

Miłowit Kuniński

Wyjaśnianie powstawania norm

Powstawanie norm ludzkiego postępowania (obyczajowych i moralnych) wyjaśniają różne modele oparte na założeniu ludzkiej racjonalności. W XX wieku w wyniku rozwoju teorii gier, model wyjaśniania zwany dylematem więźnia zyskał wielu zwolenników (Buchanan, Gauthier, Harsanyi, Nozick, Rawls, Ullmann-Margalit). Jednak pogłębione dociekania pokazują, że zastosowanie dylematu więźnia napotyka na ograniczenia (Ullmann-Margalit, de Jasay), ponieważ matryce dylematu więźnia są zdeterminowane przez normy powstałe w sposób odmienny. W artykule uzasadniane jest stanowisko, wedle którego normy determinujące ludzkie postępowanie mają charakter trójwarstwowy: normy pierwszego rzędu, czyli normy racjonalności, determinujące schematy racjonalnych wyborów w sytuacjach niepewności (dobór zachowań ze względu na cel), utrwalone w procesie ewolucyjnej selekcji grupowej, normy drugiego rzędu – moralne, powstałe w wyniku utrwalania schematów działań (wyborów) opartych na dylemacie więźnia, oraz normy trzeciego rzędu, bazujące na dwóch poprzednich, a utrwalone przez proces selekcji grupowej i krytyczną analizę ich treści i skutków.

Słowa kluczowe: wyjaśnianie, teoria gier, dylemat więźnia, norma, selekcja grupowa.

1. Społeczna struktura ontyczna.

Istnienie pojedynczych ludzi stwierdzone jest empirycznie. Poszczególne jednostki są postrzegane jako odrębne istoty, a nie jako części składowe większych całości, niezdolne, choćby na krótko, do odrębnego istnienia. Relacje między jednostkami nie są poznawane bezpośrednio empirycznie i dlatego ich istnienie ma dla nas inny status, choć nieodzielny od istnienia samych jednostek. Empiryczne kryteria istnienia mają swoje źródło w nastawieniu zdroworozsądkowym, a więc przedteoretycznym. Zatem, jeśli uznaje się występowanie relacji między jednostkami, to ze względu na dające się pośrednio im przypisać postawy wobec kogoś lub czegoś, manifestujące się we wzajemnie odniesionych działaniach, ujawnianych emocjach i wypowiedziach słownych. Mamy zatem bezpośrednią wiedzę empiryczną o poszczególnych ludziach i wiedzę pośrednią, opartą na rozumieniu ich postępowania nacechowanego regularnościami, o jego motywach i trwałych nastawieniach.

W postawach, ze względu na ich powtarzalność i regularność oraz słowne objaśnienia lub

uzasadnienia postępowania jako przejawu tych postaw, uchwytyje się normy obyczajowe, moralne i prawne¹. Zdroworozsądkowo przyjmujemy, że wyjaśniają one względną trwałość postaw i postępowania ludzi oraz nas samych, a zatem tłumaczą koordynację działań i porządek społeczny rozumiany ogólnie jako stan dynamicznej równowagi, w którym zaburzenia owej koordynacji są względnie szybko eliminowane, na ogół przy relatywnie niskich kosztach transakcyjnych. Skoro normy pełnią tak istotną funkcję w społeczeństwie, nieodparcie nasuwa się pytanie o sposób ich powstawania.

Odpowiedzi wykraczające poza doświadczenie codzienne i nastawienie zdroworozsądkowe (choć z nich wyrastające), jakich na nie udzielano, mogą być różne, ze względu na odmienne założenia teoretyczne. Napotykamy więc wyjaśnienia historyczno-narracyjne, ewolucjonistyczne, wyjaśnienia typu „niewidzialna ręka”, psychospołeczne, wreszcie nawiązujące do teorii decyzji

¹ Przez normy rozumiem zarówno reguły zachowania, w których uświadomionej treści występuje wyraźnie element normatywny (powinnościowy), jak i reguły, postrzegane jako regularności zachowań, z towarzyszącymi im zazwyczaj określonego rodzaju reakcjami ze strony innych ludzi.

i teorii gier (Ullmann-Margalit 1977, s. 1–12). Jak trafnie zauważa Edna Ullmann-Margalit, chodzi tu raczej o wyjaśnianie **typów norm** niż norm bardziej szczegółowych, których wyjaśnianie stosowane jest w podejściu historyczno-naracyjnym, i, jak się wydaje, także w psychospołecznym (ibidem, s. 8–9).

Na ogół sądzi się, że wyjaśnienia ewolucjonistyczne, tłumaczące powstawanie norm lub ich zespołów (instytucji) jako nieplannego rezultatu ludzkich działań skierowanych na cele, które nie pozostają definicyjnie tożsame z treścią owych norm utrwalonych wskutek ich funkcjonalności w dłuższym czasie, inaczej niż pozostałe sposoby wyjaśniania **nie są** oparte na założeniu racjonalności działających ludzi. Racjonalność może być pojmowana zdroworozsądkowo jako trafny dobór działań do obranych celów, jako podejmowanie decyzji przemyślnych, a nie pod wpływem emocji, czyli w sposób, który umożliwia działania skuteczne. Bywa także rozumiana jako dobór celów zgodny z pewną hierarchią wartości. W obu wypadkach racjonalność może być traktowana jako stopniowalna. Wreszcie można ją ująć z perspektywy teorii decyzji, zwłaszcza teorii gier, jako maksymalizację korzyści/użyteczności (wypląt) z zastosowaniem różnych modeli formalnych, w tym dylematu więźnia i zróżnicowanych strategii, lub z zastosowaniem zasady dominacji, jak w wypadku analizy problemu Newcomba przez Roberta Nozicka (1969, za: Miklaszewska 1994, s. 103).

Dysponując różnymi modelami wyjaśniania powstawania (typów) norm, możemy dokonać wyboru między nimi. Wydaje się, że podejście zakładające racjonalność działających ludzi, ze względu na dotychczasowe osiągnięcia, jest najlepiej uzasadnione. Z drugiej jednak strony trudno zaprzeczyć, że wyjaśnienia ewolucjonistyczne mają swoje zalety, ponieważ omijają trudność, jaką stanowi ustalanie świadomych motywów postępowania, postaw i norm, które są w nich obecne. Wyjaśnienia teoriodecyzyjne, w szczególności teoriogrowe, podejmowanych decyzji za pomocą dylematu więźnia, a także wyjaśnienia typu „niewidzialna ręka” mające za przedmiot powstawanie norm, zyskały sobie uznanie w ostatnich kilkudziesięciu latach (zob. np. Buchanan 1975; Campbell, Sowden 1985; Gauthier 1986; Harsanyi 1976; Nozick 1969,

s. 114–146, 1974; Polanyi 1980 [1951]; Rawls 1999 [1971]; Ullmann-Margalit 1977).

