

## Nierozstrzygalność i algorytmiczna niedostępność w naukach społecznych

Motto 1: *There are actually lots of threads that led to computer technology, which come from mathematical logic and from philosophical questions about the limits and the power of mathematics.*

Greg Chaitin<sup>1</sup>

Motto 2: *Computer simulations are extremely useful in the social sciences. It provides a laboratory in which qualitative ideas about social and economic interactions can be tested. This brings a new dimension to the social sciences where 'explanations' abound, but are rarely subject to much experimental testing.*

Richard J. Gaylor (University of Illinois), Louis J. D'Andria (Wolfram Research, Inc.)<sup>2</sup>

**Wprowadzenie.** Komputery stały się narzędziem tak nieodzownym w uprawianiu nauki jak wcześniej były kartka, ołówek, czy instrumenty laboratoryjne.<sup>3</sup> Stąd, pytanie o zasięg ich możliwości znalazło się w centrum uwagi filozofów nauki i metodologów. To, co przede wszystkim trzeba wziąć pod uwagę, to zaskakujące wyniki badań logiczno-informatycznych. Jak to, że nie każdy problem arytmetyczny da się rozwiązać komputerowym algorytmem. Jak to, że nie może być algorytmu, który o dowolnym problemie matematycznym potrafiłby rozstrzygnąć, czy istnieje algorytm jego rozwiązania. A nawet, że gdy istnieją algorytmy, to ich zastosowanie wymaga czasem tak kolosalnych zasobów czasu lub miejsca (pamięci), że oczekiwane od nich rozwiązanie problemu okazuje się w praktyce niedostępne; jest to przypadek nieodostępności algorytmicznej (zwanej też obliczeniową).

Czy to ograniczenie jest jedynie wewnętrzną sprawą czystej matematyki? Czy dotyczy ono również tych nauk empirycznych, a wśród nich społecznych, którym matematyka dostarcza algorytmów do komputerowego modelowania rzeczywistości?

Odpowiedź twierdząca na drugie z tych pytań nie jest dana a priori. Nie byłaby może absurdem filozoficznym wiara w dobrego demona, który tak urządził świat empiryczny, że występują w nim jedynie relacje dające się reprezentować funkcjami obliczalnymi, i do tego takimi, że rozwiązywane przez nie problemy byłyby zawsze dostępne algorytmicznie.

---

<sup>1</sup> "A Century of Controversy over the Foundations of Mathematics" in: C. Calude and G. Paun, *Finite versus Infinite*, Springer-Verlag London 2000, pp. 75-100. Chaitin, matematyk z IBM, stał się klasykiem informatyki i filozofii nauki, dzięki oryginalnym i zyskującym coraz szerszy wpływ ideom i wynikom na temat zasięgu i granic metod algorytmicznych. Jest w tym względzie wybitnym kontynuatorem dzieła Gödla i Turinga. Najsłynniejszym jego odkryciem jest liczba nieobliczalna Omega, definiowana jako prawdopodobieństwo, że program komputerowy zatrzyma się po wprowadzeniu na wejście czysto losowego ciągu binarnego; w przystępny sposób omawia tę definicję Paul Davies w *The Mind of God*, rozdział 5, odcinek "Unknowable", 128-134 (Simon and Schuster, New York 1992).

<sup>2</sup> Z wprowadzenia do książki *Simulating Society: A Mathematical Toolkit for Modeling Socioeconomic Behavior*, Springer Verlag 1998. Zob. [www.telospub.com/catalog/FINANCEECON/SimSoc.html](http://www.telospub.com/catalog/FINANCEECON/SimSoc.html)

<sup>3</sup> Praca naukowa finansowana ze środków Komitetu Badań Naukowych w latach 2003-2006 jako projekt pt. *Nierozstrzygalność i algorytmiczna niedostępność w naukach społecznych*, nr 2 H01A 030 25.

To, co już wiemy o problemach przyrodniczych i społecznych wiary tej nie potwierdza. Ale świadomość tego faktu niełatwo dociera do badaczy, zwłaszcza w naukach społecznych. Niektórzy zdają się żywić pogląd, że kwestie nierozstrzygalne pozostają w egzotycznym rezerwacie czystej matematyki, a trzymają się zdala od faktów empirycznych. Badania logiczno-informatyczne podważające ten pogląd są stosunkowo świeżej daty. Stąd potrzeba dostarczenia o nich wiadomości wraz z refleksją, co należy w tej sytuacji czynić.

Ale na tym nie koniec zaskakujących wiadomości. Fizycy donoszą, że pewne problemy rozwiązywalne od strony algorytmów czyli software'u, pod warunkiem że sprostą im rozwój hardware'u, mogą być nie do pokonania z racji ograniczeń tego rozwoju ze strony fizyki. Miniaturyzacja, wciąż zwiększająca moce obliczeniowe, jest limitowana ostatecznie ziarnistością materii, a rozbudowa komputerów w kierunku coraz większych objętości spowolni przesyłanie sygnałów, nie mogących wszak przekroczyć prędkości światła.

Są jednak i dobre wiadomości dzięki fizyce, a także dzięki pewnemu zastanowieniu się filozoficznemu. W przyrodzie, w szczególności w mózgu, rozgrywają się procesy liczenia, co do których nie ma dotąd dowodu, że muszą być posłuszne tym restrykcjom, którym podlega uniwersalna maszyna Turinga, a tym samym komputer cyfrowy. Istnieje znaczący krąg fizyków, którzy źródeł przewag nad algorytmem upatrują w przynależeniu mózgu do sfery rządzonej mechaniką kwantową. A do wyjaśnienia jest przede wszystkim ta przewaga, że mózg (żeby nie powiedzieć „umysł”) potrafi rozpoznać prawdziwość zdania gödlańskiego, odkrywać aksjomaty i wymyślać algorytmy. Żadna z tych rzeczy nie da się uzyskać z mechanicznej procedury, jest więc ich źródło gdzieś w żywej przyrodzie (lub jakichś jej okolicach penetrowanych przez filozofię).

## 1. STAN ZAGADNIENIA I KIERUNKI ROZWIĄZAŃ

**1.1.** Pod koniec 20-go wieku ukształtowała się w informatyce **teoria złożoności obliczeniowej**. Poprowadziła ona dalej główny wątek logiki od tego punktu, do którego doprowadzili ją, przejmując od Hilberta sztafetę, Gödel, Turing, Church, Post, Tarski (by wymienić postaci sztandarowe). Logika w tym punkcie zaskoczyła świat nauki odkryciem nierozstrzygalności w matematyce i w sobie samej. Wkrótce pojawiło się też zrozumienie, że w sferze zagadnień rozstrzygalnych są takie, których rozwiązania nigdy się nie dowiemy — choćby zaprząć do pracy tyle superkomputerów, ile jest elektronów w kosmosie, a na liczenie dać im tyle miliardów lat, ile wszechświat sobie dotąd liczy. Tę nierozwiązywalność praktyczną w obrębie zagadnień, które „same w sobie” są rozstrzygalne, nazywano po angielsku *intractability* z przymiotnikiem *computational* lub *algorithmic*. Po polsku oddawać to będziemy zwrotem **niedostępność algorytmiczna**.

Zwrotu tego nie należy rozumieć w ten sposób, że nie istnieje algorytm do rozwiązania problemu, lecz że rozwiązanie poszukiwane tym algorytmem jest niedostępne (z rozważanych dalej powodów, jak niedostatek zasobów czasu czy pamięci). Analogicznie, gdy powiemy, że problem jest niedostępny obliczeniowo, to nie znaczy, że nie jest możliwy sam w sobie proces obliczeniowy prowadzący do rozwiązania (jest on możliwy, gdy rozwiązanie jest liczbą obliczalną), ale że mimo prowadzenia obliczeń nie dostąpi się rozwiązania. Używać w tym tekście przymiotnika „algorytmiczna”, ponieważ termin „algorytm” jest szerzej znany niż termin „obliczanie” w technicznym sensie Turinga [1936], różnym od potocznego; potocznie można powiedzieć „skoczek dobrze obliczył odległość” choć proces ten nie wymaga (jak wymaga u Turinga) posługiwania się cyframi.

O źródłach algorytmicznej niedostępności, o których traktuje teoria złożoności obliczeniowej, będzie mowa dalej. W tym momencie wystarczy przyjąć ją do wiadomości, żeby dostrzec następujące kwestie metodologiczne odnoszące się do nauk społecznych.

- [1] Czy algorytmom potrzebnym do modelowania i symulacji zjawisk społecznych zdarza się mieć złożoność, która czyniłaby podjęty problem nierozstrzygalnym lub algorytmicznie niedostępnym?
- [2] Jeśli tak, to czy istnieją metody takiego przekształcenia problemu, żeby stał się on dostępny algorytmicznie, a zarazem odpowiedź nań była wystarczającym przybliżeniem do odpowiedzi na problem pierwotny?
- [3] Jeśli tak, to jakie to są metody?
- [4] Czy istnieją twierdzenia lub inne elementy teorii, których przyjęcie nie jest uzasadnione żadnym stosowanym w danej teorii algorytmem?

Odpowiedź na pytanie [4] jest natychmiastowa, zarówno dla nauk dedukcyjnych (logika, matematyka), jak i dla empirycznych, w tym społecznych. Zdania takie, oczywiście, istnieją. Są nimi w teorii dedukcyjnej aksjomaty i reguły, a w empirycznej zdania obserwacyjne oraz postulaty znaczeniowe w sensie Carnapa [1956], zwane postulatami języka przez Ajdukiewicza [1965]. Mimo oczywistości odpowiedzi, pytanie [4] należało postawić, wyłania się zeń bowiem następne:

- [5] Na jakiej podstawie akceptuje się w określonej teorii zdania, gdy nie upoważnia do tego żaden algorytm?

Jeśli pytania [1], [2] i [3] adresować do fizyków, to uzyska się odpowiedzi wraz z jakąś przykładową listą zagadnień fizyki — nierozstrzygalnych, niedostępnych algorytmicznie, czy takich, co mają tylko rozwiązania przybliżone. Fizyka dysponuje pokaźnym zbiorem tego rodzaju wyników limitatywnych czyli dotyczących ograniczeń rozwiązywalności.

Dogodnym punktem wyjścia do rozpatrzenia tych zagadnień jest tekst Stephena Wolframa *Undecidability and Intractability in Theoretical Physics* [1985].<sup>4</sup> Daje on przykłady kwestii nierozstrzygalnych i kwestii niedostępnych obliczeniowo w fizyce, posługując się pojęciem *reducibility*, które oddaje też termin *compressibility* (po polsku algorytmiczna upraszczalność) z algorytmicznej teorii informacji Kołmogorowa i Chaitina. Brak tej cechy polega na tym, że algorytm symulujący badany proces musi go odtwarzać krok po kroku (*explicit simulation*), bez możliwości skrócenia. Obliczenia nieupraszczalne są z racji swej długości narażone na to, że zabraknie dla nich zasobów czasu lub pamięci, czyli że okażą się algorytmicznie niedostępne. Przykłady nieupraszczalności czerpie Wolfram z pewnych procesów w automatach komórkowych, z obwodów elektrycznych, sieci reakcji chemicznych etc. W tym repertuarze problemów jedne są niedostępne algorytmicznie, inne nierozstrzygalne. Kończy autor uwagą, że nie są to sytuacje wyjątkowe, lecz powszechne.<sup>5</sup> Tytuł artykułu Wolframa [1985] dostarcza wzoru dla sformułowania tematu obecnych rozważań. Cieszy się on dużą liczbą cytowań (np. ponad pół setki wykazuje Google), nadaje się więc na szeroko akceptowalny wzorec pewnej klasy zagadnień. Pytanie postawione w tym artykule w odniesieniu do fizyki można odnieść do każdej innej nauki lub grupy nauk,

<sup>4</sup> Wolfram jest szeroko znany dzięki pracom na temat automatów komórkowych (zebrane w książce [1994]) oraz autorstwu *software'u* "Mathematica" do obliczeń i programowania w badaniach naukowych. Jego monumentalna książka [2002] głosząca, że automaty komórkowe pewnej klasy stanowią adekwatny model świata fizycznego stała się bestsellerem naukowym roku.

