KOGNITYWISTYKA – CYBERNETYKA

Wojciech Załuski

Nauki kognitywne a filozofia prawa

1. Wprowadzenie

Terminem „nauki kognitywne” określa się interdyscyplinarny program badawczy wykorzystujący osiągnięcia m.in. psychologii poznawczej, *neuroscience*, teorii sztucznej inteligencji, teorii systemów, lingwistyki i psychologii ewolucyjnej w celu wyjaśnienia procesów poznawczych człowieka. Można spotkać się z opinią, że nauki kognitywne stanowią swoistą rewolucję w myśleniu o człowieku, i że w rezultacie mogą – i powinny – zastąpić filozoficzną refleksję o człowieku, albo przynajmniej przyczynić się do rozstrzygnięcia tradycyjnych problemów filozoficznych. Opinię taką wyraża cytat z dzieła dwóch znanych kognitywistów – G. Lakoffa i M. Johnsona:

Umysł jest z natury ucieleśniony. Myślenie jest w przeważającej mierze nieświadome. Pojęcia abstrakcyjne są najczęściej metaforyczne. Oto trzy główne odkrycia, których pod koniec XX wieku dokonał kognitywizm. Na skutek tych odkryć filozofia zmieniła się nieodwołalnie. Ponad dwa tysiąclecia spekulacji na temat tych aspektów rozumu dobiegły końca. Koncepcje filozoficzne sprzeczne z odkryciami kognitywizmu – czyli większość – muszą się teraz do nich dostosować¹.

Ta dość buńczucznie wyrażona opinia jest nie do przyjęcia z dwóch powodów. *Primo*, owe trzy „odkrycia” sformułowane są w bardzo niejasny i ogólnikowy sposób, tak że można nadawać im różnoraki sens: jeśli nada się im jednoznaczny sens niestępiony niejasnymi kwalifikacjami w rodzaju „w przeważającej mierze” czy „najczęściej” (co pozwoliłoby, być może, obronić tezę o ich „rewolucyjnym”

¹ G. Lakoff, M. Johnson, *Co kognitywizm wnosi do filozofii?*, [w:] *idem, Kognitywizm – czy jest ciało w duszy?*, Kraków 1999, s. 25 (cytat pochodzi z książki G. Lakoffa i M. Johnsona, *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*, New York 1999).

znaczeniu dla filozofii), są one ewidentnie fałszywe (myślenie nie musi być zdeterminowane przez cielesność, nie każde myślenie jest metaforyczne, istnieje myślenie świadome); jeśli zaś potraktuje się je dosłownie, są one trywialne, pozbawione „rewolucyjnego potencjału”. *Secundo*, nie jest prawdą, że te trzy tezy są powszechnie przyjmowane w ramach kognitywizmu (być może poza tezą drugą, iż myślenie ma w dużej części charakter nieświadomy): np. wielu kognitywistów nie zgadza się z tezą, że umysł jest „ucieleśniony” (czyli że każdy przypadek rozumowania jest „zakotwiczony w ciele”, co chyba ma znaczyć według Lakoffa i Johnsona, że ciało niejako wyznacza granice rozumowania i że nie istnieje myślenie bezinteresowne, tj. prowadzone niezależnie od potrzeb ciała); innymi słowy, teza Lakoffa i Johnsona o ucieleśnionym umyśle jest charakterystyczna tylko dla jednego z kilku modeli umysłu (dynamicznego) zaproponowanych w ramach kognitywizmu² (ten i inne modele przedstawimy dokładniej w pkt 2). Podsumowując: zdecydowanie nie podzielamy entuzjastycznego stanowiska w kwestii doniosłości nauk kognitywnych dla filozofii. Nasz sceptycyzm spróbujemy szerzej uzasadnić w pozostałej części niniejszego opracowania, choć nie w kontekście filozofii w ogóle, lecz w kontekście filozofii prawa. Niemniej, zważywszy, iż zagadnienia istotne dla filozofii prawa (takie jak np. odpowiedzialność, podmiotowość czy wolna wola) są badane również w ramach innych gałęzi filozofii (np. filozofii umysłu czy filozofii moralności), nasze wnioski o filozoficznych implikacjach wyników nauk kognitywnych sformułowane w kontekście filozofii prawa dają się, jak sądzimy, uogólnić także na inne gałęzie filozofii. Nasze stanowisko można przedstawić w postaci trzech tez:

1) nauki kognitywne są nauką *in statu nascendi*, nie wypracowały bowiem do tej pory jednolitego paradygmatu i nie dostarczają jednolitego obrazu człowieka, w związku z czym na obecnym etapie swojego rozwoju mają niewielkie znaczenie dla filozofii prawa;

2) istnieją strategie aplikacji nauk kognitywnych w obszarze filozofii prawa, które pozwalają zminimalizować lub zneutralizować ryzyko błędów i uproszczeń związanych z odwołaniem się do nauki, która wciąż nie dysponuje jednolitym paradygmatem;

3) nawet jednak na dojrzałym etapie swojego rozwoju nauki kognitywne będą miały prawdopodobnie niewielkie znaczenie dla filozofii prawa³.

Pozostałe punkty niniejszego opracowania stanowią rozwinięcie tych trzech tez.

² Dla określenia nauk kognitywnych stosuje się także termin „kognitywistyka”. Jest on bardziej jednoznaczny niż termin „kognitywizm”, którym określa się także pewne stanowisko metafizyczne (przypisujące wypowiedziom normatywnym wartość poznawczą); przyjęcie go ma jednak tę niefortunną konsekwencję językową, że prowadzi do nazwania badaczy uprawiających nauki kognitywne „kognitywistkami” (termin „kognitywiści” jest niewątpliwie bardziej elegancki).

