jiným – řetězcem symbolů (0 a 1). Je pak dokonce možné (a to je jádro celého důkazu) předložit na pásce Turingovu stroji jeho vlastní zakódování (v analogii s větou *G*, která se vyjadřuje sama k sobě, ke své nedokazatelnosti), tím je Turingovu stroji předložena akceptovatelná množina řetězců, kterou tento stroj nemůže rozhodnout.[34] Opět se tedy jedná o překonání paradoxu[35]

[31] Čtenáři lze pro komplexnější porozumění doporučit opět uváděnou Partee et al., *Mathematical Methods*, 505–20.

[32] Např. ibid., 522–23.

[33] Obdobně je tomu v případě Gödelovy věty *G*, která tvrdí, že je *G* nedokazatelná. Pokud bychom ji prohlásili za nepravdivou, bylo by možné ve formálním systému dokázat nepravdivé tvrzení.

[34] Leda bychom připustili, že se stroj může mýlit – čili připustili kontradikci (viz níže). Tuto kontradikci Turing připouští i v rámci své nejznámější eseje Alan Turing, „Computing Machinery and Intelligence," *Mind* 59 (1950): 444–45.

[35] V případě Gödelova důkazu se jednalo o Richardův paradox, viz např. Nagel and Newman, *Gödel's Proof*, 60–67.

díky tomu, že jsme schopní diferencovat dvojí sadu – reprezentované a reprezentující.[36]

Konečně v případě von Neumannovy sondy (sebe-reprodukujícího se automatu) nacházíme analogicky poslední distinkci syntaxe a sémantiky formálního systému v odlišování její konstruovatelnosti a reprezentovatelnosti. Na rozdíl od Gödela a Turinga je tato distinkce formulována nepřímo (oba termíny zavádíme, viz výše) a navíc je součástí posmrtně vydaných textů.[37] A domníváme se, že nebyla tak patrná, protože problém autoreference převrací.

U Gödela a Turinga se ukázalo, že autoreference je ve formálním systému „neškodná" ve chvíli, kdy odlišujeme dvojí sadu (v Tarského pojetí jazyk a metajazyk)[38] reprezentujícího a reprezentovaného a získáme tak distinkci syntaxe a sémantiky. Von Neumann postupuje obráceně, ptá se, zda je možné, aby automat sám vytvořil svou vlastní kopii, která bude schopná vytvářet dále své kopie. Autoreference se nám mění v sebereplikaci, ke které jsou hledány nutné a postačující podmínky.

Jednou z těchto nutných podmínek je existence kódu – automat musí disponovat zápisem, podle kterého má provést konstrukci své kopie. Jinak řečeno z požadavku sebereplikace vychází jako nutná podmínka existence dvojí sady – reprezentace něčeho – částí automatu a jejich vzájemného uspořádání – něčím jiným – kódem. Tam kde Gödelův a Turingův důkaz začínaly konstatováním dvojí sady, která vedla k vyřešení problému autoreference, tam požadavek sebereplikace vyžaduje existenci dvojí sady.

Proto diferencujeme reprezentovatelnost a konstruovatelnost automatu. Ptáme-li se, zda může reprezentovatelný automat zkonstruovat svou kopii, pak odpověď zní, že ano, za předpokladu, že disponuje svou reprezentací

[36] Děkujeme anonymnímu recenzentovi, který nás upozornil na význam Cantorovy diagonální metody pro konstrukci Gödelova i Turingova důkazu. Viz také Haim Gaifman, „Naming and Diagonalization, from Cantor to Gödel to Kleene," *Logic Journal of the IGPL* 15, no. 5 (2006): 709–28.

[37] Von Neumann stihl dokončit pouze první část knihy a několik kapitol druhé části. První část je částečně založená na přednáškách, které zmiňujeme výše. Von Neumannův kolega Arthur W. Burks pak následně dokončil a okomentoval druhou část knihy, která se věnuje především technickým vlastnostem sondy (funkcím jednotlivých orgánů) a formálnímu popisu sondy. Von Neumann měl tuto část rozpracovanou, ale už nestihnul rozvinout některé detaily (např. problém evoluce sondy), pravděpodobně se chtěl opřít o publikovanou přednášku, srov. John von Neumann, *Collected Works: Volume 5: Design of Computers, Theory of Automata and Numerical Analysis*, ed. Abraham H. Taub (Oxford: Pergamon Press, 1963).

[38] Alfred Tarski, „The Semantic Conception of Truth: And the Foundations of Semantics," *Philosophy and Phenomenological Research* 4, no. 3 (1944): 341–76.

(zápisem, kódem). Vše, co je konstruovatelné, je tedy reprezentovatelné. Pokud ale reprezentovatelný automat nedisponuje svou vlastní reprezentací, pak je nekonstruovatelný (není schopný vytvořit svou vlastní kopii). Reprezentovatelnost automatu je tedy analogií Turingovy akceptovatelnosti množiny řetězců a Gödelovy pravdivosti věty, zatímco konstruovatelnost automatu je analogií Turingovy rozhodnutelnosti množiny řetězců a Gödelovy dokazatelnosti věty.

Ve všech třech případech zde vidíme důležitou stopu nezbytnosti distinkce syntaxe a sémantiky formálního systému. Poukaz na ni je důležitý, protože je relevantní pro všechny vědní oblasti, pro které je klíčový koncept kódu. A tato distinkce není závazná jen pro biologii a kognitivní vědy, ale také pro sémiotiku. Právě v kontextu populárnosti různých „mezioborových" výpůjček, které transformují některé problémy přírodních věd do podoby přístupné humanitnímu bádání, je cíl naší práce obzvlášť důležitý. Je totiž například důležité odlišit syntakticko-sémantickou distinkci, která je pro formální systémy závazná (a respektovaná např. kódovou biologií),[39] od metafor, které vycházejí z peirceovské sémiotiky opřené o koncept interpretace (typické pro biosémiotiku).[40]

3. Von Neumannův „Gödel" a Gödelův a von Neumannův „Turing"

Cílem této kapitoly je doplnit poznatky z předchozí kapitoly prostřednictvím analýzy: korespondence mezi von Neumannem a Gödelem (ve 30. letech 20. století), von Neumannových přednášek (pravděpodobně z roku 1945)[41] a rozhovorů, které vedl s Gödelem Hao Wang.[42] Centrální přitom bude postava von Neumanna, jak v souvislosti s jeho výkladem Gödelova důkazu, tak s ohledem na jeho pojetí Turingova stroje.

Korespondenci zahájil von Neumann po vyslechnutí Gödelovy přednášky v Královci (1930) a korespondence končí, když se z obou stávají kolegové na Princetonu (1939), doplňuje ji pouze poslední dopis adresovaný von

[39] Srov. např. Marcello Barbieri, *Code Biology: A New Science of Life* (Berlin: Springer, 2015); Marcello Barbieri, *The Semantic Theory of Evolution* (London: Harwood Academic Publishers, 1985).

[40] Srov. např. Claus Emmmeche and Kalevi Kull, eds., *Towards a Semiotic Biology: Life is the Action of Signs* (London: Imperial College Press, 2011).

[41] Jednotlivé přednášky nejsou datované, ani v rámci složky, ve které je našli autoři článku Haigh and Priestley, *Von Neumann*.

[42] Wang, *Logical Journey*.

Neumannovi v roce 1956.[43] V dopisech se von Neumann snažil přesvědčit Gödela o relevanci vlastního návrhu formulace „věty o neúplnosti". Gödel se ovšem přesvědčit nedal, a tak na to von Neumann do určité míry rezignoval (v dopise z ledna 1932). Níže představíme skrze výňatky z von Neumannových dopisů Gödelovi strukturu von Neumannovy argumentace. V posledním Gödelově dopise, je pozornost upřena na reflexi Turingova stroje, právě v tomto dopise můžeme úvahy všech tří autorů vidět nejvíce propojené.

Von Neumannovy přednášky byly věnovány problematice Turingova stroje ve vztahu k rozvíjející se teorii informace a kybernetice (a umělé inteligenci). Opomíjeného von Neumanna zde proto představíme jako zajímavý spojovník obou častěji uváděných autorů, jako iniciátora tohoto propojení. A jak uvidíme v následující kapitole, také jako někoho, kdo nám umožňuje překlenout zdánlivě nepřekonatelnou odlišnost Turingova a Gödelova pojetí mysli.[44]

Von Neumann zahájil korespondenci s Gödelem 20. 11. 1930 ve snaze přesvědčit jej o možnosti odlišného postupu při konstrukci důkazu věty o nedokazatelnosti konzistence určité axiomatické teorie v té teorii samé, který představil Gödel v Královci:

> Podařilo se mi ukázat, že konzistence matematiky je nedokazatelná. Přesněji se jedná o toto: Ve formálním systému, který obsahuje aritmetiku, lze podle Vašich úvah vyjádřit, že formule 1 = 2 nemůže být koncovým vzorcem důkazu vycházejícího z axiomů tohoto systému – ve skutečnosti je tento vzorec totiž formulí uvažovaného formálního systému. Nazvěme ji *W*.
>
> V kontradiktorickém systému je dokazatelná každá formule, tedy i *W*. Je-li konzistence [systému] stanovena intuicionisticky, pak je možné „překladem" obsahových intuicionistických úvah (*contentual intuitionistic considerations*) do formálního [systému] dokázat i *W*. (Na základě Vašeho výsledku by se snad dalo o takové „přeložitelnosti" pochybovat. Domnívám se však, že v daném případě ji lze získat, a velmi rád bych znal Váš názor na tuto věc). S nedokazatelným *W* je tedy systém konzistentní, ale konzistence je nedokazatelná. To jsem

[43] Ve chvíli, kdy byl již von Neumann smrtelně nemocný, na dopis již Gödelovi neodpověděl. Existují dohady, že se jednalo o rakovinu slinivky, způsobenou ozářením z dob, kdy von Neumann pracoval na Projektu Manhattan. Norman Macrae, *John von Neumann: The Scientific Genius Who Pioneered the Modern Computer, Game Theory, Nuclear Deterrence, and Much More* (Lexington, MA: Plunkett Lake Press, 2019).

