

## *Replikowalność wiedzy a transhumanistyczny stan osobliwości*

Paweł Kawalec  
Katedra Metodologii Filozofii KUL  
pawel.kawalec@kul.pl

Zwolennicy transhumanizmu upatrują przyszłego podmiotu wiedzy w maszynach, które miałyby być inteligentniejsze od ludzi, ich twórców. Jeżeli raz zaistniałaby taka możliwość stworzenia inteligentniejszych podmiotów, to w kolejnym kroku maszyny inteligentniejsze od ludzi stworzą następne pokolenie maszyn bardziej inteligentne od samych siebie. To zapoczątkuje dalsze etapy procesu gwałtownego wzrostu inteligencji aż do osiągnięcia stanu osobliwości (ang. *singularity*).

W niniejszym tekście omówione są wątpliwości dotyczące drugiego z kluczowych stwierdzeń uzasadniających transhumanistyczną tezę o osobliwości, a mianowicie gwałtownego wzrostu inteligencji. Te wątpliwości dotyczą możliwości replikowalności wiedzy danego podmiotu poznawczego jako warunku gwałtownego wzrostu. Inspiracją tych wątpliwości jest program epistemologiczny, który mógłby wydawać się wzorcowym protoplastą programu transhumanizmu, a mianowicie program epistemologiczny R. Carnapa, którego ucieleśnieniem jest tzw. robot Carnapa. W ujęciu synchronicznym zasadniczo taka replikowalność ograniczałaby się wyłącznie do strukturalnych własności wiedzy, co nie daje wystarczającej podstawy do stwierdzenia gwałtownego wzrostu. W ujęciu diachronicznym, możliwości dynamicznego zwiększania zasobów wiedzy robota Carnapa są w sposób nieusuwalny uwarunkowane przyjętymi subiektywnie schematami, a wybór pomiędzy nimi dokonuje się pragmatycznie. W tym przypadku stawianie pytania o wzrost jest bezpodstawne.

### **1. Transhumanistyczny stan osobliwości i jego uwarunkowania**

D. Chalmers (2010, 7),<sup>1</sup> podejmując tematykę stanu osobliwości zwięźle scharakteryzował go następująco:

---

<sup>1</sup> Chalmers jest także autorem najobszerniejszego współczesnego komentarza do *Aufbau* (2010).

Co stanie się, gdy maszyny będą inteligentniejsze od ludzi? Jeden z poglądów utrzymuje, że po takim wydarzeniu nastąpi gwałtowny wzrost (ang. *explosion*) coraz to wyższych poziomów inteligencji, gdyż każde pokolenie maszyn będzie tworzyć z kolei maszyny inteligentniejsze od samych siebie. Ta eksplozja inteligencji jest często określana jako „osobliwość”.

Teza o osiągnięciu stanu osobliwości w odniesieniu do pozaludzkiej inteligencji zakłada zatem dwie podstawowe przesłanki. Po pierwsze, możliwość stworzenia maszyn inteligentniejszych od ludzi. Po drugie, gwałtowny wzrost inteligencji, który stanowić ma proces analogiczny do gwałtownego wzrostu mocy obliczeniowej maszyn tworzonych obecnie przez ludzi.

Oprócz tych dwóch założeń, szczegółowo dyskutowanych przez Chalmersa, ważne w dalszej argumentacji okaże się jeszcze jedno. Proces osiągania stanu osobliwości milcząco zdaje się zakładać, że ze względu na wzrastający eksponencjalnie proces złożoności technologicznej możliwe stanie się odtwarzanie wiedzy posiadanej przez określoną maszynę przez inną maszynę tego samego lub następnego pokolenia. Określimy je mianem założenia o replikowalności.<sup>2</sup>

Założenie o replikowalności jest także niezależne od problemu niejednorodności poziomu inteligencji, jaki może wystąpić między inteligencją ludzką a inteligencją maszyn. Ze względu na różnicę poziomów zdolności poznawczych w obu przypadkach może okazać się, że wymagane byłyby inne miary:

Nie jest sprawą trywialną to, że miary inteligencji, które są dobrze skorelowane z określonymi zdolnościami poznawczymi ludzi, będą też dobrze skorelowane z takimi zdolnościami sztucznych systemów. (Chalmers 2010, 24)

---

<sup>2</sup> Założenie o replikowalności jest w oczywisty sposób niespełnione w przypadku ludzi. W przypadku maszyn prowadzi się takie badania co najmniej od lat 1970. (Charron-Bost i in. 2010). W obecnie prowadzonych badaniach chodzi jednak o inaczej sformułowany problem replikowalności, np. baz danych przechowywanych na różnych serwerach, niż sformułowany w odniesieniu do stanu osobliwości. Struktura baz danych i przedmiotów traktowana jest jako *constans* w tych badaniach, co więcej na ogół nie jest ona również generowana przez sztuczne systemy, których wiedza jest replikowana.

W dalszej części tekstu problem replikowalności wiedzy dyskutowany jest w aspekcie synchronicznym, przy założeniu, że dysponujemy jakimś zastanym stanem wiedzy danej maszyny w zasadzie określonym deterministycznie. W aspekcie diachronicznym dalsza komplikacja polega na dopuszczeniu możliwości niedeterministycznego rozbudowywania wiedzy, która ma być przedmiotem replikacji.

Jak podkreśla Chalmers (2010, 10), niezależnie od problemów teoretycznych osobliwość wydaje się być stanem godnym pożądania ze względu na oczekiwane pozytywne skutki. W zasadzie nieograniczona inteligencja otwiera bowiem możliwość uzyskania odpowiedzi na wszystkie pytania, które wymykały się ograniczonym ludzkim umysłom: „Gwałtowny wzrost inteligencji wiąże się z ogromnymi możliwymi korzyściami: lekarstwem na wszystkie znane choroby, zakończeniem ubóstwa, nadzwyczajnym postępem naukowym i wieloma innymi.” Jak dodaje Chalmers, „Ale niesie też ogromne potencjalne zagrożenia: koniec rasy ludzkiej, powstania rasy maszyn wojennych, mocy zniszczenia planety.”

Dyskusja założenia o replikowalności poprzedzona jest rysem najbardziej tu istotnych elementów epistemologicznego stanowiska Carnapa,<sup>3</sup> które, idąc za (Kawalec 2003), traktowane jest tu jako zasadniczo spójne, mimo zmian w jego poglądach w okresie wczesnym i późniejszym.