³ W Polsce kwestia relacji między naukami kognitywnymi i filozofią prawa nie była do tej pory przedmiotem szerszej debaty naukowej. Powstało także niewiele opracowań naukowych na ten temat. Z istniejących opracowań należy wymienić przede wszystkim książkę T. Pietrzykowskiego, *Intuicja prawnicza. W stronę zewnętrznej integracji teorii prawa*, Warszawa 2012.

2. Kognitywizm jako nauka *in statu nascendi*

Nauki kognitywne nie wypracowały jednolitego paradygmatu: chcieć wyprowadzać mocne i jednoznaczne wnioski na temat człowieka z nauk kognitywnych to tak, jakby chcieć wyprowadzać wnioski na temat świata z fizyki klasycznej przed powstaniem dzieła Newtona *Philosophiae Naturalis Principia Mathematica*. Tezę, iż nauki kognitywne nie wypracowały jednolitego paradygmatu, uzasadnimy na dwóch przykładach: pokażemy, że nauki kognitywne nie oferują jednolitej teorii umysłu oraz jednolitej teorii racjonalności.

2.1. Umysł

W ramach nauk kognitywnych można wyróżnić trzy główne modele umysłu: komputacyjny, konekcjonistyczny i dynamiczny. **Model komputacyjny** zakłada, że umysł jest swego rodzaju komputerem cyfrowym, a myślenie obliczaniem, tj. prowadzeniem w oparciu o określone reguły operacji na symbolicznych reprezentacjach. Na model komputacyjny składają się więc dwie teorie: reprezentacyjna teoria umysłu (wedle której stany mentalne odnoszą się do mentalnych reprezentacji, tj. myśli, pragnienia, przekonania, percepcje są relacjami do symbolicznie reprezentowanych treści mentalnych) oraz komputacyjna teoria myślenia (wedle której myślenie jest obliczaniem). Problemy tego modelu są wielorakie – wymienimy dwa z nich. Po pierwsze, ludzki umysł wykazuje rozumienie, a obliczanie – operowanie na symbolach zgodnie z określonymi regułami – nie jest wystarczające do rozumienia (argument z chińskiego pokoju J. Searle'a). Po drugie, zgodnie z twierdzeniem Löwenheim-Skolema-Tarskiego syntaktyka nie determinuje jednoznacznie semantyki⁴. Argumenty te są oczywiście powiązane ze sobą, tj. twierdzenie Löwenheim-Skolema-Tarskiego to formalny wynik wspierający argument z chińskiego pokoju (gdyby bowiem dało się w pełni odwzorować semantykę w syntaktyce, intuicyjny argument Searle'a straciłby wiele ze swojej siły perswazyjnej). **Model konekcjonistyczny** zakłada, że informacja nie jest magazynowana w umyśle

W książce tej autor wnikliwie bada znaczenie osiągnięć nauk kognitywnych dla filozofii prawa, dochodząc do wniosków bardziej optymistycznych niż bronione w niniejszej pracy (stanowisko Pietrzykowskiego można określić mianem „umiarkowanego entuzjazmu” w kwestii znaczenia nauk kognitywnych dla filozofii prawa).

⁴ Zgodnie z tym twierdzeniem, każda sformalizowana teoria I rzędu ma model przeliczalny wtedy i tylko wtedy, gdy ma model dowolnie nieskończony. Wynika stąd, że żadna sformalizowana teoria I rzędu nie wyznacza w sposób jednoznaczny swojego modelu – posiada zawsze model także w teorii liczb. Na „niedookreślenie” semantyki przez syntaktykę wskazują także inne twierdzenia limitacyjne, m.in. twierdzenie Tarskiego o niedefiniowalności prawdy, które mówi, że jeśli *T* jest teorią niesprzeczną, to zbiór zdań prawdziwych tej teorii nie daje się zdefiniować w *T*. Por. J. Woleński, *O filozoficznym sensie metamatematycznych twierdzeń limitacyjnych*, „Wiadomości Matematyczne” 2009, nr 45 (2), s. 195–216.

w postaci ciągu symbolicznych reprezentacji, lecz jest „rozproszona” w sieci neuronowej⁵. W związku z tym myślenie nie jest manipulowaniem symbolami zgodnie z określonymi regułami, lecz ma charakter płynny i zachodzi na poziomie sub-symbolicznym: jest modelowane przez stopniowe zmiany zachodzące w sile połączeń („wagach”) między jednostkami sieci neuronowej. W ramach tego modelu odrzuca się więc obie teorie składające się na model komputacyjny: stany intencjonalne nie są określonymi relacjami do mentalnych reprezentacji, a myślenie nie jest obliczaniem. **Model dynamiczny**⁶ zakłada, że systemy poznawcze są dynamiczne i złożone, obejmują bowiem system nerwowy, ciało i środowisko, w jakim funkcjonują. Według tego modelu umysł jest więc nie tylko „ucieleśniony” (*embodied*), lecz także „osadzony” (*embedded*) w określonym środowisku⁷. W ramach tego modelu odrzuca się tezę o komputacyjnym charakterze procesów psychicznych: procesy te nie są ciągami dyskretnych stanów symbolicznych, lecz ciągłymi, podlegającymi ewolucji stanami pewnego dynamicznego i złożonego systemu. Nie będziemy omawiać problemów obu modeli; odnotujemy jedynie, że są one nie mniej poważne niż problemy modelu komputacyjnego. Nie jest w pełni jasne, jakie są relacje między tymi modelami. Można bronić dwóch skrajnych poglądów: że modele te wzajemnie się wykluczają oraz że można je w jakiś sposób ze sobą pogodzić, np. przyjmując, iż modele komputacyjny i konekjonistyczny opisują dwa odmienne, ale komplementarne, poziomy funkcjonowania umysłu – odpowiednio: symboliczny i sub-symboliczny. W tym ujęciu sieć neuronowa byłaby implementacją na poziomie sub-symbolicznym procesów obliczeniowych dokonywanych na reprezentacjach mentalnych; umysł byłby więc jednocześnie siecią neuronową i – na wyższym poziomie opisu – instrumentem przetwarzania ciągów symboli. Można także argumentować, że model konekjonistyczny jest pewnym szczególnym wariantem modelu dynamicznego. Są to jednak czyste spekulacje; jest bowiem jasne, że nauki kognitywne są wciąż na etapie poszukiwania paradygmatycznego modelu umysłu.

