

# **Architektura Dostępu: Operacyjna Rama Pomiaru Świadomości Wcielonej**

*Wstępny Przegląd Teoretyczny*

Aleksander Bogusław Rej

**Wersja robocza 1.2 (syntetyczna)**

**18 września 2025**

**Problem:** Debata o świadomości ugrzęzła, bo obecne paradygmaty i miary mieszają trzy poziomy: treść (co jest reprezentowane), raport (co jest raportowane) i kontrolę poznawczą (jak system steruje przetwarzaniem). **Propozycja:** CAS rozdziela treść ( $Q$ ) od wymiaru kontrolno-metapoznawczego ( $\Sigma=A/R/C/Cost$ ) i traktuje wydajność jako wynik  $Q \times \Sigma$ . **Test:** pokazuję trzy falsyfikowalne predykcje (P25, P3, P52), które mają wykazać, że profil  $\Sigma$  przewiduje wyniki ponad IQ, a  $Q$  i  $\Sigma$  da się eksperymentalnie rozdzielić — to przesuwa debatę z metafizyki na mierzące wskaźniki z potencjałem zastosowań w naukach społecznych.

## Diagnoza – Dlaczego Stare Mapy Prowadzą Donikąd?

Pole badań nad świadomością doświadcza obecnie kryzysu metodologicznego, przejawiającego się przewagą rozbieżności nad konsensusem w kluczowych konstrukcjach i metodach (Mudrik i in., 2025; Chis-Ciure i in., 2024). Ten stan podważa zdolność tej dziedziny do rozstrzygania między teoriami. Najwyraźniej ilustrują to adwersaryjne testy teorii globalnej przestrzeni roboczej (GNWT) oraz teorii zintegrowanej informacji (IIT), które nie przyniosły jednoznacznego rozstrzygnięcia i zarazem znaczaco podważyły podstawowe założenia obu ujęć (Cogitate Consortium i in., 2025). Jak argumentują autorzy tej pracy (Melloni i in. 2025), wyniki te osłabiają zasadność czysto konfrontacyjnego pytania „która teoria jest prawdziwa?” i wskazują na potrzebę ram ilościowych do systematycznego porównywania teorii.

Fundamentalną słabością dotychczasowych podejść, w przeciwnieństwie do pojęcia treści, które doczekało się bardziej sprecyzowanych propozycji, jest opieranie się na monolitycznym pojęciu „dostępu”, które w świetle dekad badań klinicznych i eksperymentalnych jest nie do utrzymania. Klasyczne dane o ślepowidzeniu (*blindsight*) — gdzie, zgodnie z teorią przetwarzania rekurencyjnego (RPT), uszkodzenie kory wzrokowej prowadzi do utraty świadomego widzenia przy zachowanej zdolności do ponadprzypadkowego wskazywania atrybutów bodźca (Lamme, 2006) — oraz o pacjentach po komisurotomii (tzw. *split-brain*), u których jedna półkula może funkcjonować bez wiedzy drugiej (Sperry, 1968), pokazują, że skuteczne przetwarzanie może istnieć niezależnie od raportowalności. Konieczna jest więc dekompozycja na co najmniej dwa procesy: bogatszy treściowo, lecz bardziej krótkotrwalały *dostęp percepcyjny* (przedrefleksyjny) oraz o mniejszej pojemności, lecz stabilniejszy *dostęp refleksyjny* (raportowalny), wspierający kontrolę i metapoznawanie (Block, 2011). Heurystycznie rozróżnienie to koresponduje z podziałem na szybkie, intuicyjne i wolniejsze, analityczne tryby przetwarzania opisane jako „System 1” i „System 2” (Kahneman, 2011).