## **2. Nieograniczony postęp poznawczy**

Carnap wydaje się być, znacznie przed ruchem transhumanistów, piewą zasadniczo nieograniczonego postępu ludzkiej wiedzy. Najczęściej ilustruje się to przekonanie obszerniejszym fragmentem jego *Przedmowy do Aufbau*:

Jednostka nie podejmuje się już wzniesienia całego gmachu filozofii w jednym zamasytym akcie. Raczej każdy pracuje w określonym miejscu w obrębie zjednoczonej nauki (*Gesamtwissenschaft*). [...] można będzie pewniej spoglądać w przyszłość: w powolnej i ostrożnej konstrukcji uzyskamy jedno poznanie po

---

<sup>3</sup> Pełniejszą i odniesioną historycznie prezentację jego stanowiska epistemologicznego podaje (Kawalec 2011, XI-CIV).

drugim. Każdy wniesie tylko to, za co będzie odpowiedzialny i co będzie mógł uzasadnić przed wszystkimi współpracownikami. I tak, starannie dokładać się będzie kamień do kamienia i wznosić pewniejszy gmach, którego budowę mogą kontynuować kolejne pokolenia. [...] Odczuwamy wewnętrzne pokrewieństwo postawy leżącej u podstaw naszego uprawiania filozofii z postawą duchową, która występuje obecnie w innych dziedzinach życia. Wyczuwamy tę postawę w nurtach sztuki, zwłaszcza w architekturze i w ruchach, które poszukują znaczących kształtów ludzkiego życia osobistego i zbiorowego, wychowania i w ogóle zewnętrznych struktur. Wszędzie tam wyczuwamy przede wszystkim tę samą postawę, ten sam styl myślenia i działania. Jest to orientacja, która przede wszystkim zmierza do jasności, zdając sobie jednak przy tym sprawę, że materia życia nigdy nie zostanie poznana na wskroś. Dąży do dokładności w szczegółach, a jednocześnie do uchwycenia wielkich linii biegnących przez całość. Uwzględnia więzi, które łączą ludzi, lecz pozwala jednocześnie na swobodny rozwój każdej jednostki. Wiara, że przyszłość należy do tej orientacji, daje nam siłę do pracy. (Carnap 1928/2011, CXIII-CXIV)

Wiele komentarzy do prac R. Carnapa, zwłaszcza *Logicznej struktury świata* (w skrócie *Aufbau*; Carnap 1928/2011) i tych z zakresu logiki indukcji, wprost podejmuje zagadnienie „matematyzacji” czy „robotyzacji” wiedzy w jego poglądach. Strukturalistyczne ujęcie<sup>4</sup> poznania ludzkiego wraz z tezą o jedności nauki prowokuje pytania o transhumanistyczną podmiotowość tej wiedzy (Spielman 1976; Moulines 1981; Chalmers 2010).<sup>5</sup>

A. Richardson (2003, 174), charakteryzuje stanowisko epistemologiczne Carnapa przeciwie do dominujących interpretacji, podkreślając wpływ metod naukowych na kształtowanie jego metody filozoficznej i strukturalizmu:

---

4 Rolę w specyficznie rozumianym strukturalizmie J. Sneed'a omawia (Holger 2010).

5 Historyczny wkład Carnapa w rozwój informatyki i sztucznej inteligencji prezentuje (Bauer 2010, 73).

Carnapa znajomość nowych metod stosowanych w czystej i stosowanej geometrii doprowadziła do jego swoistych ujęć zasadniczego problemu epistemologicznego i przekonania, że logika była właściwym narzędziem do rozwiązania go.

W kształtowaniu się poglądów strukturalistycznych Carnapa zasadniczą rolę odegrały trzy tematy (Richardson 2003, 175):

sposób rozumienia matematyki jako teorii różnorodności czy struktur doprowadził do szczególnego sposobu rozumienia ogólności i obiektywności matematyki; a to z kolei wpłynęło na sposób ugruntowania rzeczywistej przestrzenności geometrii [...], a także sposób w jaki pojęcia niezmienniczości, współzmienności i odwzorowań wpłynął na sposób ujęcia obiektywności fizycznej.

Ten silny wpływ narzędzi matematycznych na sposób filozofowania Carnapa w efekcie doprowadził do tego, że jego strukturalizm miał w zasadzie charakter matematyczny (2003, 176): „to właśnie struktury matematyczne pozwalały na determinację lokalizacji indywidualów w całkowicie zdeterminowanych strukturach”. A ta struktura matematyczna miała charakter w zasadzie immanentny i samowystarczalny. Stąd brała się „Zasadnicza idea, że obiektywność jest zawsze jednoznaczną determinacją w obrębie struktury relacji” (Richardson 2003, 178).

Jak podkreśla M. Friedman (2011, 258) pozwalało to Carnapowi na skonstruowanie stanowiska, określanego tu jako „neutralizm”, które unikało trudności przeciwstawnych stanowisk prezentowanych w filozofii nauki – zarówno instrumentalizmu, jak i realizmu naukowego.

Strukturalizm prowadził w ten sposób do połączenia z charakterystyczną, „logiczną”, wersją pragmatyzmu<sup>6</sup> w poglądach Carnapa:

Sądzę, że pytanie nie powinno być stawiane w ten sposób: “Czy byty teoretyczne są rzeczywiste?”, ale w ten sposób: “Czy powinniśmy preferować język fizyki (i

---

<sup>6</sup> W przypadku większości reprezentantów logicznego empiryzmu, jak pokazuje (Scott 2009), poglądy pragmatystyczne stanowiły odzwierciedlenie stopniowej ewolucji, której początkiem były wizjonerskie poglądy polityczne i plany reform społecznych na dużą skalę.

nauki w ogóle), który zawiera terminy teoretyczne, czy język bez takich terminów?” Z tego punktu widzenia zagadnienie sprowadza się do preferencji i praktycznej decyzji. (Carnap 1974, 256)

Pragmatyczne „usunięcie” problemu, stojącego u podstaw opozycji instrumentalizm vs realizm, w odniesieniu do wiedzy naukowej pozostawiało jedynie zadanie weksplikowania logicznej (w sensie *Wissenschaftslogik*) struktury wiedzy naukowej (2011, 261).

Mimo więc sugerowanego przez przytoczony fragment *Przedmowy* nieograniczony charakter postępu wiedzy naukowej, ogranicza się ona po pierwsze do wiedzy o własnościach strukturalnych przedmiotów oraz po drugie, jest uwarunkowana w istotnych względach pragmatycznymi decyzjami człowieka.<sup>7</sup> Jak zobaczymy w dalszej części tekstu, ma to poważne konsekwencje dla oceny założeń twierdzenia o stanie osobliwości.

### 3. Robot Carnapa zmierza do osobliwości?

W literaturze przedmiotu można spotkać wprost wyrażone interpretacje Carnapa koncepcji wiedzy, które jako jej uprawniony podmiot wskazują robota, „myślącą maszynę”.

