Filozofia Nauki
Rok XXII, 2014, Nr 2(86)

Zbigniew Tworak

Logika przekonań warunkowych

Banałem jest powiedzenie, że ludzkie przekonania ulegają ciągłym zmianom oraz że zmiany te wywoływane są przez pozyskiwane informacje. Wiadomość, że firma *X* zamierza wytoczyć proces firmie *Y* w sprawie jej nowego produktu, dostarcza informacji, że firma *Y* najprawdopodobniej nie wprowadzi owego produktu na rynek w ustalonym terminie. Informacja ta może wywołać wśród inwestorów (przynajmniej ich części) przekonanie, że ceny akcji firmy *Y* w najbliższym czasie spadną. Temat zmiany przekonań łączy się z zagadnieniem przekonań warunkowych, czyli takich, do których podmiot dochodzi po uzyskaniu określonej, zwykle nowej, informacji. Logika przekonań warunkowych (CDL) dostarcza narzędzi ich analizy i rządzących nimi reguł. Odpowiedni funktor ujmowany jest w niej jako dwuargumentowy funktor modalny, a jego semantyka opiera się na semantyce wykorzystywanej w analizie nierzeczywistych okresów warunkowych.

1. Syntaktyka. Język logiki przekonań warunkowych J_{CDL} powstaje przez rozszerzenie języka klasycznego rachunku zdań o formuły dotyczące przekonań *simpliciter*, postaci $B_i\alpha$, oraz przekonań warunkowych, postaci $B_i(\alpha/\beta)$, dla $i \in G$[1]. Formułę $B_i\alpha$ można odczytywać: „Podmiot *i* jest przekonany, że α". Natomiast formuła $B_i(\alpha/\beta)$ sprzęga przekonanie, że α, z pozyskaniem przez podmiot *i* informacji, że β. Mówiąc nieco nieprecyzyjnie, funktor B_i reprezentuje przekonania podmiotu *i* sprzed pozyskania informacji, że β, a funktor $B_i(-/\beta)$ reprezentuje przekonania podmiotu po pozyskaniu informacji, że β. Formułę $B_i(\alpha/\beta)$ można odczytywać: „Podmiot *i* byłby przekonany, że α, gdyby odkrył/dowiedział się, że β" lub prościej: „Po uzyskaniu informacji, że β, podmiot *i* dojdzie do przekonania, że α". Pierwszy sposób akcentuje kontrfaktyczny lub dyspozycyjny charakter przekonań warunkowych, drugi kła-

[1] Napis $B_i(\alpha/\beta)$ nawiązuje do oznaczenia prawdopodobieństwa warunkowego.

dzie nacisk na ich rewizyjny charakter. Niezależnie od sposobu odczytania dopuszczalna jest sytuacja, w której $B_i(\alpha/\beta)$ i $B_i(\neg\alpha/\gamma)$. W pewnym sensie formuła $B_i(\alpha/\beta)$ opisuje proces budowy przekonań danego podmiotu jako wynik ich *aktualizacji* na podstawie pewnej dodatkowej pomocniczej informacji (zmieniającej punkt widzenia podmiotu). Można też powiedzieć, że wyraża ona „strategię" lub „plan" podmiotu na wypadek zmiany przekonań spowodowanej pozyskaniem nowej informacji. W szczególności, pozyskaniu pewnej nowej informacji może towarzyszyć uwiarygodnienie lub podważenie wiarygodności innej informacji, co prowadzi do modyfikacji stanu przekonań.

Przekonania *simpliciter*, tworzące wyjściowy zbiór (korpus) przekonań, stanowią szczególny przypadek przekonań warunkowych, a mianowicie, gdy $\beta = \top$ (gdzie $\top$ oznacza dowolną prawdę logiczną). Tak więc, $B_i\alpha$ może zostać wprowadzone jako skrót dla $B_i(\alpha/\top)$. Podmiot jest przekonany *simpliciter*, że α, jeśli przekonanie, że α, jest następstwem informacji pewnej (tj. o prawdopodobieństwie równym 1). Inaczej rzecz ujmując, informacje trywialnie prawdziwe nie wymuszają korekty przekonań.

Definicja 1. Niech ZZ będzie nieskończonym przeliczalnym zbiorem zmiennych zdaniowych, które zwyczajowo oznaczamy literami p, q, r (ewentualnie z indeksami). Pojęcie formuły zdaniowej języka J_{CDL} definiujemy przez indukcję:

$$\alpha, \beta \in J_{\mathrm{CDL}} := p\ (\in \mathrm{ZZ}) \mid \bot \mid \neg\alpha \mid \alpha \wedge \beta \mid \alpha \vee \beta \mid B_i\alpha \mid B_i(\alpha/\beta).$$

Spójniki implikacji $\rightarrow$ i równoważności $\equiv$ można wprowadzić za pomocą standardowych definicji. Stała $\bot$ oznacza dowolny fałsz logiczny; stałą $\top$ definiujemy standardowo, tj. $\top := \neg\bot$. Niech ponadto $P_i(\alpha/\beta)$ skraca: $\neg B_i(\neg\alpha/\beta)$. $P_i(\alpha/\beta)$ znaczy mniej więcej tyle, że informacja β nie wymusza wyrzeczenia się α. Zauważmy jeszcze, że funktor $B_i(-/-)$ dotyczy zmiany przekonań ze względu na informacje dotyczące zarówno świata zewnętrznego, jak i czyichś przekonań (np. jego własnych przekonań). Na przykład, formuła $B_i(\alpha/(B_j\alpha))$ może oznaczać, że i dostosowuje swoje przekonania do przekonań j w sprawie α.

2. Semantyka. Zacznijmy od semantycznej charakterystyki CDL[2]. Przekonania podmiotu są reprezentowane przez zbiory światów (stanów) możliwych. Niech R będzie trójargumentową relacją na zbiorze X. Piszemy $R_z(x, y)$ zamiast $R(x, y, z)$. Dla dowolnego $z \in X$, R_z jest binarną relacją (na X) taką, że $R_z(x, y)$ wtw, gdy $R(x, y, z)$.

Definicja 2. Warunkową strukturą doksastyczną nazywamy dowolny układ:

$$\boldsymbol{C} = \langle S, G, \{\leq_{i,w}: w \in S, i \in G\}\rangle,$$

w którym:
$S \neq \varnothing$ jest zbiorem możliwych światów (lub stanów rzeczy);
$G \neq \varnothing$ jest skończonym zbiorem podmiotów;

[2] Podobne podejście znajduje się w (Baltag, Smets 2006), (van Benthem, Martinez 2008: 238).

$\leq_{i, w}$, dla dowolnych $i \in G$ oraz $w \in S$, jest binarną relacją na $S_{i, w} \subseteq S$, quasi-porządkującą oraz dodatkowo mocno spójną, tj. $\forall x, y \in S_{i, w}$ $(x \leq_{i, w} y \vee y \leq_{i, w} x)$[3].

Relację $\leq_{i, w}$, interpretującą funktor przekonań warunkowych B_i(-/-), nazywamy *relacją relatywnej wiarygodności* lub *relacją preferencji*[4]. Wzór $x \leq_{i, w} y$ można odczytywać: „Świat x jest dla podmiotu i co najmniej tak wiarygodny (preferowany) — ze względu na dany świat bazowy w — jak świat y" (z powodów historycznych piszemy $x \leq_{i, w} y$ zamiast $y \leq_{i, w} x$). Zbiór $S_{i, w} = \{x \in S: x \leq_{i, w} y$ dla pewnego $y \in S\}$ — pole relacji $\leq_{i, w}$ — to zbiór światów, które podmiot i w danym świecie bazowym w rozważa jako możliwe. Przyjmujemy, że $S_{i, w} \neq \varnothing$ dla każdego $i \in G$ oraz każdego $w \in S$ (warunek normalności). Światy nienależące do $S_{i, w}$ są dla i ze względu na w tak niewiarygodne, że nie warto ich rozważać (są to np. światy „cudowne")[5]. Naturalne wydaje się założenie, że świat bazowy jest zawsze dla podmiotu i możliwy: dla każdego $i \in G$ oraz każdego $w \in S$, $w \in S_{i, w}$[6]. Warunek ten można skomentować następująco: uzyskane przez podmiot informacje nie wykluczają świata bazowego ze zbioru rozważanych przez niego możliwości. Zauważmy jeszcze, że nie czynimy tu żadnego założenia na temat osiągalności światów ze zbioru $S_{i, w}$. Dość naturalne wydaje się założenie, że zbiór $S_{i, w}$ tworzą światy, które są dla i osiągalne z w (można jednak przyjąć zależność odwrotną).

Para $\langle S_{i, w}, \leq_{i, w}\rangle$ tworzy *przestrzeń relatywnej wiarygodności*, w której światy tworzące zbiór $S_{i, w}$ są uporządkowane przez relację $\leq_{i, w}$[7]. Warunek mocnej spójności gwarantuje, że wszystkie rozważane przez podmiot możliwości zostają przez niego oszacowane jako mniej lub bardziej wiarygodne (względem świata bazowego) i porównane ze sobą pod względem wiarygodności; a dokładniej — z dwóch światów ze zbioru $S_{i, w}$ jeden jest dla podmiotu i bardziej wiarygodny niż drugi lub tak samo wiarygodny jak drugi (ze względu na dany świat bazowy). Relację bycia bardziej wiarygodnym definiujemy w zwykły sposób: $x <_{i, w} y$ wtw $x \leq_{i, w} y$ i $\neg(y \leq_{i, w} x)$. Przypuśćmy, że podmiot i (w danym świecie w) pozyskał informację, która wystarcza do stwierdzenia, iż zaktualizuje się jedna z możliwości x bądź y, ale nie pozwala roz-

---

[3] Relację nazywamy quasi-porządkującą, jeśli jest zwrotna i przechodnia. Przypomnijmy jeszcze, że mocna spójność implikuje zwrotność. W literaturze anglojęzycznej relację quasi-porządkującą i mocno spójną określa się terminem *total preorder*.