Może jednak powstać wątpliwość, czy wyjaśnienia te mają charakter ostateczny. Jeśli bowiem weźmiemy pod uwagę, że działanie ludzi polega na realizacji różnych strategii racjonalności: *minimaxu* lub *maximinu*, a bywa, że działania urzeczywistniają strategię, w której istotną rolę odgrywają normy ją determinujące, wyuczone w warunkach normatywnego oddziaływania otoczenia. Czy mamy zatem uznać, że wyjaśnienie teoriogrowe jest niewystarczające i czyni niezbędnym wyjaśnianie drugiego rzędu, które wymaga uwzględnienia w eksplanansie jakiejś normy lub norm? Wówczas należałoby uznać, że wyjaśnianie teoriogrowe ma ograniczony charakter, a wskazanie na normy określające dodatkowe warunki wyjaśniające powstanie norm niższego rzędu oznacza przesunięcie tego problemu na wyższy poziom i teoretycznie grozi regressem do nieskończoności. Albo więc możliwe jest uzyskanie wyjaśnień owych norm wyższego rzędu w kategoriach teorii gier czy ogólniej teorii decyzji, albo trzeba uznać, że przy wyjaśnianiu norm napotykamy na trudność nie do pokonania, czyli konieczność określenia norm ostatecznego rzędu, które nie poddają się dalszej eksplikacji, a mogą być jedynie zrekonstruowane. Dopuszczalne jest także wyjaśnianie ewolucjonistyczne, w którym istotną rolę odgrywa proces selekcji schematów zachowań. Istnieje wreszcie możliwość połączenia ze sobą różnych modeli wyjaśniania, przy świadomości, że nie mają one definitywnego charakteru.

Normy, jak trafnie zauważa Ullmann-Margalit, stanowią problem dla zwolenników indywidualizmu metodologicznego, gdyż mają charakter kolektywny, ponadjednostkowy. Autorka *The Emergence of Norms* wyprowadza z tej konstatacji interesujący wniosek: zgadza się z Johnem Harsanym, że wyjaśnianie norm społecznych w kategoriach indywidualistycznych, a w szczególności przy użyciu teorii gier, ma swoje zalety, ale równocześnie podkreśla ona, iż podejście teoriogrowe jest niewystarczające, bowiem pozostawia poza obszarem wyjaśniania nieformalny kontekst sytuacji, w których powstają normy – „kontekst reprezentowany przez matryce gier” (Ullmann-Margalit 1977, s. 12–16).

Nasze wcześniej sformułowane zastrzeżenia co do zakresu możliwości stosowania wyjaśnień powstawania norm za pomocą teorii gier znajdują więc potwierdzenie u autorki, która stosując w pewnym zakresie aparaturę teorii gier do wyjaśniania powstawania (typów) norm, ma równocześnie świadomość jej ograniczeń, i wprawdzie dąży do utrzymania w mocy wyjaśnień w terminach indywidualistycznej metodologii, lecz przedstawione przez nią wyjaśnianie norm wykracza poza obszar teorii gier (ibidem, s. 14–17).

Wspomniane wyżej zdroworoządkowe nastawienie, które skłania nas do uznawania za oczywiste istnienie jednostek podobnych pod wieloma względami do nas i odnoszących się wzajem do siebie, jest przekształcane przez różne stanowiska teoretyczne. Absolutyzują one owo nastawienie do postaci indywidualizmu ontologicznego, z którym powiązany jest indywidualizm metodologiczny, albo uznają jednostki za elementy agregatów społecznych i przypisują owym jednostkom status niesamoistnych genetycznie lub funkcjonalnie części większych całości w ramach stanowiska zwanego holizmem lub radykalnym holizmem.

Możliwe jest również wyróżnienie dwóch innych stanowisk, zbliżonych do powyższych, odpowiednio atomizmu i antyredukcjonizmu. Zdaniem Janusza Grygienia, jeśli uwzględnimy, że wyjaśnianie może dotyczyć relacji między jednostkami lub odnosić się do własności i relacji wiążących zagregowane całości społeczne, wówczas indywidualizm metodologiczny dopuszcza wyjaśnianie pierwszego typu, ale wyklucza typ drugi, radykalny holizm uwzględnia wyjaśnienia tylko drugiego rodzaju, antyredukcjonizm akceptuje oba rodzaje, natomiast atomizm odrzuca je (Wright, Levine, Sober 2003, za: Grygień 2012, s. 298; Grygień 2011, s. 237).

Wydaje się, że należy odróżnić kwestie ontologiczne od metodologicznych: ekspanacyjnych czy predykcyjnych, ponieważ przyjęcie określonych założeń ontologicznych nie musi przesądzać o akceptacji logicznie spójnego z nimi stanowiska metodologicznego. Może bowiem się okazać, że jego skonstruowanie napotyka na trudności z powodu niemożności przełożenia kategorii ontologicznych na pojęcia operacyjne, które mogłyby jako elementy zdań stanowić przesłanki wyjaśnień i przewidywań.

Choć można przyznać, że debata na temat indywidualizmu metodologicznego jest interesująca, to jednak podstawową kwestią pozostaje naszym zdaniem odpowiedź na pytanie, czy charakteryzując jednostki i relacje między nimi przez odwołanie się do norm, mamy na uwadze wpływ większych całości społecznych (społecznych tworców), wywierających nacisk na jednostki, czy też możemy normy traktować jako utrwalone schematy działań wytworzone bądź powstałe wskutek wzajemnych oddziaływań między jednostkami, które można podać „racjonalnej rekonstrukcji”, czyli przedstawić je jako powtarzalne działania w warunkach uproszczonych, idealizacyjnych². Możliwe są oczywiście także stanowiska pośrednie.

Opowiedzenie się za indywidualizmem metodologicznym, jak to czyni Ullmann-Margalit, przy równoczesnym uznaniu ograniczeń teorii gier, opartej na indywidualistycznych założeniach metodologicznych, wynika z dążenia do unikania hipostazowania czy „reifikacji” norm. Postulat ten wydaje się zrozumiały, powstaje jednak pytanie, czy można go w pełni zrealizować. Do tego zagadnienia powrócimy poniżej, rozważając koncepcję indywidualizmu prawdziwego i fałszywego, autorstwa Friedricha A. Hayeka.

Max Weber, któremu zawdzięczamy wczesne teoretyczne uzasadnienie stanowiska indywidualizmu metodologicznego, dowodził, że ostatecznym poziomem ontycznym, umożliwiającym wyjaśnianie tworców społecznych, są nie tyle jednostki, poszczególni ludzie, ile ich działania³. Podobnie czynił w późniejszym okresie Ludwig von Mises w swojej podstawowej pracy *Human Action. A Treatise on Economics* (1966 [1949])⁴.

² Do koncepcji „racjonalnej rekonstrukcji” Carnapa (1962, s. 301–318) nawiązuje Ullmann-Margalit (1977, s. 1–2).