U. Moulines odnosi się do konstytucyjnej teorii pojęć przedstawionej w *Aufbau* i formułuje taką interpretację, powołując się na stwierdzenie samego Carnapa:

Carnapa „konstruktor świata” może być rozumiany jako model maszyny obliczeniowej, która reprezentowałaby idealny podmiot epistemiczny, działającej na podstawie minimalnej pojęciowej bazy dla wiedzy spostrzeżeniowej. Sam Carnap dostarczył inspiracji do tego sposobu patrzenia na swoje wczesne dzieło. W swojej *Biografii intelektualnej* zasugerował, że „definicje” występujące w

---

<sup>7</sup> Przyjęty przez Carnapa woluntaryzm epistemologiczny wprost zakorzeniony jest w tradycji pragmatyzmu amerykańskiego; zob. (Kawalec 2010).

systematycznej części *Aufbau* powinny być rozumiane jako „reguły postępowania w odniesieniu do procedury konstruowania, które stosują się do kogokolwiek, nawet Kantowskiego podmiotu transcendentalnego czy maszyny obliczeniowej” (1991, 265).

Przy takiej interpretacji nieredukcyjny charakter przedsięwzięcia Carnapa staje się dość oczywisty. „Definicje” przedmiotów wyższych stopni podawane za pomocą pojęć dotyczących przedmiotów niższych stopni, traktowane przez W. Quine’a jako ich redukcje, są tutaj regułami postępowania. Określają mianowicie sposób, w jaki należy dokonać przyporządkowania pojęć wyższych stopni pojęciom wygenerowanym na podstawie tej „minimalnej bazy”:

Baza pojęciowa takiego podmiotu epistemicznego powinna być najprostsza, aby możliwie uprościć jego kontrolę [nad całą procedurą]. Zgodnie z tą interpretacją celu *Aufbau*, definicje pojawiające się w formalnej części książki, zwłaszcza w rozdziale IV/A (konstytucja pojęć dotyczących własnych przedmiotów psychicznych – „das Eigenpsychische”), zawierającej jedyną sformalizowaną część systemu, nie powinny być rozumiane jako „łańcuch redukcji”, zamierzony do skompresowania wszystkich pojęć empirycznych w minimalnej bazie fenomenalnej, lecz jako reguły postępowania sformułowane dla idealnego obserwatora, który jest „zaprogramowany” w taki sposób, aby rozpoczął od minimalnej struktury i wystarczająco bogatego i zróżnicowanego zbioru przeżyć pierwotnych („surowy materiał”) dla skonstruowania „pojęć obserwacyjnych”, które są niezbędne do sprawdzenia każdego naukowego stwierdzenia. (Moulines 1991, 265)

Taka interpretacja definicji podanych w *Aufbau* umożliwia uniezależnienie pojęć wyższych stopni od ich fenomenalnej i subiektywnej genezy w doświadczeniu.<sup>8</sup> A co za tym idzie, uniezależnia je także od rodzaju podmiotu tego doświadczenia:

---

<sup>8</sup> W późniejszym tekście Moulines (2001) wskazuje na historyczne źródła tego stwierdzenia Carnapa.

[...] te reguły korespondencji pozwalają rzekomemu idealnemu obserwatorowi powiązać dowolne zaproponowane stwierdzenie nauk fizycznych czy społecznych z określoną konfiguracją przeżyć konstruowalnych na własnej bazie fenomenalnej. (Moulines 1991, 266)

Zaproponowana przez Moulinesa interpretacja *Aufbau* wydaje się otwierać drogę do stanu osobliwości, o którym pisze Chalmers, pod warunkiem, że możliwa byłaby „eksplozja” tej wiedzy w przypadku osiągnięcia podmiotów o inteligencji wyższej niż ludzka.

Taką eksplozję zdaje się umożliwiać indukcyjna teoria uczenia się, którą Carnap zaproponował w późniejszych pracach z zakresu logiki indukcji. S. Spielman (1976) charakteryzując specyfikę reguł wnioskowania indukcyjnego, jakie zaproponował Carnap odwołał się do metafory, jaką posłużył się ten ostatni (1971, 17-21). Te reguły są „filozoficznie interesujące” tylko z uwagi na to, że nie są prostą konsekwencją probabilistycznie rozumianej koherencji przekonań, co wiąże się z przyjęciem aksjomatów prawdopodobieństwa jako minimalnych wymogów racjonalności (Kawalec 2011a). Ponadto:

[...] reguły Carnapa nie mają szerokiego zasięgu – one nie obejmują tego, co określał on jako „dorosłe” stopnie pewności. Można traktować je jako reguły rządzące funkcją przeświadczenia robota, mającego doskonałą pamięć, organy zmysłowe, zdolność do przetwarzania danych oraz podejmowania decyzji takie, jakie posiadają ludzie, wiedzę logiczną i matematyczną, zdolność do rozumienia formalnych języków pierwszego rzędu, lecz który całkowicie nie zna faktów empirycznych. Te zasady dotyczą nas tylko w odniesieniu do takich kontekstów problemowych, w których brak naszego odnośnego doświadczenia jest podobne temu, jaki ma robot Carnapa. Spielman (1976, 407)



Mimo zastrzeżeń, dotyczących interesujących przypadków, w których można spotkać takie sytuacje, Spielman samo zagadnienie reguł uznaje za doniosłe dla zrekonstruowania pojęcia wiedzy empirycznej.

#### 4. Założenie o replikowalności wiedzy w aspekcie synchronicznym

Jak wspomiane zostało to wyżej, gwałtowny wzrost inteligencji uwarunkowany jest tym, że wiedza osiągnięta przez jedno pokolenie maszyn będzie replikowana przez następne pokolenie i późniejsze. Gdyby ten warunek nie był spełniony, nie byłoby też miary do kategoryzowania wzrostu inteligencji.

Założenie o replikowalności rozważymy kolejno w perspektywie solipsystycznej, intersubiektywnie i międzypokoleniowo. Można najpierw zapytać, czy replikacja wiedzy przez daną maszynę (inteligentny system) jest zależna od tego, w jakim punkcie czasoprzestrzeni (kontekście) rozpoczyna ona swój proces poznawczy. I po drugie, czy replikacja jest uwarunkowana aparaturą pojęciową, w jaką wyposażona jest dana maszyna.