[4] Lewis w monografii poświęconej nierzeczywistym okresom warunkowym relację $\leq_w$ interpretuje jako relację relatywnego podobieństwa między światami (Lewis 1973). Porządkuje ona światy ze względu na ich odległość od danego świata bazowego w. Przy tej interpretacji od relacji $\leq_w$ wymaga się spełnienia warunku słabej koncentryczności (ang. *weak centering*): świat bazowy w jest elementem minimalnym względem relacji $\leq_w$, tj. dla każdego $x \in S_w$, $w \leq_w x$. Tutaj warunek ten nie wydaje się konieczny.

[5] Oczywiście, można przyjąć, że relacja $\leq_{i, w}$ jest *uniwersalna*, czyli dla każdego w, $S_{i, w} = S$. Warunek ten oznacza, że zbiór $S_{i, w}$ tworzą wszystkie możliwości logiczne.

[6] Lewis (1973: 48) określa światy spełniające ten warunek jako *self-accessible*.

[7] Stanowi ona jakościowy odpowiednik przestrzeni probabilistycznej.

strzygnąć, która z nich się zaktualizuje. Będzie on (warunkowo) przekonany, iż zajdzie x wtedy i tylko wtedy, gdy $x <_{i, w} y$. Dodajmy, że każda z relacji $\leq_{i, w}$ wyznacza w naturalny sposób pewną relację (doksastycznej) nieodróżnialności $\approx_{i, w}$ (a mianowicie: $x \approx_{i, w} y$ wtw $x \leq_{i, w} y$ i $y \leq_{i, w} x$), która dzieli rozważane przez podmiot światy na uporządkowane klasy abstrakcji (odpowiednik Lewisowskiego systemu sfer).

Wymienione własności relacji $\leq_{i, w}$ (zwrotność, przechodniość i mocna spójność) są warunkami minimalnymi. W szczególności dopuszczają sytuację, w której istnieje nieskończony łańcuch światów coraz bardziej i bardziej wiarygodnych (względem danego świata bazowego): $\ldots x_n <_{i, w} x_{n-1} <_{i, w} \ldots <_{i, w} x_1 <_{i, w} x_0$. Aby temu zapobiec, przyjmujemy dodatkowo warunek dobrego ufundowania: dla każdego $i \in G$ oraz każdego $w \in S$, relacja $\leq_{i, w}$ jest dobrze ufundowana. Symbolem $\mathrm{Min}_{i, w}(X)$ oznaczmy zbiór elementów $(\leq_{i, w})$-minimalnych w zbiorze $X \subseteq S$: $\mathrm{Min}_{i, w}(X) = \{x \in X: x \leq_{i, w} y$, dla każdego $y \in X\}$[8]. Tworzą go światy (spośród światów w X), które zdaniem podmiotu i są najbardziej wiarygodne względem danego świata bazowego w. Warunek dobrego ufundowania relacji $\leq_{i, w}$ stanowi, że każde niepuste przecięcie z $S_{i, w}$ posiada elementy $(\leq_{i, w})$-minimalne, czyli najbardziej wiarygodne dla i; a dokładniej — dla dowolnego $X \subseteq S$, jeżeli $X \cap S_{i, w} \neq \varnothing$, to $\mathrm{Min}_{i, w}(X \cap S_{i, w}) \neq \varnothing$[9].

Definicja 3. Modelem na strukturze $\boldsymbol{C}$ jest para $\boldsymbol{M} = \langle \boldsymbol{C}, V \rangle$, w której V: ZZ $\rightarrow 2^S$ jest funkcją wartościowania dla zmiennych zdaniowych, tj. poszczególnym zmiennym zdaniowym przyporządkowuje zbiory światów możliwych.

Dla dowolnej $p \in ZZ$, zbiór $V(p)$ jest traktowany jako zbiór tych światów, w których zmienna p jest prawdziwa. Relację spełniania $\models$ definiuje się przez indukcję po budowie formuł w następujący sposób (zakładamy, że zachowuje się ona klasycznie wobec klasycznych spójników):

Definicja 4. (a) $(\boldsymbol{M}, w) \models p$ wtw $w \in V(p)$, dla dowolnej $p \in$ ZZ;

(b) $(\boldsymbol{M}, w) \not\models \bot$, dla dowolnego w;
(c) $(\boldsymbol{M}, w) \models \neg\alpha$ wtw $(\boldsymbol{M}, w) \not\models \alpha$;
(d) $(\boldsymbol{M}, w) \models \alpha \wedge \beta$ wtw $(\boldsymbol{M}, w) \models \alpha$ i $(\boldsymbol{M}, w) \models \beta$;
(e) $(\boldsymbol{M}, w) \models \alpha \vee \beta$ wtw $(\boldsymbol{M}, w) \models \alpha$ lub $(\boldsymbol{M}, w) \models \beta$.

Niech $\|\alpha\|_{\boldsymbol{M}} = \{w \in S: (\boldsymbol{M}, w) \models \alpha\}$ oznacza zbiór tych wszystkich światów modelu $\boldsymbol{M}$, w których spełniona (prawdziwa) jest formuła α (gdy model $\boldsymbol{M}$ jest domyślny, można używać skróconej notacji $w \models \alpha$ oraz $\|\alpha\|$)[10]. Świat w, taki że $w \in \|\alpha\|_{\boldsymbol{M}}$, będziemy nazywać *$\alpha$-światem*. Gdy $\|\alpha\|_{\boldsymbol{M}} \neq \varnothing$, mówimy, że *$\alpha$ jest spełnialna w modelu $\boldsymbol{M}$*. Po tych ustaleniach możemy nadać następującą postać klauzuli dla formuł dotyczących przekonań warunkowych:

---

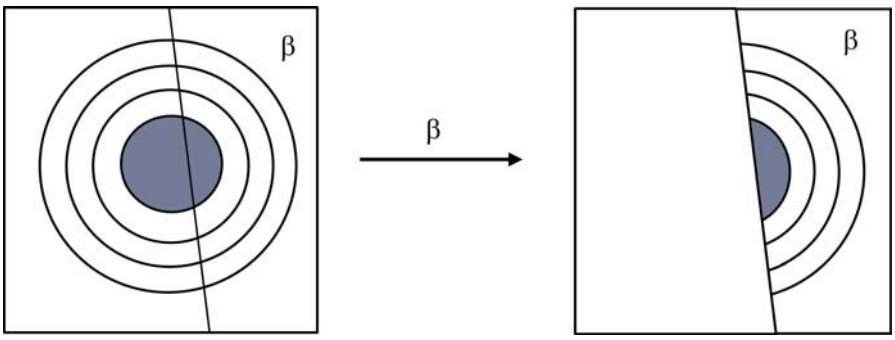
[8] Alternatywnie: $\mathrm{Min}_{i, w}(X) = \{x \in X$: nie istnieje $y \in X$, taki że $y <_{i, w} x\}$. Oczywiście, $\mathrm{Min}_{i, w}(X) \subseteq X$.

[9] Jest on spełniony automatycznie, gdy S jest zbiorem skończonym.

[10] Zwyczajowo ||α|| czyta się jako „sąd α".

(f) $(\boldsymbol{M}, w) \models B_i(\alpha/\beta)$ wtw $\mathrm{Min}_{i,w}(\|\beta\|_{\boldsymbol{M}} \cap S_{i,w}) \subseteq \|\alpha\|_{\boldsymbol{M}}$
wtw $(\boldsymbol{M}, x) \models \alpha$, dla każdego $x \in \mathrm{Min}_{i,w}(\|\beta\|_{\boldsymbol{M}} \cap S_{i,w})$,

gdzie $\mathrm{Min}_{i,w}(\|\beta\|_{\boldsymbol{M}} \cap S_{i,w}) = \{x \in \|\beta\|_{\boldsymbol{M}} \cap S_{i,w}: x \leq_{i,w} y$, dla każdego $y \in \|\beta\|_{\boldsymbol{M}} \cap S_{i,w}\}$[11]. Warunek (f) określa standard racjonalnej aktualizacji przekonań nawiązujący do pojęcia racjonalnego wyboru. Przede wszystkim ustala on strategię podmiotu na wypadek zmiany przekonań spowodowanej pozyskaniem nowej informacji (niekoniecznie spójnej z posiadanymi już informacjami). Na jego mocy o prawdziwości formuły $B_i(\alpha/\beta)$ w świecie w (danego modelu) decyduje to, czy α jest prawdziwa w każdym β-świecie należącym do $S_{i,w}$, który jest $(\leq_{i,w})$-minimalny. Mówiąc mniej precyzyjnie, $B_i(\alpha/\beta)$ jest prawdziwa w świecie w (danego modelu), jeśli α jest prawdziwa w każdym świecie, który z punktu widzenia i jest możliwy oraz potwierdza informację β, a ponadto jest dla i tak wiarygodny ze względu na świat w, jak to tylko możliwe. Tak więc gdy β jest nową nietrywialną informacją pozyskaną przez podmiot z takiego lub innego wiarygodnego źródła, zwłaszcza niespójną z jego aktualnymi przekonaniami, wówczas to, co dotychczas było dla niego najbardziej wiarygodne, zostaje podane w wątpliwość lub nawet wykluczone. W tej sytuacji powinien on w sferze tego, co możliwe, wyodrębnić β-światy i tam poszukać możliwości najbardziej wiarygodnych. Warunek dobrego ufundowania gwarantuje, że jeżeli w zbiorze $S_{i,w}$ rozważanych przez podmiot i możliwości znajdują się jakieś β-światy, to istnieje w nim β-świat (niekoniecznie jeden) minimalny ze względu na relację $\leq_{i,w}$, czyli najbardziej dla i wiarygodny. Łatwiej będzie to zrozumieć, jeśli posłużymy się diagramem (pole zacienione reprezentuje wyjściowy stan przekonań podmiotu):