<sup>5</sup> Wolfram wymienia liczne przypadki niedostępności obliczeniowej i nierozstrzygalności w fizyce. Ale odwołanie ich po polsku wymagałoby konsultacji co do polskiej terminologii, cytuję przykładowo fragment w oryginale. "Quantum and statistical mechanics involve sums over possibly infinite sets of configurations in systems. To derive finite formulas one must use finite specifications for these sets. But it may be undecidable whether two finite specifications yield equivalent configurations. So, for example, it is undecidable whether two finitely specified four-manifolds or solutions to the Einstein equations are equivalent (under coordinate reparametrization)."

Postawmy je naukom społecznym, po następującym wyjaśnieniu spraw terminologicznych związanych z tytułem artykułu Wolframa.

Terminowi *intractability* towarzyszy domyślnie przydawka *algorithmic* lub (zamiennie) *computational*; w obecnym kontekście została wybrana ta pierwsza. Po jej uwzględnieniu temat Wolframa brzmi: *Undecidability and Algorithmic Intractability in Theoretical Physics*. "Undecidability" ma w terminologii polskiej dobrze ustalony odpowiednik „nierozstrzygalność”.<sup>6</sup> Brak natomiast polskich odpowiedników dla stosunkowo nowego terminu "tractability" i jego zaprzeczenia "intractability". Pojawiające się próby tłumaczenia przez, odpowiednio, „nieoporność” i „oporność” zniekształcają oryginalną opozycję, dając termin negatywny („nieoporność”) w miejsce pozytywnego ("tractability") i vice versa, czego lepiej uniknąć. Nie ma tej usterki przykład następujący:

algorithmic tractability — dostępność algorytmiczna;

algorithmic intractability — niedostępność algorytmiczna.

Przymiotnik „rozstrzygalny” i jego negatyw „nierozstrzygalny” orzeka się w pewnych kontekstach o zdaniach (np. nazywamy zdanie gödłowskie nierozstrzygalnym), w innych kontekstach o całych teoriach (por. Tarski et al. [1968]), a kiedy indziej o problemach. Ten ostatni sposób mówienia będzie występował w tych rozważaniach; podobnie, określenia „dostępny algorytmicznie” i „złożony” będą orzekane o problemach.

**1.2.** Po ustaleniach terminologicznych wracamy do pytań [1]-[5], żeby odnotować, jak się do nich odnoszą badacze zjawisk społecznych i jak teoretycy złożoności obliczeniowej.

Jeśli odpowiadzi na [1] poszukiwać w literaturze socjologicznej, ekonomicznej etc. relacjonującej badania z użyciem modeli matematycznych, to raczej rzadko spotkamy się ze świadomością co do rodzaju złożoności stosowanych algorytmów. Jeśli natomiast szukać w pracach teoretyków złożoności obliczeniowej, to uzbiera się spory plon wyników limitatywnych, analogiczny do tego, co znajdujemy w modelowym dla obecnych rozważań studium Wolframa. W świetle tych wyników odpowiedź na [1] jest twierdząca.

Pojawia się więc potrzeba skonfrontowania obu nurtów: stosujących algorytmy badań empirycznych oraz badań logicznych nad tymi algorytmami. Potrzeba tym większa, że w niektórych badaniach empirycznych nie tylko pomija się problem ewentualnych ograniczeń mocy obliczeniowej algorytmów, ale tak się postępuje, jak gdyby było się zabezpieczonym przed ograniczeniami. Wtedy nie ma szans na pojawienie się doniosłych metodologicznie pytań [2] i [3].

Postępowanie nie liczące się z ograniczeniami algorytmów jest usprawiedliwione tylko wtedy, gdy wiadomo, iż podejmowany problem jest tak prosty, że sprostą mu zastosowany algorytm. Obserwujemy jednak coś przeciwnego. Zaniedbywanie pytań o złożoność problemu zdarza się w problemach szczególnie trudnych, o najwyższym stopniu komplikacji. Oto trzy tego rodzaju przykłady (P1-P3).

P1 • „Silna” ("strong") sztuczna inteligencja. Jest to projekt w najwyższym stopniu ambitny, chodzi bowiem o pełną (nieodróżnialną od oryginału) symulację najbardziej złożonego tworu przyrody, jakim jest mózg ludzki. Realizacja tego projektu miałaby daleko idące konsekwencje dla nauk społecznych, można by bowiem metodą symulacji, poprzez sztuczne społeczeństwa ("artificial societies") uzyskiwać wiedzę, jak tworzyć optymalne układy społeczne. Sztuczne społeczeństwo (SS) to takie, w którym każdy członek jest reprezentowany przez podprogram określający jego (sztuczny) umysł oraz interakcje z otoczeniem (ta gałąź informatyki podpada pod określenie "multi-agent simulation"). Nie tylko SS jest warunkowane przez

<sup>6</sup> Stosuję tu odmienne cudzysłowy do tworzenia nazw terminów, gdy chodzi o nazwy polskie i nazwy angielskie. Na angielską wskazuje użycie obu cudzysłówów górnych; konwencja ta zwalnia od informowania, że chodzi o termin angielski. Cudzysłowy w nazwach wyrażeń są opuszczane, gdy dany termin oprócz tego, że jest cytowany jest ponadto wyróżniony kursywą, wyłuszczeniem itp.



uprzednie stworzenie SI, ale jest też koniecznym warunkiem zaawansowanej SI, bo rozwój inteligencji wymaga odpowiedniego otoczenia społecznego. Pomimo takiej złożoności, pogłębionej przez sprzężenie zwrotne SI-SS, z frontu badań nad SI nie dochodzą meldunki o wynikach limitatywnych, płynnie natomiast strumień zapowiedzi oczekiwanego w niedługim czasie sukcesu.

P2 • Centralne planowanie socjalistyczne z pomocą komputerów ("socialist calculation") jest koncepcją głoszoną przez Oskara Langego w polemice z autriacką szkołą ekonomiczną (von Mises, Hayek). Zarzuty tej szkoły polegały na wskazywaniu złożoności faktów gospodarczych, nie do ogarnięcia dla centralnego planisty. Lange na początku lat 60-tych odpierał je argumentem, że co było niemożliwe przed wynalezieniem komputerów staje się dzięki nim bezproblemowe. Koncepcja ta jest do dziś jest podtrzymywana przez niektórych lewicujących ekonomistów. Wobec stosunkowo dużej podatności zjawisk ekonomicznych na pomiar, powinien dać się określić rząd wielkości, gdy idzie o liczbę danych wejściowych ("input data"); istnieją też propozycje modeli matematycznych np. równowagi rynkowej (Pareto, Lange etc). Są więc podstawy, żeby oszacować złożoność algorytmów niezbędnych do efektywnego centralnego planowania. Należy podjąć kwereń, żeby tego rodzaju rachunki wykryć w istniejącej literaturze, a w przypadku niepowodzenia poszukiwań, samemu określić złożoność problemu. (Więcej na ten temat – w ustępie 2.2.)

P3 • Raport Klubu Rzymskiego przewidujący pod koniec 20-go stulecia totalną klęskę ekonomiczną i ekologiczną w skali światowej powoływał się na fachowe symulacje komputerowe (wykonane w MIT). Jest oczywiste, że tak skomplikowane przedsięwzięcie jest niewykonalne dopóki nie uprości się modelu. Są uproszczenia, które nie deformują rzeczywistości, a ułatwiają i przyspieszają badanie, są też takie, które prowadzą do obrazu świata z gruntu innego niż świat realny. Takim upraszczającym założeniem Klubu Rzymskiego, które się przyczyniło do fikcyjności jego prognoz było pominięcie czynnika ludzkiej twórczości badawczej i wynalazczej, jak też twórczości w radzeniu sobie z zagrożeniami (podobnie, czynnik ten musi pomijać koncepcja centralnego planowania gospodarczego, bo nie da się planować odkryć). Głęboką trudnością jest tu brak wiedzy, co może się niepodziewanie zdarzyć w mózgach uczonych, odkrywców, reformatorów etc., ale gdyby nawet wiedzę taką objawił demon Laplace'a, to wobec bezmiernej złożoności fenomenu twórczego myślenia, trudno żeby proces ten dał się symulować za pomocą dostępnych dla komputera cyfrowego algorytmów.

Jak reagować na fakty ignorowania realnej złożoności w badaniach społecznych? Reakcja fatalistyczna polegałaby na tym, że uzna się za niemożliwe dokonanie w takich przypadkach oceny złożoności modelu, a lukę w wiedzy o modelu zrekompensuje się wiarą, że sprostą on rzeczywistości, czyli okaże się zdalny do jej wyjaśniania i przewidywania. Może i wolno by się zgodzić z taką reakcją, gdyby istotnie oszacowanie złożoności modeli w naukach społecznych były czymś przekraczającym możliwości badawcze. Tak jednak nie jest. Żeby to pokazać, wystarczy wziąć pod uwagę jakiś szeroko stosowany model zjawisk społecznych i wymienić studia nad tym modelem prowadzone w teorii złożoności obliczeniowej.

**1.3.** Dobrze nadającym się do tego celu przypadkiem jest **dylemat więźnia** — standardowy model dla rozległej klasy interakcji społecznych. O jego rozpowszechnieniu świadczy np. to, że samych stron internetowych podających odsyłacze (linki) do tej tematyki Google wykazuje (w początku roku 2003) ponad 800. Na tym przykładzie daje się prześledzić, jak wzrost liczby danych wyjściowych (liczba aktorów społecznych czyli graczy, liczba strategii, liczba rozgrywanych rund) prowadzi w pewnych przypadkach do nierozstrzygalności lub algorytmicznej niedostępności.

Nazwa „dylemat więźnia” bierze się stąd, że w ilustrującej problem opowieści stają przed dylematem dwaj podejrzani o napad (w taką fabułę ubrano pierwsze sformułowanie dylematu w 1940). Sędzia śledczy, żeby wydobyć zeznania, stawia każdemu z osobna (żeby drugi o tym nie wiedział) następujące warunki. Jeśli przyznają się obaj, dostaną po 15 lat więzienia, Jeśli obaj zaprzeczają i z tego powodu nie da się im udowodnić napadu, dostaną tylko po trzy lata (z innego paragrafu). Jeśli jeden się przyzna, a drugi zaprzeczy, ten pierwszy będzie uwolniony (jako zasłużony dla śledztwa), a drugi dostanie wyrok 20 lat więzienia.

Istotą dylematu jest pytanie, czy bardziej się opłaca lojalne współdziałanie, czy egoistyczne dążenie do własnego interesu, choćby kosztem partnera — w sytuacji, gdy nie wiadomo, jak ten

drugi postąpi; taka niewiedza o drugim należy do istoty dylematu. Będąc do wyboru sposoby postępowania nazywa się strategiami gry, określając je, odpowiednio, jako *strategię kooperacyjną* (w przykładzie z więźniami jest to wstrzymanie się od obciążających zeznań) i *strategię konkurencyjną* (dążenie do własnego zysku bez liczenia się ze stratami partnera).

Przykład z więźniami dobrze się nadaje dla podkreślenia dramatyzmu dylematu, ale gorzej ilustruje taką sytuację, ważną dla wyjaśniania procesów ewolucyjnych, gdy strony mają możliwość grania ze sobą wzajem po wiele razy. Takie gry *iterowane* ("iterated"), w których każdą grę składową nazwiemy *rundą*, dają stronom możliwość uczenia się, zmniejszając niewiedzę o partnerze i czasem harmonizując w pewien sposób działania stron. Niech tę sytuację ogólniejszą ilustruje gra z następującymi regułami.

Jeśli A i B chcą kooperować, — każdy dostaje po 3 złote talary.

Jeśli A wybiera kooperację z B, gdy B wybiera konkurencję — B dostaje 5 talarów, zaś A nic.

Jeśli obaj decydują się na konkurencję, — obaj zyskują po jednym talarze.<sup>7</sup>

Gdy gra ma jedną tylko rundę, i gdy niepewność co do zachowania drugiej strony wyrazimy jako prawdopodobieństwo 0.5 dla każdej ewentualności, to wiadomo, że większą korzyść da strategia konkurencyjna. Jeśli gra ma wiele rund, sprawa się komplikuje m.in. w przypadku, gdy każdy zdobywa wiedzę o tym, do jakich strategii skłonny jest partner. Kiedy proces rozwija się w tym kierunku, że po każdej stronie zwiększa się z czasem pewność co do kooperacyjnego nastawienia partnera, dla obu rosną szanse wygrywania po 3 talary, podczas gdy trwanie w nastawieniu konkurencyjnym przynosiłoby każdemu tylko po talarze. Mamy tu pewien model ewolucji społecznej w kierunku wzrostu kooperacyjności.