Odpowiedź na pierwsze pytanie rozpoczniemy od ogólnej charakterystyki takiej maszyny. Zgodnie z pierwszym założeniem tezy o osobliwości taka maszyna jest inteligentniejsza od człowieka. Jest więc zdolna do emulowania jego świadomości i stanów mentalnych. Taką maszynę można więc scharakteryzować schematycznie poprzez jej strumień przeżyć świadomych  $S$ , bazę  $B$ , która pozwala z  $S$  wygenerować pojęcia  $P$  danej maszyny,<sup>9</sup> określoną liczbę  $Z_n$  organów zmysłowych stanowiących  $n$ -elementowy podzbiór kombinacji wymiarów czasoprzestrzeni, zbiór pojęć kontekstowych  $P$ , uwarunkowanych lokalizacją tej maszyny i wygenerowanych na bazie jej strumienia przeżyć  $S$ , zbiór pojęć fizykalnych  $F$  oraz zbiór intersubiektywnych pojęć  $I^k$  jako teoriomnogościową kombinację pojęć fizykalnych rzędu  $k$ .

Postawione wyżej pytania mają wspólne fundamentalne założenie, że możliwe jest przyporządkowanie pojęć  $P$  wygenerowanych przez maszynę na bazie jej strumienia

---

<sup>9</sup> Baza  $B(S, P)$  wyznaczona jest przez zespół pierwotnych pojęć, odniesionych do  $S$  oraz zachodzących między nimi relacji.

przeżyć S do pojęć fizykalnych F w sposób niezależny od przebiegu jej procesu poznawczego, prowadzącego od S do P. Carnap przyjmując takie założenie traktował je jako fundamentalny warunek możliwości nauki. To wyznaczenie, zgodnie z teorią pojęć rozwijaną od H. Poincarégo przez E. Cassirera, dokonuje się jednoznacznie poprzez wskazanie relacji, w jakie wchodzi dany przedmiot z pozostałymi. Strukturalizm Carnapa dokonuje specyficznego dookreślenia – warunkiem możliwości zachowania stałości takiego wyznaczenia niezależnie do przechodzenia w replikacji na coraz wyższe stopnie pojęciowe jest posłużenie się wyłącznie własnościami strukturalnymi tych relacji.

Ilustracją roli, jaką w tym stanowisku odgrywa strukturalizm, jest spór Carnapa z R. Schrödingerem (Bitbol 2004, 115-116). Ten ostatni w artykule (1935), zatytułowanym “Quelques remarques au sujet des bases de la connaissance scientifique” (“Pewne uwagi o bazie wiedzy naukowej”) wprowadził prowokacyjną tezę, że nauka nie jest samowystarczalna, gdyż wymaga fundamentalnego założenia, które nie jest ani empirycznie testowalne, ani nie ma charakteru konwencjonalnego. To założenie Carnap zrekonstruował jako Hipotezę P następująco: “Hipoteza P: Nie tylko ja doznaję wrażeń (a w konsekwencji posiadam myśli, uczucia, pamięć, itd.); inne osoby także je posiadają” (1936, 116). Carnap starał się wykazać, że takie założenie jest testowalne eksperymentalnie pod warunkiem, że dopuści się wyłącznie precyzyjne prawa relacyjne, łączące stany mentalne z zachowaniami (1935, 119).

Pytanie, czy replikacja wiedzy, jaką posiada dana maszyna, byłaby możliwa niezależnie od kontekstu, można odnieść do sytuacji, gdy maszyna w pewnym momencie traci wszystkie dotychczas wygenerowane pojęcia w oparciu o strumień przeżyć  $S_1$ , oznaczmy je jako  $P_1$ , oraz dane uzyskane dzięki organom zmysłowym, ale nadal zachowuje zdolność generowania nowego strumienia przeżyć  $S_2$ . Wygenerowane w oparciu o ten strumień pojęcia  $P_2$  nie będą identyczne z  $P_1$  z uwagi na prosty fakt, że relacyjne ustytuowanie każdego z pojęć  $P_2$  jest trywialnie inne niż w  $P_1$ .

Jeżeli jednak przyjmie się fundamentalne założenie o możliwości przyporządkowania pojęć P do F, wówczas możliwa jest replika pewnych – w tym przypadku – strukturalnych własności  $P_1$  w  $P_2$  dzięki temu właśnie przyporządkowaniu. Przyporządkowanie do F pozwala bowiem na strukturalne wyznaczenie odniesienia pojęć

$P_2$  w taki sam sposób, jak w  $P_1$ , gdyż pozostaje ono niezależne od kontekstu ich wygenerowania w  $S_2$ .

Zagadnienia uwarunkowania tej replikowalności przez aparaturę pojęciową, jaką posiada dana maszyna, wydaje się więc dość oczywiste. Z pewnym zastrzeżeniem. Dopuszczenie korygowalności niższych poziomów pojęciowych (np.  $P$ ) z wyższych (np.  $I$ ) może prowadzić do kołowości całego schematu postępowania.

Intersubiektywna replikowalność dotyczy maszyn tego samego pokolenia, czyli o tym samym poziomie inteligencji przetwarzającej dane z organów zmysłowych o zbliżonym stopniu złożoności, które posiadają ten sam maksymalny stopień złożoności pojęć  $I^k$ .

W tym przypadku problem replikowalności w zasadzie dotyczy możliwości odtworzenia pojęć  $P^A$  danej maszyny  $A$  przez maszynę  $B$ . Taka bowiem możliwość warunkuje uzyskanie jednorodności pod względem złożoności pojęć  $I^k$  specyficznego dla tego pokolenia maszyn. Jeżeli więc zaistniałaby różnica między dwiema maszynami pod względem zbioru pojęć  $I^k$ , to dzięki możliwości korygowania niższych stopni pojęciowych, dana maszyna mogłaby skorygować swój zbiór pojęć  $P^A$  pod warunkiem jednak tego, że potrafiłaby w tym zbiorze odtworzyć pojęcia innej maszyny  $P^B$ , które pozwoliły maszynie  $B$  wygenerować swoisty dla tego pokolenia maszyn poziom pojęć  $I^k$ . Jednak, jak w przypadku solipsystycznym, taka możliwość byłaby uwarunkowana przyjęciem założenia o istnieniu przyporządkowania między  $P$  a  $F$ , a replika byłaby ograniczona do własności wyłącznie strukturalnych.

Nowe pokolenie maszyn, być może mających też bogatszy zestaw organów zmysłowych, różniłoby się bogatszym zasobem pojęć  $I^k$  w stosunku do poprzedniego pokolenia. Z zachowaniem uwarunkowań, które wystąpiły na poziomie solipsystycznym i intersubiektywnym, z założenia o wyższym stopniu inteligencji nowego pokolenia w prosty sposób wynika zdolność do repliki pojęć  $I^j$  maszyn-poprzedników w zbiorze pojęć  $I^k$  maszyn-potomków, gdzie  $k > j$ .