⁵ Por. D. E. Rumelhart, *The architecture of mind: A connectionist approach*, [w:] M. J. Posner (ed.), *Foundations of Cognitive Science*, Cambridge MA 1993, s. 133–159.

⁶ Por. T. J. van Gelder, *Dynamic approaches to cognition*, [w:] R. Wilson, F. Keil (red.), *The MIT Encyclopedia of Cognitive Sciences*, Cambridge MA 1999, s. 244–246; idem, *What might cognition be, if not computation?*, „Journal of Philosophy” 1995, vol. 92, no. 7, s. 345–381.

⁷ Teza, że umysł ludzki jest *embedded*, oznacza, że pierwotną i podstawową relacją człowieka wobec świata nie jest relacja „kartezjańskiego” obserwatora, który najpierw tworzy w umyśle obraz świata, a potem dopiero w nim działa, lecz relacja „bycia w świecie”, umiejętnego „radzenia sobie” ze światem, wykonywania w nim określonych czynności bez konieczności tworzenia jego symbolicznych reprezentacji. Ta antykartezjańska teza została na gruncie filozofii kontynentalnej najwnikliwiej i najpełniej rozwinięta przez M. Heideggera. Dlatego też filozofia Heideggera stała się źródłem inspiracji dla pewnej grupy kognitywistów (por. zwłaszcza H. Dreyfus, *Being-in-the-World: A Commentary on Heidegger's „Being and Time”, Division I*, Cambridge MA 1990).

2.2. Racjonalność

Można często spotkać się ze stwierdzeniem, że nauki kognitywne podważają założenie o racjonalności człowieka (rozumianej, zgodnie z teorią racjonalnego wyboru, jako działanie w sposób dający się matematycznie modelować jako maksymalizowanie funkcji użyteczności), gdyż ujawniają, że ludzie nie postępują zgodnie z zasadą maksymalizacji użyteczności, lecz działają niejako „na skróty”, wykorzystując w swoich procesach decyzyjnych różne heurystyki (np. dostępności, reprezentatywności, zakotwiczenia), co naraża ich na różne błędy (*biases*) i iluzje (np. błąd nadmiernego optymizmu, błąd pewności siebie, błąd opóźnionej predykcji (*hindsight bias*), iluzję pozornych korelacji, iluzję kontroli). Twierdzi się także niekiedy, że argumentem za słusznością tego – dość pesymistycznego – ujęcia poziomu racjonalności człowieka jest fakt, że pozwala ono lepiej niż tradycyjny model racjonalnego wyboru przewidywać ludzkie zachowania. Żadna z tych dwóch tez nie wydaje się przekonująca. Co do pierwszej tezy, w ramach nauk kognitywnych rozwijane są inne modele racjonalności – bardziej optymistyczne w kwestii racjonalności zachowań ludzi – niż opisany wyżej **model *Heuristics and biases*** D. Kahnemana i A. Tversky'ego. Takim modelem jest np. **model ekologicznej racjonalności** G. Gigerenzer. Gigerenzer dokonał reinterpretacji różnych eksperymentów mających przemawiać za tezą o ludzkiej nieracjonalności (np. eksperymentów dotyczących przetwarzania informacji probabilistycznych), argumentując, że eksperymenty te zostały przeprowadzone w sztucznych warunkach laboratoryjnych i jako takie nie mogą być przywoływane dla wsparcia tej tezy. Zdaniem Gigerenzer, ludzie bardzo dobrze radzą sobie z zadaniami „na racjonalność”, o ile zadania te przeprowadzane są w sytuacjach realistycznych (zwłaszcza takich, które nie odbiegają istotnie pod względem struktury od tych, w jakich w odległej przeszłości kształtowały się zdolności poznawcze naszych przodków)⁸. Ponadto, w ramach nauk kognitywnych wciąż silny pozostaje **model teorii racjonalnego wyboru**, zgodnie z którym człowiek działa w taki sposób, jakby maksymalizował swoją funkcję użyteczności. Także druga teza jest mało przekonująca, model *Heuristics and biases* nie pozwala bowiem na dokonywanie jednoznacznych predykcji ludzkich zachowań. Wystarczy zauważyć, że: heurystyki często się „znoszą” (np. heurystyka dostępności i heurystyka reprezentatywności⁹); są niekiedy sprzeczne z wynikami badań na temat człowieka prowadzonymi w ramach psychologii

⁸ Por. np. G. Gigerenzer, *Gut Feelings. Short Cuts to Better Decision Making*, London 2007.