W badaniach nad dostępnością algorytmiczną problemu więźnia naturalny model obliczeniowy stanowią automaty komórkowe.<sup>8</sup>

Rozważano m.in. układ, w którym komórka ma za partnerów gry ośmiu bezpośrednich sąsiadów. Strategia zaś, będąca stanem komórki, może być (a) czysto konkurencyjna, (b) czysto kooperacyjna, (c) mieszana typu „wet za wet”, tj. reaguje się kooperacyjnie na ruch kooperacyjny i konkurencyjnie na ruch konkurencyjny, (d) taka, że śledzi się, który sąsiad najlepiej wychodzi na swojej strategii i tę się naśladuje. W toku gry pewne strategie stają się częstsze od innych i w tym sensie uzyskują przewagę, co daje się śledzić jako zmiany konfiguracji na dwuwymiarowej planszy, na której rozgrywa się ewolucja automatów.

W pewnym zakresie potrafimy przewidywać kierunek takiej ewolucji. Powstaje problem: czy istnieje algorytm, który rozstrzygałby w każdym przypadku, czy *dana konfiguracja strategii uzyska*

<sup>7</sup> Tę wersję gry można przećwiczyć, otworzywszy interaktywną stronę internetową [serendip.brynmawr.edu/playground/pd.html](http://serendip.brynmawr.edu/playground/pd.html).

<sup>8</sup> Niniejszy opis streszcza materiał znajdujący się pod adresem: [www.sunysb.edu/philosophy/faculty/pgrim/SPATIALP.HTM](http://www.sunysb.edu/philosophy/faculty/pgrim/SPATIALP.HTM). Jest to artykuł Patricka Grima "Undecidability in the Spatialized Prisoner's Dilemma: Some Philosophical Implications". W sposób interesujący dla tych rozważań Grim akcentuje doniosłość dylematu więźnia jako modelu procesów społecznych. Oto jego sformułowanie. "This simple game-theoretic model seems to capture in miniature something of the tensions between individual acquisitiveness and the goals of collective cooperation. That is of course precisely why it has become a major focus of modelling within theoretical sociology, theoretical biology, and economics. [...] It is no simplification to say that our strongest and simplest models of the evolution of biological and sociological cooperation—and in that respect our strongest and simplest models of important aspects of ourselves as biological and social organisms—are written in terms of the Iterated Prisoner's Dilemma.

na trwałe przewagę nad innymi? Grim [1997 i cytowany tekst z WWW] podał dowód, że jest to problem nierozstrzygalny w klasycznym gödłowskim sensie: to znaczy, nie ma takiego algorytmu.<sup>9</sup>

Inny ważny nurt stanowią badania nad związkiem między złożonością obliczeniową gry i racjonalnością graczy. Teoria złożoności powinna pomóc teorii gier w zdaniu sprawy z faktu empirycznego, że strategie kooperacyjne bywają korzystne dla obu stron, podczas gdy z samej teorii gier, paradoksalnie, wynika, że gdy gracze nie zmieniają strategii (co nazywa się stanem równowagi), to korzystniejsze jest dla obu trzymać się strategii konkurencyjnej. Ta konsekwencja, która zachodzi w przypadku wyposażenia graczy w nieograniczone środki rozwiązywania problemów (tzw. nieograniczona racjonalność), przestaje obowiązywać w pewnych przypadkach np. ograniczenia pojemności pamięci; pamięć można mierzyć liczbą stanów automatu skończonego realizującego określoną strategię w grze mającej  $n$  rund. A. Neymann [1985], pionier koncepcji ograniczonej racjonalności, podał ważne twierdzenie, że gdy ograniczenie pamięci obu graczy polega na tym, że jej pojemność znajduje się w przedziale  $[n^{1/k}, n^k]$ , gdzie  $n$  jest liczbą rund, przy  $k > 1$ , to strategią korzystną dla obu graczy jest kooperacja. Udowodniono też m.in., że kooperacja dominuje nad konkurencją w grach o nieskończonej liczbie rund, jak i w grach o skończonej ale nieznanym graczom liczbie rund (zob. Papadimitriou and Yannakakis [1991]).

Te i liczne inne wyniki dotyczące szczególnie użytecznego modelu interakcji, dylematu więźnia, ilustrują, jak wielki wkład może wnieść teoria złożoności obliczeniowej do metodologii nauk społecznych. Powstaje tu jednak problem analogiczny do tego, jaki w gospodarce stanowi droga od wynalazku do jego wdrożeń przemysłowych. Wstępne rozeznanie pokazuje, jak mało z tego dorobku teoretycznego jest spożytkowane w praktyce badawczej nauk społecznych. To wstępne rozpoznanie trzeba poddać systematycznej weryfikacji, a w przypadku jego potwierdzenia podjąć pytanie o przyczyny tego stanu rzeczy. Czy są to przyczyny obiektywne, które polegałyby na tym, że wyniki teorii złożoności są zbyt subtelne lub zbyt abstrakcyjne dla realnych potrzeb praktyki badawczej? Czy może raczej powody subiektywne, biorące się z niedostatecznego nadążania badaczy za postępami teorii złożoności? Pytania te zasługują na podjęcie w badaniach z zakresu metodologii nauk i naukoznawstwa.

**1.4.** Opisane wyżej badania Grima nad rozstrzygalnością jednego z problemów w dylemacie więźnia (por. przypis 9) okazują się wielce pomocne dla kwestii sformułowanej wcześniej (ustęp 1.1) jako pytanie nr [4], mianowicie: *czy istnieją twierdzenia lub inne elementy teorii, których przyjęcie nie jest uzasadnione żadnym stosowanym w danej teorii algorytmem?* Wprawdzie odpowiedź twierdząca jest skądinąd oczywista, skoro w teorii empirycznej ani zdania obserwacyjne ani postulaty znaczeniowe nie legitymują się pochodzeniem od algorytmu, ale wyniki takie jak Grima wskazują na zdania nadające się niewątpliwie do roli aksjomatów (gdy dojdzie się do etapu aksjomatyzowania danej teorii).

Żeby skonkretyzować ten wniosek, skorzystajmy z odpowiedzi zawartej pośrednio w badaniach Grima. Na pytanie, które wedle tych badań jest nierozstrzygalne, można mieć w odpowiedzi hipotezę o charakterze intuicyjnym wychodzącą z obserwacji pewnych zjawisk społecznych. Np. ten fakt, że wojny domowe (dobrze się mieszczące w schemacie dylematu więźnia), gdy już obie strony mocno się wykrwawiają (a więc po wielu rundach gry iterowanej), często kończą się kompromisem. To znaczy, zwycięża po obu stronach strategia kooperacyjna, nastaje zatem

<sup>9</sup> Zob. Grim [WWW, op. cit.]. "There is no general algorithm [...] which will in each case tell us whether or not a given configuration of Prisoner's Dilemma strategies embedded in a uniform background will result in progressive conquest. Despite the fact that it is one of the simplest models available for basic elements of biological and social interaction, the Spatialized Prisoner's Dilemma proves formally undecidable in the classical Gödelian sense."

pewien stan równowagi. Jego nadejścia, jak okazuje wynik limitatywny Grima, nie można wnioskować z aksjomatów dostarczającej tu modelu teorii gier, co jest wskazówką, że na naszą hipotezę czeka wolne miejsce w gronie aksjomatów.

Pytaniu [4] towarzyszy zrodzone zeń następne, mianowicie [5]: *na jakiej podstawie akceptuje się w określonej teorii zdania, gdy nie upoważnia do tego żaden algorytm?*

Taką podstawą może być okoliczność, że zdania przyjęte w danej teorii bez dowodu mają dowód w innej zasługującej na akceptację teorii (a gdy dowód jest sformalizowany, mamy do czynienia z pewnym algorytmem); jest to postępowanie będące w nauce na porządku dziennym. Mniej oczywista jest supozycja, że jakieś algorytmy "krążą" poza teoriami, a pewnym teoriom dostarczają zdań zasługujących na taką akceptację jak twierdzenia danej teorii.

Dopuszczenie takiej supozycji bierze się (w obecnym kontekście) z potrzeby nawiązania dialogu z teorią tzw. silnej sztucznej inteligencji. Pomaga to wyartykułować myśl tej teorii, że twierdzenia z klasy reprezentowanej przez zdanie gödłowskie są również produktem jakiegoś algorytmu. Przy takim sformułowaniu robi się miejsce na pytanie: do jakiej teorii ów algorytm należy, skoro nie należy (na przykład) do arytmetyki? To wyznacza kolejny ruch w dyskusji, jak np. stwierdzenie, że chodzi o algorytm spoza teorii, powiedzmy, funkcjonujący w mózgu Gödla. To z kolei angażowałoby w twierdzenie, że mózg ten ma moc obliczeniową równą mocy maszyny Turinga, co byłoby znowu krokiem naprzód, przesądzając na kim spoczywa *onus probandi*.

Odnotowuję ten problem, będący w centrum uwagi sporów o sztuczną inteligencję, w tym celu, żeby się do niego odnieść z metodologicznego punktu widzenia. Istotne jest z tego punktu, że mamy do czynienia ze zdaniami, które w danej teorii nie wywodzą się z żadnych innych zdań i nie są w niej aksjomatami; nie są też zdaniami obserwacyjnymi, które rejestrują tylko spostrzeżenie zmysłowe dotyczące jakiegoś „tu i teraz”. Stosowną dla nich nazwą jest: zdania (lub sądy) **a priori**.<sup>10</sup>

Żeby uzasadnić to określenie, podkreślę raz jeszcze, co następuje. Pierwowzorem zdań zasługujących na uznanie jako prawdziwe, choć nie są w danej teorii aksjomatami ani nie wywodzą się algorytmicznie z aksjomatów jest **zdanie gödłowskie**. Wiadomo z badań metateoretycznych (referowanych wyżej), że zdania takie występują także w operujących modelami matematycznymi teoriach empirycznych, w tym społecznych. Jest to ważna metodologicznie kategoria, co wymaga wyodrębnienia jej za pomocą specjalnej nazwy. Zasługują one na miano potencjalnych aksjomatów, ale oprócz tego aspektu dotyczącego ich ewentualnego przeznaczenia powinien być odnotowany aspekt ich pochodzenia. Ten aspekt genetyczny oddaje tradycyjny, od wieków zadomowiony w filozofii zwrot „a priori”.

Spojrzenie na teorię społeczną jako zawierającą sądy aprioryczne dobrze koresponduje z poglądem takiego klasyka metodologii nauk społecznych, jakim jest Ludwig von Mises. W dziele *Human Action* [1966] charakteryzował on ekonomię jako wysoce ogólną naukę o ludzkim działaniu, którą nazywał **prakseologią** i tak opisywał od strony metodologicznej.

"Praxeology is a theoretical and systematic science. [...] It aims at knowledge valid for all instances in which the conditions exactly correspond to those implied in its assumptions and inferences. [...] Its statements are, like those of logic and mathematics, a priori." [p. 32]

"The fact that man does not have the creative power to imagine categories [of thought, action etc] at variance with the fundamental logical relations and with the principles of causality and teleology enjoins upon us what may be called *methodological apriorism*. [p. 35, italics LvM.]

<sup>10</sup> Jeśli ta kategoria wyda się komuś niepojęta lub wręcz mistyczna, to dla obrońców „naukowej trzeźwości” mamy wyjście w postaci owych algorytmów spoza wszelkiej teorii. Trudno ich nie uznać komuś, kto skądinąd reprezentuje „jedynie naukowy” pogląd, że mózg jest maszyną Turinga i produkuje aksjomaty oraz wszelką inną aprioryczność wedle jakichś sobie samemu nieznanym algorytmów.



Jakie prawidłowości społeczne są według von Misesa poznawane w taki sposób aprioryczny, to widać ze spisu treści rozpisującego na punkty tematykę dzieła *Human Action*. Mamy tu typowe tematy ekonomii, jak rachunek monetarny, rynek, ceny, kredyt, praca, płaca, rola rządu, podatki etc. Pogląd, że prawa ekonomiczne dotyczące tych spraw są wszystkie tak aprioryczne jak twierdzenia matematyki brzmi skrajnie. Można jednak uczynić go bardziej wyważonym, adaptując ideę W. V. Quine'a o stopniach aprioryczności. Najwyższy jej stopień przysługuje logice i matematyce (stąd element dopełniający, empiryczność, jest w nich w najmniejszej dawce), a dalej jest rozdzielony w różnych porcjach. Ekonomia mogłaby znaleźć się w tej hierarchii stosunkowo blisko szczytu aprioryczności. Istotnie, takie jej zasady jak zasada oczekiwanej użyteczności, prawo podaży i popytu, czy informacyjna funkcja cen (idea F. Hayeka) są silnie aprioryczne.