[44] Na rozdíl od Gödela a Turinga je von Neumannova práce do značné míry oproštěna od filosofických úvah a zaměřuje se především na technickou stránku zkoumaných problémů.

> nyní dokázal: *W* je v konzistentních systémech vždy nedokazatelné, tj. domnělý účinný důkaz *W* by se jistě mohl změnit v kontradikci.[45]

Von Neumann se nesnažil soutěžit o prvenství v nalezení důkazu neúplnosti formálních systémů, jak můžeme vidět z toho, že již v následujícím dopise (29. listopad 1930) explicitně upouští od publikace své verze důkazu: „Vzhledem k tomu, že jste již stanovil větu o nedokazatelnosti konzistence jako přirozené pokračování a prohloubení svých dřívějších výsledků, je jasné, že na toto téma nebudu publikovat."[46] Nicméně si stále stojí za svébytností vlastní formulace, když píše:

> Na základě Vašich dopisů si myslím, že mohu zrekonstruovat Váš myšlenkový postup, a proto vám mohu říci, že jsem použil poněkud jinou metodu. Vy jste dokázal $W \rightarrow A$, já nezávisle na tom prokázal nedokazatelnost *W*, a to vlastně odlišným argumentem, který rovněž napodobuje antinomie.[47]

Následně von Neumann anticipuje závěr, ke kterému Gödel dochází, a sice:

> Myslím, že Váš výsledek negativně odpověděl na základní otázku: neexistuje žádné rigorózní zdůvodnění klasické matematiky. Jaký smysl tak přisoudit naší naději, podle níž je de facto konzistentní, nevím – ale podle mého názoru to nic nemění na celkovém faktu.[48]

Gödel von Neumannovi zaslal své důkazy,[49] které von Neumann kvituje a sám navrhuje další metodu, která umožňuje: „konečné rozhodnutí v otázce efektivní dokazatelnosti (*effective provability question*) týkající se propozic, které jsou tvořeny pouze pomocí pojmů (*solely by means of the concepts*)[50] ‚ne', ‚nebo' (tedy i ‚a', ‚vyplývá' atd.), [a] ‚dokazatelné'."[51] S čím ovšem von Neumann zásadně nesouhlasí, je Gödelovo pojetí intuicionismu:

[45] Originál se nachází v Gödel Archive, Manuscript Division, Department of Rare Books and Special Collection, Princeton University. Odkazujeme ke Gödel, *Volume V*, 336–39, konkrétně dopis von Neumanna z 20. listopadu 1930, Berlín.
[46] Ibid., 339, konkrétně dopis von Neumanna z 29. listopadu 1930, Berlín.
[47] Ibid., 339.
[48] Ibid., 339–341
[49] Ibid., 341.
[50] Anonymní recenzent nás upozornil, že právě zde von Neumann zřejmě anticipuje pozdější algebraizaci či modalizaci důkazu druhé Gödelovy věty, kterou se logikové začali zabývat až v 70. letech 20. století, což by zřejmě stálo za další zkoumání.
[51] Ibid., 341. Konkrétně dopis von Neumanna z 12. ledna 1931, Berlín.

Je zřejmé, že nemohu dokázat, že každá intuicionisticky správná konstrukce aritmetiky je formalizovatelná v *A* nebo *M* nebo dokonce v *Z*[52] – protože intuicionismus je nedefinovaný a nedefinovatelný. Není tomu ale tak, že není známa ani jedna konstrukce uvedeného druhu, kterou by nebylo možné formalizovat v *A*, a že žádný žijící logik není v pozici, kdy by takovou [konstrukci] mohl jmenovat? Nebo se mýlím a znáte účinnou intuicionistickou aritmetickou konstrukci, jejíž formalizace v *A* působí potíže? Pokud by tomu tak k mému největšímu překvapení mělo být, pak by taková formalizace měla zajisté fungovat v *M* nebo *Z*!

Byl bych Vám velmi vděčný, kdybyste mi sdělil, zda existenci takových příkladů skutečně předpokládáte, nebo zda dokonce nějaké znáte?[53]

Zde vzájemná výměna mezi oběma graduje, z dopisů je patrné, že byli oba autoři navzájem fascinováni svými výsledky. V průběhu 30. let se oba několikrát setkali osobně a kromě uvedené debaty spolupracovali na jiných tématech.[54] V dopisech stále vidíme, že von Neumann hájí svou verzi intuicionismu a opakovaně podotýká, že disponuje kratším důkazem nedokazatelnosti konzistence.[55] Důkaz samotný předkládá a my jej nabízíme čtenáři v Apendixu. Debata nekončí von Neumannovou rezignací, spíše se začne soustředit na jiná témata, více související s technickou stránkou umělé inteligence.

Von Neumannova konstrukce alternativního důkazu je zajímavá také v kontextu práce, kterou odvedl při budování Hilbertova programu. Hilbert ještě na konci 20. let věřil, že díky práci Ackermanna a von Neumanna je možné naleznout důkaz konzistence.[56] Von Neumannova znalost problema-

[52] „označme aritmetický axiomatický systém, který obsahuje číselné proměnné, ale neobsahuje funkce ani množinové proměnné a pro číselné proměnné volně používá kvantifikátory ((*x*), (*Ex*)). Jsou-li k dispozici i funkční proměnné prvního řádu (například funkce jen jedné proměnné) spolu s jejich kvantifikátory ((*f*), (*Ef*)), pak tento systém můžeme nazvat *M*. A konečně, můj systém množinově teoretických axiomů se může nazývat *Z*." Ibid., 341.

[53] Ibid., 343

[54] Vyplývá to z jejich vzájemné korespondence, kde jsou však tato témata pouze nastíněna. Ibid., 341.

[55] Ibid., 343–45.

[56] „Ve svém projevu ‚Problémy základu matematiky' na Mezinárodním kongresu matematiků v Bologni v roce 1928 (1929) Hilbert optimisticky prohlásil, že práce Ackermanna a von Neumanna stanovila konzistenci teorie čísel a že důkaz pro analýzu již provedl Ackermann, do té míry, že jediný zbývající úkol spočívá v důkazu elementární věty o konečnosti (*elementary finiteness theorem*), která je čistě aritmetická.'" Richard Zach, „Hilbert's Program," in *Stanford Encyclopedia of Philosophy*, navštíveno 4. března 2022, https://plato.stanford.edu/archives/fall2019/entries/hilbert-program.

tiky také vysvětluje, že si byl jako jeden z mála schopen všimnout Gödelova vystoupení v Královci. A to, že docenil dosah jeho příspěvku, svědčí o tom, že již svou verzi důkazu sám promýšlel.

Zatímco ve 30. letech sledujeme von Neumannův zájem o teoretické problémy základů matematiky, ve 40. a 50. letech se von Neumannova pozornost přesouvá k teorii automatů, opřené o koncept Turingova stroje, se snahou o její technickou aplikaci.[57] Ta je také hlavním obsahem již připomínané knihy (von Neumann, *The Theory of Self-Reproducing Automata*), která je z velké části shrnutím práce, kterou provedl během poslední dekády svého života.

Cenné jsou přednášky, které von Neumann pronesl ve 40. letech,[58] v nichž se věnuje uplatnění univerzálního automatu v kybernetice a v programu umělého života včetně umělých neuronových sítí.[59] Dalším důležitým materiálem je přednáška „The General and Logical Theory of Automata", kterou von Neumann pronesl v roce 1948 v Pasadeně (viz pozn. 36 výše). Von Neumann se opírá o Turingův koncept univerzálního automatu:[60]

> Tento automat, který je sestrojen tak, aby přečetl popis a napodobil popisovaný objekt, je pak univerzálním automatem v Turingově smyslu. K tomu, aby duplikoval jakoukoli operaci, kterou může provést jakýkoli jiný automat, mu stačí poskytnout popis daného automatu a navíc instrukce, které by tento přístroj pro zamýšlenou operaci potřeboval.[61]

Hlavním von Neumannovým cílem přitom je navrhnout sebe-reprodukující se automat.[62] A pro tento cíl je podle von Neumanna Turingovo pojetí příliš úzké:

[57] Von Neumann uplatňuje Turingovu koncepci univerzálního automatu také v rámci designu sebe-reprodukujícího se automatu (sondy). Nevyjadřuje se definitivně, zda lze fyzickou sondu sestrojit, ve většině případů o ní hovoří čistě hypoteticky, jako o mechanistickém modelu umělého života.