⁹ Wyobraźmy sobie, że Łukasz gra w ruletkę. Pięć razy z rzędu wypadło czarne. Na co Łukasz postawi za szóstym razem? Jeśliby kierował się heurystyką reprezentatywności (a więc oczekiwał, że sekwencja kolorów będzie odzwierciedlać losowy charakter mechanizmu generującego kolory), postawiłby na czerwone; jeśli natomiast kierowałby się heurystyką dostępności (a więc przewidywał wynik na podstawie najbardziej „dostępnych” w pamięci wyników), to na czarne. Model *Heuristics and biases* nie mówi, która z tych heurystyk zaktywizuje się w tej sytuacji, nie pozwala więc sformułować jednoznacznej predykcji zachowania Łukasza.

ewolucyjnej (np. błąd nadmiernego optymizmu jest sprzeczny z tezą psychologów ewolucyjnych mówiącą o skłonności ludzi do pesymizmu, zaś oba te zjawiska w kontekście zachowań inwestorów na giełdzie są sprzeczne z heurystyką reprezentatywności i wynikającym z niej efektem gracza); nikt nie oszacował „wag” poszczególnych skonfliktowanych heurystyk lub błędów, a jeśli da się to w ogóle zrobić, to zapewne wyłącznie *de casu ad casum*, tj. dla konkretnych sytuacji decyzyjnych, w jakich znajdują się konkretne osoby, nie zaś dla typów sytuacji (a tylko wtedy, gdyby dało się je wyznaczyć dla typów sytuacji, można byłoby je wykorzystać do dokonywania predykcji). Nie wiadomo także, jak mocną tezę na temat ludzkiej nieracjonalności implikuje model *Heuristics and biases*. Kahneman przyznaje, że procesy poznawcze mają dualną naturę¹⁰: można je podzielić na procesy Systemu 1 (w którym zakorzenione są heurystyki i błędy) – szybkie, holistyczne, automatyczne, nieświadome, ewolucyjnie starsze, niewymagające wiele mentalnego wysiłku – oraz procesy Systemu 2 – relatywnie powolne, oparte na regułach, kontrolowane, wymagające więcej wysiłku mentalnego, deliberatywne¹¹. Taki podział implikuje jednak, że ludzie mogą dokonywać korekt błędów wygenerowanych przez System 1. Nie wiadomo więc, jak głęboko owa – mniej lub bardziej silna – nieracjonalność ujawniona przez *Heuristics and biases* tkwi w ludzkiej naturze. Poza tym, psychologia kognitywna pokazuje, iż ludzie mogą istotnie różnić się między sobą pod względem racjonalności. Warto wspomnieć, że rezultaty badań eksperymentalnych prowadzonych w ramach modelu *Heuristics and biases* są często przywoływane jako argument na rzecz paternalizmu prawnego. Naszym zdaniem, taki sposób wykorzystania tego modelu w rozważaniach filozoficznoprawnych jest nieuprawniony. Po pierwsze, *Heuristics and biases* jest jednym z trzech modeli w ramach nauk kognitywnych wyodrębnionych w kontekście pytania o ludzką racjonalność; pozostałe dwa modele (model racjonalnego wyboru i model ekologicznej racjonalności) nie implikują tezy o ludzkiej nieracjonalności – tezy mówiącej (jak ujął to Gigerenzer w swoim krytycznym omówieniu paradygmatu *Heuristics and biases*), że „the untutored mind is running on a shoddy software, that is, on programs that work only with a handful of heuristics”¹²; nie mogą być więc wykorzystane w uzasadnianiu paternalizmu prawniczego (jeśli ludzie są racjonalni, sami najlepiej potrafią zatroszczyć się o własne interesy). Po drugie, nawet jeśli model ten okazałby się prawdziwy, wciąż nie byłoby jasne, jak głęboko nieracjonalność jest zakorzeniona w naszej naturze, tj. w jakim stopniu System 1 podlega korekcie ze strony Systemu 2. Jeśli zakorzenienie to nie okazałoby się zbyt głębokie, to być może odpowiednia edukacja wystarczyłaby do tego, aby ją „zneutralizować” i paternalistyczna

¹⁰ Por. np. D. Kahneman, *Thinking, Fast and Slow*, New York 2011.

¹¹ Szczegółową analizę tych dwóch systemów można znaleźć w: K. E. Stanovich, *Who Is Rational? Studies of Individual Differences in Reasoning*, New York 1999.

¹² G. Gigerenzer, *On cognitive illusions and rationality*, „Poznan Studies in the Philosophy of the Sciences and the Humanities” 1991, vol. 21, s. 235.

ingerencja prawna nie byłaby konieczna. Po trzecie, jak już wspomnieliśmy, istnieją różnice indywidualne między ludźmi, jeśli chodzi o stopień ich racjonalności. Czy miałyby stąd wynikać, że w stosunku do niektórych paternalizm byłby pożądany, a w stosunku do innych nie? Jest jednak jasne, że prawo nie może wprowadzać dystynkcji między dorosłymi ludźmi na mniej i bardziej racjonalnych, i adresować określone reguły paternalistyczne tylko do tych pierwszych. Sam pomysł tego rodzaju wydaje się absurdalny¹³. Podsumowując: nauki kognitywne wciąż nie udzielają jednoznacznej odpowiedzi na pytanie o racjonalność człowieka, dlatego ich wykorzystywanie w celu uzasadnienia takiej czy innej polityki prawnej jest co najmniej ryzykowne, jeśli nie zupełnie nieuprawnione.