Z warunku (f) otrzymujemy łatwo warunek prawdziwości dla przekonań *simpliciter* $B_i\alpha$. Ponieważ $\|\top\|_{\boldsymbol{M}} = S$, więc:

(g) $(\boldsymbol{M}, w) \models B_i\alpha$ wtw $\mathrm{Min}_{i,w}(S_{i,w}) \subseteq \|\alpha\|_{\boldsymbol{M}}$

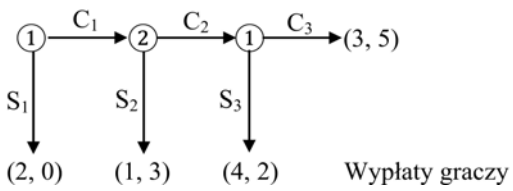
[11] Z warunku tego otrzymujemy: $\|B_i(\alpha/\beta)\|_{\boldsymbol{M}} = \{w \in S: \mathrm{Min}_{i,w}(\|\beta\|_{\boldsymbol{M}} \cap S_{i,w}) \subseteq \|\alpha\|_{\boldsymbol{M}}\}$. Istnieje wyraźne podobieństwo między warunkiem (f) a warunkiem prawdziwości dla nierzeczywistych okresów warunkowych.

wtw $(\boldsymbol{M}, x) \models \alpha$, dla każdego $x \in \mathrm{Min}_{i,w}(S_{i,w})$.

Intuicyjnie rzecz ujmując, w świecie w podmiot i jest przekonany, że α wtedy i tylko wtedy, gdy α zachodzi w każdym świecie najbardziej wiarygodnym wśród światów rozważanych jako możliwe przez i w w[12]. Zauważmy przy okazji, że przekonania mogą być zawodne, jako że nie wymaga się, by świat w był elementem $\mathrm{Min}_{i,w}(S_{i,w})$.

Dla ilustracji posłużmy się następującym przykładem pochodzącym z teorii gier.

Przykład (teoria gier — Stonoga). „Stonoga" jest grą ekstensywną z pełną informacją, autorstwa Rosenthala[13]. Oto jedno z wielu jej ujęć (dla uproszczenia ograniczmy ją do trzech ruchów). W grze biorą udział dyrektorzy dwóch instytutów — gracze 1 i 2 — którym urzędnik ministerstwa składa oferty budżetu na przyszły rok (wypłaty). Każdy gracz może ją przyjąć (akcja S) lub odrzucić (akcja C). Przyjęcie oferty kończy grę. Skutkiem jej odrzucenia jest kontynuacja gry i nowa oferta. Oto zapis gry w postaci drzewa (węzły to punkty decyzyjne poszczególnych graczy, krawędzie reprezentują możliwe akcje, wyniki opisane są za pomocą par liczb: pierwsza liczba to wypłata gracza 1, druga — wypłata gracza 2):



[12] Warunek ten odbiega od zwykle przyjmowanego w logice epistemicznej warunku prawdziwości dla zdań tej postaci: $B_i\alpha$ jest prawdziwe w świecie w (modelu $\boldsymbol{M}$) wtedy i tylko wtedy, gdy α jest prawdziwe w każdym świecie dostępnym z w (możliwym ze względu na w) dla podmiotu i. Źródłem owej różnicy jest podwójna relatywizacja przekonań: do światów, które — w danym świecie — podmiot rozważa jako możliwe oraz do uporządkowania owych możliwości. Zauważmy jednak, że im więcej światów podmiot rozważa jako dostępne (możliwe ze względu na dany świat), tym skromniejszy jest jego zbiór przekonań (i na odwrót). W skrajnym przypadku jest to zbiór złożony wyłącznie z prawd logicznych.

[13] Przypomnijmy, że w grach tego typu gracze podejmują decyzje sekwencyjnie (tj. na przemian) oraz każdy gracz w każdym punkcie decyzyjnym zna cały dotychczasowy przebieg gry (tj. ma informacje o wcześniejszych decyzjach pozostałych graczy). Gdy składa się ona ze skończonej liczby ruchów, gracze (i analitycy) mogą za pomocą określonych procedur przewidzieć wynik gry (jej rozwiązanie). Racjonalny gracz wykonuje swój pierwszy ruch, rozważając każdą z sekwencji reakcji i kontrreakcji, które będą konsekwencjami wybranego ruchu. Następnie ustala preferowany wynik spośród możliwych wyników takich sekwencji i wybiera ruch rozpoczynający sekwencję ruchów prowadzącą do ustalonego rezultatu. Taki proces nazywa się *indukcją wsteczną*, ponieważ wnioskowanie działa wstecz, tj. od ewentualnych wyników sekwencji decyzji. Opisany proces reprezentuje się często pod postacią drzew.

Strategia gracza to, z grubsza rzecz biorąc, plan akcji na wszystkie możliwe sytuacje. Gracz 1 ma cztery strategie: C_1C_3, C_1S_3, S_1C_3, S_1S_3 (na przykład, C_1S_3 oznacza wybór C_1 w pierwszym punkcie decyzyjnym i S_3 w trzecim punkcie decyzyjnym, o ile zostanie on osiągnięty, czyli po akcjach C_1C_2). Gracz 2 ma tylko dwie strategie (po C_1): C_2 i S_2. Standardowo przyjmuje się założenie, że gracze postępują *racjonalnie*, co w języku teorii gier oznacza, że używają oni strategii dominujących, tzn. każdy z graczy stara się zmaksymalizować swoją własną wypłatę, niezależnie od tego, co zrobią inni gracze. To, czy gracz jest racjonalny, jest całkowicie zdeterminowane przez wybór strategii i jego przekonania. Ogólnie, gracz *i* postępuje racjonalnie wtedy i tylko wtedy, gdy w każdym osiągniętym punkcie decyzyjnym (węźle drzewa) jest przekonany, że wybrana przez niego akcja prowadzi do wyższej wypłaty niż dowolna inna. Przekonanie to opiera się na indukcji wstecznej przeprowadzonej przez podejmującego decyzję gracza. Dla przykładu, w pierwszym węźle decyzyjnym gracz 1 powinien wybrać S_1 na podstawie następującego wnioskowania: „(a) Gdyby osiągnięty został węzeł trzeci, wybrałbym S_3. (b) Mój partner w węźle drugim — gdyby został on osiągnięty — wybrałby S_2: jego przekonanie o mojej racjonalności pociąga przekonanie, że (a), a ponadto jest on racjonalny. (c) Skoro uważam, że mój partner w węźle drugim wybierze S_2, powinienem — jako osoba racjonalna — w węźle pierwszym wybrać S_1”[14].

Struktura warunkowa tej gry zawiera po jednym świecie dla każdej kombinacji strategii (lub akcji) wybieranych przez obu graczy, tj. profilu strategii. Aby nie komplikować notacji, będziemy utożsamiać światy z trójkami: ⟨C, C, C⟩, ⟨C, C, S⟩, …, ⟨S, S, S⟩. Na przykład ⟨C, C, S⟩ reprezentuje profil strategii, w myśl której gracz 1 w węźle pierwszym decyduje się kontynuować grę (C), a w węźle trzecim — o ile zostanie on osiągnięty — decyduje się ją zakończyć (S), gracz 2 w węźle drugim — o ile zostanie on osiągnięty — decyduje się kontynuować grę (C)[15]. Podobnie zinterpretujemy trójkę ⟨S, S, S⟩, która reprezentuje profil strategii odpowiadający kolejnym krokom indukcji wstecznej przeprowadzonej przez gracza 1 przed podjęciem pierwszej decyzji: obaj gracze w każdym osiągniętym węźle decydują się zakończyć grę (jest on profilem strategii, mimo że wybór S w węźle pierwszym sprawia, iż pozostałe węzły nie zostaną osiągnięte i gracze nie będą mieli możliwości dokonać w nich wyboru akcji). Przyjmijmy, że ⟨C, C, S⟩ jest światem aktualnym. Relacje preferencji (relatywnej wiarygodności) właściwe dla obu graczy są określone przez następujące warunki:

$\langle C, C, S\rangle <_{1, \langle C, C, S\rangle} x$, dla dowolnego $x \neq \langle C, C, S\rangle$;
$\langle S, S, S\rangle <_{1, \langle S, C, S\rangle} x$, dla dowolnego $x \neq \langle S, S, S\rangle$;

[14] Zauważmy, że przesłanki tego wnioskowania mają postać nierzeczywistych okresów warunkowych.

[15] Reprezentowanie profilów strategii w ten sposób nie jest standardowe. Dla naszych celów jest jednak wygodne. W szczególności pozwala pominąć indeksy przy symbolach reprezentujących wybrane przez graczy akcje.