Wspólnym uwarunkowaniem replikowalności w różnych omówionych odmianach są strukturalne własności pojęć. A zatem każda maszyna, niezależnie od pokolenia, posiadałaby własny zasób kontekstowego poznania, który nie jest replikowalny w opisany wyżej sposób.

Pozostaje jednak kluczowe pytanie o możliwość replikowalności własności niestrukturalnych wiedzy maszyny A. Skoro nie istnieje możliwość replikowania tej wiedzy przez jedną maszynę, nawet wyższego pokolenia, to warto rozważyć taką możliwość w odniesieniu do wielu maszyn pokolenia co najmniej równego z pokoleniem A.

Pojęcia  $P^A$  wygenerowane na bazie niepowtarzalnego strumienia  $S^A$  nie są replikowalne. Odpowiednio modyfikując bazę  $B^A(S, P)$  można starać się wyczerpać wszystkie możliwe kombinacje rodzajów pojęć bazowych i zachodzących między nimi relacji, generując następnie ich przypisanie do pojęć F. Oznaczmy tę liczbę przypisań przez  $i$ . W przypadku, gdyby słuszne było pierwsze założenie stwierdzenia stanu osobliwości o możliwości stworzenia maszyny inteligentniejszej od człowieka, można by przyjąć, że w danym pokoleniu istnieje co najmniej  $i$  maszyn, z których każda realizowałaby inne z  $i$  kombinatorycznie wygenerowanych przyporządkowań. Zestaw tych maszyn  $i$  wzięty łącznie stanowiłby najbliższą z możliwych aproksymacji repliki pojęć  $P^A$ .

Taka *śródpokoleniowa* replika jednak nie potwierdza drugiego z podstawowych założeń tez o osobliwości, jakim jest gwałtowny wzrost inteligencji. Replika śródpokoleniowa nie jest bowiem związana z tworzeniem maszyn o wyższej inteligencji. W związku z tym nie istnieje możliwość określenia, na czym miałyby polegać ów „gwałtowny wzrost” inteligencji. Maszyny nowego pokolenia, które strukturalnie replikowałyby wiedzę maszyn-przodków, każdorazowo dokonywałyby śródpokoleniowo tylko jednej z możliwych replikacji. Wybór między śródpokoleniowymi replikacjami poszczególnych pokoleń może mieć co najwyżej charakter pragmatyczny i nie ma jednoznacznie wyznaczonej „właściwej” śródpokoleniowej replikacji. O faktycznie zrealizowanej replikacji w maszynach-potomkach nie można więc powiedzieć, że jest najlepsza, optymalna czy jedyna, lecz jedna z wielu możliwych. Wzięty w oderwaniu żadna śródpokoleniowa replikacja nie mogłaby zostać uznana za najlepszą. Wydaje się zatem, że konieczne jest zachowanie pluralizmu ścieżek replikacji, co wykluczałoby tezę o gwałtownym wzroście inteligencji kolejnych pokoleń.

Takie przekonanie obecne jest, nie bez wpływu asymilacji zasady tolerancji, wśród inżynierów faktycznie pracujących nad tworzeniem inteligentnych systemów:

Logiczny pluralizm faktycznie dominuje we wszystkich obszarach inżynierii wiedzy; jego szeroka akceptacja jest faktem, który należy uznać w sposób oczywisty. W rzeczy samej, jesteśmy przekonani, że z perspektywy metodologicznej, a zwłaszcza ze względu na czystą praktyczną użyteczność, zalety przyjęcia pluralizmu logicznego w inżynierii ontologicznej trudno byłoby podważyć w sposób poważny. (Kutz i in., 2010, 262)

## 5. Założenie o replikowalności wiedzy w aspekcie diachronicznym

Klasyk brytyjskiego empiryzmu, D. Hume w radykalny sposób zakwestionował racjonalny charakter replikacji wiedzy zdobywanej na podstawie doświadczenia. Regularności, które obserwujemy w otaczającym świecie nie są nawinięte na sprężynę konieczności pojęciowej. Związek zatem dwóch zdarzeń, które dotychczas zawsze występowały łącznie, np. poranek i wschód słońca, nie jest związkiem logicznym, racjonalnym czy deterministycznym. Po podważeniu podstaw indukcyjnych teorii uczenia się z doświadczenia, pozostawałoby jedynie miejsce dla replikowalności związków dedukcyjnych w wiedzy.

Odpowiedzi na indukcyjny sceptycyzm Hume'a podjął się angielski pastor T. Bayes. Wyprowadził on w rachunku prawdopodobieństwa twierdzenie,<sup>10</sup> które – jak zostanie to szerzej omówione niżej – miało precyzyjnie określać, jak należy aktualizować dotychczasową wiedzę w oparciu o nowe informacje, których dostarcza doświadczenie. Powrócił do niego w latach 1940. Carnap, tworząc indukcyjną teorię uczenia się, która w zamierzeniu miała być uogólnieniem logiki dedukcji.

O słuszności zarzutów Hume'a podważających racjonalność replikowania wiedzy zdobytej w indukcyjnym procesie uczenia się w dużej mierze przesądza to, jaką rolę w tych teoriach przypisuje się kontekstowi. Carnap nakładał restryktywne warunki, które powinien spełniać kontekst, zwłaszcza pod względem jego artykulacji językowej. Diametralnie inaczej rzecz przedstawia się w ujęciu bayesowskim w wersji subiektywnej

---

<sup>10</sup> Przystępne nieformalne omówienie podane jest w (Kawalec 2011a), techniczne w (Kawalec 2003).

tzw. personalistycznej (Maher 2010). Mimo że bayesowska personalistyczna teoria uczenia się dzieli z Carnapem przekonanie o centralnej roli twierdzenia Bayesa w uczeniu się na podstawie doświadczenia, to jednak – w przeciwieństwie do Carnapa – radykalnie minimalizuje wszelkie ograniczenia nakładane na kontekst.

Załóżmy, że robot Carnapa zastanawia się, jakie prawdopodobieństwo przypisać przewidywaniu, że w kolejnym obrocie kołem fortuny wypadnie pole „bankrut”, biorąc pod uwagę wyniki dotychczasowych obrotów tym kołem. Jest to przykład tzw. *instance confirmation*, czyli wnioskowania indukcyjnego, gdzie na podstawie danych staramy się przewidzieć następny wynik obserwacji. Zaznaczmy na marginesie, że właśnie ten rodzaj wnioskowań indukcyjnych Carnap uznawał za podstawowy dla wszystkich innych rodzajów.