3. Strategie aplikacji nauk kognitywnych w obszarze prawa

Można wyróżnić dwa rodzaje strategii „bezpiecznego” wykorzystywania nauk kognitywnych na obecnym etapie ich rozwoju w filozofii prawa. Pierwszy rodzaj strategii pozwala zminimalizować ryzyko uproszczeń związanych z aplikacją „nauki bez paradygmatu” w obszarze filozofii prawa, drugi pozwala owe uproszczenia zneutralizować. **Strategia minimalizacji uproszczeń** polega na prowadzeniu dokładnej analizy metodologicznej szczegółowych wyników nauk kognitywnych i na badaniu ich możliwych implikacji dla filozofii prawa. W ramach tej strategii nie podejmuje się próby budowania ogólnego obrazu człowieka w oparciu o nauki kognitywne i analizowania przez pryzmat tego obrazu fundamentalnych problemów filozofii prawa. Wnioski na temat prawa prowadzone na podstawie tego rodzaju strategii są, by tak rzec, mało efektowne: zazwyczaj dotyczą raczej drugorzędnych problemów filozofii prawa, są mało wyraziste, opatrzone wieloma metodologicznymi zastrzeżeniami. Przykładem takiej strategii może być przeprowadzona w poprzednim punkcie, szkicowa analiza znaczenia modelu *Heuristics and biases* dla filozoficznoprawnego pytania o dopuszczalność paternalizmu prawnego. Innym przykładem takiej aplikacji jest analiza prawniczych negocjacji w oparciu o teorię perspektyw Kahnemana i Tversky’ego (stanowiącą część modelu *Heuristics and biases*). Teoria perspektyw zakłada, że ludzie nie doszacowują średnich i wysokich prawdopodobieństw oraz przeszacowują niskie prawdopodobieństwa. Otóż na podstawie tej teorii można zbudować interesujący model negocjacji prawniczych, określający warunki, w jakich istnieje szansa na zawarcie ugody między stronami¹⁴. Kolejny przykład wiąże

¹³ Na temat problemów z wykorzystywaniem wyników nauk kognitywnych w uzasadnianiu paternalizmu prawnego por. w szczególności: R. H. Thaler, C. R. Sunstein, *Nudge. Improving Decisions about Health, Wealth, and Happiness*, New Haven–London 2008.

¹⁴ Por. Ch. Guthrie, *Prospect theory, risk preference and the law*, „Northwestern University Law Review” 2003, vol. 97, no. 3.

się z testowaniem ekonomicznego modelu składania pozwu. Wedle tego modelu poszkodowany złoży pozew, jeśli iloczyn prawdopodobieństwa jego zwycięstwa w sądzie i wysokości odszkodowania jest wyższy niż koszty procesu. Model ten, jak wiadomo, jest predykcyjnie słaby. Próbując wyjaśnić predykcijną nietrafność tego modelu, można odwołać się do osiągnięć nauk kognitywnych, które oferują kilka tłumaczących ją hipotez. Jedną z hipotez mówi, że ludzie kierują się retrybutywną złośliwością, która skłania ich do składania pozwów, nawet jeśli powyższa nierówność jest niespełniona, ponieważ w ten sposób budują reputację osoby, której nie można bezkarnie krzywdzić. Inną hipoteza odwołuje się do błędu nadmiernego optymizmu, który sprawia, że ludzie składają pozwy nawet wtedy, gdy obiektywne prawdopodobieństwo wygranej jest niskie. Można wreszcie argumentować, że ludzie troszczą się o status relatywny, a nie absolutny, jeśli więc poprzez złożenie pozwu wyrządzą większą szkodę przeciwnikowi niż sobie samym, złożenie pozwu będzie z ich perspektywy decyzją racjonalną¹⁵. **Strategia neutralizacji uproszczeń** ma inny charakter: polega ona na wyprowadzeniu wszystkich możliwych konsekwencji filozoficznych z określonego modelu umysłu czy racjonalności i zbadaniu ich znaczenia dla zasadniczych problemów dyskutowanych w ramach filozofii prawa. W ramach tej strategii wybiera się więc jeden z modeli, mając świadomość, że jest on jednym z kilku możliwych i że istnieją poważne argumenty naukowe przeciw niemu (a więc, że nie ma on statusu paradygmatu), i próbuje się w oparciu o niego budować pewną antropologię filozoficzną. Strategia ta ma tę przewagę nad pierwszą, że prowadzi do zbudowania pewnej wizji prawa (strategia minimalizacji uproszczeń jest skoncentrowana na problemach szczegółowych), jej słabością jednak jest oczywiście to, że generuje wizję człowieka i prawa, która ma charakter hipotetyczny (mówi więc: „jeśli model *x* jest prawdziwy, to antropologia filozoficzna *i*, w konsekwencji, ujęcie określonych problemów filozoficznoprawnych wyglądałoby mniej więcej tak”)¹⁶. Strategia jest więc swego rodzaju ćwiczeniem dedukcyjnym.

¹⁵ Por. O. D. Jones, *Evolutionary psychology and the law*, [w:] D. M. Buss (ed.), *The Handbook of Evolutionary Psychology*, New Jersey 2005, s. 953–974.