[58] Dochovaly se v archivu v písemnostech Hermana H. Goldstina, von Neumannova kolegy z Los Alamos, viz Macrae, *John von Neumann*.

[59] Srov. též John von Neumann, *The Computer and the Brain* (New Haven: Yale University Press, 2012).

[60] „Nyní je možné popsat automat, který […] když je nakrmen funkcemi (*fed the functions*), definujícími ve výše popsaném smyslu specifický automat, bude poté fungovat jako popsaný objekt." Von Neumann, *Collected Works*, 314.

[61] Ibid., 314–15.

[62] Viz „Universal Constructor" in von Neumann, *Theory of Self-Reproducing Automata*, 271–96.

> Jeho automaty jsou čistě výpočetní stroje. Jejich výstupem je kus pásky s nulami a jedničkami. Pro konstrukci, na kterou jsem odkazoval, je zapotřebí automat, jehož výstupem jsou jiné automaty. V zásadě však zde není problém pracovat s tímto širším pojmem a odvodit z něj ekvivalent Turingova výsledku.[63]

Z úryvků je patrné, že von Neumann Turingově práci – jeho velkému myšlenkovému experimentu – rozuměl natolik, že ji dokázal rozvinout do vlastní koncepce sebe-reprodukujícího se automatu, pro kterou usiluje o možnou technickou realizaci. Rozvíjí potencionální využití takového univerzálního stroje v otázkách umělého života, kterým Turing nevěnoval takovou pozornost.

V našem zkoumání jsme zatím viděli dvojici von Neumann a Gödel (korespondence), dvojici von Neumann a Turing (přednášky), nyní nám zbývá prozkoumat vztah Gödela k Turingově koncepci univerzálního automatu. Vyváženě se tomuto problému věnoval Wang,[64] který zachycuje také řadu rozhovorů, které vedl s Gödelem o otázce umělé inteligence. Podle Wanga lze z Gödelovy analýzy Turingova stroje vyvodit tři důležité závěry:

> (1) Turingovy stroje plně vystihují intuitivní pojem mechanických (nebo výpočetních) postupů – a ekvivalentně tak i pojem formálního systému – v přesné definici, čímž odhalují plnou obecnost Gödelových vět o neúplnosti; (2) Turingovy stroje jsou důležitým důkazem Gödelova přesvědčení, že ostré pojmy (sharp concepts) existují a že jsme schopni je jasně vnímat; (3) Turingův argument pro adekvátnost jeho definice obsahuje chybný důkaz silnějšího závěru, že mysli a stroje jsou ekvivalentní.[65]

Wang připomíná kritiku Turinga obsaženou v Gödelově „Some Remarks on the Undecidability Results",[66] z Gödelovy analýzy Turinga nicméně vyplývá, že představu Turingova stroje i ve vztahu k vlastním zjištěním (Gödelovy věty) přijímá. Tvrdíme, že odlišnost obou autorů se zakládá na rozdílných filosofiích mysli, které, jak ukážeme v poslední kapitole, nejsou pro zachování distinkce syntaxe a sémantiky podstatné. Gödel přijímal Turingovu práci vstřícně, vadilo mu „pouze" to, že Turing nebere

[63] Von Neumann, *Collected Works*, 315.
[64] Wang, *Logical Journey*.
[65] Ibid., 194.
[66] Jedná se přitom o drobnou poznámku pod čarou, Kurt Gödel, *Collected Works, Volume II*, 306.

v potaz některé filosofické aspekty problému, o nichž byl Gödel přesvědčen, že jsou nezbytné pro správné vymezení lidské mysli.

Turinga zajímalo, zda odlišné – strojové – myšlení je stále myšlením, zda imitační hra, kterou stroj předvádí, obstojí při konfrontaci s člověkem. Turing netvrdil, že dokáže sestrojit lidskou mysl, pouze to, že některé funkce mysli lze pojmout mechanisticky. Podle Hao Wanga chtěl Gödel Turingovu definici procedury mechanického výpočtu využít k podpoře svého platonismu v matematice[67] (viz i výše bod (2)), navíc ovšem: „začal obhajovat svou tezi, že mysl nebo duch není hmota (a ani jí není ekvivalentní) a je nadřazena počítačům.“[68]

Tradiční polarita obou myslitelů[69] získává mnohem přesnější podobu, když zjišťujeme, že Gödel Turingovu práci nejen velmi dobře znal, ale také ji aplikoval pro vlastní výklady povahy matematiky a úlohy matematiků. Tato mnohem vytříbenější reflexe Turinga vyvstává také při analýze posledního dopisu, který Gödel napsal nemocnému von Neumannovi do nemocnice v roce 1956. Uvažuje v něm o konstrukci Turingova stroje, který by umožňoval rozhodovat speciální úlohu: pro každou formuli F (speciálního kalkulu)[70] a každé přirozené číslo n, zda existuje pro F důkaz o délce n (v počtu symbolů).[71] Z existence takového stroje,[72] podle Gödela vyplývá, že:

> myšlení matematika v případě otázek typu ‚ano‘, nebo ‚ne‘ by bylo možné zcela nahradit stroji, a to navzdory neřešitelnosti problému rozhodnutelnosti (the Entscheidungsproblem). *n* by pouze muselo být zvoleno tak velké, aby v případě, že stroj neposkytne výsledek, nemělo smysl o problému vůbec přemýšlet.[73]

[67] Viz Wang, *Logical Journey*, 138.

[68] Ibid.

[69] Viz Lucas, „Minds, Machines and Gödel“; Penrose, *Emperor's New Mind*.

[70] „Je zřejmé, že je snadné zkonstruovat Turingův stroj, který nám umožní pro každou formuli F omezeného funkcionálního kalkulu 3 a každé přirozené číslo n rozhodnout, zda F má důkaz délky n [délka = počet symbolů].“ Gödel, *Volume V*, 375, konkrétně dopis Gödela z 20. března 1956, Princeton.

[71] Ibid., 375

[72] Dále specifikuje, že: „Nechť $\psi(F, n)$ je počet kroků, které stroj potřebuje k tomu, aby to udělal, a nechť $\phi(n) \geq \max_F \psi(F, n)$. Otázka zní, jak rychle roste $\phi(n)$ pro optimální stroj? Je možné ukázat, že $\phi(n) \geq Kn$. Pokud by skutečně existoval stroj $\phi(n) \sim Kn$ (nebo dokonce jen Kn^2), důsledky jeho existence by byly nesmírně důležité.“ Ibid., 375.

[73] Gödel přidává vlastní „odhad“ vlastností navrhovaného algoritmu: „Zdá se mi, že je docela dobře možné, že $\phi(n)$ roste tak pomalu. Neboť 1.) $\phi(n) \geq Kn$ se zdá být jediným odhadem, který lze získat zobecněním důkazu neřešitelnosti problému rozhodnutelnosti (the *Entscheidungsproblem*); 2.) $\phi(n) \sim Kn$ (nebo $\sim Kn^2$) prostě znamená, že počet kroků ve srovnání s klasickým přístupem pokus a omyl lze snížit z N na $\log N$ (nebo $\log N^2$). O takové výrazné

Gödelovo tvrzení podle nás ukazuje, že v zásadě uznává celou koncepci Turingova stroje a to do té míry, že je schopný na ni převést celou konceptualizaci matematiky a matematického myšlení.[74] Gödel v podstatě může přistoupit i na Turingovo pojetí imitační hry, pouze s dodatkem, že k samotnému mechanismu myšlení (Turingem přesně popsanému) je třeba přiřadit ještě extra kvalitu „nehmotné" mysli.

Proto jsme přesvědčeni o tom, že tradiční spory nejdou k meritu věci, protože sémantika formálního systému jako podmínka fungující mysli je obsažena i u Turinga. Gödel si dostatečnost Turingovy koncepce nepřipouštěl, protože byl směřovaný svou platonskou vizí matematiky. Nicméně „syntakticko-sémantický mechanismus" objevený všemi třemi uváděnými mysliteli je na filosofické koncepci nezávislý. Jinak řečeno, nezpochybnitelná nezbytnost sémantiky při formálním modelování mysli nevyžaduje platonskou oporu. A to ani v interpretaci matematiky. Domníváme se, že nezávislost mysli nevyžaduje transcendentní povahu matematiky, tak jak se to snažili tvrdit autoři, kteří rozpracovávali Turingovu vlastní matematickou námitku proti koncepci Turingova testu.[75]

3. Mysl: stroj, který pozná, že má pravdu

Závěrečnou kapitolu věnujeme přehodnocení tradičního posuzování významu Turingova testu ve filosofii mysli. Cestu k tomu jsme si připravili již v předchozí kapitole, když jsme se snažili ukázat koncept Turingova stroje jako nástroj akceptovaný nejen von Neumannem, ale i Gödelem. Nyní ukážeme, jaké přesnější vymezení mysli nalezneme kromě Turinga také u von Neumanna a Gödela. Hlavním cílem je opět zdůraznění společné syntakticko-sémantické distinkce formálního systému nezbytné pro vymezení základní podmínky pro konceptualizaci mysli. A také to, jak jednotlivá

snížení se rozhodně jedná i v případě jiných finitistických problémů, např. při výpočtu kvadratického zbytku symbolu opakovanou aplikací zákona reciprocity. Bylo by zajímavé vědět, o jaký případ by se jednalo např. při určování, zda je číslo prvočíslo, a jak výrazně lze *obecně* u finitistických kombinatorických problémů snížit počet kroků ve srovnání s přístupem pokus a omyl." Ibid., 375.