$\langle S, S, S\rangle <_{1, \langle S, S, S\rangle} x$, dla dowolnego $x \neq \langle S, S, S\rangle$;
$\langle S, C, S\rangle <_{2, \langle C, C, S\rangle} \langle C, C, C\rangle <_{2, \langle C, C, S\rangle} x$, dla dowolnego $x \neq \langle S, C, S\rangle$ i $\langle C, C, C\rangle$;
$\langle S, C, S\rangle <_{2, \langle S, C, S\rangle} \langle C, C, C\rangle <_{2, \langle S, C, S\rangle} x$, dla dowolnego $x \neq \langle S, C, S\rangle$ i $\langle C, C, C\rangle$;
$\langle S, S, S\rangle <_{2, \langle S, S, S\rangle} \langle C, S, S\rangle <_{2, \langle S, S, S\rangle} x$, dla dowolnego $x \neq \langle S, S, S\rangle$ i $\langle C, S, S\rangle$.

Niech $At = \{p_1, p_2, p_3\}$ będzie zbiorem zdań, w którym:

p_1 skraca zdanie: 1 wybiera C_1;
p_2 skraca zdanie: 2 wybiera C_2;
p_3 skraca zdanie: 1 wybiera C_3.

Negacje ich będą wówczas skracać zdania dotyczące dokonania wyboru przez podmiot w danym węźle strategii opozycyjnej (S); na przykład, $\neg p_1$ skraca zdanie: „1 wybiera S_1" (wybór ów stanowi urzeczywistnienie wyniku indukcji wstecznej przeprowadzonej przez gracza 1). W ten sposób każdy świat czy profil strategii można wyczerpująco opisać, łącząc koniunkcją odpowiednie zdania i ich negacje; na przykład, $p_1 \wedge p_2 \wedge \neg p_3$ opisuje $\langle C, C, S\rangle$. Funkcję wartościowania V definiujemy następująco: dla dowolnego $j \in \{1, 2, 3\}$, $\langle d_1, d_2, d_3\rangle \in V(p_j)$ wtw $d_j = C$. W konsekwencji:

$\langle d_1, d_2, d_3\rangle \models p_j$ wtw $d_j = C$;
$\langle d_1, d_2, d_3\rangle \models \neg p_j$ wtw $d_j = S$.

Racjonalność gracza 1 jest opisywana przez zdanie $r_1 := (p_1 \wedge B_1 p_2 \wedge \neg p_3) \vee (\neg p_1 \wedge B_1 \neg p_2 \wedge \neg p_3)$, racjonalność gracza 2 jest zaś opisywana zdaniem $r_2 := (p_2 \wedge B_2(p_3/p_1)) \vee (\neg p_2 \wedge B_2(\neg p_3/p_1))$. Oba zdania są spełnione w świecie aktualnym $\langle C, C, S\rangle$ oraz w światach $\langle S, C, S\rangle$ i $\langle S, S, S\rangle$. Rozważmy świat $\langle S, C, S\rangle$. Łatwo sprawdzić, że $\langle S, C, S\rangle \models \neg p_1 \wedge B_1 \neg p_2 \wedge \neg p_3$ (co przesądza o prawdziwości r_1). Warunek spełniania dla $B_1 \neg p_2$ wygląda następująco:

$\langle S, C, S\rangle \models B_1 \neg p_2$ wtw $x \models \neg p_2$, dla każdego $x \in \mathrm{Min}_{1, \langle S, C, S\rangle}(S_{1, \langle S, C, S\rangle})$.

Potwierdzenie prawej strony tej równoważności uzyskujemy na podstawie ustalenia, że (a) $\langle S, S, S\rangle <_{1, \langle S, C, S\rangle} x$, dla dowolnego $x \neq \langle S, S, S\rangle$, oraz (b) $\langle S, S, S\rangle$ spełnia $\neg p_1$.

Z drugiej strony, $\langle S, C, S\rangle \models p_2 \wedge B_2(p_3/p_1)$ (co przesądza o prawdziwości r_2). Warunek spełniania dla $B_2(p_3/p_1)$ wygląda następująco:

$\langle S, C, S\rangle \models B_2(p_3/p_1)$ wtw $x \models p_3$ dla każdego $x \in \mathrm{Min}_{2, \langle S, C, S\rangle}(\|p_1\| \cap S_{2, \langle S, C, S\rangle})$.

Potwierdzenie prawej strony równoważności uzyskujemy na podstawie ustalenia, że (a) $\langle S, C, S\rangle <_{2, \langle S, C, S\rangle} \langle C, C, C\rangle <_{2, \langle S, C, S\rangle} x$ dla dowolnego $x \neq \langle S, C, S\rangle$ i $\langle C, C, C\rangle$ oraz (b) $\langle C, C, C\rangle$ spełnia zarówno p_1, jak i p_3[16].

[16] W świecie $\langle S, C, S\rangle$ dla gracza 2 w węźle drugim — tj. węźle, w którym dokonuje wyboru (rzecz jasna, w tej sytuacji gracz 1 w węźle pierwszym musiał wybrać C) — możliwy i zarazem najbardziej preferowany jest świat $\langle C, C, C\rangle$.

W podobny sposób możemy sprawdzić pozostałe przypadki. Z pewnego punktu widzenia interesujące są następujące dwie zależności: $\langle C, C, S\rangle \not\models r_1 \wedge B_1(r_2 \wedge B_2 r_1) \rightarrow \neg p_1$ oraz $\langle C, C, S\rangle \models \neg B_2(r_1/p_1)$. Zajmiemy się tą kwestią w paragrafie 6.

Pojęcie prawdziwości w modelu i strukturze definiuje się w zwykły sposób:

Definicja 5. Formuła α jest prawdziwa w modelu $\boldsymbol{M}$ (symbolicznie: $\boldsymbol{M}. \models \alpha$) wtw dla każdego $x \in S$, $(\boldsymbol{M}, x) \models \alpha$. Formuła α jest prawdziwa w strukturze $\boldsymbol{C}$ (symbolicznie: $\boldsymbol{C} \models \alpha$) wtw jest ona prawdziwa w każdym modelu na tej strukturze (tzn. jest ona prawdziwa przy dowolnym wartościowaniu V określonym dla tej struktury).

3. Aksjomatyka. Logikę CDL określają następujące aksjomaty i reguły inferencji:[17]

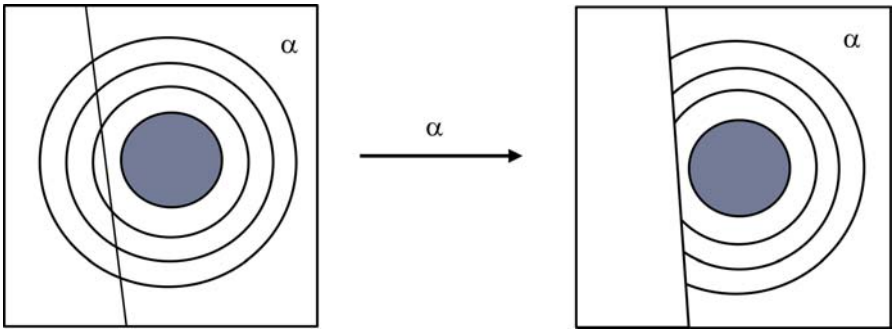
A0. Dowolne α będące schematem tautologii KRZ.
A1. $B_i((\alpha \rightarrow \beta)/\gamma) \rightarrow (B_i(\alpha/\gamma) \rightarrow B_i(\beta/\gamma))$.
A2. $B_i(\alpha/\alpha)$.
A3. $\alpha \rightarrow P_i(\top/\alpha)$.
A4. $B_i(\alpha/\beta) \wedge B_i(\beta/\alpha) \rightarrow (B_i(\gamma/\alpha) \equiv B_i(\gamma/\beta))$.
A5. $B_i(\alpha/\beta) \rightarrow (B_i(\gamma/(\alpha \wedge \beta)) \equiv B_i(\gamma/\beta))$.
A6. $P_i(\alpha/\beta) \rightarrow (B_i(\gamma/(\alpha \wedge \beta)) \equiv B_i((\alpha \rightarrow \gamma)/\beta))$.
A7. $B_i\alpha \equiv B_i(\alpha/\top)$.

MP. Jeżeli $\vdash \alpha \rightarrow \beta$ i $\vdash \alpha$, to $\vdash \beta$.
BRN. Jeżeli $\vdash \alpha$, to $\vdash B_i(\alpha/\beta)$.