¹⁶ Jako przykład (nie w pełni świadomej) realizacji takiej strategii widzę dziś swoją książkę *Ewolucyjna filozofia prawa* (Warszawa 2009), będącą próbą zbudowania wizji prawa w oparciu o model komputacyjny (wzbogacony o założenia psychologii ewolucyjnej, zakładającej, że umysł ludzki ma strukturę silnie modułową, a moduły – darwinowskie algorytmy wyspecjalizowane w rozwiązywaniu różnego rodzaju problemów, wobec których stawiali nasi przodkowie w epoce plejstocenu – są biologicznymi adaptacjami). Niestety, w książce nie zaznaczyłem wyraźnie, że realizuję taką strategię, co wynikało stąd, że w okresie jej pisania byłem pod silnym wpływem ewolucyjnego myślenia o człowieku i sądziłem, że model ten jest w swoich zasadniczych punktach prawdziwy. Gdybym tę książkę pisał dziś, wskazałbym wyraźniej, że jest ona przykładem zastosowania takiej strategii: byłbym w niej bowiem dużo bardziej krytyczny wobec psychologii ewolucyjnej (w książce tej pisałem wprawdzie o niekompletności obrazu człowieka, którego dostarcza psychologia ewolucyjna, ale czyniłem to w sposób zbyt mało stanowczy).

4. Implikacje nauk kognitywnych dla filozofii prawa

Stwierdziliśmy, że nauki kognitywne nawet na dojrzałym etapie ich rozwoju, tj. takim, w którym wykrystalizuje się ich paradygmat¹⁷, będą mieć jedynie niewielkie znaczenie dla filozofii prawa, tzn. nie doprowadzą do jakichś zasadniczych czy choćby nawet po prostu istotnych zmian w sposobie myślenia o prawie. Można tę tezę uzasadniać na dwa różne sposoby, w zależności od przyjmowanego stanowiska w kwestii wpływu rezultatów nauk empirycznych na rozstrzyganie problemów filozoficznych. Według pierwszego stanowiska wyniki badań naukowych nie mają bezpośredniego znaczenia dla rozstrzygania problemów filozoficznych – aby móc wykorzystać te wyniki w argumentacji na rzecz określonego poglądu filozoficznego, trzeba poddać je interpretacji filozoficznej (niejako przełożyć na język, w jakim dany problem filozoficzny jest sformułowany), a przyjęta ostatecznie interpretacja jest zawsze jedną z kilku równo uprawnionych. Przykładowo: wyniki słynnych eksperymentów Libeta nie mają (wbrew często głoszonej tezie) bezpośrednich konsekwencji dla filozoficznego sporu o istnienie wolnej woli – można je interpretować zarówno jako wspierające tezę o istnieniu wolnej woli, jak i tezę o jej niestnieniu¹⁸. Tak więc wedle tego stanowiska wszelkie problemy filozoficzne mające znaczenie dla filozofii prawa (obok problemu wolnej woli można wymienić tutaj także problem podmiotowości, tj. pytanie, czy człowieka można uznawać za autora własnych działań, za które może ponosić odpowiedzialność, oraz problem relacji umysł – mózg) nie mogą być rozstrzygnięte bezpośrednio przez odwołanie się do wyników nauk empirycznych; nauki te w istotnym sensie (tj. dlatego, że wymagają interpretacji, a te mogą być zwykle bardzo odmienne) pozostają irrelewantne dla tych sporów. Według drugiego stanowiska wyniki nauk empirycznych mają bezpośrednie znaczenie dla rozstrzygania problemów filozoficznych – nawet więc jeśli jest tak, iż wyniki te należy poddać interpretacji, aby można je było odnieść do tych problemów,

¹⁷ Bądź też paradygmaty, jeśli przyjmiemy, że dwa fundamentalne obszary badań w ramach nauk kognitywnych – umysł i racjonalność – wymagają odrębnych paradygmatów (choć, oczywiście, paradygmaty te muszą być spójne, aby można było mówić o „dojrzałym” etapie rozwoju nauk kognitywnych).

¹⁸ Za tym, że stanowisko to jest częściowo słuszne, przemawia także fakt, iż przedstawiciele nauk kognitywnych różnią się w opisie konsekwencji filozoficznych omówionych wyżej modeli umysłu. Zwolennicy jednego modelu mogą przyjmować różne stanowiska w podstawowych kwestiach z zakresu filozofii umysłu. Przykładowo, J. Fodor jest – podobnie jak P. Churchland i D. Dennett – zwolennikiem komputacyjnego modelu umysłu, ale o ile Fodor przyjmuje przekonanie o istnieniu „wewnętrznej (*intrinsic*) intencjonalności” i o trafności *folk psychology*, o tyle Churchland i Dennett przekonanie to odrzucają. Z kolei J. Searle, który odrzuca komputacyjny model umysłu, dzieli z Fodorem przekonanie o istnieniu „wewnętrznej intencjonalności” i o trafności *folk psychology*. Zapewne jest tak, że niektóre wyniki nauk empirycznych mają bardziej bezpośrednie implikacje filozoficzne, inne zaś mniej; prawdziwe mogą być więc oba stanowiska (w odniesieniu do różnych rezultatów badań naukowych).