[74] Připomeňme, že Turing se nespokojil s konceptem deterministického automatu, původní koncept stroje upravuje a později zavádí tzv. *oracle-machine* (viz výše), přičemž využívá podobného argumentu jako Gödel při úvahách nad problémem Cantorova kontinua. Srov. Turing, *Systems of Logic,* 172–73; Kurt Gödel, „What is Cantor's Continuum Problem?," *The American Mathematical Monthly* 54, no. 9 (1947): 515–25.

[75] Opět Lucas, „Minds, Machines and Gödel"; Penrose, *Emperor's New Mind.*

specifika tří autorů nehrají pro zásadní úpravu této základní podmínky určující roli.

Von Neumann se snaží využít poznatků matematické teorie automatů (ale také statistiky a vznikající teorie informace) k porozumění nervovému systému.[76] Automaty chápe jako reálné stroje, soustředěnost na technickou stránku věci je pro něj typická vždy, ať už uvažuje o umělých neuronech nebo o umělých živých systémech. Zároveň je von Neumann přesvědčen, že matematické zkoumání nervového systému povede k proměně samotné matematiky:

> Domnívám se, že hlubší matematické studium nervové soustavy […] ovlivní naše chápání aspektů samotné matematiky, které se na ní podílejí. Ve skutečnosti to může změnit způsob, jakým nahlížíme na vlastní matematiku a logiku.[77]

Centrálním tématem ve von Neumannových úvahách o umělém i přirozeném nervovém systému je kód, jako nutná podmínka jejich fungování. Von Neumann jej vymezuje následovně: „Systém logických instrukcí, které může automat provádět a které způsobí, že automat provede daný organizovaný úkol […].“[78] Musíme mít na paměti, že von Neumann uvažuje o kódu paralelně pro umělé i přirozené systémy (automaty). Tato paralela jej vede také k závěru, že logické a aritmetické procesy musí existovat separovaně nejen v případě automatů, ale i v případě přirozeného nervového systému:

> Je to dáno tím, že při našem způsobu myšlení a vyjadřování myšlenek je velmi obtížné vyjádřit jakoukoli skutečně komplikovanou situaci, aniž bychom se uchýlili ke vzorcům a číslům. […] Tyto poznámky ukazují, že nervová soustava, nahlížíme-li na ni jako na automat, musí mít rozhodně jak aritmetickou, tak logickou část a že potřeby aritmetiky jsou v ní stejně důležité jako potřeby logiky.[79]

[76] Von Neumann, *Computer and the Brain,* 1–2.
[77] Ibid., 2.
[78] Von Neumann o kódu hovoří jako o principu, který organizuje „logické příkazy“ (*logical orders*) (ibid., 70–71). Přičemž: „Logickými příkazy mám na mysli věci jako nervové impulzy objevující se na příslušných axonech, vlastně cokoli, co přiměje digitální logický systém, jako je nervová soustava, aby fungoval reprodukovatelným, účelným způsobem.“ Ibid., 70–71.
[79] Von Neumann s tím spojuje ještě jednu důležitou otázku: „když na nervový systém nahlížíme jako na výpočetní stroj, jakou lze očekávat přesnost fungování jeho aritmetické části?“ Ibid., 76.

Předešlé von Neumannovy úvahy se stanou pochopitelnými v kontextu kapitoly „The Language of the Brain Not the Language of Mathematics“,[80] ve které von Neumann relativizaci našich konceptualizací matematiky a logiky vyjadřuje explicitně:

> Stejně tak jako jsou jazyky jako řečtina nebo sanskrt historickými fakty, a nikoli absolutní logickou nutností, je jen rozumné předpokládat, že logika a matematika jsou podobně historickými, náhodnými formami vyjadřování.[81]

Jazyk, který používá centrální nervová soustava, je podstatně odlišný, ať už se jedná o logiku nebo aritmetiku, od našeho běžného pojetí jazyka. S ohledem na matematiku, jak jí běžně rozumíme, má podle von Neumanna také „menší hloubku“.[82] V závěru kapitoly pak doplňuje:

> když mluvíme o matematice, můžeme mluvit o sekundárním jazyce, který je postaven na primárním jazyce skutečně používaném centrálním nervovým systémem. Tyto vnější formy naší matematiky tedy nejsou zcela relevantní z hlediska posouzení toho, jaký matematický či logický jazyk je skutečně používaný centrální nervovou soustavou. Výše uvedené poznámky o spolehlivosti a logické a aritmetické hloubce však dokazují, že ať už je tento systém jakýkoli, nemůže se výrazně nelišit od toho, co vědomě a explicitně považujeme za matematiku.[83]

Uvedená kapitola podle nás ukazuje, že von Neumann rozlišoval mezi myslí a nervovou soustavou a nepokoušel se sestrojit mozek nebo inteligenci v lidské podobě. Zamýšlel se nad způsobem, jakým se přenáší a zpracovává informace v nervovém systému, a jak funguje kód a jazyk takového procesu. Toto pojetí se potkává s Turingovým přesvědčením, že když něco myslí jinak než člověk, neznamená to, že to nemyslí.[84] Jak von Neumann tak Turing byli velmi zaujati představou umělého života a umělé inteligence, ale ani jeden z nich netvrdil, že se tím koncept mysli (a života), jak jej spojujeme s člověkem (a přirozenou biosférou) zcela vyčerpal. Oba se snažili poukázat na to, že tradiční představy o životě a inteligenci jsou a priori příliš úzké a současně po technické stránce vágní a je zde prostor pro upřesnění obou konceptů.

[80] Ibid., 81–83.
[81] Ibid., 82.
[82] Ibid., 82–83.
[83] Ibid., 83.
[84] Srov. Turing, *Computing Machinery.*

Otázka zpracování informací umělým nervovým systémem automatu vyžaduje kalkulovat také se zpracováním chyb, které mohou v systému vznikat. Zásadní obtíží pro nakládání s chybami je ovšem klíčový rozdíl mezi přirozenými mozky a umělými nervovými systémy. Von Neumann o tom pojednává v kapitole “Effects of the Lack of a Logical Theory of Automata on the Procedures in Dealing with Errors”. Základní rozdíl je v tom, že organismy dokáží chyby odstraňovat autonomně, zatímco u umělých systémů (navzdory částečné automatizaci) je třeba „zásahu zvenčí“.[85]

Kromě tohoto rozdílu je navíc dopad chyb na chod organismů mírnější, tyto chyby mohou být odstraňovány postupně a neznemožňují další fungování organismu. Na rozdíl od toho chod automatů je chybami ohrožen bezprostředně, tyto chyby musejí být odstraňovány okamžitě, když se objeví, protože hrozí okamžitou disfunkcí automatů.[86] Von Neumann to shrnuje následovně:

> Umělý automat by jistě mohl být navržen tak, aby byl schopen normálně fungovat navzdory omezenému počtu poruch v určitých omezených oblastech. Každá porucha však představuje značné riziko, že v automatu již začal probíhat nějaký obecně degenerativní proces. Je proto nutné okamžitě zasáhnout, protože stroj, který začal selhávat, má jen zřídka tendenci se obnovit a spíše se bude ještě zhoršovat.[87]

Opět zdůrazňujeme, že pro von Neumanna je tato distinkce důležitá z technického hlediska, pro konstrukci sebe-reprodukujících se automatů, nevyvozuje z ní nějaké principiální (či filosofické) omezení umělých systémů.