Aksjomat A1 (aksjomat dystrybucji) jest odpowiednikiem aksjomatu K; czyni on z przekonań warunkowych zbiór dedukcyjnie domknięty. Aksjomat A2 można odczytać jako żądanie, aby informacja pozyskana przez podmiot została dołączona do jego wyjściowych przekonań (jako prawdziwa) i w ten sposób je przekształcała. W pewnym sensie stanowi on *klauzulę najwyższego uprzywilejowania*: pozyskaną informację podmiot postrzega jako tak wiarygodną, że ma ona pierwszeństwo nad wszystkimi innymi informacjami (przekonaniami), niezależnie od tego, co to za in-

[17] Istnieje pewna odpowiedniość między wyróżnionymi aksjomatami i regułami a postulatami charakteryzującymi operację rewizji w ujęciu AGM, z jednej strony, oraz niektórymi aksjomatami logiki okresów warunkowych z drugiej. Na związek między postulatami AGM a semantyką nierzeczywistych okresów warunkowych wskazywali Gärdenfors (1998) i Grove (1998). Nie sposób tu dokładniej porównać przedstawianą aksjomatykę z postulatami wymienionych teorii. Pewne uwagi na ten temat znajdują się w podsumowaniu. Dodajmy, że następujące aksjomaty i reguły są wspólne dla prezentowanej tu logiki przekonań warunkowych oraz dla logiki opisanej przez Baltaga i Smets (2006): A0, A1, A2, A6, A7, MP, BRN. Aksjomat A3 jest tam obecny w postaci postulatu charakteryzującego wiedzę: $K_i\alpha \rightarrow \alpha$ (zob. paragraf 7).

formacje. Aksjomat ów jest przez to problematyczny[18]. Z semantycznego punktu widzenia stanowi on jednak, że α zachodzi we wszystkich najbardziej wiarygodnych α-światach, które z punktu widzenia podmiotu są możliwe: $\mathrm{Min}_{i,w}(||\alpha||_{\boldsymbol{M}} \cap S_{i,w}) \subseteq ||\alpha||_{\boldsymbol{M}}$. Nie budzi to żadnych wątpliwości. Tak więc w sposób nieproblematyczny aksjomat A2 można odczytać następująco: pozyskanie przez podmiot informacji, że α, nie wpłynie na zmianę jego przekonania, że α. Innymi słowy, jak długo podmiot będzie dysponował informacją, że α, tak długo będzie przekonany, że α. Graficznie:



Zastępując α przez $\top$, otrzymujemy formułę: $B_i(\top/\top)$, która na podstawie aksjomatu A7 jest równoważna formule $B_i\top$ (odpowiednik aksjomatu N). Stanowi ona, że w korpusie przekonań podmiotu znajdują się (wszystkie) prawdy logiczne[19]. Aksjomat A3, będący ograniczonym postulatem niesprzeczności, można zapisać również w postaci formuł: $\alpha \rightarrow \neg B_i(\bot/\alpha)$ lub $\alpha \rightarrow \neg(B_i(\beta/\alpha) \wedge B_i(\neg\beta/\alpha))$. Zatem stwierdza on, że dopóki pozyskiwane przez podmiot informacje są prawdziwe, dopóty rewizje przekonań dokonywane ze względu na owe informacje nie prowadzą do sprzeczności[20]. Konsekwencją A3 jest formuła: $\neg B_i\bot$ (odpowiednik aksjomatu D), która sta-

[18] Można to pokazać, wstawiając za α do A2 zdanie z paradoksu Moore'a: $p \wedge \neg B_i p$. Oznaczmy je przez φ. Zdanie φ jest niesprzeczne, co oznacza, że jest ono prawdziwe w pewnym świecie. Po uzyskaniu informacji, że φ, podmiot i nie może niesprzecznie sądzić, że φ, ponieważ wtedy i jest przekonany, że p, co jest niezgodne z φ (jego drugim czynnikiem). A zatem po uzyskaniu informacji, że φ, podmiot i jest przekonany o fałszywości φ. Podważa to wspomnianą interpretację aksjomatu A2.

[19] Zauważmy, że odrzucenie A2 dopuszcza światy sprzeczne, tj. spełniające $\alpha \wedge \neg\alpha$.

[20] Postulat ten wydaje się zbyt silny dla „zwykłych" podmiotów. Znakiem szczególnym racjonalności przekonań jest jednak dążenie do utrzymania niesprzeczności przekonań. Rezygnacja z postulatu niesprzeczności przekonań oznaczać mogłaby nadmierną uległość wobec sprzeczności, a w rezultacie pozbawiłaby je wartości heurystycznej. Zauważmy na marginesie, że opierając się na A0, A1 i A7, można udowodnić implikację $B_i\alpha \wedge B_i\neg\alpha \rightarrow B_i\bot$, która wyklucza istnienie sprzecznych, a zarazem nietrywialnych przekonań. Oznacza to, że dopuszczenie istnienia sprzecznych przekonań wymaga głębszej modyfikacji logiki. Ponadto A3 wiąże się z naturalnym warunkiem, zgodnie z którym świat bazowy w znajduje się w zbiorze $S_{i,w}$. Uzasadnienie tego faktu poprzedźmy podaniem warunku spełniania dla formuł postaci $P_i(\beta/\alpha)$: $(\boldsymbol{M}, w) \models P_i(\beta/\alpha)$ wtw $(\boldsymbol{M}, x) \models \beta$, dla

nowi, że w korpusie przekonań podmiotu nie występują logiczne fałsze (sprzeczności). Aksjomat A4 stwierdza, że jeżeli podmiot *i* dochodzi do przekonania, że α, pozyskując informację, że β, i na odwrót (co sugeruje, że α i β są dla *i* równoważne), to przekonania uzyskane na podstawie α i uzyskane na podstawie β są takie same. Jest to aksjomat ekstensjonalności.

Aksjomaty A5 i A6 są zaś postulatami *informacyjnej ekonomii* (lub *minimalnych zmian*) opartymi na tzw. *kryterium informacyjnej ekonomii*: zmiana przekonań nie jest — ogólnie rzecz biorąc — darmowa, jeśli więc już do niej dochodzi ze względu na pewną nową informację, to zmiana ta powinna być minimalna, tzn. nie powinna być większa niż to konieczne, aby tę nową informację dopasować do przekonań wyjściowych (Gärdenfors 1988: 49). A5 stanowi, że pozyskanie informacji należącej do przekonań podmiotu (choćby *implicite*) nie powinno powodować ich zmiany. Dokładniej, jeśli przekonania podmiotu zostały zaktualizowane o α na podstawie β, to zmiana przekonań zmierzająca do przyłączenia γ na podstawie α i β redukuje się do zmiany zmierzającej do przyłączenia γ na podstawie samego β. Aksjomat A6 głosi, że gdy pozyskana informacja nie jest sprzeczna z przekonaniami wyjściowymi podmiotu, wtedy powinien on po prostu włączyć ją do swych przekonań, a uzyskany zbiór domknąć na *modus ponens*. Treść tego aksjomatu stanie się może jaśniejsza, gdy zastąpimy β przez $\top$. Uzyskana w ten sposób formuła: $P_i\alpha \rightarrow (B_i(\gamma/\alpha) \equiv B_i(\alpha \rightarrow \gamma))$ stanowi, że jeżeli pozyskana przez podmiot informacja jest niesprzeczna z jego wyjściowymi przekonaniami, to może ją wykorzystać do rozszerzenia swych przekonań tylko wtedy, gdy przyłączane zdanie będzie (w jego mniemaniu) konsekwencją owej informacji. Wreszcie aksjomat A7 definiuje przekonania *simpliciter*. Reguła MP nie wymaga komentarza. Reguła BRN jest odpowiednikiem reguły ukonieczniania. Jej działanie jest jednak ograniczone do funktorów $B_i(\text{-}/\beta)$. Kryje się za nią następująca intuicja: jeżeli α jest tezą, to żadna pozyskana przez podmiot informacja nie skłoni go do porzucenia przekonania, że α.

4. Podmiot introspekcyjny. „Podmiot introspekcyjny" to taki, który ma pełny dostęp do własnych stanów przekonaniowych: poza przekonaniami dotyczącymi świata zewnętrznego może formułować przekonania dotyczące swych własnych przekonań, zarówno tych już posiadanych, jak i ich (ewentualnej) zmiany. Jako formalną charakterystykę takiego podmiotu można przyjąć następujące dwie zasady (i dołączyć je jako kolejne aksjomaty):

A8. $B_i(\alpha/\beta) \rightarrow B_i(B_i(\alpha/\beta))$ (pozytywna introspekcja)
A9. $\neg B_i(\alpha/\beta) \rightarrow B_i(\neg B_i(\alpha/\beta))$ (negatywna introspekcja)

---

pewnego $x \in \text{Min}_{i,w}(||\alpha||_{\boldsymbol{M}} \cap S_{i,w})$. Przypuśćmy teraz, że dla dowolnego $w \in S$, $(\boldsymbol{M}, w) \models \alpha$, czyli $w \in ||\alpha||_{\boldsymbol{M}}$. Skoro $w \in S_{i,w}$, to $||\alpha||_{\boldsymbol{M}} \cap S_{i,w} \neq \varnothing$. Z warunku dobrego ufundowania relacji $\leq_{i,w}$ wnosimy, że $\text{Min}_{i,w}(||\alpha||_{\boldsymbol{M}} \cap S_{i,w}) \neq \varnothing$, czyli istnieje świat x, taki że $x \in \text{Min}_{i,w}(||\alpha||_{\boldsymbol{M}} \cap S_{i,w})$. Ponieważ $||\top||_{\boldsymbol{M}} = S$, to $(\boldsymbol{M}, x) \models \top$. Na podstawie podanego wyżej warunku spełniania wnosimy, że $(\boldsymbol{M}, w) \models P_i(\top/\alpha)$, co kończy dowód.

Zastępując w nich β przez $\top$, otrzymujemy zasady introspekcji odnoszące się do przekonań *simpliciter*: $B_i\alpha \rightarrow B_iB_i\alpha$, $\neg B_i\alpha \rightarrow B_i\neg B_i\alpha$. Można też zaproponować bardziej ogólną (pełną) postać zasad introspekcji:

A8'. $B_i(\alpha/\beta) \rightarrow B_i((B_i(\alpha/\beta))/\gamma)$ (pełna pozytywna introspekcja).
A9'. $\neg B_i(\alpha/\beta) \rightarrow B_i((\neg B_i(\alpha/\beta))/\gamma)$ (pełna negatywna introspekcja).