Daje to asumpt do przyswojenia na użytek tych rozważań proponowanego przez von Misesa określenia **metodologiczny aprioryzm**. Ujmuje ono fakt, że liczne prawa nauk społecznych zachowują się jak aksjomaty matematyki czy logiki. Stosowne będzie dla nich określenie **postulaty znaczeniowe** wprowadzone w ustępie 1.1, obejmuje ono bowiem aksjomaty, ale odnosi się do klasy pojemniejszej, w której znajdują się zdania podzielające z aksjomatami aprioryczność, a nie podzielające ich roli pierwszych przesłanek systemu.

Postulaty znaczeniowe nie są produktem algorytmu działającego wewnątrz teorii. Można powiedzieć, że wytwarza je jakiś algorytm spoza teorii funkcjonujący w naszych mózgach w celu produkcji zasad apriorycznych albo też że wytwarza je jakiś proces nie-algorytmiczny. W każdym razie, nie będą one przyprawiać badaczy o ból głowy z powodu nadmiernej złożoności algorytmicznej, skoro w ogóle nie powstaje co do nich problem postępowania algorytmicznego.

Jak się zapatrywać na taką sytuację w naukach społecznych? Co warte są poznawczo owe zasady aprioryczne – cieszące się swoistym immunitetem, skoro nie są poddawane ani sprawdzeniu empirycznemu ani kontroli co do niesprzeczności z innymi zdaniami systemu (możliwej dopiero w systemie aksjomatycznym)? Jest to doniosły problem badawczy. Z jednej strony, uznanie tak znaczącej roli elementu apriorycznego narusza szeroko akceptowany paradygmat empirystyczny, z drugiej jednak praktyka modelowania matematycznego (gdy równania matematyczne stają się twierdzeniami teorii empirycznej) wprowadza znaczący składnik aprioryzmu; niezbędność postulatów znaczeniowych jako konstytuujących język teorii to drugi taki składnik.

Ów problem badawczy da się atakować z szansą na sukces przy następującej strategii. Trzeba poddać analizie metodologicznej funkcjonowanie w naukach społecznych tak płodnych i szeroko akceptowanych modeli, jakimi są teoria gier czy automaty komórkowe. W toku zastosowań model matematyczny, jako wyrażający pewne założenia o racjonalności (inteligencji) działań, jest konfrontowany z obserwacjami, które nie zawsze taką racjonalność potwierdzają, a czasem skłaniają do jej negowania. Jaką wtedy badacz ma mieć regułę preferencji? Czy zmienić założenia aprioryczne, czy je pozostawić i w ich świetle interpretować dane obserwacyjne? Nie ma na to gotowej odpowiedzi ogólnej. Trzeba badać każdy przypadek z osobna, a ostateczny werdykt zapadnie w toku rozwoju nauki, gdy przyjęte dla danego przypadku rozwiązanie potwierdzi się lub, przeciwnie, zaniknie po pewnym czasie, pozostając eksponatem w archiwum teorii, które wypadły z obiegu.

**1.5.** Obecny ustęp stanowi rodzaj aneksu erudycyjnego. Nie wysuwa się w nim problemów czy hipotez, a tylko komentuje niektóre wcześniej użyte pojęcia (dla tych, co mieliby potrzebę dokładniej z nimi się zapoznać).

Pojęcie rozstrzygalności wywodzi się z logiki, gdzie pojawiło się explicite w kontekście Programu Hilberta wraz z przekonaniem, że każdy poprawnie sformułowany problem matematyczny jest rozstrzygalny (por. Hilbert i Ackermann [1928]). Udowodnienie tej hipotezy określa się jako pozytywne, zaś obalenie jako negatywne, rozwiązanie **problemu rozstrzygalności** (*Entscheidungs-*

sproblem). W oryginalnym, ważkim historycznie, sformułowaniu Hilberta i Ackermanna [1928, s. 73] brzmi on, jak następuje (wyróżnienie kursywą przez HiA).

*Das Entscheidungsproblem ist gelöst, wenn man ein Verfahren kennt, das bei einem vorgelegten logischen Ausdruck durch endlich viele operationen die Entscheidung über die Allgemeinheit bzw. Erfüllbarkeit erlaubt.*

*Die Lösung des Entscheidungsproblem ist für die Theorie aller Gebiete, deren Sätze überhaupt einer logischen Entwickelbarkeit aus endlich vielen Axiomen fähig sind, von grundsätzlicher Wichtigkeit.*

Pomimo relatywizacji do jakiejś określonej aksjomatyki i określonego zbioru formalnych reguł wnioskowania, występujące tu pojęcie procedury (Verfahren) ma zakres równie szeroki jak pojęcie algorytmu czy programu komputerowego. Każdy bowiem krok w wykonywaniu algorytmu przez komputer, będąc operacją obliczeniową, tym samym jest wnioskowaniem z aksjomatów arytmetyki prowadzonym według praw logiki (forma implikacyjna tych praw pozwala je wykorzystywać w roli reguł wnioskowania).

Pozytywne rozwiązanie problemu rozstrzygalności dla logiki pozwoliłoby na posiadanie takiego algorytmu, że poprawność każdego kroku dałaby się wykazać przez powołanie się na odpowiednią formułę logiczną, a tautologiczność czyli ważność (*Allgemeinheit*) tej formuły dałaby się zawsze wykazać dzięki rozstrzygalności logiki. Algorytm taki zapewniłby rozstrzygalność każdej teorii zaksjomatyzowanej, a przy tym sformalizowanej przez prawa logiki.

Gdy Turing [1936] i Church [1936] wykazali nierozstrzygalność logiki, czyli udowodnili negatywne rozwiązanie problemu rozstrzygalności, powstały przesłanki do stawiania nowych pytań pod adresem algorytmów, a wraz z tym narodziła się nowa teoria. Nazywa się ją teorią **złożoności obliczeniowej** (*computational complexity*) i traktuje jako dział informatyki; widać z historii zagadnienia, że jest to strefa na pograniczu informatyki z logiką, będąca w polu uwagi obu stron.

Trzeba tu zwrócić uwagę na następujące niuanse terminologiczne. Terminu "algorithmic" używa się zamiennie z "computational" w kontekstach takich, jak "algorithmic intractability" i "computational intractability". Nie przyjęła się jednak podobna zamiennność w kontekście „complexity”. Co innego oznacza zwrot "algorithmic complexity", a co innego "computational complexity". Pierwszy dotyczy miary złożoności określonej (niezależnie przez Kołmogorowa, Chaitina i Solonoffa) ze względu na stosunek między długością ciągu symboli wyprodukowanego przez algorytm a długością tego algorytmu (zob. np. Chaitin [2002]). Drugi dotyczy tego, jakie zasoby pamięci, czasu (liczba kroków) etc. są niezbędne do rozwiązania problemu przez dany algorytm; wielkość koniecznych zasobów jest miarą złożoności (do pionierskich na tym polu należy studium: Hartmanis and Stearns [1965]).

Dostępność lub niedostępność algorytmiczną, podobnie jak rozstrzygalność lub nierozstrzygalność, orzeka się o problemach. Problem jest **dostępny algorytmicznie**, gdy jest rozstrzygalny, a ponadto nie jest aż tak złożony obliczeniowo, żeby rozwiązujący go algorytm (program) nie dał się wykonać przy osiągalnych dla użytkowników komputera zasobach przestrzeni (czyli pamięci) i czasu (takie pojęcie osiągalności jest jawnie nieostre, ale jest to nieostrość praktycznie mało szkodliwa). Jak złożoność problemu wiąże się z ograniczeniami przestrzeni (pamięci) oraz czasu wykonywania algorytmu, będzie mowa w odcinku 3.

## 2. IDEA RACJONALNOŚCI I POJĘCIE INTELIGENCJI W NAUKACH SPOŁECZNYCH

**2.1.** Z rozmysłem mówi się w tym tytule racjonalności w kontekście słowa „idea”, zaś o inteligencji w kontekście słowa „pojęcie”. Z tych dwóch słowo „idea” więcej ma w sobie ładunku normatywnego czy aksjologicznego (zapewne z powodu bliskości znaczeniowej z „ideał”), a to właśnie różnicuje te dwa skądinąd bliskie sobie znaczenia.

Inteligentnym jest ten, kto skutecznie i z możliwie małymi nakładami rozwiązuje stojące przed nim problemy oraz umie odróżniać problemy ważne od mniej ważnych. Mniej więcej to samo wypadnie powiedzieć w definicji racjonalności, a więc różnica jest sprawą asocjacji, akcentu i kontekstu, a nie jakiejś wyraźnie odmiennej treści. To uzasadnia łączne podjęcie obu tematów, mające przyczynić się do tego, że te dwie sfery rozważań będą się wzajem wspierać i z bogacać. A przykład problemu społecznego, który podam w drugiej części tego odcinka może być równie dobrze dyskutowany w kategoriach racjonalności jak i w kategoriach inteligencji.

Pojęcie racjonalności jest nieodłączne od standardowego w naukach społecznych modelu gier. W grze chodzi o to, żeby wygrać, naturalne jest więc zdefiniować jako racjonalne postępowanie prowadzące do zysku, a jako nieracjonalne to, które prowadzi do straty. Wyrazi się i w tym kontekście podobną myśl, gdy „racjonalne” zastąpi się przez „inteligentne”. Ale oprócz zamienności mamy w tej teorii także z bogacanie jednej treści przez drugą. Problematyka sztucznej inteligencji wiąże inteligencję z mocą obliczeniową, a ta należy do głównych tematów teorii złożoności obliczeniowej. I oto, jak widać z tekstów wspomnianych w ustępie 1.3, ograniczenia mocy obliczeniowej, a więc także inteligencji, bywają nazywane przez autorów z tego kręgu ograniczeniami racjonalności ("bounded rationality").

Tak te dwa pojęcia zaczynają schodzić się w jedno, co prowadzi też do zbliżenia nauk społecznych i teorii inteligencji. Warto odnotować kilka kierunków tego zbliżenia. Wchodzą tu w grę, między innymi, następujące fakty.

- Intensywny proces zbliżania się nauk społecznych do SI zaczął się we wczesnych latach 90-tych, gdy postęp SI doprowadził do programów umożliwiających interakcje między sztucznymi umysłami reprezentowanymi przez odpowiednie programy. Nazwano to **rozproszoną SI** ("distributed AI"). Następny krok stał się możliwy dzięki zaistnieniu interakcji sieciowych (Internet etc.); stała się osiągalna interakcja między programami funkcjonującymi w różnych komputerach. Podmioty takich interakcji nazwano "agents"; stąd termin "multi-agent models". Wyłonił się z tych rezultatów nowy kierunek badań — **sztuczne społeczeństwo** ("AS" tj. "artificial society") — kontynuacja SI w kierunku nauk społecznych. Tak więc, programy funkcjonujące jako sztuczne umysły występują w modelach obliczeniowych do komputerowej symulacji zjawisk społecznych.

- Teoria automatów komórkowych (pochodząca od Johna von Neumanna i Stanisława Ulama) to bogate źródło modeli obliczeniowych m.in. dla procesów zachodzących w społeczeństwach. Automat komórkowy jest zbiorem obiektów rozmieszczonych w regularnie podzielonej na komórki przestrzeni. Stany tych obiektów zmieniają się w zależności od tego, gdzie i jakie obiekty występują w ich najbliższym otoczeniu. Dostarcza to modeli interakcji społecznych, takich np. jak rozprzestrzenianie się plotek czy tworzenie się izolowanych skupisk etnicznych. Proste reguły określające zależności prowadzą nieraz do bardzo złożonych i nieprzewidywalnych procesów, skąd powstają problemy nierozstrzygalne lub algorytmicznie niedostępne.

- Dział SI polegający na konstruowaniu uczących się maszyn dostarcza modeli obliczeniowych nie tylko do śledzenia ewolucji indywidualnych umysłów lecz także ewolucji struktur społecznych. Typowym przykładem uczenia się takiej struktury jest jej adaptowanie się do nowych warunków. Mamy więc w uczących się maszynach kolejny model obliczeniowy do symulacji społecznych.