Problem ten jest analogiczny do dobrze rozpoznanych ograniczeń psychometrycznych miar inteligencji. Testy IQ pomijają istotne aspekty myślenia racjonalnego: mierzą przede wszystkim sprawność obliczeniową (tzw. „umysł algorytmiczny”), zaniedbując komponent refleksyjny odpowiedzialny za samoregulację (Stanovich, 2009). Są one jak testowanie silnika w sterylnych warunkach laboratoryjnych: mierzą jego maksymalną moc, ignorując kluczowe dla realnej adaptacji wymiary. Ignorują "kierowcę" (kalibrację metapoznawczą), "wydajność paliwową" (koszt zasobowy) oraz, co fundamentalne, "niezawodność" (trwałość wiedzy, czyli retencję). To ostatnie zaniedbanie odpowiada zjawisku znanemu w uczeniu maszynowym (ML) jako katastroficzne zapominanie — gwałtowna utrata uprzednio przyswojonej wiedzy po przyswojeniu nowej — jedno z głównych ograniczeń paradygmatu uczenia ciągłego (continual learning) w sztucznych sieciach neuronowych (ANN) (McCloskey & Cohen, 1989). W tym sensie obecne miary świadomości są „ślepe” na jej dynamiczne komponenty.

Wobec tych luk — nierostrzygniętych sporów teoretycznych, empirycznie problematycznego pojęcia „dostępu” i niekompletnych narzędzi pomiarowych — dalszy postęp wymaga zmiany paradygmatu. Potrzebujemy nie kolejnej ontologicznej „wielkiej teorii”, lecz — na wzór dojrzałych nauk — pragmatycznego, wielowymiarowego i falsyfikowalnego instrumentarium pomiarowego. Takie ujęcie odpowiada podejściu funkcjonalistycznemu (Dennett, 1991), według którego nauka powinna wyjaśniać, jak powstają nasze subiektywne raporty z doświadczenia, zamiast poszukiwać enigmatycznego rozwiązania „trudnego problemu” świadomości.

## Propozycja Rozwiązań

W odpowiedzi na te luki proponujemy Wskaźnik Dostępu Świadomego (CAS – *Conscious Access Score*) — ramę pomiarową, która porzuca jednowymiarowe ujęcia i ocenia dwa ortogonalne wymiary: wymiar treści (Q) oraz wymiar kontrolno-metapoznawczy ( $\Sigma$ ). Należy podkreślić, że Q i  $\Sigma$  to wymiary funkcjonalne, a nie ściśle anatomiczne, mimo ich oczywistych korelatów neuronalnych. Ten podział jest operacyjnym rozwinięciem klasycznego rozróżnienia na świadomość fenomenalną (P-consciousness) i świadomość dostępu (A-consciousness) (Block, 1995). Centralna hipoteza (i postulat) brzmi: dostęp świadomy ma charakter „koniunkcyjny”, co formalizujemy heurystycznym, multiplikatywnym indeksem:  $CAS = Q \times \Sigma$ . Q kwantyfikuje bogactwo treści fenomenalnej, czerpiąc z teorii treści (IIT/RPT) i opierając się na wskaźnikach mierzonych niezależnie od raportu (Tononi i in., 2016; Lamme, 2006)<sup>1</sup>. Indeks  $\Sigma$  operacyjalizuje monitorowanie metapoznawcze oraz kontrolę poznawczą, inspirowany ujęciami dostępu (GNWT, teorie wyższego rzędu – HOT), koncepcją lewopółkulowego „interpretera”, wczesnymi modelami architektur kognitywnych oraz krytyką pomiaru inteligencji (Dehaene, 2014; Brown i in., 2019; Gazzaniga, 2000; Duch, 2005; Stanovich, 2009).

Centralnym elementem jest dekompozycja indeksu  $\Sigma$ , który mierzy synergiczne współdziałanie czterech mierzalnych komponentów. W przeciwieństwie do klasycznych miar skupionych na chwilowej sprawności, nasz indeks ocenia dynamiczną, ucieleśnioną adaptację poprzez: (A) samokorektę — rozumianą jako skuteczność iteracyjnej korekty błędów, inicjowanej przez mechanizm monitorowania konfliktu poznawczego (Botvinick i in., 2001); (R) regulację uczenia i retencję — trwałość nabytej wiedzy i odporność na katastroficzne zapominanie (McCloskey & Cohen, 1989); (C) kalibrację metapoznawczą — trafność samooceny pewności (Fleming & Lau, 2014); oraz (Cost) koszt zasobowy — czasowy, energetyczny i obliczeniowy, który uziemia funkcjonowanie w realnych ograniczeniach (Lennie, 2003). Podobnie jak kompozytowe wskaźniki w naukach społecznych, np. Wskaźnik rozwoju społecznego (HDI), CAS celowo agreguje różne wymiary, by dostarczyć użytecznego standardu porównania — pozostając przy tym pragmatycznym indeksem heurystycznym, a nie ostateczną miarą. Taka kolejność komponentów nie jest przypadkowa: odzwierciedla ona funkcjonalną pętlę adaptacji A → R → C → A, stabilizowaną przez ograniczenia zasobowe (Cost) i sprzężenia zwrotne ze światem (Q).