Odpowiedź robota będzie w tym przypadku zdeterminowana kontekstem. Przedmiotem rozważań robota jest przewidywanie wyniku obrotu kołem, opis dotychczasowych wyników uzyskanych na tym kole oraz relacja logiczna, jaka zachodzi między tymi zdaniami. Do kontekstu należy więc zaliczyć wszystkie pozostałe elementy, które wpływają na odpowiedź, jakiej udzieli robot, a które nie są bezpośrednio przedmiotem jego uwagi. Takim zasadniczym elementem jest język, w którym sformułowane są zdania rozważane przez robota, logiczna struktura języka, scharakteryzowana przez Carnapa oraz w specyficzne dla tego autora uzależnienie relacji logicznej między obydwoma zdaniami od struktury języka. Krótco teraz omówione zostaną poszczególne elementy tworzące kontekst.

Język jest niezwykle prosty, gdyż zasadniczo dopuszcza jako predykaty proste tylko predykaty jednoargumentowe, np. „jest polem bankrut”, a jego dziedzina może obejmować skończoną liczbę indywiduów, np. pole wylosowane w kolejnym rzucie.

Przyjmijmy, że łącznie mamy do dyspozycji 5 obrotów kołem fortuny, a do tej pory wykonano 4. Odnosząc każdy z predykatów prostych występujących w języku kolejno do jakiegoś indywiduum otrzymamy wszystkie możliwe opisy tego indywiduum, z których jeden jest prawdziwy. Tę czynność powtarzamy w stosunku do każdego indywiduum z dziedziny i dla każdej jednostki otrzymujemy zestaw wszystkich możliwych jej opisów.

Następnie dla każdego indywiduum wybieramy jeden taki opis i łącząc je w jedno otrzymujemy możliwy „przekrój” świata – nazywany przez Carnapa opisem stanu. Wyczerpując wszystkie możliwe zestawienia opisów indywiduowych poszczególnych jednostek otrzymujemy wszystkie możliwe opisy stanów. A zatem pełną przestrzeń logiczną, w której jeden opis stanu musi być prawdziwy. W naszym przykładzie opis stanu musi więc określić, jaki wynik otrzymano w każdym z 5 obrotów kołem, np. bankrut w przypadku pierwszym, bankrut w drugim, nie bankrut w trzecim, bankrut w czwartym, nie bankrut w piątym.

Dla każdego zdania sformułowanego w tym języku jesteśmy też w stanie określić, w jakim obszarze tej przestrzeni jest ono prawdziwe. A dla dowolnych dwóch zdań jesteśmy w stanie określić czy mają one obszar wspólny, w którym oba są prawdziwe.

Opisy stanu, które zawierają dokładnie taki sam zestaw opisów indywiduowych – niezależnie od tego, jakim przysługują one indywiduom – Carnap traktował jako symetryczne i grupował je w tzw. opis strukturalny. Wracając do podanego przykładu: do tego samego opisu strukturalnego zaliczymy te wszystkie opisy stanu, które zawierają trzy razy „bankrut” i dwa razy „nie bankrut” – niezależnie od tego, w jakiej kolejności pojawiały się te wyniki w opisie stanu (czy nie bankrut był w pierwszym i drugim czy pierwszym i czwartym).

Każdy z opisów strukturalnych Carnap uznawał za równo prawdopodobny. W naszym przykładzie jest ich 6 (5 razy bankrut, 4 razy, 3 razy, 2 razy, 1 raz, 0 razy), a zatem każdy opis strukturalny ma prawdopodobieństwo  $1/6$ . Na każdy z opisów strukturalnych składa się jednak inna liczba opisów stanu. I tak przypadek, gdy wystąpi 5 razy bankrut może zdarzyć się tylko raz, podobnie gdy pole bankrut nie wystąpi ani razu. Przypadek, gdy bankrut pojawia się czterokrotnie, a raz tylko pojawia się inne pole – może wystąpić na 5 różnych sposobów, np. inne pole w pierwszym obrocie, drugim, trzecim, czwartym lub piątym.

Także, gdy pole bankrut wystąpi tylko raz – może być zrealizowane w jednym z pięciu opisów stanu.

W przypadku, gdy pole bankrut występuje dwa razy oraz gdy występuje trzy razy – mamy 10 opisów stanu, które składają się na opis strukturalny. Jeśli w dotychczasowych 4 obrotach otrzymano za każdym razem pole bankrut, to wywnioskuje

on, że prawdopodobieństwo wylosowania tego pola w następnym obrocie wzrosło z  $1/30$  do  $1/5$ .

W później wprowadzonych modyfikacjach pierwotnej teorii Carnapa robot ma więcej swobody. Wprowadzony zostaje parametr, który pozwala robotowi poruszać się pomiędzy dwiema skrajnymi wartościami. Jedną ze skrajnych wartości oparta jest wyłącznie na doświadczeniu: zlicza się ilość interesujących wyników, np. wystąpienia pola bankrut, w stosunku do liczby wykonanych eksperymentów – obrotów kołem fortuny. Np. jeśli w dotychczasowych 4 obrotach zawsze wystąpiło pole bankrut, to stosunek równy jest 1.

Drugą skrajną wartość jest z kolei zależna wyłącznie od struktury języka i jest całkowicie niezależna od doświadczenia. Jest jedynie funkcją ilości predykatów zawartych w języku.

Robot w swoim uczeniu się z doświadczenia nie jest już jednoznacznie zdeterminowany kontekstem. Posługuje się natomiast funkcją ważoną obu skrajnych wartości: czysto empirycznej oraz czysto apriorycznej. To, który z czynników ma większą wagę wyraża się za pomocą parametru  $\lambda$ , który przyjmuje wartości w zakresie od 0 do  $\infty$ . W przypadku, gdy ma wartość 0, czynnik aprioryczno-językowy zostaje całkowicie wyeliminowany i pozostaje wyłącznie wartość czysto empiryczna. Gdy natomiast  $\lambda$  zmierza do  $\infty$ , wówczas mamy sytuację odwrotną – pozostaje wyłącznie czynnik aprioryczny, a doświadczenie jest całkowicie ignorowane.

W tej fazie sposób nabywania wiedzy z doświadczenia przez robota Carnapa nie jest już zdeterminowany kontekstem. Ma on do wyboru kontinuum różnych sposobów uczenia się, które mieszczą się wszakże między czysto empirycznym a czysto apriorycznym uczeniem się.

Równoległe do prac Carnapa powstawały podstawy indukcyjnej teorii, która także za centralne dla mechanizmu uczenia się z doświadczenia traktuje twierdzenie Bayesa. Prace Bruno de Finetti'ego we Włoszech i Franka Ramseya w Cambridge zmierzały jednak w przeciwną stronę – usunięcia z tej teorii wszystkich elementów poza tymi, które były konieczne do wprowadzenia w niej twierdzenia Bayesa.