Turingův text „Computing Machinery and Intelligence“ je dobře známý, přesto jeho reflexí v kontextu úvah Johna von Neumanna a Kurta Gödela můžeme odhalit některé nové skutečnosti. Turing v něm, jak je známo, navrhuje imitační hru (později označenou jako Turingův test), která by sloužila jako kritérium pro připsání inteligence strojům a současně se vyrovnává s řadou námitek, kterými lze takovou koncepci konfrontovat.[88] Podobně jako von Neumann i Turing jasně odlišuje vágní vymezení lidské mysli od přesného specifikování potenciální umělé mysli. To se ukazuje i v Turin-

[85] Von Neumann, *Collected Works*, 305.
[86] Ibid., 305.
[87] Ibid., 305–6.
[88] Podrobnosti čtenář nalezne poutavě představené např. v knize Tvrdý, *Turingův test*.

gově pronikavé odpovědi na námitku, která se opírá o existenci nesmrtelné lidské duše. Turing pokorně píše:

> Tím, že se pokoušíme sestrojit takové stroje, bychom si neměli neuctivě přivlastňovat Jeho moc tvořit duše, stejně jako to neděláme při plození dětí: v obou případech jsme spíše nástroji Jeho vůle, které poskytují příbytky duším, jež stvořil.[89]

Turing zde ukazuje irelevanci argumentů, které vycházejí z náboženských nebo metafyzických premis, je jedno, zda věříme v existenci Boha, nesmrtelného intelektu (jako Gödel) nebo v určitou verzi metafyziky ve vztahu k matematickým entitám, závěry myšlenkového experimentu Turingova testu jimi zůstanou netknuty.[90]

Turing má dva stejně důležité úkoly, konceptualizovat přirozené myšlení[91] skrze výpočetní proces a současně (analogicky k von Neumannovi) navrhnout, jak by mohlo vypadat umělé myšlení. Napětí mezi výkladem Turingova testu jako myšlenkového experimentu (ke konceptualizaci myšlení) a výkladem, který se soustředí na budoucí potenciál strojů, není nutné potlačovat. Na rozdíl od von Neumanna se Turing neomezuje ve svých úvahách o „stroji“ tím, jaké „*digital computers*“ zná:

> neptáme se, zda by si ve hře vedly dobře všechny digitální počítače, ani zda by si vedly dobře počítače, které jsou v současnosti k dispozici, ale spíše zda existují představitelné počítače, které by v této situaci obstály.[92]

Centrální matematickou námitku týkající se omezení digitálních strojů (*discrete-state machines*) Turing spojuje s Gödelovými větami,[93] přitom víme, že Turingova analýza univerzálního automatu jej dovedla k ekvivalentnímu zjištění.[94] A jako už i v Turing (1937) poukazuje na to, že v abs-

[89] Turing, *Computing Machinery*, 443.

[90] Do této skupiny patří pro Turinga také „problém vědomí“. Ibid., 445–47.

[91] Toto je užitečný způsob čtení typický pro analytickou filosofii, vztahuje se k tomu, jak je mysl pojímána ve funkcionalismu, například u Gilberta Rylea, srov. Gilbert Ryle, *The Concept of Mind* (London: Routledge, 2009). V českém prostředí tento „analyticko-filosofický význam“ Turingova testu vykresluje Peregrin, srov. Jaroslav Peregrin, *Kapitoly z analytické filosofie* (Praha: Filosofia, 2014).

[92] Turing, *Computing Machinery*, 436.

[93] Ibid., 444–45.

[94] Svůj ekvivalentní přínos skromně uvádí až na posledním místě: „Nejznámější z těchto výsledků je známý jako Gödelův důkaz (1931) a ukazuje, že v každém dostatečně silném logickém systému lze formulovat tvrzení, která nelze v rámci systému ani dokázat, ani vyvrátit, pokud případně není systém sám nekonzistentní. Existují i další, v některých ohledech po-

traktním matematickém případě (myšlenkovém experimentu) potřebujeme stroj s nekonečnou kapacitou (respektive pamětí).[95]

Ovšem Turinga zajímá také reálný stroj a většinu argumentů staví právě na vlastnostech reálného stroje a (možná překvapivě) na analýze chybovosti stroje (lidského i umělého). Turing odmítá nezdůvodněný předpoklad, že na lidský intelekt se nevztahují omezení identifikovaná pro stroje:

> a to v nás vyvolává určitý pocit nadřazenosti. Je tento pocit pouhou iluzí? Nepochybně je zcela upřímný, ale nemyslím si, že bychom mu měli přikládat příliš velký význam. Sami příliš často odpovídáme na otázky špatně, než abychom byli oprávněni mít z takového důkazu omylnosti některých strojů velkou radost. […] O vítězství nad všemi stroji současně by však nemohla být řeč. […] mohli by existovat lidé chytřejší než kterýkoli daný stroj, ale naopak by zase mohly existovat jiné stroje, které by byly zase chytřejší, a tak dále.[96]

Analýza chybovosti stroje nás přivádí k odlišení myšlenkového experimentu (sloužícího k definici myšlení) a úvahy o reálném stroji, který obstojí v imitační hře. Turing poukazuje na to, že stroj bude v průběhu hraní dělat chyby v řešení aritmetických problémů, aby soupeře zmátl.[97] Současně předpokládá, že argumenty proti schopnosti strojů dělat chyby jsou založeny na nerozlišování dvojího typu chyb: „chyby fungování a chyby závěru“ (*errors of functioning and errors of conclusion*).[98] K jejich vymezení Turing uvádí:

> Chyby fungování jsou způsobeny nějakou mechanickou nebo elektrickou závadou, která způsobuje, že se stroj chová jinak, než byl zkonstruován. […] Chyby závěru mohou vzniknout pouze tehdy, když je výstupním signálům ze stroje přikládán nějaký význam.[99]

V případě obou typů chyb jsou matematici (a filosofové) zatíženi předsudky, respektive klamy. První typ chyby nepřipouští, protože uvažují o stroji jako o abstraktní entitě, která z definice nemůže dělat tento druh chyb.[100] Druhý typ chyby spojuje Turing se schopností nabídnout překva-

dobné výsledky, které mají na svědomí Church (1936), Kleene (1935), Rosser a Turing (1937).“ Ibid., 444–45.

[95] „Předmětný výsledek se týká typu stroje, který je v podstatě digitálním počítačem s nekonečnou kapacitou […].“ Ibid.

[96] Ibid.

[97] Ibid., 448.

[98] Ibid., 449.

[99] Ibid.

[100] Ibid.

pivý závěr, který si opět matematici a filosofové klamně představují jako abstraktní proceduru (která nezabere žádný čas):

> Jedná se o předpoklad, že jakmile je mysli předložen nějaký fakt, všechny důsledky tohoto faktu vyvstanou v mysli současně s ním. Je to za mnoha okolností velmi užitečný předpoklad, ale člověk příliš snadno zapomíná, že je mylný. Přirozeným důsledkem takového počínání je pak předpoklad, že pouhé vyvozování důsledků z dat a obecných principů není žádnou ctností.[101]

Vidíme, že zásadní Turingova argumentace se opírá přesně o to, neuvažovat o stroji jako o abstraktní matematické entitě, ale jako o reálném výpočetním prostředku (jímž je i mozek), který může občas špatně fungovat (první typ chyby), a jehož schopnost poskytovat závěry je založena na tvůrčí schopnosti (stejně jako u mozků) v časové řadě zpracovat data podle řídících principů a poskytnout, třeba mylný, závěr.

Vidíme, že von Neumannovy a Turingovy myšlenky se ubírají v tomto ohledu podobným směrem, oba autoři uvažují o reálných strojích a mozcích a všímají si jejich podobností, či strukturních analogií, ale současně i jejich rozdílů.[102] Zároveň, ve světle těchto zjištění můžeme nově nahlížet Gödelův návrh použití Turingova stroje v jeho posledním dopise von Neumannovi.

Souhrnné propojení řešení matematického problému skrze chybovost a důvěru ve schopnost strojů napodobovat chování člověka vyjadřuje Turing takto:

> Tvrdím, že lze sestrojit stroje, které budou velmi věrně simulovat chování lidské mysli. Občas se budou dopouštět chyb a občas mohou učinit nová a velmi zajímavá tvrzení a celkově bude jejich výstup stát za pozornost ve stejném rozsahu jako výstup lidské mysli.[103]

Problém je v antropomorfické tendenci běžného uvažování o mysli, a to i přes to, že je předložen čistě mechanistický model některých aspektů toho, čemu vágně říkáme mysl a myšlení. Filosofové mají tendenci interpretovat Turinga a von Neumanna tak, jako kdyby ani jeden z nich nedělal rozdíl

[101] Ibid., 451.
[102] Sem patří Turingovo konstatování, ve vazbě na von Neumanna, že: „Nervová soustava rozhodně není diskrétní stavový stroj (*discrete-state machine*)." Ibid., 451.
[103] Copeland, *The Essential Turing*, 472. To souvisí také s Turingovou reflexí problému svobodné vůle: „Pokud dáme stroji k dispozici program, jehož výsledkem je to, že stroj sám udělá něco zajímavého, co jsme nepředpokládali, měl bych spíše tendenci říci, že stroj něco vytvořil, než tvrdit, že jeho chování bylo implicitně obsaženo v programu, a že tedy originalita spočívá výhradně na nás." Ibid., 485.

mezi mozkem (strojem) a myšlením. Přitom oba dva jasně poukazují nejen na rozdíl „člověk a stroj“, ale i „mozek a mysl“. Podobně jako von Neumann i Turing si stojí za tím, že sestrojit mozek jakožto stroj lze a lze do určité míry napodobit jednak biologické procesy neuronů, ale i kognitivní procesy myšlení. Ale nemusí to znamenat, že stroj myslí jako člověk a že vlastně člověk plně rozumí tomu, co mysl je.