<sup>1</sup> Paradygmaty *no-report* mają separować sygnały treści od aktów raportowania; metodologia pozostaje dyskutowana (m.in. ryzyko resztowej kognicji). Por. Block (2019); nowsze dane — np. Hirao i in. (2025) — sugerują możliwość izolacji korelatów oceny podobieństwa qualów barw od procesów decyzyjnych.

Aby dokładnie umiejscowić CAS względem głównych paradygmatów:

- I. Teorie treści (IIT, RPT): CAS przyjmuje intuicję, że bogactwo doświadczenia ( $Q$ ) wiąże się z integracją informacji (Tononi i in., 2016). W odróżnieniu od IIT świadomość unikamy roszczeń ontologicznych i problemów obliczalności  $\Phi$ , traktując wskaźniki inspirowane IIT jako pragmatyczne przybliżenia (proxy) poziomu  $Q$ . Przykładem jest Perturbational Complexity Index (PCI; Casali i in., 2013), który mierzy złożoność odpowiedzi mózgu na zewnętrzną perturbację, przy czym nie utożsamiamy  $Q$  z żadną pojedynczą miarą. Podejście to jest zgodne z rosnącym konsensem empirycznym, wskazującym na miary złożoności jako na wiarygodne wskaźniki stanu świadomości (Sarasso i in., 2021).
- II. Teorie dostępu (GNWT, HOT, AST — teoria schematu uwagi):  $\Sigma$  porządkuje i ujmuje w indeks to, co te ujęcia opisują jakościowo (Dehaene, 2014; Brown i in., 2019; Graziano & Webb, 2015). CAS idzie dalej: zamiast binarnego, monolitycznego „dostępu”, proponuje stopniowalny wskaźnik, który ocenia nie tylko jego istnienie, ale także jakość jego dwóch funkcjonalnie odrębnych procesów (percepcyjnego i refleksyjnego), ich trwałość ( $R$ ) i koszt zasobowy ( $Cost$ ).
- III. Przetwarzanie predykcyjne i zasada wolnej energii (PP/FEP): CAS stanowi instrumentarium operacyjne do pomiaru procesów postulowanych w PP (Clark, 2013):  $Q$  — treści na różnych poziomach hierarchii predykcyjnej;  $\Sigma$  — mechanizmy metapoznawczej inferencji wyższego rzędu.

## **Nowość i Program Badawczy**

Fundamentalna nowość ramy CAS nie polega na odkrywaniu jej komponentów na nowo, lecz na ich rygorystycznej syntezie i bezprecedensowej operacyjonalizacji. Jej unikalność wyraża się na trzech, komplementarnych poziomach:

- ❖ Po pierwsze, wprost przekształca jakościowe idee monitorowania metapoznawczego i kontroli poznawczej w ilościowy, wielowymiarowy indeks  $\Sigma$  ( $A/R/C/Cost$ ) – wspólny język porównań dla systemów biologicznych (ludzie, zwierzęta) i docelowo, dla zaawansowanych sieci sztucznych.
- ❖ Po drugie, jako pierwsza jawnie integruje z pomiarem  $\Sigma$  te wymiary, które są kluczowe dla ucieleśnionej adaptacji, obejmując samokorektę ( $A$ ), retencję ( $R$ ), kalibrację metapoznawczą ( $C$ ) oraz koszt zasobowy ( $Cost$ ) — ze szczególnym naciskiem na  $R$  i  $Cost$ , często pomijane w dotychczasowych ujęciach.
- ❖ Po trzecie, CAS nie jest celem samym w sobie, lecz generatywną platformą badawczą, która dostarcza odważnych, falsyfikowalnych predykcji i umożliwia systematyczne testowanie założeń.