Zwróćmy na koniec uwagę na korzyści z tej ewolucji pojęciowej dla interpretacji tradycyjnych problemów socjologicznych. Oto u klasyka socjologii Maxa Webera, centralnym tematem jest racjonalność struktur społecznych, np. pewnego typu cywilizacji, ale nie przyjęło się używanie w tym kontekście terminu „inteligencja”. Dostrzeżenie, że chodzi o to samo pojęcie w różnych szatach słownych pozwoli korzystać z teorii inteligencji do modelowania wspomnianych struktur za pomocą jej metod.

**2.2.** Pojęcie inteligencji czyli racjonalności orzekane o jakiejś strukturze społecznej ma egzemplifikację historyczną, która ukazuje rolę metodologiczną pojęć rozstrzygalności i algorytmicznej dostępności. Jest to słynny spór zainicjowany w latach dwudziestych 20-go wieku przez Ludwiga von Misesa dotyczący możliwości rachunku w centralnym planowaniu socjalistycznym (zob. też von Mises [1966]). Spór ten miał kulminację latami trzydziestych, gdy w szranki wstąpili Friedrich Hayek i Oskar Lange. Przez tych samych polemistów był kontynuowany w okresie powojennym, do śmierci Langego (1965); obecnie idee Langego z ich kontekstem informatycznym też bywają podejmowane przez niektórych autorów (np. Cottrell and Cockshott [1993]).

Lange spierał się z Friedrichem A. Hayekiem (1899-1992), czy inteligentniejszym regulatorem gospodarki jest wolny rynek, czy centralne planowanie. Gdy pojawiły się komputery, Lange nabrał przekonania, że to on definitywnie w tym sporze zwyciężył. Traktował bowiem wolny rynek jedynie jako instrument kalkulacyjny do obliczania prawidłowych cen, to znaczy takich, które by zapewniały równowagę podaży i popytu. Nie przeczył, że rynek jakoś się z tej roli wywiązuje, ale powoli i z błędami. Tymczasem komputer w Centralnej Komisji Planowania wyliczy bezbłędnie "w jednej sekundzie" (własny zwrot Langego) to, co rynek liczyłby z właściwą sobie powolnością.

Z drugiej strony, rozumiał Lange, że moc obliczeniowa komputerów nie zawsze sprostaa złożoności gospodarki; dopuszczał więc pomocniczą rolę rynku w sterowaniu gospodarką na krótszych dystansach jako instrumentu centralnej władzy ekonomicznej. Absolutną natomiast przewagę uzbrojonego w komputer centralnego planisty upatrywał w rozwiązywaniu długofalowych problemów wzrostu gospodarczego. Podczas gdy rynek nadaje się, jak sądził, co najwyżej do regulowania na bieżąco równowagi ekonomicznej, nie potrafi on wytyczać dalekosiężnych celów rozwoju.

Hayek przeciwstawiał się tym poglądom wychodząc z rozważań z zakresu przetwarzania informacji. Jego myśl da się wyrazić krótko we współczesnej terminologii złożoności obliczeniowej, jak następuje. Teza o wykonalności centralnego planowania przy zastosowaniu komputerów implikuje, że stosowane do tego celu algorytmy są dostatecznie szybkie do tego, żeby nie czekać na wynik obliczeń latami lub setkami lat. Złożoność staje się tym bardziej monstrualna że centralny planista potrzebowałby wszystkich danych z całego państwa o podaży, popycie etc, w odniesieniu do każdego produktu, żeby na tej podstawie wyliczyć optymalną cenę, a do tego, aktualizować ją, gdy trzeba, z godziny na godzinę. Tymczasem system obliczeniowy, jakim jest wolny rynek, na dwa sposoby radykalnie ogranicza ten zalew danych. Każdy uczestnik gry rynkowej potrzebuje tylko danych o cenie towarów i tylko tych towarów, które są w polu jego działalności. Jest to podobne do przetwarzania danych, które jest równoległe, rozproszone, a przy tym ma cechy analogowe.

Twierdzenia te Hayek uzasadniał w sposób intuicyjny. Obecnie rysują się możliwości ich dokładniejszego sprawdzenia dzięki aparaturze pojęciowej teorii złożoności. Jedną z dróg postępowania mogłaby być następująca. Ponieważ nadal są zwolennicy poglądu Langego operujący pojęciami informatycznymi, należy od nich oczekiwać dowodu, że problemy ekonomiczne centralnego planowania rozwiązuje się algorytmami pracującymi w czasie wielomianowym, a nie wykładniczym, że istnieją wystarczające dla nich, także mierzone wielomianowo, zasoby pamięci itd. Z drugiej strony, badanie zachowań uczestników rynku powinno wykazać, czy problemy przez nich rozwiązywane dadzą się modelować jakąś teorią rozstrzygalną, a jeśli tak, to czy są bardziej dostępne obliczeniowo od problemów, przed jakimi staje centralny planista.

Ten spór nie da się potraktować jako opowieść wyłącznie historyczna. Nabiera on nowej aktualności w obecnych latach z dwóch powodów, równie doniosłych, choć pochodzących z różnych stron. Jest powód polityczny, gdyż obserwuje się w świecie narastającą falę skłonności socjalistycznych, obecnych choćby w żywiołowym ruchu antyglobalistycznym czy w pewnych standardach „politycznej poprawności”. Potrzeba więc możliwie jak najbardziej precyzyjnej analizy porównawczej gospodarki socjalistycznej z wolnorynkową. Idealnym narzędziem wskazanym



przez Lengego i Hayeka jest teoria złożoności obliczeniowej w jej obecnym stanie zaawansowania. Gdyby nawet nie było zamówienia praktycznego, ten stan wyostrenia narzędzi zachęca do ich wypróbowania na tak interesującym teoretycznie polu. To jest ten drugi powód do kontynuacji tamtego sporu, teoretyczny, a wzmacniający się wzajem z praktycznym.

### 3. ATAKOWANIE ZŁOŻONOŚCI

**3.1.** Na różne sposoby można próbować radzić sobie ze złożonością procesów przyrodniczych, umysłowych, społecznych: upraszczać problemy, wzmacniać środki obliczeniowe, sięgać dalej niż maszyna Turinga, tworzyć interakcję intuicji z algorytmem. Cywilizacja informatyczna polega na coraz lepszym radzeniu sobie z tym zadaniem. Istnieją conajmniej dwa fronty zmagania ze złożonością. Na jednym operuje teoria chaosu deterministycznego (dynamicznych układów niestabilnych), na drugim teoria złożoności obliczeniowej, o której mowa w tych rozważaniach,

Zaczyna się od rozpoznania, jakie problemy są osiągalne dla algorytmów. W sferze nieosiągalnej znajdują się zarówno zagadnienia nierozstrzygalne, jak i te, które będąc rozstrzygalne, wymagają niedostępnych praktycznie zasobów czasu i przestrzeni. Czas to liczba kroków niezbędnych do rozwiązania, a przestrzeń to pojemność pamięci, która może nie wystarczyć przy jakiejś gigantycznej liczbie danych wejściowych (mogą też wchodzić w grę inne zasoby, np. liczba współdziałających procesorów, ale te dwa są najczęściej rozważane). Za linię demarkacyjną oddzielającą strefę algorytmicznej niemożności od tego, co osiągalne, uważa się rozróżnienie dwóch kategorii czasu: wielomianowej i wykładniczej.

**Czas wielomianowy** określa np. funkcja  $n^3$ , a **czas wykładniczy** funkcja  $2^n$ , gdzie  $n$  jest liczbą danych wejściowych. Niech przy  $n$  danych wejściowych maksymalną liczbę kroków określa wielomian  $7n^3 + 5n^2 + 27$ . Dla oszacowania złożoności algorytmu wielomianowego wystarczy wziąć jego składnik o najwyższym wykładniku potęgowym, pomijając przy tym współczynnik (jak 7 w  $7n^3$ ) jako wielkość zaniedbywalną. Ten wyróżniony składnik określa rząd ("order") złożoności algorytmu; ; powiada się, że dany algorytm wymaga np. czasu  $O(n^3)$  (notacja z „O” wskazuje na ograniczenie się do rzędu wielkości, z pominięciem wielkości zaniedbywalnych). Przykładem algorytmu pracującego w czasie wielomianowym jest algorytm sortowania, który ma rząd złożoności  $O = n \log n$ , a więc mniejszy niż  $O(n^2)$ .

Do klasy zagadnień wymagających czasu wykładniczego należy problem spełnialności formuły rachunku zdań, zwany skrótowo SAT (od "satisfiability"). Mając daną formułę rachunku zdań, np. w koniunkcyjnej postaci normalnej (tj. koniunkcji alternatyw) należy rozpoznać, czy istnieje taki układ przyporządkowań wartości logicznych symbolom zmiennym, który czyni tę formułę prawdziwą. Załóżmy, że formuła ma 300 zmiennych. W najgorszym przypadku, gdy np. tylko jedno przyporządkowanie czyni formułę prawdziwą, a napotka się je dopiero przy końcu, rozwiązanie będzie wymagać  $2^{300}$  kroków.

Inny przykład niewyobrażalnie wielkiego zapotrzebowania na czas, nawet większy niż wykładniczy, bo silniejszy, to problem komiwojażera: mając dane położenia  $n$  miast, objechać je wszystkie najkrótszą trasą bez odwiedzania któregośkolwiek więcej niż raz. Niech do odwiedzenia będzie 20 miast (nie licząc miejsca startu). Liczba tras wynosi wtedy  $20!$ , bo tyle jest możliwych uporządkowań w zbiorze 20 elementów. Nie znaleziono dotąd algorytmu innego niż tego, który polega na wyliczeniu wszystkich kombinacji, zsumowaniu w każdej z kombinacji długości odcinków i rozpoznania najmniejszej z tych sum. Ponieważ mamy do czynienia z faktem, że

$$20! = 2,432,902,008,176,640,000$$

można sobie na tym przykładzie uprzytomnić, na czym polega algorytmiczna niedostępność. Jeśli nasz komputer potrafi sprawdzić milion kombinacji w ciągu sekundy, to sprawdzenie wszystkich

musiałoby zająć 77.000 lat, a dorzucimy jeszcze kilka miast, to na liczenie nie starczyłoby dotychczasowego wieku wszeświata. Mamy tu do czynienia z algorytmem posługującym się „ślepą siłą” („brute force”) czyli takim, który polega na mechanicznym zrealizowaniu wszystkich możliwości. Nie ma dla tego zagadnienia szybszego algorytmu, który dawałaby równie pewny i dokładny wynik, ale jeśli zgodzimy się na wyniki przybliżone, czas rozwiązywania problemu komiwojażera da się wydatnie skrócić.

**3.2.** Klasę problemów rozwiązywalnych przez algorytmy wielomianowe przyjęto oznaczać symbolem  $P$  (od „polynomial”). Ta zwięzła notacja ułatwia rozważanie, które stało się źródłem imponujących wyników w teorii złożoności. Określiwszy inną klasę problemów symbolem  $NP$ , formułuje się fundamentalne pytanie, czy klasy te są sobie wzajem równe:  $P=NP$ ?

Rozważanie to wychodzi od spostrzeżenia, że istnieją algorytmy wielomianowe, które na pytania rozstrzygnięcia (odpowiedź „tak” lub „nie”) dają potwierdzenie (certyfikat – „certificate”), wtedy i tylko wtedy, gdy odpowiedź „tak” jest tą prawdziwą. Na pytanie, czy dana formuła logiki zdań jest spełnialna (jeśli istotnie jest spełnialna) odpowie twierdząco np. algorytm wielomianowy. Podobnie jest w przypadku komiwojażera, gdy pytanie brzmi: czy długość danej trasy jest nie większa od takiej to a takiej liczby?