³ Weber zajął się problemem działania społecznego i metodologii wymaganej do jego analizy oraz jego rozumiejącego wyjaśniania w §1 (ks. 1, r. I) swego głównego dzieła *Gospodarka i społeczeństwo. Zarys socjologii rozumiejącej* (2002, s. 6). Natomiast po raz pierwszy użył terminu „metodologiczny indywidualizm” uczeń Webera – Joseph Schumpeter (1908, s. 88), gdzie r. VI, cz. I nosi tytuł „Der methodologische Individualismus”, a następnie w wersji angielskiej w artykule „On the concept of social value” (1909, s. 213–232).

⁴ Pośród innych autorów, którzy opowiadali się za indywidualizmem metodologicznym warto wymienić

Znana Weberowska typologia działań i zachowań wyraźnie na to wskazuje. Weber wyróżnił działanie racjonalno-celowe (*zweckrationales Handeln*), racjonalno-wartościowe (*wertrationales Handeln*), zachowanie afektywne (emocjonalne) (*affektuelles Verhalten*) oraz tradycyjne (*traditionelles Verhalten*) (Weber 1951, s. 551–553 za: Kuniński 2006, s. 4–5; Weber 2002, s. 6–13, 17–20).

1. Działanie racjonalno-celowe (*zweckrationales Handeln*) skierowane jest na odrębnie postrzegane cele i obejmuje postrzeganie związków między środkami realizacji celów a celami oraz skutkami działań a celami. Ten typ działania zakłada posiadanie wiedzy o celach, środkach i skutkach.

Gdy cele postrzegane przez podmiot działania mają charakter alternatywny (nie mogą być zrealizowane równocześnie), podmiot ów może postąpić w dwojaki sposób:

- a) dokonać wyboru celu ze względu na pewne wartości absolutne: etyczne, estetyczne lub inne – działanie racjonalno-celowe zbliża się wówczas do działania racjonalno-wartościowego; jedynie dobór środków dokonuje się w sposób racjonalno-celowy, to znaczy ze względu na wiedzę o skuteczności środków do osiągnięcia danego celu;
- b) podmiot działania może potraktować cele alternatywne jako subiektywne pragnienia i uszeregować je wzdłuż skali świadomie ustalonych potrzeb czy preferencji (**zasada krańcowej użyteczności**), by następnie dobrać odpowiednie środki ich realizacji.

2. Działanie racjonalno-wartościowe (*wertrationales Handeln*) skierowane jest na wartości absolutne, postrzegane w sposób świadomy. Podmiot działania tworzy spójny plan doboru środków do realizacji tych wartości. Nie uwzględnia przy tym kosztów realizacji wartości i skutków działania. Podmiot działa konsekwentnie ze względu na swoje przekonania traktowane jako powinności (obowiązek, honor, powołanie religijne itp.). Niesprzyjające okoliczności nie prowadzą do rezygnacji z realizacji wartości i nie modyfi-

kują w istotny sposób środków jej osiągnięcia. Działanie racjonalno-wartościowe z punktu widzenia działania racjonalno-celowego może mieć więc charakter irracjonalny. Dążenie do wartości absolutnych zakłóca racjonalne rozpoznanie środków właściwych dla osiągnięcia celu i utrudnia racjonalne oszacowanie skutków działania zorientowanego na wartość.

3. Zachowanie afektywne (emocjonalne) (*affektuelles Verhalten*) jest to najczęściej gwałtowna reakcja na bodziec o szczególnym charakterze. Znajduje się ono na granicy automatycznych reakcji na bodźce zewnętrzne i działań świadomych, tzn. takich, w których w wyniku sublimacji następuje świadome rozładowanie stanu napięcia emocjonalnego. Zachowanie afektywne może wówczas zbliżyć się do działań racjonalno-wartościowych, a nawet racjonalno-celowych.

4. Zachowanie tradycyjne (*traditionelles Verhalten*) znajduje się na pograniczu działań, z którymi podmiot wiąże pewne uświadomione znaczenia, i zachowań naśladowczych, powtarzanych pod wpływem bodźców. Podmiot zachowuje się zgodnie z ustalonymi normami zwyczajowymi, przy czym im większy stopień uświadamiania sobie zależności reakcji od bodźców, tym bardziej zachowania tradycyjne zbliżają się do działań racjonalno-wartościowych.

W istocie Weberowska typologia wyróżnia trzy rodzaje działań racjonalnych, mimo że te zgodne z normami obyczajowymi określa mianem „zachowań tradycyjnych”. Z perspektywy późniejszych rozróżnień związanych z teorią decyzji w powiązaniu z teoriami ekonomicznymi i socjologicznymi, można przyjąć, że tzw. działania racjonalno-celowe odpowiadają modelowi tych opartych na zasadzie krańcowej użyteczności albo, gdy w grę wchodzi jakieś wartości traktowane jako cele samoistne lub są to tzw. działania racjonalno-wartościowe i zachowania tradycyjne, model wyjaśniający powstawanie norm rzeczywistnianych w działaniach i w zachowaniach oparty byłby na schemacie eksplanacyjnym dylematu więźnia, wzbogaconym o wyjaśnianie w indywidualistycznych kategoriach norm determinujących matryce dylema-

Watkinsa (1952), Hayeka (1955; polskie wydanie 2002), Poppera (1957; polskie wydanie 1999 [1989]) i Elstera (1982, s. 453–482).

tu więźnia, a więc wykraczający poza wyjaśnienia teoriogrowe.

Traktując działanie jako elementarny poziom struktury społecznych stosunków i instytucji, autor *Gospodarki i społeczeństwa* chciał wyeliminować tendencję do traktowania zbiorowości jako podmiotów działań, co wiązało się z negatywnym – w jego opinii – wpływem na myśl niemiecką zarówno filozofii Hegla, jak i prawa, które również dzisiaj posługuje się pojęciem osoby prawnej (kolektywnej) (Weber 2002, s. 12). Według niego, aby można było zastosować w sposób zadowalający **rozumienie wyjaśniające** działań jednostek, trzeba wniknąć w ich znaczenie, zrekonstruować intencje działań, planowane lub zawczasu uznane środki ich realizacji itp. Przy czym, jak wiadomo, nie chodzi tu o odtworzenie zjawisk psychicznych zachodzących w umysłach działających podmiotów – aktorów, lecz właśnie o swego rodzaju racjonalną rekonstrukcję, opartą na zastosowaniu typów idealnych, które na podstawie danych empirycznych, uzyskiwanych z różnych źródeł (obserwowane działania, deklarowane motywacje, dokumenty, biografie itp.) tworzą spójny schemat motywów i norm(y) działania, nie w pełni urzeczywistniany w realnych działaniach ludzi, ale pozwalający je wyjaśnić, gdyż uwzględnia ich podstawowe właściwości. Na ile rozumienie wyjaśniające spełnia kryteria rzetelnych wyjaśnień, a więc unika błędu krzyżowania się zakresów znaczeniowych eksplanansu i eksplanandum, to odrębny problem, wymagający obszerniejszej analizy, która wykracza poza ramy niniejszego artykułu.