V souvislosti s Turingovou analýzou chybovosti stroje je vhodné připomenout jeho snahu (vyjádřenou v dizertaci[104]) rozšířit koncept automatu (Turingova stroje) tak, aby se do určité míry eliminovaly důsledky Gödelových vět[105] a dosáhlo se rovnováhy mezi intuicí a vynalézavostí (*ingenuity*), kterými vládne stroj. V tomto ohledu vytvořil Turing koncept „*oracle-machine*“ (*o-machine*).[106]

Gödelův přístup k mysli působí na první pohled jako přímá opozice k pohledům Turinga a von Neumanna. Podle Gödela z jeho teorému vyplývá, že lidská mysl předčí všechny myslitelné stroje nebo, že je matematika transcendentní (možné je obojí).[107] Nicméně z některých jiných Gödelových vyjádření až taková opozice nevyplývá. Například jeho Gibbsovská přednáška v roce 1951 obsahuje i toto vyjádření:

> Je představitelné (i když daleko za hranicemi současné vědy), že fyziologie mozku pokročí tak daleko, že bude s empirickou jistotou známo, (1) že mozek postačuje k vysvětlení všech mentálních jevů a je strojem v Turingově smyslu;

[104] Turing, *Systems of Logic*.

[105] Eliminovat důsledky Gödelových vět, tj. neformálně řečeno „obnovit úplnost formálních systémů“, standardním (tj. finitistickým) způsobem nelze. Nedokáže to ani „*oracle-machine*“ (viz níže): „Účelem zavedení ordinální logiky bylo co nejvíce se vyhnout důsledkům Gödelova důkazu. Důsledkem tohoto důkazu, vhodně modifikovaného, je nemožnost získat úplnou logickou formuli nebo (nyní zhruba řečeno) úplný logický systém. Byli jsme však schopni z daného systému získat úplnější systém tím, že jsme jako axiomy dosadili formule, které jsme intuitivně pokládali za správné, ale které, jak ukazuje Gödelova věta, jsou v původním systému nedokazatelné; z nich jsme opakováním procesu získali ještě úplnější systém atd. Zjistili jsme, že opakováním procesu jsme získali nový systém pro každou C-K ordinální formuli.“ Ibid., 198.

[106] Ibid., 172–73.

[107] Wang, *Logical Journey*, 3. Pro srovnání Turingova reflexe Gödelových vět: „Známý Gödelův důkaz […] ukazuje, že každý logický systém je v jistém smyslu neúplný, ale zároveň naznačuje prostředky, kterými lze z logického systému *L* získat úplnější systém *L′*.“ Turing, *Systems of Logic*, 161.

> (2) že taková a taková je přesná materiální struktura a fyziologické fungování části mozku, která provádí matematické myšlení.[108]

Ovšem tato empiristická perspektiva se týká, jak vyplývá z dalšího čtení (viz níže), pouze mozku. Mentální jevy jsou spjaty s mozkem, stejně jako část(i) mozku jsou zodpovědné za matematické myšlení. Nicméně mysl je zřejmě něco jiného než mozek (jak uvidíme níže), Gödel zastává určitou variantu dualismu (opřenou o jeho verzi matematického platonismu).[109] Proto, když tvrdí, že: „Lidská mysl není schopna formulovat (nebo zmechanizovat) všechny své matematické intuice."[110] působí to jako tvrzení, že „mysl není mechanismus".

Navzdory tomu, ale Gödel připouští existenci „stroje na matematickou intuici":

> zůstává možné, že může existovat (a dokonce být empiricky objeven) stroj na dokazování tvrzení, který je reálně ekvivalentní matematické intuici, ale nelze dokázat, že tomu tak je, a dokonce ani dokázat, že výsledkem budou pouze správné teorémy konečné teorie čísel (*finitary number theory*).[111]

Vidíme, že ačkoliv je Gödelova motivace jiná, než motivace Turinga a von Neumanna, přesto je tak užitečným nástrojem jako je koncept Turingova stroje uchvácen natolik, aby zkoušel meze jeho uplatnitelnosti. Viděli jsme to i v Gödelově úvaze o Turingově stroji v posledním dopisu von Neumannovi. Ačkoliv se filosoficky autoři rozcházeli, jejich paralelní objev (1931/1937) je stále udržoval v hranicích, které vedly k paralelním výsledkům (např. Turingova *o-machine*).

Vraťme se nyní ke Gödelově shrnujícímu verdiktu:

[108] Rodriguez-Consuegra, *Kurt Gödel*, 150, pozn. 12.

[109] V duchu tohoto platonismu („možnost empiricky evidovat matematické objekty") interpretuje také úlohu jazyka: „Jazyk je užitečný, a dokonce nezbytný pro upevnění našich myšlenek (fixing our ideas). Ale to je čistě praktická záležitost. Naše mysl více inklinuje ke smyslovým objektům, které pomáhají zaměřit naši pozornost na abstraktní předměty. To je jediný význam jazyka. Je směšné [očekávat], že bychom měli mít nějaké primitivní intuice ohledně jazyka, který je jen spojením symbolů s pojmy a jinými entitami." Wang, *Logical Journey*, 180.

[110] „The human mind is incapable of formulating (or mechanizing) all its mathematical intuitions." Ibid., 184.

[111] Ibid., 184–85. Obdobně to vyjadřuje v rozhovoru s Wangem: „Výsledky neúplnosti nevylučují možnost, že existuje počítač schopný dokazovat teorémy (*theorem-proving computer*), který je ve skutečnosti ekvivalentní matematické intuici, ale naznačují, že v takovém – z jiných důvodů velmi nepravděpodobném – případě buď neznáme přesnou specifikaci počítače, nebo nevíme, že funguje správně." Ibid., 186.

> Buď lidská mysl překonává všechny stroje (přesněji řečeno: dokáže rozhodnout více číselně-teoretických otázek než jakýkoli stroj), nebo existují číselně-teoretické otázky nerozhodnutelné pro lidskou mysl. [Není také vyloučeno, že obě alternativy mohou být správné.][112]

Gödel z něj vyvozuje závěr, že mysl nemůže být redukována na mozek.[113] Nicméně stále se zamýšlí nad touto implikací a používá alternativní formulaci, která připouští ještě další výklad: „Vzhledem k mé větě o neúplnosti je pravděpodobné, že mysl není mechanická, jinak řečeno, že *mysl nemůže porozumět svému vlastnímu mechanismu* [zvýrazněno autory].“[114] Mohli bychom rozumět této druhé formulaci jako návratu ke zjištění, že syntax a sémantika formálního systému nejsou vzájemně redukovatelné? Pokud ano, pak ovšem tento neproblematický, ale centrální závěr neimplikuje nějakou podobu dualismu.

Nemoci rozumět vlastnímu mechanismu znamená, že není možné redukovat pravdivost na dokazatelnost. Vlastně bychom tvrzení mohli přeformulovat tak, že: „Není možné mechanicky uchopit fungování mechanismu mysli.“ Gödel se ještě na jiném místě opět připojuje k této druhé (výše zvýrazněné) formulaci: „Když hovoříme o mysli, nemáme na mysli stroj (v nějakém obecném smyslu), ale *stroj, který pozná, že má pravdu* [zvýrazněno autory].“ Mysl je tedy stroj, který dokáže odlišit koncept pravdivosti od konceptu dokazatelnosti.[115]

[112] „Either the human mind surpasses all machines (to be more precise: it can decide more number-theoretical questions than any machine), or else there exist number-theoretical questions undecidable for the human mind. [It is not excluded that both alternatives may be true.]“ Ibid., 185. Alternativně se vyjadřuje Gödel také takto: „Dalším důsledkem mého důkazu je disjunkce dvou tvrzení: (a) matematika je neúplná v tom smyslu, že její evidentní axiomy nemohou být vtěleny do konečného pravidla (*cannot be embodied in a finite rule*), a proto lidská mysl překonává konečné stroje, nebo (b) pro lidskou mysl existují absolutně nerozhodnutelné diofantovské úlohy (*Diophantine problems*). Tento důsledek mé věty je stejně jako ten předchozí výrazný (sharp). Obě alternativy jsou v rozporu s materialistickou filosofií. Alternativa (a) je proti ztotožnění mozku s myslí. Alternativa (b) vyvrací názor, že matematické objekty jsou naším výtvorem.“ Ibid., 187.

[113] Přesněji: „fungování lidské mysli nelze redukovat na fungování mozku, který je podle všeho konečným strojem s konečným počtem částí, totiž neuronů a jejich spojení.“ Ibid., 186.

[114] „My incompleteness theorem makes it likely that mind is not mechanical, or else *mind cannot understand its own mechanism* [zvýrazněno autory].“ Ibid., 186.

[115] „When one speaks of mind one does not mean a machine (in any general sense) but *a machine that recognizes itself as right* [zvýrazněno autory].“ Ibid., 189. Ke konceptualizaci důkazu Gödel uvádí: „jakmile pochopíme obecný pojem důkazu, máme také důkaz mysli o její vlastní konzistenci (*proof by the mind of its own consistency*). Tak, jak to je, můžeme z obecného pojmu důkazu skutečně odvodit rozpory, včetně vlastní aplikace důkazu (*self-application of proof*).