Rozpatrzmy dokładniej przykład rozstrzygania o jakiejś liczbie czy jest pierwsza, posłużymy się pomocnym do zrozumienia kontekstem psychologicznym ze wspomnień piszącego te słowa. Andrzej Mostowski, światowej klasy badacz problematyki rozstrzygalności, dał mi kiedyś żartobliwie do rozstrzygnięcia bardzo proste zadanie. Gdy przy pewnej sprawie organizacyjnej poprosił o mój adres i dowiedział się numeru mieszkania 917, z miejsca zapytał, czy jest to liczba pierwsza. Zdziwiłem się nieco, będąc przekonany, że ktoś taki powinien znać odpowiedź od razu (choć sam jej nie znałem), ale dziś sądzę, że pytanie miało charakter dydaktyczny; prof. Mostowski testował, jak sobie z tym poradzę. Otóż ze spóźnionym refleksem (minęło 30 lat), ale poradziłem sobie – w sposób typowo niedeterministyczny – dając na próbę pierwszą z odpowiedzi, jaka się nasuwa po eliminacji tych, które są z pewnością nietrafne. Eliminuję liczby parzyste i liczbę 3, bo szybko obliczam, że  $9+1+7$  nie jest podzielne przez 3, wreszcie eliminuję 5, bo 917 nie kończy się na 0 ani na 5. Dochodzę do 7, mając też na widoku następne kandydatury (11, 13 etc). Wchodzę więc na tę ścieżkę, gotów zarazem do wycofania się i próbowania następnych, gdy ta zawiedzie. Stosuję teraz prosty algorytm wielomianowy, w którym liczba kroków (czas wykonania) zależy liniowo od długości ciągu cyfr oznaczającego liczbę dzieloną. Mam szczęście, bo po pierwszej próbie otrzymuję wynik bez reszty, mianowicie 132. Mam więc rozstrzygnięcie: 917 *nie* jest liczbą pierwszą. Można tak zgadywać dowolnie wielkie liczby, to już jest kwestia talentów obliczeniowych. Ktoś szczególnie uzdolniony mógłby np. od razu odgadnąć, że 226107 dzieli się bez reszty przez 777 (dając 291).

Ten rodzaj algorytmów wielomianowych określa się mianem **niedeterministycznych** („non-deterministic”), ponieważ służą do weryfikacji zdań, których uzyskanie nie jest zdeterminowane jakąś procedurą (nie należy ich mylić z algorytmami probabilistycznymi); od „non” bierze się litera  $N$  w oznaczeniu  $NP$ . Jeśli problem nie mieści się w klasie  $NP$ , to nawet rozwiązanie ograniczone do takiego potwierdzenia może się okazać skrajnie trudne. Założenie, że zawsze możemy dysponować trafnym odgadnięciem jest fikcją, której naprawdę nie realizuje żadna maszyna. Jest to jednak fikcja wielce użyteczna, ponieważ pozwala postawić wspomniane fundamentalne pytanie ( $P=NP?$ ).

Klasa  $P$  zawiera się w  $NP$  w tym sensie, że jeśli mamy algorytm wielomianowy dla rozwiązania problemu w sposób ogólny, to posłuży on także w tych wszystkich przypadkach, które sprowadzają się do pytania, czy takie a takie konkretne rozwiązanie jest poprawne. Nie ma natomiast udowodnionego twierdzenia o zawieraniu się  $NP$  w  $P$ . Gdybyśmy je mieli, to z faktu, że np. problem certyfikatu w przypadku komiwojażera jest wielomianowy wynikałby wniosek, że problem komiwojażera w całej ogólności jest wielomianowy czyli należy do  $P$ . Szeroko uznawana jest hipoteza, że odpowiedź na pytanie  $NP \subset P?$  jest negatywna, a więc że  $NP \neq P$ .

W klasie  $NP$  wyróżnia się podzbiór problemów  $NP$ -zupełnych („NP-complete”). Nazywamy problem  **$NP$ -zupełnym**, gdy należy on do  $NP$  oraz ma następującą własność: jeśli dla jakiegoś

problemu NP-zupełnego istnieje algorytm wielomianowy, to istnieje algorytm wielomianowy dla każdego problemu w NP. Wynika stąd, że jeśliby bodaj jeden problem z tej klasy dał się rozwiązać algorytmem wielomianowym, to wobec owej wzajemnej przekształcalności dotyczyłoby to wszystkich pozostałych, a więc zachodziłaby równość  $P=NP$ . Tego rodzaju przekształcenia między algorytmami dokonują się w czasie wielomianowym, dzięki czemu są one algorytmicznie dostępne.

Tak więc, relacje między rozważanymi klasami złożoności rysują się, jak następuje:  $P$  i klasa problemów NP-zupełnych są obie podzbiorami właściwymi klasy NP, zaś między sobą wzajem są (wedle powszechnego mniemania) rozłączne.

W klasie NP-zupełnych znajduje się problem spełnialności, pierwszy rozpoznany pod względem tej własności. Stał się on miarą dostępności dla pozostałych zagadnień z klasy NP: jeśli on dałby się rozwiązać wielomianowo, dotyczyłoby to całej klasy NP. Do NP-zupełnych należy także problem komiwojażera i setki innych zagadnień z wielu dziedzin, jak teoria grafów, badania operacyjne, kryptografia, teoria gier, teoria wyboru społecznego.

**3.3. Niepokonalność NP-problemów** skłania do prac nad rozwiązaniami przybliżonymi. Dokonuje się stratyfikacji tej klasy według stopni złożoności, a więc zabiegów bardziej wyrafinowanych niż podstawowe dystynkcje omawiane wyżej. Rozwija się też teoria NP-zupełności w kierunku rozmaitych zagadnień aproksymacji. Tworzy się w tym celu algorytmy aproksymacji; jak subtelnej teorii wymagają te badania, może świadczyć następujący cytat ze studium [Impag-WWW, s. 2].

"Define SNP to be the class of properties expressible by a series of second order existential quantifiers, followed by a series of first order universal quantifiers, followed by a basic formula (a boolean combination of input and quantified relations applied to the quantified element variables). This class is considered for studying approximability of optimization problems". Autorzy ([WWW-Impag]) powołują się na pozycję Papadimitriou and Yannakakis [1991].

Przykładem na inne sposoby radzenia sobie ze złożonością planowania jest zakodowanie planu w rachunku zdań, a po przyporządkowaniu zmiennym wartości logicznych przetłumaczenie tego znów na wyjściowy problem planowania. Jest to metoda, którą Ernst et al. [1997] opisują na wstępie swego studium, jak następuje.

"Recent work by Kautz et al. [1992] provides tantalizing evidence that large, classical planning problems may be efficiently solved by translating them into propositional satisfiability problems, using stochastic search techniques, and translating the resulting truths assignments back into plans for original problems."

Dowiadujemy się z tegoż tekstu, że stosując opisane przez autorów metody przekształcania formuł rachunku zdań da się, co stwierdzono eksperymentalnie, ograniczyć liczbę zmiennych o połowę, a długość formuł o 80%. To zaś wydatnie upraszcza problemy planowania rekonstruowane w wyniku ich zdekodowania z tak uproszczonych formuł.

Powyższe przykłady wpisują się w ogólną strategię aproksymacji i uproszczeń, w której mogą występować następujące kierunki:

- Gdy problem jest zbyt skomplikowany przy danym modelu matematycznym, upraszczamy model, dbając jednak o to, żeby był on aproksymacją rzeczywistości na tyle bliską, że nie przekreśli to trafności przewidywań.
- Zachowując model bez uproszczeń, co z powodu zbyt dużej złożoności czyni niemożliwym dokładne rozwiązanie, kontentujemy się rozwiązaniem przybliżonym. Ten kierunek jest z powodzeniem reprezentowany przez algorytmy genetyczne, to jest, naśladujące proces darwinowskiej ewolucji w określonej populacji (np. formuł matematycznych czy programów) z jego prawami doboru naturalnego, dziedziczenia cech, walki o byt (giną osobniki nie spełniające zadanych kryteriów) i losowych mutacji (które się utrwalają, gdy prowadzą do spełnienia kryteriów). Algorytmy genetyczne radzą sobie dobrze np. z problemem komiwojażera.

— Szukamy rozwiązania dokładnego, ale bez pewności, czy uda się je uzyskać; wtedy kontentujemy się dostatecznie wysokim prawdopodobieństwem trafności rozwiązania, co wymaga metod szacowania prawdopodobieństwa.

Ograniczenia wyników związane z takimi metodami nie muszą być jakimś istotnym uszczerbkiem poznawczym. W nauce, jak w codziennym życiu wciąż dokonujemy uproszczeń i przybliżeń, nie jest więc czymś zaskakującym, że podlega temu także sfera badań operujących algorytmami. Ma ona tę przewagę nad tradycyjnymi metodami badawczymi, że dadzą się precyzyjnie oszacować odchylenia od precyzji oraz ich konsekwencje poznawcze.

**3.4.** W radzeniu sobie ze złożonością istnieje kierunek odwrotny do takiej przemyślanej zgody na ograniczenia, kierunek zdecydowanie ofensywny. Jego punktem wyjścia jest także uznanie ograniczenia, ale tylko jednego, określającego ramy pozostałych działań, będących już samą ekspansją. Fundamentalnym ograniczeniem, którego nie da się nie respektować jest nierozstrzygalność arytmetyki i logiki. Wiąże się z tym określenie zasięgu mocy algorytmów; ortodoksyjnym stanowiskiem w informatyce jest teza Churcha-Turinga, że każde urządzenie zdolne do algorytmicznego rozwiązywania problemów pokrywa się co do swej mocy z uniwersalną maszyną Turinga (UMT).

Przy tak jasnym określeniu granic możliwości rysuje się pole, na którym można dążyć do przewyższenia UMT. Powstaje. mianowicie, pytanie, czy te same problemy, które ona rozwiązywałaby w czasie nie do przyjęcia długim dałyby się rozwiązywać znacząco szybciej. Odpowiedź jest twierdząca. Powstały różne konkurencyjne względem UMT systemy obliczeniowe, dzięki którym coraz lepiej radzimy sobie ze złożonością problemów. Oto ich przykładowy przegląd.

**Przetwarzanie równoległe** ("parallel computing") zachodzi wtedy, gdy pewien zbiór procesorów wykonuje jedno zadanie, rozdzielone między poszczególne procesory. **Przetwarzanie rozproszone** ("distributed computing") zachodzi wtedy, gdy proces obliczeń jest rozdzielony w pewnym zbiorze komputerów tworzących sieć i wymieniających między sobą dane. Choć różne są w każdym przypadku elementy zbiorów (w jednym są to procesory tego samego komputera, w drugim niezależne komputery), systemy te łączy pewna analogia, co znajduje m.in. wyraz w tytule elektronicznego czasopisma *Journal of Parallel and Distributed Computing*. Jeden i drugi system w oczywisty sposób przyspiesza procesy obliczeniowe. Przetwarzanie rozproszone zasługuje na zbadanie pod kątem tego, na ile nadaje się ono na model wolnego rynku odwzorowujący ten jego aspekt, który Friedrich Hayek nazywał rozproszoną (lub lokalną) wiedzą ekonomiczną w odróżnieniu od scentralizowanej wiedzy wymaganej przez system centralnego planowania.

**Przetwarzanie interaktywne** polega na interakcji systemu z otoczeniem i uczeniu się przez system w wyniku tej interakcji. Istotą uzyskanego usprawnienia jest to, że nie ma potrzeby wyposażania układu w wysoce złożone algorytmy przygotowujące go na wszelkie ewentualności. Zamiast tego jest on wyposażony w program sterujący uczeniem się na podstawie informacji uzyskiwanych od otoczenia, co jest strategią nieporównanie bardziej ekonomiczną. Przykładem takiego systemu jest pocisk samosterujący, który zachowuje się odpowiednio do uzyskanych obserwacji. Takie reakcje na otoczenie wymagają wyposażenia układu w odpowiednie organy (urządzenia wejścia i wyjścia).

**Automaty komórkowe** ("cellular automata"), w skrócie AK (wspomniane wcześniej w opisie dylematu więźnia) nazywają się tak dlatego, że składają się z prostych obiektów zlokalizowanych w komórkach przypominających układ szachownicy. Każdy obiekt ma pewną liczbę możliwych stanów (np. żywy lub martwy; albo, biały, czarny, czy jeszcze w innym kolorze, itd). Obiekty te zmieniają stany, a więc przechodzą jakąś ewolucję, wedle nałożonych na nie reguł, które uzależniają przyjęcie takiego lub innego stanu od sytuacji w otoczeniu (np. ginie na skutek przygniecenia przez obecność dookoła innych obiektów). Jednym z procesów zachodzących w obiektach jest samoreprodukcja. Skonstruowanie automatów zdolnych do reprodukcji samych siebie z użyciem ma-



teriałów znajdujących w otoczeniu było pierwotnym zamysłem von Neumanna przyświecającym mu w projekcie konstrukcji AK. Pomimo prostoty reguł kierujących zachowaniem AK, staje się ono często nieprzewidywalne; pozwala to stosować AK jako modele do symulacji układów niestabilnych (chaotycznych), co w szczególności badał Stephen Wolfram. Potwierdza się też, że AK ma moc UMT (tj. ten sam zakres problemów rozwiązywalnych) przy nieporównywalnie większej sprawności.