2. Problem koordynacji (działań) a powstawanie norm

Współcześnie tzw. podejście rozumiejąco-wyjaśniające zostało zastąpione wspomnianą racjonalną rekonstrukcją, choć wydaje się, że jest ona rozwinięciem koncepcji Webera, który uważał, iż rozumienie wyjaśniające, jeśli ma charakter idealnotypologiczny, pozwala formułować „prawo” czystej ekonomii teoretycznej przedstawiające przebieg działań ludzkich jako ściśle racjonalno-celowych, bez wpływu uczuć (Weber 2002, s. 9).

Edna Ullmann-Margalit i Anthony de Jasay, nawiązując do teorii gier, wykorzystują w swych analizach schemat dylematu więźnia (de Jasay 1989, 2010). Koordynację działań ludzi można wyjaśnić po określeniu elementów, jakie mają być ze sobą skoordynowane, i środków, które temu służą. De Jasay ze względów metodologicznych i ontologicznych przyjmuje, iż uczestnikami gier są osoby, i to zarówno „naturalne” (jednostki i reprezentowane przez nie rodziny), jak i prawne, ewentualnie koalicje, które trwają, nie rozpadając się, a podejmowane decyzje odnoszą się do wszystkich jej członków (bez względu na sposób ich podejmowania). Jak widać, de Jasay nie uważa, by podejście teoriogrowe musiało być ściśle powiązane z indywidualizmem metodologicznym, a właściwie z atomizmem.

Wyjaśnienia koordynacji działań za pomocą dylematu więźnia są zarazem wyjaśnieniami powstawania norm. W tym względzie de Jasay podąża śladem Ullmann-Margalit. Czy wyjaśniając powstawanie norm, a tym samym koordynacji działań, należy analizować sytuacje rzeczywiste, czy „wypreparowane”, zbliżone do rzeczywistych, lecz przebiegające w warunkach uproszczonych, hipotetycznych? Jeśli bowiem będziemy badać podejmowanie decyzji i działań w warunkach rzeczywistych, wówczas wyjaśniając powstawanie norm, musimy uwzględnić wpływ norm już istniejących.

a. Hipotetyczny stan natury i prosty system normatywny

Aby uniknąć tego rodzaju wyjaśnień „mieszanych”, należy brać pod uwagę wyłącznie interakcje między osobami i ich dążenie do uzyskania korzystnych rezultatów. De Jasay odwołuje się do pojęcia stanu natury, rozumianego jako idealna równościowa dystrybucja siły (prowadzi to do jednostronnej współpracy lub owocuje kontraktem przestrzegany bez zewnętrznego przymusu [*self-enforcing contract*]). Można go zdefiniować następująco:

„Każda ze stron decyduje jednostronnie, czy kontynuacja lub zerwanie relacji z drugą stroną jest dla niej korzystne. Dana strona zrywa relację wtedy i tylko wtedy, gdy bieżący zysk z przerwania relacji przekracza oczekiwaną aktualną wartość zysków z jej kontynuacji. Nie występuje przy tym żadna zewnętrzna interwencja jakiegokolwiek strony, mająca na celu narzucenie porozumienia,

ustalenie, czy zaszły jego naruszenia i ukierunkowana na nałożenie kar⁵.

Koordinacja według Ullmann-Margalit i de Jasaya ma charakter doskonały (*pure*), gdy korzyści czy „wypłaty” dla jednego z uczestników interakcji nie mogą być powiększone bez strat ponoszonych przez innego uczestnika. Koordinacja niedoskonała (*impure*) oznacza korzyści dla jednego z uczestników i stratę dla innego. Doskonała ma tylko jedno rozwiązanie optymalne w sensie Pareto, niedoskonała zaś nieskończenie wiele, w zależności od udziałów uczestników gry w jej dodatnim rezultacie (de Jasay 1989, s. 48–49).

Zamiast traktować normy jako składowe eksplanansu przy wyjaśnianiu działań, które są skoordynowane i umożliwiają kooperację społeczną (w tym ekonomiczną), w ramach teorii gier, za pomocą dylematu więźnia, można wyjaśniać koordynację jako uogólniony i utrwalaony rezultat interakcji czy działań. Normy powstają w wyniku powtarzania się sytuacji, w których pewne działania czy zachowania przynoszą korzyści uczestnikom lub przynajmniej w dłuższej perspektywie są częściej korzystniejsze. W ten sposób normy stanowią efekt samorzutnego i nieprzewidywalnego dla podmiotów działań procesu powtarzania (jednego lub więcej) rozwiązań problemu koordynacji. Jeśli normy i ich zespoły, ze względu na ich utrwalenie jako schematów działań wzajemnie powiązanych, ulegają instytucjonalizacji, wówczas, jak wskazuje de Jasay, wzrasta prawdopodobieństwo pojawienia się pieczeniarzy (*free riders*), którzy, nie podporządkowując się normom, uzyskują korzyść bez własnego wkładu polegającego na podporządkowaniu się normie.

Początek procesu tworzenia się norm może mieć (wedle niektórych ujęć zazwyczaj ma), charakter przypadkowy. W tradycji szkockiego oświecenia (David Hume, Adam Ferguson,

Adam Smith), w koncepcji Carla Mengera instytucji organicznych w odróżnieniu od pragmatycznych, czy w teorii społecznej i ekonomicznej Hayeka, traktującej instytucje jako normy lub ich zespoły, podkreśla się przypadkowy, nieplanowy i niezamierzony sposób ich powstawania. Ludzie – jak to ujął Ferguson – natykają się (*stumble upon*) na instytucje, które choć przez nich wytworzone, nie zostały przez nich zaprojektowane⁶.

Wyjaśnienie koordynacji w kategoriach teorii gier i procesów samorzutnych można uznać za obarczone błędem naturalistycznym. Normatywny charakter reguł umożliwiających koordynację działań powstaje w dalszych fazach procesu ich utrwalania się, kiedy wskutek internalizacji zaczynają być postrzegane jako obowiązujące. Ocena trafności tego zarzutu nie jest łatwa.