zakres uprawnionych interpretacji jest wąski, bo często ogranicza się do jednej interpretacji, wspierającej określony pogląd na dany problem filozoficzny. Ponieważ wielu najbardziej znanych przedstawicieli nauk kognitywnych (np. Daniel Dennett, Paul Churchland, Patricia Churchland) w swoich tekstach filozoficznych przywołuje wyniki nauk kognitywnych jako argument na rzecz monistyczno-materialistycznych¹⁹ rozstrzygnięć wspomnianych problemów filozoficznych (wolnej woli, podmiotowości, relacji umysł – mózg), mogłoby się wydawać, że drugie stanowisko jest słuszne i, co ważniejsze, że prowadzi ono właśnie do monistyczno-materialistycznych rozstrzygnięć tych problemów. Gdyby rzeczywiście rzeczy tak się miały, to nasza teza o niewielkim znaczeniu wyników nauk kognitywnych dla filozofii prawa byłaby nietrafna, albowiem monistyczno-materialistyczne rozstrzygnięcia tych problemów (oznaczające w istocie odrzucenie klasycznego rozumienia pojęcia wolnej woli, pojęcia podmiotowości, kauzalnej efektywności stanów mentalnych, oraz uznanie moralności za biologiczną adaptację, tzn. za pewien zespół dyspozycji służący niczemu więcej niż utrzymaniu leżących u jego podstaw genów w puli genowej populacji) wymagałyby zasadniczej przebudowy siatki pojęciowej, na jakiej wspiera się tradycyjne rozumienie prawa. Uważamy jednak, że teza, iż nauki kognitywne prowadzą do monistyczno-materialistycznych rozstrzygnięć wspomnianych problemów jest nie do utrzymania, nawet jeśli przyjmie się, że słuszne jest stanowisko drugie, tj. że wyniki nauk kognitywnych mogą bezpośrednio wspierać określone stanowisko filozoficzne. Istnieją bowiem silne argumenty naukowe na rzecz tezy idącej pod prąd dominującego obecnie monistyczno-materialistycznego „klimatu opinii”, tzn. tezy mówiącej, iż nauki kognitywne wspierają raczej tradycyjne rozwiązania wspomnianych problemów filozoficznych. Przykładowo: badania prowadzone na pograniczu mechaniki kwantowej i psychologii poznawczej pokazują, że pewne zjawiska (np. skuteczność różnych technik kognitywno-behawioralnych, takich jak ukierunkowana uwaga czy poznawcze redefiniowanie sytuacji, w regulowaniu naszych reakcji emocjonalnych, leczeniu zaburzeń obsesyjno-kompulsywnych, depresji czy fobii) można przekonująco tłumaczyć poprzez przyjęcie jakiejś wersji dualizmu umysł – mózg²⁰, i w efekcie poprzez uznanie kauzalnej efektywności stanów mentalnych. Budowane są wyrafinowane modele oparte na formalizmie mechaniki kwantowej, które opisują szczegółowo, w jaki sposób sfera mentalna (umysł) może oddziaływać na sferę fizyczną (mózg, ciało)²¹. Nie jest oczywiście tak,

¹⁹ Użycie tego mało zgrabnego terminu jest konieczne z uwagi na to, że monizm może być również niematerialistyczny (np. Berkeley był monistą niematerialistycznym).

²⁰ Stojąc na gruncie naukowym, trudno oczywiście bronić radykalnej – kartezyjskiej – wersji dualizmu, która zakłada całkowitą ontologiczną niezależność substancji umysłowej (*res cogitans*) od substancji cielesnej (*res extensa*). Jako akceptowalne wersje dualizmu pozostają więc niekartezyjskie odmiany dualizmu substancji i tzw. dualizm własności.

²¹ Por. np. J. Schwartz, H. Stapp, M. Beauregard, *Quantum physics in neuroscience and psychology: a neurophysical model of mind-brain interaction*, „Philosophical Transactions

że modele te negują występowanie kauzalnego oddziaływania sfery fizycznej na mentalną; niemniej pokazują, że – wbrew panującym dogmatom monistyczno-materialistycznym – teza o odwrotnym wpływie kauzalnym (sfery mentalnej na fizyczną) i o pewnej autonomiczności sfery mentalnej względem fizycznej jest w świetle współczesnej nauki (zwłaszcza mechaniki kwantowej) nie tylko uprawniona, lecz – co ważniejsze – bardziej uprawniona niż teza przeciwna (która może przybrać wiele różnych postaci, od najbardziej radykalnych: materializmu eliminacyjnego i epifenomenalizmu, do bardziej umiarkowanych: różnych wersji materializmu emergentystycznego). Teza ta ma tę dodatkową zaletę, iż, w przeciwieństwie do monistyczno-materialistycznego ujęcia relacji umysł – mózg, nie jest kontrintuicyjna, współbrzmi bowiem z naszymi „spontanicznymi” przekonaniami: że stany mentalne nie są tożsame ze stanami pewnego systemu fizycznego (mózgu czy – szerzej – mózgu i ciała), że stany mentalne są kauzalnie skuteczne, że wolna wola istnieje i że podmiotowość nie jest iluzją. Podsumujmy, trawestując słynną myśl Pascala: wydaje się, że monizm materialistyczny jest oznaką siły myśli, ale tylko do pewnego stopnia²². Oczywiście, powyższe uwagi mają jedynie szkicowy charakter (szczegółowe rozwinięcie tezy o dualistycznych implikacjach nauk kognitywnych wykracza poza zakres niniejszego tekstu); niemniej wydają się wystarczające, aby pokazać, że teza, iż nauki kognitywne prowadzą do monistyczno-materialistycznych rozstrzygnięć wspomnianych problemów filozoficznych, nie tylko nie jest oczywista, ale jest prawdopodobnie fałszywa. Wydaje się bowiem, że nauki te (w zależności od tego, jakie stanowisko przyjmujemy w kwestii znaczenia wyników nauk empirycznych dla rozstrzygania problemów filozoficznych) albo nie mają żadnych wyraźnych implikacji filozoficznych, albo mają raczej implikacje dualistyczne, i w rezultacie potwierdzają, a nie podważają, siatkę pojęciową, na której wspiera się tradycyjne myślenie o człowieku (i w efekcie o prawie). Rzecz jasna, jeśli prawdziwa jest teza druga (oparta na stanowisku drugim w kwestii znaczenia wyników nauk przyrodniczych dla rozstrzygania problemów filozoficznych), nasz pogląd o niewielkim znaczeniu nauk kognitywnych dla filozofii prawa musimy złagodzić, ale tylko nieznacznie: wyniki nauk kognitywnych mają znaczenie dla filozofii prawa, ale tylko o tyle, że potwierdzają trafność rozstrzygnięć filozoficznych leżących u podstaw „prawniczej siatki pojęciowej” (jest to modyfikacja nieznaczna także dlatego, że główne argumenty za tą siatką mają filozoficzny, a nie naukowy charakter).