Według A8' podmiot ma wgląd nie tylko w wyjściowy zbiór przekonań (tj. w to, o czym jest aktualnie przekonany), lecz także rejestruje proces aktualizacji przekonań, do którego dochodzi w następstwie pozyskania nowej informacji. Analogicznie należy zinterpretować aksjomat A9'. Uszczegółowieniami obu wymienionych zasad, oprócz A8 i A9, są następujące implikacje:

$B_i(\alpha/\beta) \rightarrow B_i((B_i(\alpha/\beta))/\beta)$,
$\neg B_i(\alpha/\beta) \rightarrow B_i((\neg B_i(\alpha/\beta))/\beta)$,
$B_i\alpha \rightarrow B_i((B_i\alpha)/\beta)$,
$\neg B_i\alpha \rightarrow B_i((\neg B_i\alpha)/\beta)$.

Dwie pierwsze tezy dotyczą introspekcji odnoszącej się w pewnym sensie do aktualnych zmian przekonań podmiotu. Według pierwszej z nich ta sama informacja, która spowodowała zmianę przekonania, jest daną, dzięki której podmiot rejestruje ową zmianę: jeżeli podmiot dochodzi do przekonania, że α, po uzyskaniu informacji, że β, to jest o tym przekonany w następstwie tej samej informacji. Sens drugiej jest analogiczny. Następne dwie tezy stanowią, że podmiot ma doskonałą pamięć. Nie tylko nie „gubi" tego, o czym jest aktualnie przekonany, lecz także potrafi przywołać lub zrekonstruować „historię" owych przekonań.

Aksjomaty A8' i A9' są spełnione we wszystkich modelach mających własność absolutności (tj. w których relacja $\leq_{i,w}$ jest niezależna od wyboru w):

(Ab) $\forall w, x, y, z \in S\,(x \in S_{i,w} \rightarrow (y \leq_{i,x} z \equiv y \leq_{i,w} z))$.

Na jej mocy we wszystkich światach spośród światów rozważanych przez podmiot i jako możliwe zachowywany jest taki sam porządek relatywnej wiarygodności[21].

5. Wspólne przekonania grupy. Warunkiem koniecznym powodzenia różnych działań zespołowych, tj. działań wymagających współpracy wszystkich członków

[21] Najpierw pokażemy, że A8' jest spełniony w modelach absolutnych, a następnie, że zachodzi to też dla A9'. Niech więc $\boldsymbol{M}$ będzie (Ab)-modelem. (1) Przypuśćmy, że dla dowolnego $w \in S$, $(\boldsymbol{M}, w) \models B_i(\alpha/\beta)$. Wtedy $(\boldsymbol{M}, x) \models \alpha$ dla każdego $x \in \mathrm{Min}_{i,w}(\|\beta\|_{\boldsymbol{M}} \cap S_{i,w})$. Na mocy założenia, że relacja $\leq_{i,w}$ ma własność (Ab) mamy: dla każdego $y \in \mathrm{Min}_{i,w}(\|\gamma\|_{\boldsymbol{M}} \cap S_{i,w})$, $z \leq_{i,y} u$ wtw $z \leq_{i,w} u$, co pociąga, że $\mathrm{Min}_{i,y}(\|\beta\|_{\boldsymbol{M}} \cap S_{i,y}) = \mathrm{Min}_{i,w}(\|\beta\|_{\boldsymbol{M}} \cap S_{i,w})$. Stąd $(\boldsymbol{M}, y) \models B_i(\alpha/\beta)$, a w konsekwencji $(\boldsymbol{M}, w) \models B_i((B_i(\alpha/\beta))/\gamma)$. (2) Przypuśćmy, że dla dowolnego $w \in S$, $(\boldsymbol{M}, w) \models \neg B_i(\alpha/\beta)$. Wtedy $(\boldsymbol{M}, x) \not\models \alpha$ dla pewnego $x \in \mathrm{Min}_{i,w}(\|\beta\|_{\boldsymbol{M}} \cap S_{i,w})$. Na mocy założenia, że relacja $\leq_{i,w}$ ma własność (Ab) mamy: dla każdego $y \in \mathrm{Min}_{i,w}(\|\gamma\|_{\boldsymbol{M}} \cap S_{i,w})$, $z \leq_{i,y} u$ wtw $z \leq_{i,w} u$, co pociąga, że $\mathrm{Min}_{i,y}(\|\beta\|_{\boldsymbol{M}} \cap S_{i,y}) = \mathrm{Min}_{i,w}(\|\beta\|_{\boldsymbol{M}} \cap S_{i,w})$. Stąd $(\boldsymbol{M}, y) \models \neg B_i(\alpha/\beta)$, a w konsekwencji $(\boldsymbol{M}, w) \models B_i((\neg B_i(\alpha/\beta))/\gamma)$.

danej grupy, jest posiadanie pewnej puli wspólnych przekonań (*common beliefs*). Mówiąc ogólnie, wspólne przekonanie jest to informacja publiczna, tj. posiadana przez wszystkich członków danej grupy. W takim jego określeniu kryje się jednak pewna dwuznaczność. W logice utożsamia się zwykle zwroty „każdy" i „wszyscy". Jednak w języku potocznym nie mają one tego samego znaczenia. Słowo „każdy" występuje często w znaczeniu „każdy z osobna", natomiast słowo „wszyscy" występuje w znaczeniu „wspólnie" lub „razem" (tj. wskazuje na zbiór jako pewną całość, a nie jego poszczególne elementy). Można łatwo wyobrazić sobie sytuację, w której powiedzenie, że wszyscy to a to zrobiliśmy (możemy zrobić) będzie prawdziwe, powiedzenie zaś, że każdy owo coś zrobił (może zrobić) będzie fałszywe. Zatem z jednej strony przekonanie, że α, może być „rozłożone" między członkami danej grupy, tzn. każdy, lecz z osobna, jest przekonany, że α. Z drugiej strony może ono być wspólne dla danej grupy, tzn. wszyscy razem są przekonani, że α. Oznacza to, że:

— każdy członek grupy jest przekonany, że α,
— każdy członek grupy jest przekonany, że każdy członek grupy jest przekonany, że α,
— każdy członek grupy jest przekonany, że każdy członek grupy jest przekonany, że każdy członek grupy jest przekonany, że α,

i tak *ad infinitum*. Na przykład w teorii gier, analizując różne sytuacje współpracy lub konfliktu, zwykle przyjmuje się, że gracze mają wspólne przekonanie o swej racjonalności. Nawiązując do przykładu z paragrafu drugiego, w świecie ⟨S, S, S⟩ wspólnym przekonaniem obu graczy jest to, że użyty zostanie profil strategii ⟨S, S, S⟩.

Niech $A = \{1, 2, \ldots, n\} \subseteq G$ będzie grupą osób wyróżnioną wewnątrz G. Dla uproszczenia przyjmijmy, że $A = G$ (tj. wszystkie podmioty tworzą grupę). W celu reprezentowania przekonań grupy G wprowadza się dwa osobne funktory. Mianowicie funktor E (lub E_G) reprezentuje przekonanie „rozłożone" między członkami grupy G. Odpowiada on wyrażeniu „każdy [w grupie G] jest przekonany, że". Jego definicja przyjmuje postać koniunkcji:

E. $$E\alpha := B_1\alpha \wedge B_2\alpha \wedge \ldots \wedge B_n\alpha.$$

Prawą stronę E można skrócić, pisząc: $\wedge_{i \in G} B_i\alpha$. Z kolei funktor C (lub C_G) odpowiada wyrażeniu „jest wspólnym przekonaniem [grupy G]" lub „wszyscy [razem w grupie G] są przekonani, że". Korzystając z funktora E, możemy funktor C zdefiniować następująco:

$$C\alpha := E\alpha \wedge EE\alpha \wedge EEE\alpha \wedge \ldots$$

Definicja ta wymaga jednak jakieś logiki infinitarnej, tj. dopuszczającej formuły o nieskończonej długości. W logice finitarnej funktor C trzeba scharakteryzować aksjomatycznie, dodając np.:

FP.	$C\alpha \rightarrow E(\alpha \wedge C\alpha)$	(Fixed-Point Axiom)[22]
RC.	Jeżeli $\vdash \alpha \rightarrow E(\alpha \wedge \beta)$, to $\vdash \alpha \rightarrow C\alpha$	(reguła indukcji)

Warunek prawdziwości dla formuł postaci $E\alpha$ przyjmuje następującą postać:

(h) (M, w) $\models E\alpha$ wtw (M, w) $\models B_i\alpha$, dla każdego $i \in G$.

Z warunku tego otrzymujemy: $\|E\alpha\|_M = \cap\{\|B_i\alpha\|_M: i \in G\}$, gdzie $\|B_i\alpha\|_M = \{w \in S: \mathrm{Min}_{i, w}(S_{i, w}) \subseteq \|\alpha\|_M\}$. Intuicyjnie rzecz ujmując, w świecie w każdy w grupie G jest przekonany, że α, wtedy i tylko wtedy, gdy dla każdego członka i grupy G jest tak, że α zachodzi w każdym świecie najbardziej wiarygodnym wśród światów rozważanych jako możliwe przez i w w. Gwarantuje on, że w każdym modelu spełniona jest równoważność: $E\alpha \equiv \wedge_{i \in G} B_i\alpha$. Nie jest natomiast spełniona zasada pozytywnej introspekcji: $E\alpha \rightarrow EE\alpha$. Istotnie, z tego, że każdy w grupie jest przekonany, iż α, nie wynika, że każdy w grupie jest przekonany, iż każdy w grupie jest przekonany, że α (nawet jeśli założymy, że każdy poszczególny członek grupy jest doskonałym logikiem oraz że uświadamia sobie to, o czym jest przekonany). Symbolem E^k oznaczmy k-krotną iterację funktora E[23]. Warunek dla formuł postaci $C\alpha$ przyjmuje wtedy postać:

(i) (M, w) $\models C\alpha$ wtw (M, w) $\models E^k\alpha$, dla każdego $k \geq 1$.