Gdyby wziąć pod uwagę stanowisko Hume’a, to z jednej strony okaże się, że uważał on, iż oprócz tzw. cnót naturalnych, zwanych społecznymi ze względu na ich sprzyjanie dobru społecznemu, opartych na oddźwięku uczuciowym (łagodność, miłosierdzie, szczodrobliwość, miękkość, umiarkowanie) i odczuwaniu aprobaty oraz dezaprobaty, a niepowstałych w wyniku „sztucznej konstrukcji i wychowania”, istnieją cnoty sztuczne: sprawiedliwość i własność, zaistniałe jako konwencje czy ugody w wyniku dostrzegania korzyści, jakie w dalszej perspektywie (w dłuższym czasie) przynoszą całej społeczności (Hume 1952 – ks. III, cz. III, r. II i ks. III, cz. II, r. II). Hume twierdzi więc, że tzw. sztuczne cnoty to normy wytworzone w sposób nieplanowy i niezamierzony, lecz przynoszące korzyści w ramach społeczności, co wskutek powtarzania staje się widoczne i sprzyja utrwalaniu norm. Z drugiej strony, na długo przed *Principia Ethica* George’a E. Moore’a (1903), gdzie pojawiło się pojęcie błędu naturalistycznego, wykazywał, że nie można wywieść logicznie zdań preskryptywnych (zawierających słowa **powinien** lub **nie powinien**) ze zdań deskryptywnych (zawierających

⁵ „[E]ach party decides unilaterally whether he is better off continuing or stopping his relationship with the other party. He stops if and only if the current gain from stopping exceeds the expected present value of his gains from continuing. No outside party intervenes to enforce the agreement, to determine whether there have been violations, and to impose penalties” (Telser 1980, s. 27, za: Stump <http://www.impgroup.org/uploads/papers/262.pdf>).

⁶ „Every step and every movement of the multitude, even in what are termed enlightened ages, are made with equal blindness to the future; and nations stumble upon establishments, which are indeed the result of human action, but not the execution of any human design” (Ferguson 1782, s. 90).

spójki **jest** lub **nie jest**) (Hume 1952 – ks. III, cz. I, r. I). Ale to oznaczałoby, że konwencje jako normy określające powinność wywodzą się z powtarzalnych zachowań. Zatem, początkowo ich obowiązywalność oparta jest na przyzwyczajeniu wytworzonym dzięki powtarzaniu określonych zachowań, w wyniku czego wytwarza się schemat zachowań, a w następstwie pojawia się świadomość obowiązywania, przybierająca postać uświadomionego obowiązku, a zatem normy.

Jeśli ze względu na pojęcie stanu natury używane przez de Jasaya mielibyśmy sięgnąć do Thomasa Hobbesa, okaże się, że według niego w stanie natury zagrożenie „gwałtowną śmiercią”, gdy każdy dla każdego jest wrogiem, może być znacznie osłabione, ponieważ możliwe jest sprzymierzanie się dla obrony i uzyskanie pomocy od sprzymierzonych. A więc w stanie natury istnieją formy współpracy – nawet, jeśli nie są tak trwałe jak w stanie społeczno-państwowym – oparte na przyrzeczeniach, które w razie ich spełnienia przez jedną ze stron pociągają za sobą spełnienie ich przez drugą stronę, ponieważ „nie jest to przeciwne rozumowi”, a „to znaczy: przeciwne korzyści drugiej strony” (Hobbes 1954, s. 129; 2005, s. 229)

Roztropność człowieka, oparta – według Hobbesa – na doświadczeniu, skłania go do wywiązania się z przyrzeczenia (które nie musi mieć formy słownej), mimo iż korzyść krótkoterminowa może polegać na wycofaniu się z niego. Oznacza to, że w stanie natury powstają pewne, choć nietrwałe, załączki więzi społecznych i norm moralnych, a ludzie nie są wyłącznie maksymalizatorami użyteczności (korzyści własnej). Roztropność (przezorność), czyli nagromadzone doświadczenie, choć jest źródłem wiedzy niepewnej, umożliwia rozróżnianie między rywalizacją a współpracą, i na tej podstawie przyszłe częstsze wybieranie kooperacji jako warunku realizacji długoterminowych dążeń ludzi. To skłania niektórych interpretatorów Hobbesa do uznania, iż już w stanie natury ludzie są globalnymi, a nie tylko lokalnymi, maksymalizatorami użyteczności w dążeniu do samozachowania⁷.

⁷ Pojęcia globalnego i lokalnego maksymalizatora użyteczności wprowadził Elster, a Hampton posłużyła się nimi, analizując stanowisko Hobbesa; zob. Hampton 1988 (1986), s. 16.

Można wszakże przyjąć, że w związku z występowaniem przyrzeczeń dla uzyskania pomocy sprzymierzonych, lokalni maksymalizatorzy, w wyniku powtarzania się tego rodzaju sytuacji, uczą się przewidywania w dłuższym horyzoncie czasowym, stając się globalnymi maksymalizatorami (zob. Kuniński 2010). To tłumaczy, dlaczego możliwa była interpretacja teorii Hobbesa przez Dawida Gauthiera za pomocą dylematu więźnia. Podczas gdy Gauthier pokazywał, że przejście ze stanu natury do stanu społecznego/politycznego można wyjaśnić za pomocą uogólnionego dylematu więźnia, uważamy, że dylemat ów dobrze charakteryzuje narodziny norm w stanie natury, gdy nie istnieją jeszcze złożone i dwuwarstwowe systemy normatywne.

b. Hipotetyczny stan natury w warunkach dwuwarstwowego systemu normatywnego

Pojęcie dwuwarstwowych systemów normatywnych nawiązuje do rozróżnienia, jakiego dokonał Herbert L.A. Hart w swojej znanej pracy *Pojęcie prawa* między regułami pierwotnymi i wtórnymi, których związek charakteryzuje prawo (1998)⁸. Wątpliwości, jakie budzi przeniesienie opisu struktury prawa na inne dziedziny norm, można rozwiać, wskazując na stanowisko samego Harta, argumentującego na rzecz tezy, iż prawo ma podstawy moralne, które wolno pojmować jako minimalną treść prawa naturalnego. Przyjmijmy zatem, że normy pierwotne jako określające obowiązki mają charakter najbardziej trwały, gdyż nie tylko stanowią one podstawę reguł wtórnych, lecz są regułami najstarszymi, występującymi w różnych społecznościach, o różnym stopniu komplikacji struktury społecznej, i mogą być – jak stwierdza Hart – „całkowicie zwyczajowe, gdy idzie o ich genezę”, czyli nie muszą pozostawać wynikiem stanowienia przez jakiś organ (jednoosobowy czy wieloosobowy). Zatem normy pierwotne można uznać za powstałe zgodnie ze schematem dylematu więźnia jako przynoszące korzyści jednostkom wchodzącym w powtarzające się interakcje i sprzyjające koordynacji działań i kooperacji wskutek utrwalania się. Przy czym koordynacja uzyskiwana za pomocą reguł pierwotnych ma charakter mieszany, z przewagą koordynacji czystej.

⁸ Ullmann-Margalit, nawiązując do Harta, rozróżnia konwencje i rozporządzenia (zob. 1977, s. 97).