**Sieci neuronowe** to fizyczne („hardwarowe”) lub logiczne („softwarowe”) wielce uproszczone imitacje systemu nerwowego. Ich zasadnicza przewaga nad UMT polega na zdolności uczenia się. Różni je też od UMT w sposób istotny to, że ich działanie tylko w części jest cyfrowe, a w części analogowe (co imituje analogowe stany chemiczne w organizmie, np. funkcjonowanie neuroprzekazników). Gdy idzie o porównanie z mocą obliczeniową UMT, to pogląd ortodoksyjny głosi, że nie jest ona większa, większa jest natomiast sprawność i tempo rozwiązywania problemów czyli obliczeń. Istnieje jednak grupa dysydentów, którzy głoszą zasadniczą wyższość sieci, czyli zdolność rozwiązywania problemów nierozwiązywalnych dla UMT.

**3.5.** Wraz z tą kontrowersją przechodzimy do ostatniego punktu w zagadnieniu atakowania złożoności, pominąwszy, dla skrócenia wywodów, parę innych odmian przetwarzania danych, w tym obliczenia kwantowe. Ostatnim punktem do podjęcia stanowi pytanie: czy jest możliwy atak tak frontalny, że zostałaby przekroczona bariera możliwości maszyny Turinga? Przekroczenie to nazywano **hiperkomputacją** ("hypercomputation").

Problem hiperkomputacji rozgałęzia się na dwie kwestie. Jedną zarysowała się (choć jeszcze nie pod tą nazwą) po pojawieniu się twierdzenia Gödla w sprawie nierozstrzygalności arytmetyki i po wynikach roku 1936 dotyczących nierozstrzygalności logiki. Jest to pytanie, czy umysł ludzki ma jakąś zasadniczą przewagę nad maszyną Turinga; na to Gödel skłonny był odpowiadać twierdząco, sam Turing od 1950 przecząco. Odpowiedź przecząca oprócz wariantu Gödlańskiego, w którym umysł pojmowany jest w oderwaniu od fizyki, ma wariant fizyczny, który rozwija Penrose [1989, 1994].

Pogląd Penrose’a został zradykalizowany w ideach i projektach kręgu badaczy, w którym szczególną inicjatywę badawczą i pisarską przejawia Jack Copeland. W tym kręgu funkcjonuje pojęcie hiperkomputacji. Podczas gdy Penrose broni hipotezy, że istnieją w przyrodzie układy zdolne rozwiązywać problemy nieozwiązywalne dla maszyny Turinga, mianowicie mózgi, hiperkomputacjoniści idą dalej, głosząc też możliwość naśladowania przyrody przez konstruowanie urządzeń technicznych przewyższających mocą obliczeniową uniwersalną maszynę Turinga, a więc i układy z nią równoważne (jak komputery cyfrowe).

Hiperkomputacjonizm, paradoksalnie, próbuje nawiązywać do Turinga (jego nazwisko, dopełnione fotografią, wzięto za nazwę witryny internetowej hiperkomputacjonistów). Nawiązanie to zasługuje na uwagę, pozwoli bowiem przedyskutować ważne a zarazem zagadkowe pojęcie **wyroczeni** ("oracle") wprowadzone przez Turinga w pracy [1938].<sup>11</sup> Praca ta skupia na sobie uwagę historyków, ponieważ była rozprawą doktorską Turinga, której promotorem był Alonzo Church; stoją więc za nią dwa wielkie nazwiska twórców problematyki obliczalności.

Ponieważ sprawa jest przedmiotem sporów interpretacyjnych, nie da się jej przedstawić (o ile nie ograniczymy się do cytatów) bez określonej interpretacji. Przyłączam się do stanowiska A. Hodgesa, autora słynnej biografii Turinga [1983], który w odczycie [2002] wrócił do sprawy, polemizując z

<sup>11</sup> Przydać się może ostrzeżenie przed zaistniałą w literaturze dwuznacznością terminu "oracle". Oprócz oryginalnego znaczenia, nawiązującego do Turinga [1938], pojawia się drugie, w kontekście NP-problemu, w którym metafora zgadywania wyniku bywa też ubierana w słowo „wyrocznia” (jako coś, co potrafi odgadywać rzeczy dla innych nieodgadnione). W tym drugim sensie wyrocznia nie ma nic wspólnego z liczbami nieobliczalnymi.

Copelandem. Otóż Turing [1938] podjął próbę sformalizowania pojęcia intuicji, rozumianej jako ten czynnik, który umożliwia rozpoznanie prawdziwości zdania gödłowskiego. Nie podejmując się rozpatrzenia, na ile próba się powiodła i jakie były inne wyniki tej rozprawy, wystarczy odnotować, że Turing wprowadził pojęcie wyroczni jako czegoś, co potrafi obliczać funkcje nieobliczalne, traktując to jednak nie jako hipotezę o istnieniu takiego realnego obiektu (co przypisuje mu Copeland) lecz jako fikcję pomocną w rozważaniach teoretycznych. W tej sytuacji nie jest uprawnione ogłaszanie Turinga patronem superkomputacjonistów, nie powstaje też problem uzgodnienia Turinga [1938] z Turingiem [1950].

Szanse projektu hiperkomputacjonistów zależą od tego, czy prawdą jest pogląd filozoficzny, że istnieją realnie, a nie tylko jawią się nam jako takie, procesy ciągłe czyli nie-dyskretne, do których należą obliczenia analogowe (przeciwnikami tego poglądu są fizycy „dyskrecjoniści”, jak Ed Fredkin czy Frank Tipler). Nie musi być jednak tak, że będzie się bronić projektu na podstawie przesłanek filozoficznych. To filozofia ciągłości może zyskać potwierdzenie eksperymentalne, jeśli uda się skonstruować urządzenie analogowe, o którym się udowodni, że przewyższa mocą obliczeniową maszynę Turinga. Do tej misji pretenduje system Havy Siegelman [1999] opisany przez jego autorkę w książce: *Neural Networks and Analog Computation: Beyond the Turing Limit*. Streszczenie tych idei, zawarte w pewnym odczycie (Siegelman [2000]), tak wyraziście oddaje alternatywę względem poglądu o nieprzekraczalności bariery Turinga, że najlepiej zacytować je dosłownie.

"[...] The theory of computational complexity requires the assumption of discrete computation and does not allow for other types of computational paradigms.

We consider a basic neural network model: finite number of neurons, recurrent interconnection, continuous activation function, and real numbers as weights. This model is considered "analog" for both the use of real numbers as weights which makes the phase space continuous, and the continuity of its activation function. In computational terms, this type of continuous flow (due to its activation function) is definitely a restriction on our model. There are no discrete flow commands such as "If  $z$  is greater than 0, compute one thing, otherwise continue in another computation path".

We show that the network [...] can compute anything that a digital machine does. Furthermore, due to its real-valued weights, it is inherently richer than the digital computational model and can compute functions which no digital computer can. The network, although transcending the Turing model, is still sensitive to resource constraints and still computes only a small subclass of all possible functions. It thus constitutes a well-defined super-Turing computational model."

Jak widać z tego opisu, sieć Siegelman ma się lokować co do mocy obliczeniowej między UMT a wyrocznią w sensie Turinga [1938]. Czy sieć taka istotnie da się dobrze zdefiniować jako obiekt teoretyczny i czy doświadczalnie wykaże, jako urządzenie fizyczne, swą wyższość nad UMT, pozostaje sprawą otwartą (jak można wnosić, śledząc toczącą się w tej sprawie dyskusję).

**3.6.** Nim znając jeszcze rozstrzygnięcia rzeczonego sporu, możemy w pewnym zakresie polegać na naszej wiedzy o nas samych jako ludziach, konfrontując ją z wiedzą o komputerach. Zasługują na szczególną uwagę dwie ludzkie dyspozycje: zdolność do przyjmowania aksjomatów i zdolność do stawiania pytań. Wykorzystując te dyspozycje, a zarazem zlecając maszynom cyfrowym zadania, w których są one od nas lepsze, możemy najsukuteczniej maksymalizować wyniki badawcze. Mówiąc w przenośni, jeśli te dwie dyspozycje zaliczyć do funkcji pełnionej przez wyrocznię w sensie Turinga [1938], to optymalna strategia badawcza polega na kooperacyjnym dodatnim sprzężeniu zwrotnym między wyrocznią i maszyną.

Co do zdolności znajdowania aksjomatów i rozpoznawania ich prawdziwości, to jest to punkt notorycznie pomijany w dyskusjach o SI, jeśli nie brać pod uwagę wzmianek czynionych jakby „na boku”. Zdaje się to świadczyć, że nawet najwięksi obrońcy tezy o redukcji umysłu ludzkiego do UMT nie wyobrażają sobie jak dotąd procedury algorytmicznej prowadzącej do odkrycia np. ak-

sjomatu wyboru.<sup>12</sup> Pomijają tę sprawę rzecznicy silnej SI, choć ciążyłby na nich onus probandi, że wszelkie aksjomaty powinna wyprodukować maszyna Turinga. Wprawdzie można by sobie wyobrazić produkcję aksjomatów przez algorytmy genetyczne, stosujące m.in. wymóg niesprzeczności jako probabilistyczne kryterium ewolucyjne. Ostatecznie jednak chodzi o to, żeby aksjomaty były prawdziwe, a rozpoznawanie prawdy to domena poznawcza umysłu ludzkiego.

Drugim zdaniem egzaminem SI byłoby skonstruowanie maszyny zdolnej do stawiania pytań. Wokół tej sprawy istnieje w kręgach badaczy SI cisza podobna do tej, jaka otacza kwestię dochodzenia do aksjomatów. Cisza tym bardziej zastanawiająca, że wobec popularności tzw. testu Turinga wielu ludziom powinien by być nasunąć się pomysł takiego jego rozwinięcia, żeby bliskość umysłu i maszyny mierzyć nie tylko (jak proponował Turing) podobieństwem co do trafności odpowiedzi na zadawane przez człowieka pytania, lecz także podobieństwem co do zasadności, oryginalności i wnikliwości stawianych pytań.

Dopóki maszyna taka nie powstanie, dorównanie inteligencji naturalnej przez sztuczną pozostanie poza horyzontem marzeń. Do udzielania odpowiedzi wystarczy mieć wpisana do pamięci bazę danych i program do ich logicznego przetwarzania, podczas gdy do stawiania pytań niezbędna jest, jak to zauważył Ch. S. Peirce, zdolność do irytacji czyli do rozdrażnienia. Irytacji z powodu własnej niewiedzy, niejasności pojęć, czy napotkanej gdzieś sprzeczności. Albo (uzupełnijmy myśl Peirce'a), w innych sytuacjach, trzeba ciekawości. Jedno i drugie jest stanem umysłu nieodłącznym od pewnego stanu emocjonalnego, do którego – na ile nam wiadomo – zdolne jest żywe białko, a nie są zdolne kostki krzemu czy inne materialne nośniki sztucznej inteligencji (o tym, że system nerwowy ma swoistą logikę nieosiągalną dla maszyny cyfrowej, pisał von Neumann [1958]; zob. komentarz Marciszewskiego [1996b]).

Powyższe uwagi służą do wzmocnienia postawionej wyżej tezy, że optymalna strategia badawcza przyszłości nie będzie polegać na sukcesywnym zastępowaniu ludzi przez komputery lecz na równoczesnym wzroście potencjału jednej i drugiej strony. Możliwości obliczeniowe komputerów rosną w miarę tworzenia coraz szybszych algorytmów i doskonalenia sprzętu, a przyspieszony w wyniku tego wzrost wiedzy przyspiesza proces narastania u ludzi nowych pytań i nowych aksjomatów, co prowadzi do nowych zadań badawczych dla komputerów. I tak dalej, in infinitum — o ile ktoś nam zagwarantuje nieskończony czas istnienia żywej i mechanicznej inteligencji.

## Literatura cytowana

Ajdukiewicz, Kazimierz, *Logika pragmatyczna*, PWN, Warszawa 1965.