Siła CAS leży w zdolności do generowania konkretnych, często kontrintuicyjnych hipotez. Poniżej trzy przykłady z szerszego programu badawczego, ilustrujące interdyscyplinarny potencjał:

- **Predykcja psychometryczna (P25: Podwójna dysocjacja IQ i  $\Sigma$ )**

W badaniu populacyjnym z baterią CAS przewidujemy dwie stabilne subpopulacje:  
(1) osoby o bardzo wysokim IQ przy przeciętnym/niskim  $\Sigma$ ;  
(2) osoby o przeciętnym IQ przy bardzo wysokim  $\Sigma$ .

To rozdzieliłoby „moc algorytmiczną” od wymiaru kontrolno-metapoznawczego i pokazałoby wyższą trafność profilu  $\Sigma$  dla adaptacji (mierzonej np. wynikami w zadaniach decyzyjnych lub wskaźnikami życiowego sukcesu)

*Fals.: brak obu subpopulacji przy odpowiedniej mocy badania.*

- **Predykcja fundamentalna (P3: Podwójna dysocjacja neuronalna)**

Rama CAS postuluje funkcjonalną odrębność między wymiarem treści (Q) a mechanizmem kontrolno-metapoznawczym ( $\Sigma$ ). Predykcja ta, stanowiąca *experimentum crucis* dla całego modelu, idzie o krok dalej: możliwa jest eksperymentalna, podwójna dysocjacja tych systemów (np. TMS<sup>2</sup>)<sup>3</sup>, co stanowiłoby mocny test przeciwko teorii unitarnym.

*Fals.: żadna manipulacja nie zmienia Q bez  $\Sigma$  ani  $\Sigma$  bez Q.*

- **Predykcja metapoznawcza (P52: Paradoksalne poczucie wyboru)**

W zadaniu wymuszonego wyboru dwualternatywnego (2AFC) z manipulacją rywalizacji bodźców (niska vs wysoka; trudność dopasowana) przewidujemy podwójną dysocjację w raportach metapoznawczych:

(1) kalibracja metapoznawcza (C/confidence) będzie niższa w warunku wysokiej rywalizacji;  
(2) poczucie sprawstwa wyboru (*sense of agency over choice*, mierzone np. skalą subiektywną) będzie paradoksalnie wyższe w warunku wysokiej rywalizacji, nawet przy kontroli czasu reakcji i poprawności odpowiedzi.

*Fals.: po wyrównaniu trudności confidence nie spada, a poczucie sprawstwa nie rośnie w warunku wysokiej rywalizacji.*

W tej perspektywie, CAS nie jest ostatecznym celem, lecz pierwszym, niezbędnym krokiem ku prawdziwie zunifikowanej, ucielesnionej i dynamicznej nauce o umyśle, zdolnej do dialogu z filozofią i wyzwaniami ery AI.

---

<sup>2</sup> Przezczaszkowa stymulacja magnetyczna (*Transcranial Magnetic Stimulation*).

<sup>3</sup> Warto przy tym zauważyc, że ta fundamentalna predykcja już teraz znajduje nieintencjonalne, dwukierunkowe wsparcie w kluczowych badaniach. Z jednej strony, Brascamp i in. (2015), realizując paradygmat *no-post-perceptual-cognition* (Block, 2019), wykazali, że zmiana w treści percepcji (Q) może zachodzić bez zaangażowania sieci czołowo-ciemieniowej, kojarzonej z procesami  $\Sigma$ . Z drugiej strony, wyniki projektu COGITATE (Melloni i in., 2025) pokazały, że maszyneria kojarzona z  $\Sigma$  w korze przedczoloowej operuje na abstrakcyjnej kategorii, nie dekodując pełnego, bogatego Q.

- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2), 227–247. <https://doi.org/10.1017/S0140525X00038188>
- Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, 15(12), 567–575. <https://doi.org/10.1016/j.tics.2011.11.001>
- Block, N. (2019). What Is Wrong with the No-Report Paradigm and How to Fix It. *Trends in Cognitive Sciences*, 23(12), 1003–1013. <https://doi.org/10.1016/j.tics.2019.10.001>
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624–652. <https://doi.org/10.1037/0033-295X.108.3.624>
- Brascamp, J., Blake, R., & Knapen, T. (2015). Negligible fronto-parietal BOLD activity accompanying unreportable switches in bistable perception. *Nature Neuroscience*, 18(11), 1672–1678. <https://doi.org/10.1038/nn.4130>
- Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the Higher-Order Approach to Consciousness. *Trends in Cognitive Sciences*, 23(9), 754–768. <https://doi.org/10.1016/j.tics.2019.06.009>
- Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., Casarotto, S., Bruno, M.-A., Laureys, S., Tononi, G., & Massimini, M. (2013). A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior. *Science Translational Medicine*, 5(198). <https://doi.org/10.1126/scitranslmed.3006294>
- Chis-Ciure, R., Melloni, L., & Northoff, G. (2024). A measure centrality index for systematic empirical comparison of consciousness theories. *Neuroscience & Biobehavioral Reviews*, 161, 105670. <https://doi.org/10.1016/j.neubiorev.2024.105670>
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204. <https://doi.org/10.1017/S0140525X12000477>
- Cogitate Consortium, Ferrante, O., Gorska-Klimowska, U., Henin, S., Hirschhorn, R., Khalaf, A., Lepauvre, A., Liu, L., Richter, D., Vidal, Y., Bonacchi, N., Brown, T., Sripad, P., Armendariz, M., Bendtz, K., Ghafari, T., Hetenyi, D., Jeschke, J., Kozma, C., ... Melloni, L. (2025). Adversarial testing of global neuronal workspace and integrated information theories of consciousness. *Nature*, 642(8066), 133–142. <https://doi.org/10.1038/s41586-025-08888-1>
- Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Penguin Publishing Group.
- Dennett, D. C. (1991). *Consciousness explained* (1st ed). Little, Brown and Co.
- Duch, W. (2005). Brain-Inspired Conscious Computing Architecture. *The Journal of Mind and Behavior*, 26(1/2), 1–21. JSTOR.

- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00443>
- Gazzaniga, M. S. (2000). Cerebral specialization and interhemispheric communication: Does the corpus callosum enable the human condition? *Brain*, 123(7), 1293–1326. <https://doi.org/10.1093/brain/123.7.1293>
- Graziano, M. S. A., & Webb, T. W. (2015). The attention schema theory: A mechanistic account of subjective awareness. *Frontiers in Psychology*, 06. <https://doi.org/10.3389/fpsyg.2015.00500>
- Hirao, T., Miyamae, M., Matsuyoshi, D., Inoue, R., Takado, Y., Obata, T., Higuchi, M., Tsuchiya, N., & Yamada, M. (2025). A neuroimaging dataset during sequential color qualia similarity judgments with and without reports. *Scientific Data*, 12(1), 389. <https://doi.org/10.1038/s41597-025-04511-0>
- Kahneman, D. (2011). *Thinking, fast and slow* (1st ed). Farrar, Straus and Giroux.
- Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, 10(11), 494–501. <https://doi.org/10.1016/j.tics.2006.09.001>
- Lennie, P. (2003). The Cost of Cortical Computation. *Current Biology*, 13(6), 493–497. [https://doi.org/10.1016/S0960-9822\(03\)00135-0](https://doi.org/10.1016/S0960-9822(03)00135-0)
- McCloskey, M., & Cohen, N. J. (1989). Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. In *Psychology of Learning and Motivation* (T. 24, s. 109–165). Elsevier. [https://doi.org/10.1016/S0079-7421\(08\)60536-8](https://doi.org/10.1016/S0079-7421(08)60536-8)
- Mudrik, L., Boly, M., Dehaene, S., Fleming, S. M., Lamme, V., Seth, A., & Melloni, L. (2025). Unpacking the complexities of consciousness: Theories and reflections. *Neuroscience & Biobehavioral Reviews*, 170, 106053. <https://doi.org/10.1016/j.neubiorev.2025.106053>
- Sarasso, S., Casali, A. G., Casarotto, S., Rosanova, M., Sinigaglia, C., & Massimini, M. (2021). Consciousness and complexity: A consilience of evidence. *Neuroscience of Consciousness*, 2021(2), niab023. <https://doi.org/10.1093/nc/niab023>
- Sperry, R. W. (1968). Hemisphere disconnection and unity in conscious awareness. *American Psychologist*, 23(10), 723–733. <https://doi.org/10.1037/h0026839>
- Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought*. Yale University Press.
- Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7), 450–461. <https://doi.org/10.1038/nrn.2016.44>