Za sprawą Ramseya wprowadzono interpretację prawdopodobieństwa w kategoriach siły przekonania, którą z kolei behawioralnie ujawnia gotowość



podejmowania zakładów z określoną wysokością stawek. Jedynym ograniczeniem siły przekonań jest koherencja z aksjomatami rachunku prawdopodobieństwa. W teorii bayesowskiej personalistycznej więc jest ogromna dowolność w określaniu siły przekonania. Prześledźmy to na uproszczonym przykładzie. Załóżmy, że pediatra ma ocenić, czy uporczywy suchy kaszel (dane) jest wynikiem alergii wdechowej (hipoteza pierwsza) czy bezobjawowego zapalenia płuc (hipoteza druga).

Robot Carnapa ustali stopień potwierdzenia hipotezy przez dane po wprowadzeniu wchodzących w grę zdań do języka o określonej strukturze i skończonej dziedzinie. Prawdopodobieństwo pierwotne hipotezy jest dla niego zdeterminowane strukturą języka – najpierw traktuje się jako równo prawdopodobne opisy strukturalne, a następnie – w ich obrębie jako równo prawdopodobne poszczególne opisy stanu.

Aby zatem kontekst wpływał na uczenie się robota musi: 1) dać się wyśłowić w danym języku (mocno ograniczonym) oraz 2) zawarte w nim informacje muszą bez reszty dać się rozłożyć na proste składniki treściowe odpowiadające „niewielkiej liczbie” predykatów pierwotnych.

Odmienne rolę kontekstu ujmuje personalista bayesowski. Prawdopodobieństwo pierwotne traktuje on bowiem jako siłę *swojego* przekonania o prawdziwości danej hipotezy. Siła przekonania jest ukształtowana przez niezliczoną liczbę czynników: np. wykształcenie tej osoby, jej życiowe doświadczenie oraz wiedza ogólna itp. Zatem różnica między dwoma osobami – z których każda jest równie racjonalna i poprawnie stosuje bayesowski mechanizm uczenia się – może być niemal dowolna. Np. jedna osoba może uznać – nie biorąc pod uwagę symptomów obserwowanych u dziecka, że prawdopodobieństwo wystąpienia u niego zapalenia płuc jest bliskie 1, a druga osoba – może uznać je za bliskie 0.

Personalista bayesowski odrzuca zarzut subiektywizmu na podstawie dwóch twierdzeń, charakteryzujących uczenie się z doświadczenia na podstawie twierdzenia Bayesa: 1) przy odpowiednio dużej liczbie obserwacji siła przekonania w przypadku hipotezy prawdziwej zmierza do jedności 2) niezależnie od tego, jak duża jest różnica prawdopodobieństw pierwotnych między dwoma personalistami – przy odpowiednio dużej liczbie obserwacji ich siła przekonania będzie zmierzać do tej samej wartości.

Bayesowska teoria uczenia się jest jedną z odmian formalnej indukcyjnej teorii uczenia się. Jedną z podstawowych różnic między poszczególnymi odmianami teorii bayesowskich dotyczy warunków, które muszą być spełnione przez kontekst, aby wykorzystać je w twierdzeniu Bayesa. Kontekst jest w tym twierdzeniu uwzględniany w postaci prawdopodobieństwa pierwotnego. Carnap żądał, by było ono zdeterminowane logiczną strukturą języka, w którym przeprowadza się wnioskowania. Personalista bayesowski przeciwnie – przeciwstawia się jakiegokolwiek ujęciu kontekstu. Funkcję obiektywizacji – którą Carnap dostrzegał w języku i jego strukturze – pełni twierdzenie o zbieżności, które pokazuje, że rola wiedzy kontekstowej wraz z napływem informacji z doświadczenia maleje – pod warunkiem konsekwentnego stosowania mechanizmu bayesowskiego w asymilacji tych informacji.

Bayesowska teoria uczenia się, niezależnie czy w wersji obiektywizującej Carnapa, czy personalistycznej, w celu zagwarantowania replikowalności uzyskanej w takim procesie uczenia się wiedzy, musi posłużyć się założeniami dotyczącymi kontekstu. Jak podkreślał Carnap, takie założenia mają charakter pragmatyczny<sup>11</sup> z uwagi na kontekst aplikacji. Nie są one elementem replikowanej wiedzy, lecz czynnikiem ją warunkującym, zwykle w dość jednoznaczny sposób.

Problemu tego nie rozwiązuje zmiana poziomu inteligencji maszyn. Żeby to pokazać wyraźniej, trzeba byłoby za Carnapem odróżnić dwa procesy uczenia się (Zabell 2010). W pierwszym zakładamy, że ilość kategorii pojęciowych  $I^m$  jest stała, w drugim – proces uczenia obejmuje także możliwość rozszerzenia ilości kategorii pojęciowych  $m$  o pewną liczbę dotychczas nieznaną  $n$ .

Nawet w przypadku prostszym, ze względu na probabilistyczny charakter zależności włączonych do wiedzy maszyny, warunkiem replikowalności tej wiedzy jest odtworzenie założeń determinujących – zgodnie z sekwencyjnym stosowaniem twierdzenia Bayesa – te zależności probabilistyczne.

Zwięźle wyraził to jeden z bliższych współpracowników Carnapa w ostatniej fazie jego prac nad logiką indukcji, R. Jeffrey (1984, 84):

---

<sup>11</sup> Jak zwraca uwagę T. Uebel (2011, 12), Carnap podkreślał, że to, w jaki sposób konceptualizujemy uwarunkowane językowo zjawiska w składni i semantyce nie jest determinowane przez nasze doświadczenie “zwierząt używających znaki”.

nikt nie uważa, że w mózgu istnieją zlokalizowane tam potencjały, które konstytuują fizjologiczne podłoże prawdopodobieństw subiektywnych, które ujawniają się w naszym zachowaniu. Posługując się analogią z komputerami, to program, a nie układy scalone, „mieści” takie prawdopodobieństwa. Jeśli jest to sprawa instynktu, to znaczącym czynnikiem są także artefakty kulturowe. Przypisania prawdopodobieństw są dalekie od bycia parametrami stanów ludzkiego organizmu, które można byłoby mierzyć za pomocą słownych wiarometrów.

Podobnie więc jak w przypadku wiedzy synchronicznej, tak i w tym przypadku replikowalność wiedzy uzależniona jest od założeń, które są niezależne od stopnia inteligencji maszyny-podmiotu tej wiedzy.