of the Royal Society of London. Series B: Biological Sciences” 2005, vol. 29, no. 360 (1458), s. 1309–1327; J. Schwartz, S. Begley, *The Mind and the Brain: Neuroplasticity and the Power of Mental Force*, New York 2002; H. Stapp, *Mind, Matter and Quantum Mechanics*, Berlin–Heidelberg 2004; H. Stapp, *Mindful Universe: Quantum Mechanics and the Participating Observer*, Berlin–Heidelberg 2007.

²² Chodzi oczywiście o myśl następującą (nr 225): „Athéisme, marque de force d’esprit, mais jusqu’à un certain degré seulement”.

Powróćmy na zakończenie do kwestii badań nad racjonalnością, prowadzonych w ramach nauk kognitywnych, i zastanówmy się nad jej związkiem z przedstawionym wyżej poglądem o niewielkim znaczeniu nauk kognitywnych dla filozofii prawa. Otóż zarysowane w pkt 2.2 trzy deskryptywne modele racjonalności, zaproponowane w ramach tych nauk (tj. model teorii racjonalnego wyboru, model *Heuristics and biases* oraz model ekologicznej racjonalności), mają w istocie umiarkowane znaczenie filozoficzne, tzn. zwycięstwo któregośkolwiek z nich nie będzie wymuszało dokonania żadnej zmiany w tradycyjnym obrazie człowieka jako istoty wolnej i racjonalnej (obraz ten można nazwać „humanistycznym”). Jest tak dlatego, że modele te opisują inne pojęcie racjonalności niż pojęcie racjonalności przyjmowane w tym obrazie. Modele te odpowiadają na pytanie, czy człowiek jest racjonalny w sensie instrumentalnym, tj. czy właściwie dobiera środki do celów (a więc, czy maksymalizuje swoją funkcję użyteczności), natomiast w tradycyjnym obrazie człowieka przez racjonalność rozumie się zdolność formułowania racji i argumentów, reagowania na racje i argumenty wysuwane przez innych oraz podejmowania działań pod wpływem racji i argumentów (racjonalność tego rodzaju można nazwać „argumentacyjną”). Obraz ten byłby więc zagrożony tylko wtedy, gdyby nauki kognitywne podważały tezę, iż człowiek jest racjonalny w takim – fundamentalnym i charakterystycznym wyłącznie dla niego²³ – sensie, a więc gdyby wykazywały, że kierowanie się racjami i argumentami jest wyłącznie iluzją czy „epifenomenem”, gdyż zachowania ludzkie dają się całkowicie wyjaśnić na poziomie czysto fizykalnym, tj. na poziomie relacji kauzalnych między stanami mózgu. Próbowaliśmy jednak wcześniej wykazać, iż nauki kognitywne albo nic takiego nie wykazują (ponieważ w kontekście tego zagadnienia nie mają wyraźnych implikacji filozoficznych), albo (jeśli uznamy, że takie implikacje dają się jakoś zrekonstruować) wspierają raczej tezę przeciwną, tj. tradycyjny obraz człowieka jako istoty wolnej i racjonalnej. Badania nad racjonalnością prowadzone w ramach nauk kognitywnych mogłyby do pewnego stopnia podważyć ten obraz, gdyby dowodziły, że człowiek jest **zupełnie** nieracjonalny w sensie instrumentalnym, można bowiem uznać, iż założenie o przynajmniej umiarkowanym poziomie racjonalności instrumentalnej jest częścią tego obrazu. Jest jednak oczywiste, że nauki kognitywne nie mogą dowieść tej tezy, gdyż jest ona sprzeczna ze zdrowym rozsądkiem i – co ważniejsze – z elementarnymi zasadami biologii ewolucyjnej: nasi przodkowie musieli wykazywać przynajmniej umiarkowany poziom racjonalności instrumentalnej, aby móc przetrwać i przekazać nam swoje geny. Nawet model *Heuristics and biases*, najbardziej sceptyczny względem tezy o racjonalności instrumentalnej człowieka, nie neguje tezy, że człowiek jest umiarkowanie racjonalny w sensie instrumentalnym. Zatem gdyby nawet ten właśnie model miał uzyskać status paradygmatu, nie miałoby to większego znaczenia dla dyskusji nad trafnością tradycyjnego – humanistycznego – obrazu człowieka, leżącego u podstaw myślenia o prawie.

²³ Zwierzęta są (albo przynajmniej mogą być) racjonalne w sensie instrumentalnym, ale nie w sensie „argumentacyjnym”.