Z warunku tego otrzymujemy: $\|C\alpha\|_M = \cap\{\|E^k\alpha\|_M: k \geq 1\}$. Związki między wyróżnionymi rodzajami przekonań można wyrazić następująco: jeżeli $i \in G$, to:

$$C_G\alpha \rightarrow E_G\alpha \rightarrow B_i\alpha.$$

Z pojęciem wspólnych przekonań na terenie teorii gier związane jest pytanie, czy w grach z pełną informacją założenie o wspólnym przekonaniu graczy dotyczącym ich racjonalności pociąga za sobą wynik indukcji wstecznej (tj. akcję S)? Odpowiedzi pozytywnej na to pytanie patronuje Aumann, odpowiedzi negatywnej zaś Stalnaker. Zauważmy, że w rozważanym wcześniej przykładzie koniunkcja zdań $r_1 \wedge r_2 \wedge C_{\{1, 2\}}(r_1 \wedge r_2)$ oraz p_1 nie prowadzi do sprzeczności. Załóżmy, że obaj gracze są racjonalni i stanowi to ich wspólne przekonanie w chwili przystąpienia do gry. Wtedy ich wspólnym przekonaniem jest, że 1 w węźle trzecim — o ile zostanie on osiągnięty — wybierze S_3. Gracz 2 w węźle drugim — o ile zostanie on osiągnięty — nie powinien jednak wykluczać, że 1 wybierze w węźle trzecim C_3. Istotnie, już wybór C_1 przez gracza 1 jest zaskakujący. Gdyby 1 był racjonalny — tak, jak 2 sądził

[22] Aksjomat FP można wzmocnić do równoważności. Głosi on wówczas, że α jest wspólnym przekonaniem członków danej grupy wtedy i tylko wtedy, gdy każdy w owej grupie jest przekonany, że α zachodzi i że α jest wspólnym przekonaniem członków jego grupy. Formuła $C\alpha$ stanowi punkt stały funkcji $f(x) = E(\alpha \wedge x)$ przyporządkowującej formule x formułę $E(\alpha \wedge x)$ i w tym sensie modeluje nieskończoną koniunkcję $E\alpha \wedge EE\alpha \wedge EEE\alpha \wedge \ldots$

[23] $E^0\alpha = \alpha$, $E^1\alpha = E\alpha$, $E^2\alpha = EE\alpha$, …, $E^k\alpha = EE^{k-1}\alpha$, ... Na przykład E^2 znaczy: każdy [w grupie G] jest przekonany, że każdy [w grupie G] jest przekonany, że α.

— węzeł drugi w ogóle nie powinien być osiągnięty: 1 nie powinien wybierać C_1, lecz po przeprowadzeniu indukcji wstecznej powinien zakończyć grę, wybierając S_1. Tak więc wybór C_1 przez gracza 1 każe graczowi 2 zwątpić w trafność przekonania o racjonalności partnera. W rezultacie w swoim węźle decyzyjnym zmienia on strategię i wybiera C_2. Z kolei gracz 1 żywi nadzieję, że 2 — jako osoba racjonalna — tak właśnie postąpi. Z tego powodu w węźle pierwszym wybiera C_1, podstępnie sugerując, że jest nieracjonalny[24].

6. Wiedza. Tradycyjne (platońskie) rozumienie wiedzy przeciwstawia ją przekonaniu (mniemaniu). Wśród warunków, które musi spełniać wiedza wymienia się niezawodność. Podczas gdy przekonania mogą być błędne, tj. można być o czymś fałszywie przekonanym, zwrot „wiedza błędna (fałszywa)" wydaje się oksymoronem takim samym jak „wirtualna rzeczywistość" czy „żywy trup". Choć w literaturze przedmiotu można znaleźć wiele definicji wiedzy[25], w tym paragrafie ograniczę się do dwóch sugerowanych przez Stalnakera propozycji, w których wiedzę utożsamia się z mocnym (lub trwałym) przekonaniem. W rezultacie funktor wiedzy jest rodzajem funktora doksastycznego.

Zgodnie z pierwszą z nich wiedza to prawdziwe, nierewidowalne lub niepodważalne przekonanie (Stalnaker 2006). Dokładniej rzeczy ujmując, podmiot wie, że α, wtedy i tylko wtedy, gdy (1) α jest prawdziwe, (2) podmiot jest przekonany, że α, oraz (3) pozostanie on w przekonaniu, że α, w obliczu dowolnej prawdziwej informacji (tj. żadna pozyskana przez niego prawdziwa informacja nie skłoni go do wyrzeczenia się przekonania, że α)[26]. Z uwagi na warunek (3) definicja ta zdaje się wymagać jakieś logiki infinitarnej:

DEFK1. $K_i\alpha := \alpha \wedge (\beta_1 \rightarrow B_i(\alpha/\beta_1)) \wedge (\beta_2 \rightarrow B_i(\alpha/\beta_2)) \wedge \ldots$[27]

W myśl drugiej propozycji podmiot wie, że α, jeśli pozostanie w przekonaniu, że α, (nawet) w obliczu informacji, że $\neg\alpha$. Innymi słowy, podmiot wie, że α, jeśli nie-

[24] Dodajmy, że różnice między stanowiskami Aumanna i Stalnakera można sprowadzić do kwestii, czy gracze mogą sądzić, że partnerzy zmienią strategię po osiągnięciu swoich punktów decyzyjnych.

[25] Wielość ta ma swe źródło w krytyce tzw. *warunku uzasadnienia* klasycznej definicji wiedzy. Wymaga on, aby podmiot wiedzy posiadał racje przemawiające za przyjęciem danego zdania jako prawdziwego.

[26] (1) można nazywać *warunkiem prawdziwości*, (2) — *warunkiem przekonania*, a (3) — *warunkiem nierewidowalności*. Warunek (3) niekiedy zostaje wzmocniony przez opuszczenie wymagania, by pozyskiwana informacja była prawdziwa. Wiedza w tym sensie jest *absolutnie* nierewidowalnym przekonaniem, czyli nierewidowalnym nawet w obliczu fałszywej informacji (dezinformacji).

[27] Warunek przekonania $B_i\alpha$ został pominięty, ponieważ wynika z $\beta \rightarrow B_i(\alpha/\beta)$ po podstawieniu $\top$ za β.

możliwe jest anulowanie przekonania, że α, (nawet) po uzyskaniu informacji, że $\neg\alpha$[28]. Formalnie:

DEFK2. $K_i\alpha := B_i(\alpha/\neg\alpha) := B_i(\bot/\neg\alpha)$.

Warunek prawdziwości dla formuł postaci $K_i\alpha$ przyjmuje wtedy następującą postać:

(j) $(\boldsymbol{M}, w) \models K_i\alpha$ wtw $\mathrm{Min}_{i,w}(\|\neg\alpha\|_{\boldsymbol{M}} \cap S_{i,w}) \subseteq \|\alpha\|_{\boldsymbol{M}}$.

Intuicyjnie rzeczy ujmując, w świecie w podmiot i wie, że α, wtedy i tylko wtedy, gdy zbiór możliwości rozważanych przez podmiot i w w nie zawiera żadnych $(\neg\alpha)$-światów, tj. $\|\neg\alpha\|_{\boldsymbol{M}} \cap S_{i,w} = \varnothing$.

Charakterystykę wiedzy w sensie DEFK2 dopełniają następujące dwie tezy:

KB.	$K_i\alpha \rightarrow B_i\alpha$	(warunek przekonania)
T.	$B_i(\bot/\neg\alpha) \rightarrow \alpha$, czyli $K_i\alpha \rightarrow \alpha$	(warunek niezawodności)

Pierwszą z nich można uzasadnić następująco: niech w będzie dowolnym światem. Gdy $(\boldsymbol{M}, w) \models K_i\alpha$, wtedy $\|\neg\alpha\|_{\boldsymbol{M}} \cap S_{i,w} = \varnothing$, co daje inkluzję $S_{i,w} \subseteq \|\alpha\|_{\boldsymbol{M}}$, z której otrzymujemy: $(\boldsymbol{M}, x) \models \alpha$, dla każdego $x \in S_{i,w}$. Ponieważ $\mathrm{Min}_{i,w}(S_{i,w}) \subseteq S_{i,w}$, to $(\boldsymbol{M}, x) \models \alpha$, dla każdego $x \in \mathrm{Min}_{i,w}(S_{i,w})$, czyli $(\boldsymbol{M}, w) \models B_i\alpha$.

Druga teza jest prostą konsekwencją aksjomatu A3, a tym samym założenia, iż $w \in S_{i,w}$[29]. Niech więc w będzie jakimkolwiek światem. Gdy $(\boldsymbol{M}, w) \models K_i\alpha$, wtedy $(\boldsymbol{M}, x) \models \alpha$ dla każdego $x \in S_{i,w}$. Ponieważ $w \in S_{i,w}$, to $(\boldsymbol{M}, w) \models \alpha$.