Reguły wtórne nadbudowane nad pierwotnymi („pasożytnicze” [Hart] albo „satelitarne” [de Jasay]) pozwalają modyfikować pierwotne, wprowadzać nowe, określać ich zasięg i działanie (Hart 1998, s. 116–117). De Jasay przypisuje normom satelitarnym ochronę norm pierwotnych przez ich sankcjonowanie, wskazując na trudności związane z ich ustanowieniem. Można więc uznać, że normy wtórne to sposoby interpretacji norm pierwotnych i określanie form ich egzekwowania. Nie jest jasne, w jakiej mierze są one z racji swych funkcji koordynacyjnych rezultatem rozwiązywania dylematu więźnia. Gdyby przyjąć stanowisko zwolenników teorii samorzutnego, nieplanowego procesu powstawania instytucji jako systemów norm, wówczas należałoby uznać, że z pewnością część z nich pojawia się w wyniku rozwiązywania dylematu więźnia i prowadzi do mieszanej koordynacji. W duchu teorii Mengera czy Hayeka można też uznać, że część powstaje w rezultacie stanowienia, a więc w sposób świadomy i celowy⁹. Mechanizm selekcji, który eliminuje reguły koordynacji niepowtarzające się systematycznie lub przynoszące straty jednostkom wchodzącym w interakcje, jest istotnym czynnikiem wpływającym na powstawanie norm pierwotnych. Funkcjonuje on w ograniczonym zakresie na poziomie norm wtórnych albo – ujmując rzecz ostrożniej – działa w dłuższym horyzoncie czasowym.

Koncepcja Harta, z której zaczerpnęliśmy ideę dwuwarstwowego systemu normatywnego, w nieznacznym tylko stopniu rozpoznaje samorzutność procesu powstawania norm pierwotnych i wtórnych. Świadczy o tym analiza minimalnej treści prawa naturalnego (Hart 1998, r. IX). Hart, ewidentnie nawiązując do Hobbesa, wskazuje m.in. na „przybliżoną równość” ludzi, którzy różnią się siłą fizyczną, zręcznością i zdolnościami intelektualnymi – jednak nikt nie jest tak potężny, by dominować nad innymi lub podporządkować ich sobie na długo¹⁰. „Przybliżona

równość, bardziej niż cokolwiek innego, przekazuje o konieczności wzajemnego powstrzymywania się od agresji i niezbędności kompromisu, co jest podstawą dla prawnych jak i moralnych obowiązków” (Hart 1998, s. 263). Owa przybliżona równość ma mieć, według Harta, „bardzo poważne znaczenie dla zrozumienia rozmaitych postaci prawa i moralności”. Jeśli, jak wskazuje na to jego wywód, miałyby to być wyjaśnienie „powstrzymywania się od agresji i niezbędności kompromisu”, czyli koordynacji, wówczas można by tę koncepcję uznać za upatrującą w normach podstawy koordynacji i społecznej kooperacji. Wychodząc z tej samej konstatacji „truizmu” o przybliżonej równości, byłibyśmy również w stanie wywieść normy nieagresji i kompromisu z „gry” przynoszącej jej uczestnikom przewagę korzyści nad stratami.

c. Racjonalna rekonstrukcja norm w stanie społecznym

Wyjaśnianie norm pierwotnych za pomocą dylematu więźnia, przy założeniu, iż normy funkcjonują w warunkach stanu społecznego, w którym działania podporządkowane są regułom pierwotnym i wtórnym, oznacza przeprowadzenie analizy określającej charakter uzyskiwanej koordynacji. W rzeczywistym stanie społecznym, ze względu na obecność dwóch warstw norm i ich dużej liczby, mamy do czynienia z rozległym systemem normatywnym. Koordynacja urzeczywistniana dzięki systemowi norm obu rodzajów ma charakter niedoskonały. Lecz uzyskiwana za sprawą norm pierwotnych ma charakter w znacznej mierze doskonały. Natomiast koordynacja będąca wypadkową norm pierwotnych i wtórnych przyjmuje charakter mieszany.

Zastosowanie dylematu więźnia jest ograniczone, gdyż w jego rozwiązywaniu nie tylko interweniuje racjonalność zorientowana na korzyść, czyli dodatni wynik „gry” (koordynacja doskonała), lecz również utrwalone i przeka-

⁹ Jak w instytucjach społecznych zwanych przez Mengera „pragmatycznymi”, a przez Hayeka „organizacjami”.

¹⁰ Zob.: „Natura uczyniła wszystkich ludzi tak równymi, co się tyczy władz cielesnych i duchowych, że chociaż znajdzie się czasem człowiek wyraźnie silniejszy na ciele lub bystrzejszy umysłem niż inny, to przecież, gdy weźmie się wszystko w rachubę, różnica między człowiekiem a człowiekiem nie jest tak znaczna, ażeby na tej podsta-

wie jeden człowiek mógł mieć roszczenie do korzyści, do której inny nie mógłby mieć **równego roszczenia**. Co się tyczy siły cielesnej, to najsłabszy człowiek ma dość siły, ażeby zabić najmocniejszego albo tajemną machinacją, albo łącząc się z innym ludźmi, którzy są w takim samym niebezpieczeństwie jak on” [podkr. – M.K.] (Hobbes 1954, s. 107; 2005, s. 204).

zywane kulturowo (przez naśladownictwo, wychowanie i edukację) normy pierwotne i wtórne. To skłania do modyfikacji zastosowań dylematu więźnia przez wprowadzenie dodatkowego założenia o działaniu czynnika normatywnego wzmacniającego motywację do powtarzania (utrwalania) działań/zachowań korzystnych dla stron interakcji. Taką funkcję w analizowanym przez Ullmann-Margalit przykładzie dwóch żołnierzy obsługujących niezależnie dwa moździerze pełnią: „matryca dyscypliny” i „matryca honoru”, które wyjaśniają, dlaczego pozostają oni na pozycjach, choć najbardziej racjonalnym działaniem z punktu widzenia korzyści jest opuszczenie pola walki (Ullmann-Margalit 1977, s. 33–37). Okazuje się, zgodnie ze zdrowym rozsądkiem, że na racjonalność jako dążenie do korzyści nakładają się (i modyfikują ją) normy moralne, ukształtowane w długim okresie. Ich geneza być może mogłaby zostać wyjaśniona za pomocą dylematu więźnia jako reguł rozwiązujących problem koordynacji działań na polu walki. Jak wspominaliśmy wyżej, Ullmann-Margalit wprawdzie dostrzega ograniczenia zastosowania wyjaśnień za pomocą dylematu więźnia, lecz nie rezygnuje z podstawowego rozstrzygnięcia metodologicznego, jakim jest indywidualizm metodologiczny.

Rozważmy jednak, czy inna jego wersja, powiązana ze słabą wersją indywidualizmu ontologicznego, nie byłaby bardziej zasadnym podejściem przy badaniu genezy norm.