## **Bibliografia**

- Bauer F. L., (2010), *Origins and Foundations of Computing*, Berlin: Springer.
- Bitbol M., (2004), “The problem of other minds: A debate between Schrödinger and Carnap”, *Phenomenology and the Cognitive Sciences* 3: 115–123.
- Carnap R., (1936), “Existe-t-il des prémisses de la science qui soient incontrôlables?”, *Scientia* 60: 129–135.
- Carnap R., (1974). *An introduction to the philosophy of science*. New York: Basic Books.
- Camap R., Jeffrey R., (1971), *Studies in Inductive Logic and Probability*, Berkeley: Berkeley University Press.
- Carnap R., (1928/2011), *Logiczna struktura świata*, tłum. i oprac. P. Kawalec, Warszawa: PWN.
- Chalmers D. J., (2010), “The Singularity. A Philosophical Analysis”, *Journal of Consciousness Studies*, 17: 7–65.
- Charron-Bost B., Pedone F., Schiper A., *Replication: theory and practice*, Berlin: Springer 2010.
- Friedman M., (2011), “Carnap on theoretical terms: structuralism without metaphysics”, *Synthese* 180: 249–263.
- Holger A., (2010), “New account of empirical claims in structuralism”, *Synthese* 176: 311-332.
- Jeffrey R., (1984), “De Finetti's probabilism”, *Synthese* 60: 73-90.
- Jeffrey R., (1991), “After Carnap”, *Erkenntnis* 35: 255-262.
- Kawalec P., (2003), *Structural Reliabilism: Inductive Logic as a Theory of Justification*, Dordrecht: Kluwer Academic Publishers.

- Kawalec P., (2010), „Woluntaryzm epistemologiczny Williama Jamesa. Próba rekonstrukcji na gruncie probabilizmu B. van Fraassena”, *Przegląd Filozoficzny* 19: 283-295.
- Kawalec P., (2011a), „Bayesianizm w polskiej tradycji probabilizmu – studium stanowiska Kazimierza Ajdukiewicza”, *Ruch Filozoficzny*, w opracowaniu.
- Kawalec P., (2011b), „Wstęp”, w: Carnap R., *Logiczna struktura świata*, tłum. i oprac. P. Kawalec, Warszawa: PWN, s. XI-CIV.
- Kutz O., Mossakowski T., Lücke D., (2010), “Carnap, Goguen, and the Hyperontologies: Logical Pluralism and Heterogeneous Structuring in Ontology Design”, *Logica Universalis* 4: 255–333.
- Maher P., (2010), “Bayesian probability”, *Synthese* 172:119-127.
- Moulines C. U., (1991), “Making sense of Carnap's *Aufbau*”, *Erkenntnis* 35: 263-286.
- Moulines C. U., (2001), “Die Mathematisierung der Erfahrung: Vorgänger zu Carnaps *Aufbau*”, *Erkenntnis* 54: 105–120.
- Richardson A., (2003), “The geometry of knowledge: Lewis, Becker, Carnap and the formalization of philosophy in the 1920s”, *Studies in the History of Philosophy of Science* 34: 165–182.
- Schrödinger E., (1935), “Quelques remarques au sujet des bases de la connaissance scientifique”, *Scientia* 57: 181–191.
- Scott E., (2009), “Logical Empiricism, Politics, and Professionalism”, *Science & Education* 18:177–189.
- Uebel T., (2011), “Pragmatics in Carnap and Morris and the Bipartite Metatheory Conception”, *Erkenntnis*, DOI 10.1007/s10670-011-9352-5, ss. 24.
- Zabell S. L., (2010), Carnap and the logic of inductive inference, w: D. Gabbay, S. Hartmann, J. Woods (red.), *Handbook of the History of Logic. Volume 10: Inductive Logic*. Amsterdam: Elsevier, s. 265-305.

## Abstrakt

Zwolennicy transhumanizmu upatrują przyszłego podmiotu wiedzy w maszynach, które miałyby być inteligentniejsze od ludzi, ich twórców. Jeżeli raz zaistniałaby taka możliwość stworzenia inteligentniejszych podmiotów, to w kolejnym kroku maszyny inteligentniejsze od ludzi stworzą następne pokolenie maszyn bardziej inteligentne od samych siebie. To zapoczątkuje dalsze etapy procesu gwałtownego wzrostu inteligencji aż do osiągnięcia stanu osobliwości (ang. *singularity*).

W niniejszym tekście omówione są wątpliwości dotyczące drugiego z kluczowych stwierdzeń uzasadniających transhumanistyczną tezę o osobliwości, a mianowicie gwałtownego wzrostu inteligencji. Te wątpliwości dotyczą możliwości replikowalności wiedzy danego podmiotu poznawczego jako warunku gwałtownego wzrostu. Inspiracją tych wątpliwości jest program epistemologiczny, który mógłby wydawać się wzorcowym protoplastą programu transhumanizmu, a mianowicie program epistemologiczny R. Carnapa, którego ucieleśnieniem jest tzw. robot Carnapa. W ujęciu synchronicznym zasadniczo taka replikowalność ograniczałaby się wyłącznie do strukturalnych własności wiedzy, co nie daje wystarczającej podstawy do stwierdzenia gwałtownego wzrostu. W ujęciu diachronicznym, możliwości dynamicznego zwiększania zasobów wiedzy robota Carnapa są w sposób nieusuwalny uwarunkowane przyjętymi subiektywnie schematami, a wybór pomiędzy nimi dokonuje się pragmatycznie. W tym przypadku stawianie pytania o wzrost jest bezpodstawne.

Słowa kluczowe: replikowalność wiedzy, stan osobliwości, sztuczna inteligencja, Rudolf Carnap, David Chalmers

## Abstract

Advocates of transhumanism project the future epistemic subjects to be machines that are more intelligent than the people who created them. Having the ability to create intelligent entities, these machines will create the next generation of more intelligent ones than themselves. This will open a reiteration of further steps in the process of explosion of intelligence ultimately into the state of Singularity.

The paper finds problematic one of key assumptions of transhumanists justifying the thesis of singularity, namely the intelligence explosion. These concerns stem from a

program that may seem exemplary embodiment of the transhumanism, namely the so-called Carnap's robot. These concerns relate to possible replicability of cognitive knowledge of the entity as a condition of rapid growth. The inspiration for these doubts is the epistemological program, which would seem to be the ancestor of transhumanism model, namely, R. Carnap's epistemological program, embodied in the so-called Carnap's robot. In synchronous terms replicability basically would be confined to the structural properties of knowledge, which does not provide a sufficient basis for the explosion of intelligence. In the diachronous approach, the ability to dynamically increase the stock of knowledge of Carnap's robot are indelibly determined by subjectively accepted patterns, and the choice between them is made on pragmatic ground. In this case, putting the question of explosion would seem unfounded.

Keywords: replicability of knowledge, the state of singularity, artificial intelligence, Rudolf Carnap, David Chalmers