Można też pokazać, że:

MON. $B_i(\alpha/\neg\alpha) \rightarrow B_i(\alpha/\beta)$, czyli $K_i\alpha \rightarrow B_i(\alpha/\beta)$.

Intuicyjnie rzeczy ujmując, jeżeli podmiot wie, że α, to pozostanie w przekonaniu, że α, w obliczu dowolnej informacji β. Niech w będzie dowolnym światem. Gdy $(\boldsymbol{M}, w) \models K_i\alpha$, wtedy $S_{i,w} \subseteq \|\alpha\|_{\boldsymbol{M}}$. Oczywiście, $S_{i,w} \cap \|\beta\|_{\boldsymbol{M}} \subseteq S_{i,w} \subseteq \|\alpha\|_{\boldsymbol{M}}$, dla dowolnego β. Ponieważ dla dowolnego X, $\mathrm{Min}_{i,w}(X) \subseteq X$, to $\mathrm{Min}_{i,w}(S_{i,w} \cap \|\beta\|_{\boldsymbol{M}}) \subseteq \|\alpha\|_{\boldsymbol{M}}$, skąd dostajemy: $(\boldsymbol{M}, w) \models B_i(\alpha/\beta)$.

7. Podsumowanie. W artykule przedstawiona została pewna prosta logika przekonań warunkowych (CDL). Opisuje ona stany przekonaniowe jako układy dynamiczne, tj. zmieniające się wraz z pozyskiwanymi przez podmiot informacjami. CDL przypomina, z jednej strony, teorię zmian stanów przekonaniowych (AGM), z drugiej zaś — logikę okresów warunkowych. Formułę $B(\alpha/\beta)$ można uznać za odpowiednik metajęzykowego wyrażenia $\alpha \in \mathrm{B} * \beta$, w którym $\mathrm{B} * \beta$ reprezentuje na gruncie AGM stan przekonań (teorię) będący wynikiem rewizji stanu wyjściowego

[28] Określenie to nawiązuje do definicji funktora konieczności Stalnakera: konieczne jest, że α, wtw $\neg\alpha$ implikuje kontrfaktycznie α (Stalnaker 1968: 105).

[29] Właściwie A3 i T są równoważne na gruncie DEFK2 i definicji $\top$.

B przez dołączenie do niego zdania β. Przypomnijmy, że w paradygmacie AGM zakłada się, iż reprezentacjami stanów przekonań są Cn-teorie. Wyróżnia się trzy typy zmian przekonań: ekspansję, kontrakcję i rewizję. *Ekspansja*, oznaczana przez +, polega na dodaniu informacji: jest funkcją, która stanowi przekonaniowemu (teorii) B oraz formule α przypisuje nowy — o ile to możliwe, niesprzeczny — stan przekonaniowy (teorię) $B + \beta = Cn(B \cup \{\beta\})$, powstały na skutek uznania β za prawdziwe. *Kontrakcja*, oznaczana przez ÷, jest funkcją działającą na odwrót: polega na usunięciu informacji z danego stanu przekonaniowego. Daje się ona zredukować do rewizji $*$ za pomocą tzw. identyczności Harpera: $B \div \beta := B \cap (B * \neg\beta)$. *Rewizja* jest funkcją, która stanowi przekonaniowemu B oraz formule β, w najbardziej interesującym wypadku sprzecznej z B, przypisuje nowy stan $B * \beta$, taki że: (1) zawiera on β, (2) jest niesprzeczny (chyba że samo β jest wewnętrznie sprzeczne) oraz (3) jest maksymalnie podobny do zbioru wyjściowego B.

Postulaty dla rewizji:		
B∗1.	$B * \beta$ jest teorią	domknięcie
B∗2.	$\beta \in B * \beta$	sukces
B∗3.	$B * \beta \subseteq Cn(B \cup \{\beta\}) = B + \beta$	inkluzja
B∗4.	$\neg\beta \notin B \Rightarrow B + \beta \subseteq B * \beta$	pustość
B∗5.	$\bot \in B * \beta \Leftrightarrow \neg\beta \in Cn(\varnothing)$	niesprzeczność
B∗6.	$(\alpha \equiv \beta) \in Cn(\varnothing) \Rightarrow (B * \alpha) = (B * \beta)$	ekstensjonalność
B∗7.	$B * (\alpha \wedge \beta) \subseteq (B * \beta) + \alpha$	
B∗8.	$\neg\alpha \notin B * \beta \Rightarrow (B * \beta) + \alpha \subseteq B * (\alpha \wedge \beta)$	

Aksjomaty A0, A1 oraz obie reguły odpowiadają postulatowi (B∗1), a aksjomaty A2, A3 i A4 postulatom (B∗2), (B∗5) i (B∗6). Aksjomat A6 odpowiada bezpośrednio postulatom (B∗7) i (B∗8), a pośrednio (B∗3) (B∗4). Aksjomat A5 nie posiada swego odpowiednika w wymienionych postulatach, ale można z nich wyprowadzić paralelną własność: $\alpha \in B * \beta \Rightarrow B * (\alpha \wedge \beta) = B * \beta$[30].

Na poziomie semantycznym formuła $B(\alpha/\beta)$ odpowiada epistemicznie zinterpretowanemu nierzeczywistemu okresowi warunkowemu $\beta > \alpha$ (tj. opisującemu dyspozycję podmiotu do zmiany przekonań w obliczu nowej informacji). CDL może być wykorzystana na przykład do epistemicznej analizy gier, które z jednej strony wymagają planowania działań, a z drugiej — rewizji wcześniejszych przekonań.

---

[30] Niech β będzie formułą taką, że $\neg\beta \notin Cn(\varnothing)$ (tzn. β nie jest wewnętrznie sprzeczna) i załóżmy, że $\alpha \in B * \beta$. Wtedy $B * \beta$ jest niesprzeczny (wobec (B∗5)), co pociąga, że $\neg\alpha \notin B * \beta$. Opierając się na (B∗7, 8) oraz (B+4) (tj. $\alpha \in B \Rightarrow B + \alpha = B$), dostajemy: $B * (\alpha \wedge \beta) = (B * \beta) + \alpha = B * \alpha$.

BIBLIOGRAFIA

Aumann R. J. (1995), *Backward Induction and Common Knowledge of Rationality*, „Games and Economic Behavior" 8(1), 6-19.

Baltag A., Smets S. (2006), *Conditional Doxastic Models. A Qualitative Approach to Dynamic Belief Revision* [w:] *Proceedings of WOLLIC'06, Electronic Notes in Theoretical Computer Science*, t. 165, 5-21.

Baltag A., van Ditmarsch H. P., Moss L. S. (2008), *Epistemic Logic and Information Update* [w:] *Handbook of the Philosophy of Science*, P. Adriaans, J. van Benthem (red.), t. 8, *Philosophy of Information*, Amsterdam: Elsevier/North-Holland, 369-463.

Board O. J. (2003), *Algorithmic Characterization of Rationalizability in Extensive Form Games*, Oxford: Department of Economics, University of Oxford.

Chellas B. F. (1980), *Modal Logic. An Introduction*, Cambridge: Cambridge University Press.

Gärdenfors P. (1988), *Knowledge in Flux. Modeling the Dynamics of Epistemic States*, Cambridge (MA): MIT Press.

Grove A. (1988), *Two Modellings for Theory Change*, „Journal of Philosophical Logic" 17(2), 157-170.

Halpern J. Y. (1999a), *Set-Theoretic Completeness for Epistemic and Conditional Logic*, „Annals of Mathematics and Artificial Intelligence" 26(1-4), 1-27.

Halpern J. Y. (1999b), *Hypothetical Knowledge and Counterfactual Reasoning*, „Game Theory" 28(3), 315-330.

Lechniak M. (2011), *Przekonania i zmiana przekonań. Analiza logiczna i filozoficzna*, Lublin: Wydawnictwo KUL.

Leitgeb H., Segerberg K. (2007), *Dynamic Doxastic Logic. Why, How, and Where To?*, „Synthese" 155(2), 167-190.

Lewis D. (1973), *Counterfactuals*, Oxford: Basil Blackwell.

Nute D., Cross C. B. (2002), *Conditional Logic* [w:] *Handbook of Philosophical Logic*, D. M. Gabbay, F. Guenthner (red.), t. 4, 2nd ed., Dordrecht: Reidel, 1-98.

Spohn W. (1975), *An Analysis of Hansson's Dyadic Deontic Logic*, „Journal of Philosophical Logic" 4(2), 231-252.

Stalnaker R. (1968), *A Theory of Conditionals* [w:] *Studies in Logical Theory*, N. Rescher (red.), Oxford: Blackwell, 98-112.

Stalnaker R. (1996), *Knowledge, Belief and Counterfactual Reasoning in Games*, „Economics and Philosophy" 12(2), 133-163.

Stalnaker R. (1998), *Belief Revision in Games. Forward and Backward Induction*, „Mathematical Social Science" 36(1), 31-56.

Stalnaker R. (2006), *On Logics of Knowledge and Belief*, „Philosophical Studies" 128(1), 169-199.

Szymanek K. (1999), *Formalna teoria zmiany przekonań*, Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Van Benthem J., Martinez M. (2008), *The Stories of Logic and Information* [w:] *Handbook of the Philosophy of Science*, t. 8: *Philosophy of Information*, P. Adriaans, J. van Benthem (red.), Amsterdam: Elsevier/North-Holland, 225-288.