Hayek w znanym artykule „Indywidualizm prawdziwy i fałszywy” charakteryzuje relacje, jakie zachodzą między jednostką a otoczeniem społecznym, co uzasadnia indywidualistyczną teorię społeczeństwa i założenia metodologicznego indywidualizmu, określającego sposób wyjaśniania zjawisk i procesów społecznych, w tym w szczególności norm postępowania (Hayek 1998, s. 7–41). Przeciwwstawiając indywidualizm prawdziwy fałszywemu, krytykuje ten ostatni, który pojmuje jednostki ludzkie jako podstawowe elementy większych całości społecznych, będące w tym ujęciu agregatami składającymi się z ludzkich monad (zob. Kuniński 1999, s. 214–222; w poniższych uwagach wykorzystuję niektóre tezy i sformułowania zawarte w tej książce). Pojęcie „jednostek izolowanych i samoistnych” (*isolated or self-contained individu-*

als) Hayek uważa za jeden z najpoważniejszych, a przy tym pospolitych, błędów teoretycznych mających także konsekwencje praktyczne w życiu społecznym.

W zależności od sposobu rozumienia człowieka i jego związków z otoczeniem społecznym ulegamy złudzeniu samowystarczalności poznawczej, moralnej i społecznej poszczególnych jednostek, przyjmując, że mają one zdolność kierowania sobą, czerpią z samych siebie, i dlatego są w stanie świadomie kierować agregatami społecznymi. Możemy wszakże, jak to ujmuje Hayek, „za punkt wyjścia [brać] ludzi, których cała natura i charakter są określone ich egzystencją w społeczeństwie” (1998, s. 12).

Hayek przeciwstawia się pojmowaniu człowieka w oderwaniu od kontekstu społecznego, od pełnionych przezeń ról, od utrwalonych w tradycji obyczajów, wzorów czy norm postępowania. W tym sensie człowiek jest jednostką, ale ze względu na swoją niesamowystarczalność w naturalny sposób zmuszoną do konformizmu. Aby stać się zdolnym do stanowienia o sobie, w tym do korzystania z wolności, musi przyswoić sobie normy postępowania, bez których jego życie w społeczeństwie nie byłoby możliwe.

Indywidualizm ontologiczny Hayeka oparty został na założeniu, że jednostka jest wytworem skomplikowanego splotu czynników, z których część ma charakter wrodzony, część wiąże się z doświadczeniem osobniczym, a znacząca część z udziałem w życiu różnych wspólnot, od których przejmuje ona normy, standardy, obyczaje, język itp. Jak podkreśla John Gray, normy i reguły postępowania nie dlatego są użyteczne, że ludzie dostrzegając ich instrumentalność względem obranych przez siebie celów, stosują się do nich, lecz przeciwnie, przejmując je bezinteresownie i podporządkowując im się „bezkrytycznie”, osiągają skuteczność działań zgodnych z regułami. „W większości przypadków (...) – stwierdza Gray – musimy rozpatrywać ludzką osobowość jako ukonstytuowaną przez społeczne reguły i zarazem jako wytwór kultury. Istotnie, nawet w przypadku osobowości opowanej przez anomie, analiza Hayeka sugeruje, iż nie byłoby możliwych do rozpoznania regularności zachowania się, czy stabilnego procesu poznania, o ile przynajmniej niektóre z rozpo-

wszechnionych reguł społecznych nie zostałyby skutecznie zinternalizowane” (Gray 1986, s. 43).

Na poziomie metodologicznym indywidualizm w wersji Hayeka polega na przyjęciu założenia, że ludzi cechuje ograniczona wiedza o okolicznościach, w jakich działają, a koordynacja ich działań z działaniami innych jest wynikiem podporządkowania swojego postępowania regułom, które przejęte bezwiednie w procesie socjalizacji i po utrwaleniu uświadamiane jako normy wykazują funkcjonalność i skuteczność, umożliwiając działania na ogół przynoszące spodziewane rezultaty. Normy te nie muszą być w pełni uświadamiane lub kryją się pod postacią wzorów postępowania uświęconych tradycją i utrwalonych dzięki rozpowszechnionym konwencjom społecznym. Jeśli przyjąć ten model wyjaśniania pojawiania się norm, wówczas należy uznać, że znaczna ich część jednostki zastają w otoczeniu. Nowe normy powstają w rezultacie modyfikacji tych zastanych lub pojawiają się wskutek powtarzania pewnych działań, które utrwalają się dzięki swojej skuteczności w dłuższym horyzoncie czasowym.

Niezbędnym elementem tego modelu wyjaśniania jest selekcja schematów działań, które są eliminowane w rezultacie ich niefunkcjonalności w dalszej perspektywie. Selekcja ta ma charakter grupowy, a utrwała te powtarzalne schematy działań, które przyczyniają się do sukcesów grup społecznych, gdzie ich członkowie podporządkowali się owym ustabilizowanym schematom, czyli normom.

Jak wobec tego ma się propozycja wyjaśnień ewolucjonistycznych powstawania norm do wyjaśnień w kategoriach teorii gier? Odpowiedź na to pytanie wymagałaby odrębnej, pogłębionej analizy, uwzględniającej różne modele wyjaśniania oparte na założeniach ewolucjonistycznych i teoriodecyzyjnych, a w szczególności teoriogrowych.

Wydaje się, że wyjaśnienia ewolucjonistyczne prowadzą do wniosku, że racjonalność podejmowania decyzji i dokonywania wyborów jest efektem przyswojenia sobie pewnych reguł zachowania/postępowania w wyniku uczenia się ich pod wpływem otoczenia społecznego i dzięki kumulacji doświadczenia. Dylemat więźnia jako model wyjaśniania powstawania norm należałoby wówczas uznać za możliwy do zasto-

sowania, gdyż opiera się on na wyuczonych regułach racjonalności (prawdopodobnie część z nich jest wrodzona). Zatem dylemat więźnia może służyć wyjaśnieniom powstawania norm przy założeniu, że racjonalne decyzje, jak wiadomo niekoniecznie najkorzystniejsze, do których się odnosi, uznamy za podstawowe normy powstałe w wyniku procesu ewolucyjnej (grupowej) selekcji. Następnie można przyjąć, iż nad normami racjonalności nadbudowują się normy moralne wskutek utrwalonych schematów działań (wyborów); w istocie są to normy najstarsze i najtrwalsze w grupach społecznych. Dopiero na tej podstawie mogą być tworzone normy zaprojektowane, które również poddawane są zarówno procesowi selekcji grupowej, ze względu na ich funkcjonalność w dłuższym czasie, jak i analizie krytycznej, badającej skutki bezpośrednio obserwowalne.