4. Závěr

V naší stati jsme se pokusili přinést několik drobných nových vhledů do vzájemných vztahů mezi konceptualizacemi myslí a strojů u Alana Turinga, Kurta Gödela a Johna von Neumanna. Hlavní oporou nám v tom byla naše hlavní teze, že všichni tři autoři rozkryli paralelně klíčovou distinkci syntaktické a sémantické roviny pojímání formálního systému. Domníváme se, že jsou tato zjištění důležitým přínosem především pro filosofii mysli a umělé inteligence. Tradiční přístupy stavějí pojetí mysli u Turinga a Gödela do nepřekonatelné dichotomie, ulpívají tak na metafyzických distinkcích, které jsou ex definitione (a zde se odvoláváme na naše motto) nepřekonatelné a které zastírají společnou podstatu konceptualizace mysli a stroje u obou (všech tří) autorů.

Hlavní přínos první kapitoly spočívá v tom, že jsme ukázali souvislost mezi ve filosofickém kontextu známou distinkcí dokazatelnosti a pravdivosti matematické věty u Kurta Gödela (kde dokazatelnost implikuje pravdivost ale ne a vice versa), ve stejném kontextu méně známou distinkcí rozhodnutelnosti a akceptovatelnosti množiny řetězců u Alana Turinga (kde rozhodnutelnost implikuje akceptovatelnost, ale ne a vice versa) a konečně celkově nereflektovanou distinkcí mezi konstruovatelností a reprezentovatelností automatu u Johna von Neumanna (kde konstruovatelnost implikuje reprezentovatelnost, ale ne a vice versa).

Von Neumannovi přitom přísluší zásluha za to, že skrze vztah k teorii informace a kybernetice zdůraznil význam kódu, možnosti reprezentovat tvrzení o struktuře systému prvky systému samotného, v němž je rozdíl syntaxe a sémantiky formálního systému explicitně přítomný. Díky němu se tvorba dvojí sady (explicitně Gödelovo číslování), tedy odlišení reprezentujícího a reprezentovaného, která stojí jako důležitý předpoklad důkazu Gödelových vět a také důkazu Turingova teorému, stává hmatatelným předpokladem sebe-replikace automatu. Tam kde vynález dvojí sady pomůže vyřešit paradox autoreference, tam kód umožňuje řešit problém sebe-replikace.

V druhé kapitole se nám podařilo na základě korespondence mezi von Neumannem a Gödelem ukázat, že von Neumann již při Gödelově vystoupení v Královci (1930) disponoval znalostmi a pravděpodobně i variantou důkazu, které Gödel v přednášce nastiňoval. To, že tento důkaz sám nepublikoval, nesvědčí o tom, že by byl chybný (o jeho správnosti opakovaně

Na základě našeho špatného chápání obecného konceptu důkazu můžeme potenciálně dospět k závěru, že důkaz je prostě nekonzistentní. To ukazuje, že s našimi logickými představami není něco v pořádku, což by mělo být zcela evidentní.“ Ibid., 187–88.

přesvědčuje Gödela), ale svědčí to pouze o jeho profesionalitě (dal přednost kolegovi, který byl s přípravou mnohem dál). Náčrt von Neumannova důkazu nabízíme v Apendixu. Domníváme se, že je existence alternativního von Neumannova důkazu hodná pozornosti historiků a filosofů logiky a matematiky a zasloužila by si další bádání.[116]

Ukázali jsme také, že jak von Neumann, tak Gödel přijali přínos koncepce Turingova stroje s velkým vděkem. Von Neumann především z hlediska jeho využití v potenciálních dalších technických aplikacích (umělý život, umělé myšlení) a Gödel z hlediska možnosti celkově s jeho pomocí konceptualizovat práci matematika. Dokonce jsme v posledním dopise von Neumannovi (1956) viděli Gödelovu vlastní úpravu Turingova stroje.

V nejrozsáhlejší třetí kapitole jsme ukázali, že von Neumannova reflexe rozdílů mezi přirozeným a umělým myšlením není kladena na filosofické rovině (jako u Gödela), ale s ohledem na jejich praktické dopady pro konstrukci umělých mozků (a umělého života). Také Turinga zajímá analýza mysli, jako komplexního fenoménu, s ohledem na možnost modelovat nějakou její klíčovou charakteristiku – právě k tomu mu slouží model stroje. Nedomnívá se, že by bylo možné konstruovat umělou mysl zcela v té podobě, jako je ta lidská a pokorně se přiznává, že není schopný popsat mysl v celé její komplexitě.[117]

Ačkoliv Turing ani von Neumann nehovoří explicitně o distinkci syntaxe a sémantiky formálního systému, ukázali jsme, že je pro ně stejně závazná, jako pro Gödela. Proto všechny Gödelovy specifikace povahy mysli, které jej odlišují od Turinga (a von Neumanna) nepřinášejí nic, co by toto základní syntakticko-sémantické odlišení překonávalo. Proto je možné prohlásit, že to, co je implikováno Gödelovými větami, je přirozeně splněno i v Turingových (a von Neumannových) konceptualizacích, proto tradiční námitky[118] nepotřebujeme znovu tematizovat.

Gödel přináší řadu příkladů k odlišení mozku a mysli. Mluví o tom, že mozek (jako *computing machine*) je spojen s duchem (*spirit*), že hlavní úlohu hraje vědomí, že vědomí je celostní, na rozdíl od stroje, odlišuje aktivní

[116] Z hlediska jednoho z anonymních recenzentů se dokonce jedná o nejdůležitější přínos naší stati.

[117] Takto i všechny námitky, které uvádí (Turing, „Computing Machinery and Intelligence“) směřují především k tomu, aby otestovaly funkčnost Turingova modelu mysli.

[118] Viz Lucas, „Minds, Machines and Gödel“; Penrose, *Emperor's New Mind*.

a pasivní intelekt.[119] Všechna tato dualistická vyjádření[120] můžeme připojit jako metafyzické finesy k základnímu odlišení dokazatelnosti a pravdivosti (respektive syntaxe a sémantiky formálního systému, jak ji potkáváme u všech tří autorů), které pokládáme za centrální.

Připomeňme ještě na úplný závěr, že Gödel není učebnicovým racionalistou, neustále se odvolává na možnost empirického prokazování existence mysli (podobně jako matematických objektů, jak již bylo řečeno). Tvrdí, že: „Existuje zde logická možnost, že existence mysli [oddělené od hmoty] je empiricky řešitelná otázka."[121] Nicméně se v tomto případě jedná o návrhy, které (domníváme se) rozvoj neurověd vyvrátil.[122]

Apendix[123]

Lze argumentovat zhruba takto:

1. Nechť a je jakákoliv rekurzivní propozice. Pak tedy lze dokázat

$$a \rightarrow B(a)$$

(kde B znamená dokazatelný, ve vašem smyslu). (Není to přibližně Váš Teorém číslo 5?)

2. Jestliže $b = (Ex)a$, můžeme usuzovat na

$$b \rightarrow B(b)$$

[119] Wang, *Logical Journey*, 189.
[120] V tomto duchu jasně vyznívá i Gödelovo tvrzení: „Hranice vědy: Je možné, že všechny činnosti mysli – například nekonečné, stále se měnící a tak dále – jsou činnostmi mozku? Na tuto otázku může existovat věcná odpověď. Říci ne myšlení jako vlastnosti specifické povahy vyžaduje říci ne také elementárním částicím. Hmota a mysl jsou dvě různé věci." Ibid., 191.
[121] Ibid., 191.
[122] Například toto tvrzení: „Samotná možnost, že k výkonu funkce mysli není dostatek nervových buněk, vnáší do problému mysli a hmoty empirickou složku. Například podle některých psychologů je mysl schopna vybavit si všechny detaily, které kdy zažila. Zdá se pravděpodobné, že k tomu není dostatek nervových buněk, pokud by empirický mechanismus ukládání, jak se zdá pravděpodobné, zdaleka nevyužíval plnou kapacitu paměti. Samozřejmě si lze představit i jiné možnosti empirického vyvrácení, zatímco ve filosofických diskusích o mysli a hmotě je celá otázka obvykle opomíjena." Ibid., 191.
[123] Gödel, *Volume V*, 343–44, konkrétně dopis von Neumanna z 12. ledna 1931, Berlín.

z

$$a \to B(a).$$

3. Vzhledem k tomu, že každé $B(a)$ má tento tvar, dostáváme

$$B(a) \to B(B(a)),$$

pro libovolné a.

4. Zkonstruoval jste a pomocí

$$\bar{a} \sim B(a)$$

Vzhledem k 3. dostáváme

$$\bar{a} \to B(\bar{a}).$$

Nechť O je absurdita a W konzistence; pak dostáváme vzhledem k $\bar{a} \to B(a)$ a $\bar{a} \to B(\bar{a})$, že

$$\bar{a} \sim B(a) \& B(\bar{a}) \sim B(a \& \bar{a}) \sim \mathrm{B}(O) \sim \overline{W}$$

[a tudíž]

$$W \sim a$$

W je tedy stejně nedokazatelné jako a.
(Určitě si dokážete doplnit mezery v mé prezentaci.)