[Buss-WWW] Buss, R. Samuel, *On Gödel's Theorem on the Lengths of Proofs II: Lower Bounds for Recognizing Symbol Provability*, [math.ucsd.edu/~sbuss/ResearchWeb/publ.ps](http://math.ucsd.edu/~sbuss/ResearchWeb/publ.ps)

<sup>12</sup> O tym, jak uwaga logiki i teorii złożoności koncentruje się na problemach dowodzenia z aksjomatów, a nie znajdowania aksjomatów, świadczy pewien szczegół w znanym liście Gödla do von Neumanna z roku 1956, prekursorskim względem teorii złożoności obliczeniowej i problemu „P=NP?”. Gödel zapytuje von Neumanna, co sądzi o pewnym algorytmie, który jego zdaniem mógłby w pełni wyręczyć matematyka (cytowane w przekładzie z niemieckiego na angielski za [Buss-WWW]). "It [Ten wynik] would obviously mean that in spite of the undecidability of the Entscheidungsproblem, the mental work of a mathematician concerning Yes-or-No questions could be completely<sup>2</sup> replaced by a machine." Przypis 2 po słowie „completely” brzmi (akcent kursywą WM): *except for the setting up of axioms*. Potraktowanie sprawy tylko w przypisie ręcznie pisanego tekstu pozwala przypuszczać, że niemożność uzyskania aksjomatów na drodze algorytmicznej była dla Gödla i zapewne (w przekonaniu Gödla) dla jego adresata poza dyskusją. Pełny tekst listu i komentarz historyczny daje Hartmanis [1989].

Carnap, Rudolf, *Meaning and Neceesity*, University of Chicago Press, 1958.

Chaitin, Gregory J., *The Limits of Mathematics, The Unknowable, Exploring Randomness, Conversations with a Mathematician*, Springer-Verlag, Berlin etc. 1998, 1999, 2001, 2002.

Church, Alonzo, „An unsolvable problem of elementary number theory”. *Am. J. Math.* 58, 345-363, 1936.

Church, Alonzo, „A note on the Entscheidungsproblem”, *J. of Symbolic Logic* 1, 40-41 and 101-102, 1936.

Copeland, Jack and Proudfoot, Diane, *Introduction to Hypercomputation: Computing the Uncomputable*, Copyright by the authors, July 2000.

[www.alanturing.net/turing\\_archive/pages/Reference](http://www.alanturing.net/turing_archive/pages/Reference)

Copeland. Jack, *Turing's O-Machines of 1938*, May 2000.

[www.alanturing.net/turing\\_archive/pages/Reference%20Articles/hypercomputation/](http://www.alanturing.net/turing_archive/pages/Reference%20Articles/hypercomputation/)

Allin Cottrell and W. Paul Cockshott. „Calculation, Complexity and Planning: The socialist calculation debate once again”. / [www.dcs.gla.ac.uk/wpc/reports/](http://www.dcs.gla.ac.uk/wpc/reports/) / Por. tych samych autorów: „Socialist Planning after the Collapse of the Soviet Union”, *Revue European des Sciences Sociales*, 167-186. Publisher: Academic Press, 1993.

Davies, Paul, *The Mind of God* (rozdział 5, odcinek „Unknowable”, 128-134), Simon and Schuster, New York 1992.

Grim, Patrick „The Undecidability of the Spatialized Prisoner's Dilemma,” *Theory and Decision* 42, 53-80, 1997. Zob. też [www.sunysb.edu/philosophy/faculty/pgrim/SPATIALP](http://www.sunysb.edu/philosophy/faculty/pgrim/SPATIALP)

Ernst, Michael D.; Millstein Todd D.; Weld, Daniel S., „Automatic SAT-Compilation of Planning Problems”, *Proc. the 15th Intern. Joint Confer. on Art. Intell.*, Nagoya, Japan, August 23-29, 1169-1176, 1997. [pag.lcs.mit.edu/mernst/pubs/satcompile-ijcai97-abstract.html](http://pag.lcs.mit.edu/mernst/pubs/satcompile-ijcai97-abstract.html)

Gödel, Kurt, „Die Vollständigkeit der Axiome des logischen Funktionenkalküls”, *Monatshefte für Mathematik und Physik* 37, 349-360, 1930

Gödel, Kurt, „Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme – I”, *Monatshefte für Mathematik und Physik* 38, 173-198, 1931.

Gödel, Kurt, Über die Länge der Beweisen, *Ergebnisse eines mathematischen Kolloquiums* Heft 7,, Franz Deuticke, Leipzig und Wien 1936.

Hartmanis J. and Stearns R., „On the computational complexity of algorithms”, *Transactions of the AMS* 117, 285-306, 1965.

Hartmanis J., „Gödel, von Neumann and the P=?NP problem”, *Bulletin of the European Association for Theoretical Computer Science* 38, 101-107, 1989.

Heller, Michał, *Wszechświat u schyłku stulecia*, Znak, Kraków, 1994.

Hilbert D. und Ackermann W., *Grundzüge der theoretischen Logik*, Springer, Berlin 1928.



Hodges, Andrew, *What would Alan Turing have done after 1954? Lecture at the Turing Day*, Lausanne, 2002, [www.turing.org.uk/philosophy/lausanne2.html](http://www.turing.org.uk/philosophy/lausanne2.html)

Impag-WWW] Impagliazzo, Russell; Paturi, Ramamohan, and Zane, Francis, *Which Problems have Strongly Exponential Complexity?* (Research supported by a NFI grant). [cm.bell-labs.com/cm/ms/who/francis/papers/focs98-subexp.pdf](http://cm.bell-labs.com/cm/ms/who/francis/papers/focs98-subexp.pdf)

Kautz H. and Selman B, „Planning as satisfiability”, *Proc. 10th Eur. Conf. on AI*, 359-363, Wiley, Vienna 1992.

Marciszewski, Witold Murawski, Roman, *Mechanization of Reasoning in a Historical Perspective*, Editions Rodopi, Amsterdam 1995.

Marciszewski, Witold, „Leibniz’s Two Legacies. Their Implications for Knowledge Engineering”, *Knowledge Organization* vol. 23, no. 2, 1996-a.

Marciszewski, Witold, „Hätte Leibniz’s von Neumanns logischen Physikalismus geteilt?”, *Beiträge zur Geschichte der Sprachwissenschaft*, 6.2, 1996-b.

Neyman, A., „Bounded complexity justifies cooperation in the finitely repeated prisoners’ dilemma”, *Economic Letters* 19, pp. 227-229, 1985.

Papadimitriou C. and Yannakakis M., „Optimization, approximation and complexity classes”, *Journal of Computer and System Sciences*, 43, 425-440, 1991.

Penrose, Roger. *The Emperor’s New Mind: Concerning Computers, Minds, and the Laws of Physics*, Oxford Univ. Press, Oxford etc. 1989.

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

Siegelmann, Hava T., *Neural Networks and Analog Computation: Beyond the Turing Limit*, 1999.

Siegelman, Hava T., „Computing Beyond Digital Computers by Neural Models” Odczyt na konferencji *Hypecomputation Workshop*, 2000, University College, London.  
[www.alanturing.net/turing\\_archive/conference/hyper/abstracts.html](http://www.alanturing.net/turing_archive/conference/hyper/abstracts.html)

[Sirridney-WWW] <http://www.sirrodney.nu/public/english/Misc>

Tarski, Alfred; Mostowski, Andrzej; Robinson, Raphael M.: *Undecidable Theories*, North Holland, Amsterdam, 1968.

Turing, Alan, „On computable numbers, with an application to the Entscheidungsproblem”, *Proc. of the London Math. Society*, Series 2, 42, pp. 230-265, 1936.

Turing, Alan, „Systems of logic based on ordinals”, *Proc. of the London Math. Society*, Series 2, 45, pp. 161-228, 1939. Cytowane jako Turing [1938].

Turing, Alan „Intelligent machinery: Raport, National Physics Laboratory”. In: B. Meltzer and D. Michie (eds), *Machine Intelligence*, 5, Edinburgh Univ. Press, Edinburgh 1969.

Turing, Alan, „Computing machinery and intelligence”, *Mind* 59, 433-460, 1950.

von Mises, Ludwig, *Human Action. A Treatise on Economics*, Contemporary Books, Chicago, 1966.

von Neumann, John, *The Computer and the Brain*, Yale Univ. Press, New Haven 1958.

Wolfram, Stephen: „Undecidability and Intractability in Theoretical Physics”, *Physical Review Letters* 54, pp. 735-738, 1985. Dostępne też pod adresem:

[www.stephenwolfram.com/publications/articles/physics/85-undecidability/index.html](http://www.stephenwolfram.com/publications/articles/physics/85-undecidability/index.html)

Wolfram, Stephen: *Cellular Automata and Complexity*, Addison-Wesley, 1994.

Wolfram, Stephen: *A New Kind of Science*, Wolfram Media, Inc., 2002.

---

## Streszczenie

Temat artykułu jest skonstruowany na wzór tytułu znanej publikacji S. Wolframa *Undecidability and Intractability in Theoretical Physics*. Nazwy rozważanych własności oddaje się terminami „nierozstrzygalność” i „niedostępność algorytmiczna”, a człon „fizyka teoretyczna” zastępuje się członem „nauki społeczne”. Celem artykułu jest włączenie tej problematyki do kanonu metodologicznego nauk społecznych. O takiej możliwości świadczy literatura logiczno-informatyczna dotycząca m.in. podstawowego dla nauk społecznych modelu, jakim jest dylemat więźnia; przynosi ona w tej sprawie ważne wyniki limitatywne.

Drugim celem (wychodzącym poza naśladowanie Wolframa) jest zbadanie, jak włączenie wymienionej problematyki miałoby się do tradycyjnego programu badawczego socjologii rozumiejącej (Max Weber i in). Odpowiedź nawiązuje do idei L. von Misesa o dominującej roli poznania apriorycznego w naukach społecznych, co implikuje możliwość przybliżania się teorii społecznych do postaci aksjomatycznej. Potencjalne aksjomaty (postulaty znaczeniowe w sensie Carnapa i Ajdukiewicza) uzyskuje się dzięki intuicji czyli rozumieniu (pojęcie wiązane w artykule z informatyczną koncepcją wyroczeni), a pozostałe twierdzenia teorii — przy pomocy algorytmów użytych do modelowania i symulacji. Wyniki limitatywne mają tu rolę kluczową, wskazując na potencjalne aksjomaty (krok podobny do potraktowania zdania gödłowskiego jako nowego aksjomatu).

## Konspekt

**Wprowadzenie:** czy nierozstrzygalność i algorytmiczna niedostępność przenosi się z matematyki na nauki empiryczne, w tym społeczne, jako czerpiące z matematyki algorytmy do modelowania i symulacji?

### 1. Stan zagadnienia i kierunki rozwiązań.

1.1. Sformułowanie zagadnień rozstrzygalności i dostępności w naukach społecznych oraz zagadnień od niego pochodnych.

1.2. O braku zainteresowania dla w/w zagadnień, w szczególności dla wyników limitatywnych, w praktyce modelowania i symulacji w naukach społecznych.

1.3. Problemy rozstrzygalności i algorytmicznej dostępności modeli matematycznych w naukach społecznych na przykładzie wyników limitatywnych dotyczących dylematu więźnia.

1.4. Określenie klasy zdań apriorycznych w teorii empirycznej w nawiązaniu do wyników limitatywnych i do stanowiska L. von Misesa.

1.5. Podstawowe informacje historyczne.

## **2. Idea racjonalności i pojęcie inteligencji w naukach społecznych.**

2.1. Wzajemnie zbliżanie się pojęć inteligencji i racjonalności w naukach społecznych i w teorii inteligencji.

2.2. Spór między teorią socjalistyczną w ujęciu Langego i teorią rynku w ujęciu Hayeka jako pole do wypróbowania pojęć z zakresu złożoności obliczeniowej.

## **3. Atakowanie złożoności**

3.1. Czas wielomianowy a czas wykładniczy.

3.2. Algorytmy wielomianowe (P), niedeterministyczne wielomianowe (NP), NP-zupełne. NP-trudne. Problem  $P=NP$ ?

3.3. Zaradzić złożoności przez zmniejszenie trudności problemu.

3.4. Zaradzić złożoności przez efektywniejsze środki obliczeniowe.

3.5. Spór o możliwość hiperkomputacji czyli o przekroczenia bariery maszyny Turinga za pomocą wyroczni.

3.6 Maksymalizacja wyników badawczych przez interakcję rozumienia (wyroczni) z algorytmem. Miejsce dla socjologii rozumiejącej w świetle wyników limitatywnych.