Bibliografia

- Buchanan J.M. (1975). *The Limits of Liberty: Between Anarchy and Leviathan*. Chicago: The University of Chicago Press.
- Campbell R., Sowden L. (red.) (1985). *Paradoxes of Rationality and Cooperation: Prisoner's Dilemma and Newcomb's Problem*. Vancouver: University of British Columbia Press.
- Carnap R. (1962). „The aim of inductive logic”, w: E. Nagel, P. Suppes, A. Tarski (red.), *Logic, Methodology and Philosophy of Science*. Proceedings of the 1960 International Congress, Stanford: Stanford University Press.
- De Jasay A. (1989). *Social Contract, Free Ride. A Study of the Public Goods Problem*. Indianapolis: Liberty Fund Indianapolis.
- De Jasay A. (2010). *Political Philosophy Clearly. Essays on Freedom and Fairness, Property and Equalities*. Indianapolis: Liberty Fund Indianapolis.
- Elster J. (1982). „The case for methodological individualism”, *Theory and Society*, nr 11.
- Ferguson A. (1782). *An Essay on the History of Civil Society* (wyd. 5). London: Printed for T. Cadell, http://files.libertyfund.org/files/1428/Ferguson_1229_EBk_v6.0.pdf [dostęp: 17.11.2012].
- Gauthier D. (1986). *Morals by Agreement*. Oxford: Clarendon Press.
- Gray J. (1986). *Hayek on Liberty* (wyd. 2). Oxford: Basil Blackwell.

- Grygięć J. (2011). *‘Wola powszechna’ w brytyjskiej myśli politycznej. Thomas Hill Green, Francis Herbert Bradley, Bernard Bosanquet*. Praca doktorska. Toruń: Uniwersytet Mikołaja Kopernika.
- Grygięć J. (2012). *Wola powszechna w filozofii politycznej*. Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika.
- Hampton J. (1986). *Hobbes and the Social Contract Tradition*. Cambridge–London–New York: Cambridge University Press.
- Harsanyi J.C. (1976). *Essays on Ethics, Social Behaviour, and Scientific Explanation*. Dordrecht–Boston: D. Reidel Publishing Company.
- Hart H.L.A. (1998). *Pojęcie prawa*, tłum. J. Wołęński. Warszawa: Wydawnictwo Naukowe PWN.
- Hayek F.A., von (1955). *The Counter-Revolution of Science*. New York: Free Press.
- Hayek F.A., von (1998). „Indywidualizm prawdziwy i fałszywy”, w: *Indywidualizm i porządek ekonomiczny*, tłum. G. Łuczkiwicz. Kraków: Wydawnictwo Znak.
- Hayek F.A., von (2002). *Nadużycie rozumu*, tłum. Z. Siembierowicz. Warszawa: Oficyna Wydawnicza Volumen.
- Hobbes T. (1954). *Lewiatan, czyli materia, forma i władza państwa kościelnego i świeckiego*, tłum. Cz. Znamierowski. Warszawa: Państwowe Wydawnictwo Naukowe.
- Hobbes T. (2005). *Lewiatan, czyli materia, forma i władza państwa kościelnego i świeckiego*. Ze wstępem i przypisami J.C.A. Gaskina, tłum. Cz. Znamierowski. Warszawa: Fundacja Aletheia.
- Hume D. (1952). *Traktat o naturze ludzkiej*, tłum. Cz. Znamierowski. Kraków: Polska Akademia Umiejętności.
- Kuniński M. (1999). *Wiedza, etyka i polityka w myśli F.A. von Hayeka*. Kraków: Księgarnia Akademicka.
- Kuniński M. (2010). „Problem stanu natury i stanu politycznego w teorii umowy społecznej Thomasa Hobbesa”, w: Z. Rau, M. Chmieliński (red.), *Umowa społeczna i jej krytycy w myśli politycznej i prawnej*. Warszawa: Wydawnictwo Naukowe Scholar.
- Miklaszewska J. (1994). *Libertariańskie koncepcje wolności i własności*. Rozprawy habilitacyjne nr 280. Kraków: Uniwersytet Jagielloński.
- Mises L., von (1966 [1949]). *Human Action. A Treatise on Economics* (wyd. 3). Chicago: Contemporary Books, Inc.
- Nozick R. (1969) „Newcomb’s problem and two principles of choice”, w: N. Rescher (red.), *Essays in Honor of Carl G. Hempel*. Dordrecht: D. Reidel.
- Nozick R. (1974). *Anarchy, State and Utopia*. New York: Basic Books Inc.
- Polanyi M. (1980 [1951]). *The Logic of Liberty. Reflections and Rejoinders*. Chicago: The University of Chicago Press.
- Popper K.R. (1957). *The Poverty of Historicism*. Boston: Beacon Press.
- Popper K.R. (1999 [1989]). *Nędza historycyzmu*, pod red. S. Amsterdamskiego. Warszawa: Wydawnictwo Naukowe PWN.
- Rawls J. (1999 [1971]). *A Theory of Justice*. Cambridge: The Belknap Press of Harvard University Press.
- Schumpeter J.A. (1908). *Das Wesen und der Hauptinhalt der theoretischen Nationalökonomie*. München–Leipzig: Duncker und Humblot.
- Schumpeter J.A. (1909). „On the concept of social value”, *Quarterly Journal of Economics*, nr 23(2).
- Telser L.G. (1980). „A theory of self-enforcing agreements”, *Journal of Business*, nr 53(1), za: R.L. Stump „A model of self-enforcing agreement use in business-to-business exchange relationships”, <http://www.impgroup.org/uploads/papers/262.pdf> [dostęp: 17.11.2012].
- Ullmann-Margalit E. (1977). *The Emergence of Norms*. Oxford: Oxford University Press.
- Watkins J.W.N. (1952). „The principle of methodological individualism”, *The British Journal for the Philosophy of Science*, nr 3(10).
- Weber M. (1951). *Gesammelte Aufsätze zur Wissenschaftslehre* Tübingen, za: M. Kuniński (2006). „Demokracja i demokratyczny człowiek”, w: tegoż, *O cnotach, demokracji i liberalizmie rozumnym*. Kraków: Ośrodek Myśli Politycznej.
- Weber M. (2002). *Gospodarka i społeczeństwo. Zarys socjologii rozumiejącej*, tłum. D. Lachowska. Warszawa: Wydawnictwo Naukowe PWN.
- Wright E.O., Levine A., Sober E. (2003). „Marxism and methodological individualism”, w: D. Matravers, J. Pike (red.), *Debates in Contemporary Political Philosophy. An Anthology*. London–New York: Routledge.

Explaining the emergence of norms

Since the second half of the 20th century, the game theory, and particularly the Prisoner's Dilemma, has been widely used to explain the emergence of norms (Buchanan, Gauthier, Harsanyi, Nozick, Rawls, Ullmann-Margalit). Nevertheless, in light of deeper analyses, it has been demonstrated that the Prisoner's Dilemma is not enough, since its matrix itself is determined by norms which could not be explained in its own terms (Ullmann-Margalit, de Jasay). Following E. Ullmann-Margalit, the author argues that the explanation of norms involves three normative levels: norms of rational decision making, solidified in the process of group selection; moral norms based on the action patterns determined by the rational decision making of the Prisoner's Dilemma structure; and finally norms designed and selected in the process of group selection and critical assessment of their meaning and consequences.

Keywords: explanation, game theory, Prisoner's Dilemma, norm, group selection.