Bibliografie:

Allo, Patrick, Donald W. Loveland, Richard E. Hodel, and S. G. Sterrett. *Three Views of Logic: Mathematics, Philosophy and Computer Science*. Princeton: Princeton University Press, 2005. https://doi.org/10.1007/s11023-015-9375-9.

American Philosophical Society Library. *John Von Neumann – Folder 5*, https://diglib.amphilsoc.org/islandora/object/john-von-neumann-folder-5#page/1/mode/1up.

Barbieri, Marcello. *Code Biology: A New Science of Life*. Berlin: Springer, 2015. https://doi.org/10.1007/s12304-012-9147-3.

Barbieri, Marcello. *The Semantic Theory of Evolution*. London: Harwood Academic Publishers, 1985. https://doi.org/10.1201/9780429290039-10.

Carter, Matt. *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence*. Edinburgh: Edinburgh University Press, 2007. https://doi.org/10.1111/j.1468-2265.2009.00484_50.x.

Copeland, Jack B., ed. *The Essential Turing*. Oxford: Clarendon Press, 2004. https://doi.org/10.1080/01611190508951318.

Davis, Martin. „What Did Gödel Believe and When Did He Believe It?" *Bulletin of Symbolic Logic* 11 (2005): 194–206. https://doi.org/10.2178/bsl/1120231630.

Davis, Martin. *The Universal Computer. The Road from Leibniz to Turing.* New York: W. W. Norton & Company, 2000. https://doi.org/10.2307/2695463.

Dreyfus, Hubert L. *Alchemy and Artificial Intelligence*. Santa Monica, CA: Rand Corporation, 1965.

Emmmeche, Claus, and Kalevi Kull, eds. *Towards a Semiotic Biology: Life is the Action of Signs*. London: Imperial College Press, 2011. https://doi.org/10.1142/p771.

Faltýnek, Dan. *Sémiotické primitivy v konstrukci gramatik*. Olomouc: Univerzita Palackého v Olomouci, 2011.

Feferman, Solomon. „Are There Absolutely Unsolvable Problems? Gödel's Dichotomy." *Philosophia Mathematica* 14, no. 2 (2006): 134–52. https://doi.org/10.1093/philmat/nkj003.

Franzén, Torkel. *Gödel's Theorem: An Incomplete Guide to its Use and Abuse*. Wellesley: A. K. Peters, 2005. https://doi.org/10.1201/b10700.

Gaifman, Haim. „Naming and Diagonalization, from Cantor to Gödel to Kleene." *Logic Journal of the IGPL* 15, no. 5 (2006): 709–28. https://doi.org/10.1093/jigpal/jzl006.

Gödel, Kurt. „Some Basic Theorems on the Foundation of Mathematics and Their Philosophical Implications." In *Kurt Gödel: unpublished philosophical essays*, edited by Francisco A. Rodriguez-Consuegra, 129–70. Basel: Birkhäuser, 1995. https://doi.org/10.1007/978-3-0348-9248-3_6.

Gödel, Kurt. „Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I." *Monatshefte für Mathematik und Physik. Akademische Verlagsgesellschaft, Leipzig* 38 (1931): 173–98. https://doi.org/10.1007/BF01700692.

Gödel, Kurt. *Collected Works, Volume I (1986, Publications 1929–1936) Volume II (1989, Publications 1938–1974), Volume III (1995, Unpublished Essays and Lectures), Volume IV (2003, Selected Correspondence, AG), Volume V (2003, Selected Correspondence, HZ).* Edited by Solomon Feferman, John W. Dawson Jr., Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort. New York: Oxford University Press, 1986–2003.

Gödel, Kurt. „What is Cantor's Continuum Problem?" *The American Mathematical Monthly* 54, no. 9 (1947): 515–25. https://doi.org/10.1080/00029890.1947.11991877.

Haigh, Thomas, and Mark Priestley. „Von Neumann Thought Turing's Universal Machine Was ‚Simple and Neat': But That Didn't Tell Him How to Design a Computer." *Communications of the ACM* 63, no. 1 (2019): 26–32. https://doi.org/10.1145/3372920.

Havlík, Vladimír. „Kurt Gödel a AI." In *Meze formalizace, analytičnosti a prostoročasu*, editovali Tomáš Čana a Vladimír Havlík, 161–77. Praha: Filosofia, 2007.

Hofstadter, Douglas R. *Gödel, Escher, Bach: Existenciální gordická balada: metaforická fuga o mysli a strojích v duchu Lewise Carrolla*. Praha: Argo/Dokořán: 2012.

Igamberdiev, Abrir U., and Joseph E. Brenner. „Mathematics in Biological Reality: The Emergence of Natural Computation in Living Systems." *BioSystems* 204 (2021): 104395. https://doi.org/10.1016/j.biosystems.2021.104395.

Lucas, John R. „Minds, Machines and Gödel." *Philosophy* 36, no. 137 (1961): 112–27. https://doi.org/10.1017/S0031819100057983.

Macrae, Norman. *John von Neumann: The Scientific Genius Who Pioneered the Modern Computer, Game Theory, Nuclear Deterrence, and Much More*. Lexington, MA: Plunkett Lake Press, 2019.

Nagel, Ernest, and Newman, James Roy. *Gödel's Proof*. Edited by Douglas R. Hofstadter. New York: New York University Press, 2002.

Papineau, David. *Philosophical Devices*. Oxford: Oxford University Press, 2012.

Partee, Barbara Hall, Alice ter Meulen, and Robert E. Wall. *Mathematical Methods in Linguistics*. Dordrecht: Kluwer Academic Publishers, 1990.

Pattee, Howard. „The Physics and Metaphysics of Biosemiotics." *Journal of Biosemiotics* 1 (2005): 281–301.

Pattee, Howard, and Kalevi Kull. „A Biosemiotics Conversation: Between Physics and Semiotics." *Sign System Studies* 37, no. 1/2 (2009): 331–21. https://doi.org/10.12697/SSS.2009.37.1-2.12.

Penrose, Roger. *Shadows of the Mind*. Oxford: Oxford University Press, 1994.

Penrose, Roger. *The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics*. Oxford: Oxford University Press, 1990. https://doi.org/10.1093/oso/9780198519737.001.0001.

Peregrin, Jaroslav. *Kapitoly z analytické filosofie*. Praha: Filosofia, 2014.

Rédei, Miklós. *John von Neumann: Selected Letters*. Providence, RI: American Mathematical Society, 2005. https://doi.org/10.1090/hmath/027.

Russell, Stuart, and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Hoboken: Pearson, 2021.

Ryle, Gilbert. *The Concept of Mind*. London: Routledge, 2009. https://doi.org/10.4324/9780203875858.

Smullyan, Raymond. *Gödel's Incompleteness Theorems*. Oxford: Oxford University Press, 1992. https://doi.org/10.1093/oso/9780195046724.001.0001.

Tarski, Alfred. „The Semantic Conception of Truth: And the Foundations of Semantics." *Philosophy and Phenomenological Research* 4, no. 3 (1944): 341–76. https://doi.org/10.2307/2102968.

Turing, Alan. „Computing Machinery and Intelligence." *Mind* 59 (1950): 444–45.

Turing, Alan. „On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society* s2-42, no. 1 (1937): 230–65. https://doi.org/10.1112/plms/s2-42.1.230.

Turing, Alan. „Systems of Logic Based on Ordinals." *Proceedings of the London Mathematical Society* s2-45, no. 1 (1939): 161–228. https://doi.org/10.1112/plms/s2-45.1.161.

Tvrdý, Filip. *Turingův test: Filozofické aspekty umělé inteligence*. Praha: Togga, 2014.

von Neumann, John. *Collected Works: Volume 5: Design of Computers, Theory of Automata and Numerical Analysi*s. Edited by Abraham H. Taub. Oxford: Pergamon Press, 1963.

von Neumann, John. *The Computer and the Brain*. New Haven: Yale University Press, 2012.

von Neumann, John. *The Theory of Self-Reproducing Automata*. Urbana: University of Illinois Press, 1966.

Wang, Hao. *A Logical Journey: From Gödel to Philosophy*. Cambridge, MA: MIT Press, 1996. https://doi.org/10.7551/mitpress/4321.001.0001.

Whitehead, Alfred North, and Bertrand Russell. *Principia Mathematica. Volume I*. Cambridge: Cambridge University Press, 1963.

Zach, Richard. "Hilbert's Program." In *Stanford Encyclopedia of Philosophy*. Stanford University, 1997–. Article revised May 24, 2019. https://plato.stanford.edu/archives/fall2019/entries/hilbert-program.

Zámečník, Lukáš H., and Jaroslav Krbec. „Describing Life: Towards the Conception of Howard Pattee." *Linguistic Frontiers* 2, no. 1 (2019): 1–9. https://doi.org/10.2478/lf-2018-0012.

Zlatoš, Pavol. *Ani matematika si nemôže byť istá sama sebou: Úvahy o množinách, nekonečne, paradoxoch a Gödelových vetách*. Bratislava: IRIS, 1995.