Marek Tokarz

Logika świadomych przekonań

Nowa Krytyka 2, 53-61

1992

NOWA KRYTYKA 2 Rok 1992

Marek Tokarz

Logika świadomych przekonań

Poniższy tekst stanowi rozpaczliwą próbę przetłumaczenia na język możliwie strawny dla czytelnika czysto formalnego i dość zaawansowanego technicznie referatu, który wygłosiłem w kwietniu 1989 r. na konferencji „Dialektyka a problem subiektywności" w Polanicy.

Ideologia

Logiki przekonań – jeden z wariantów logiki epistemicznej – zajmują się różnymi formalnymi objawami fenomenu tzw. postaw zadaniowych *(propositional attitudes)*. Analizie poddaje się tu więc pewne zjawisko o charakterze pragmatycznym, a zatem **subiektywnym**, zjawisko polegające na tym, że ktoś *wierzy*, że *p* (gdzie *p* jest sądem, lub po prostu zdaniem oznajmującym), albo *wątpi*, że *p*, wie, że *p*, *dopuszcza możliwość*, że *p*, *odrzuca p* itd. Subiektywność postaw manifestuje się pewną ich niezależnością od obiektywnej **prawdziwości**. Przypuśćmy, że zdanie *p* brzmi *Idzie ku lepszemu*; nie wdając się w merytoryczną treść tego zdania zauważmy jedynie, że kilka lat temu dziennikarze „Trybuny Ludu" byli głęboko przekonani o jego prawdziwości, zaś dziennikarze „Gazety Bankowej" – równie głęboko – o jego fałszywości.

Poczynając od lat pięćdziesiątych (von Wright, Carnap), przez sześćdziesiąte (Hintikka), do osiemdziesiątych (Gärdenfors i in.), próbowano dokonać formalnej rekonstrukcji procesu tworzenia przekonań, owych „prywatnych teorii", posługując się w tym celu językami o różnym stopniu złożoności, od języka zerowego rzędu (zdaniowego) poczynając.

Celem niniejszego opracowania jest prezentacja podstawowych zasad logicznych rządzących postawą wyrażaną zwrotem *ktoś wierzy, że p*. Mówiąc, że **ktoś** wierzy, mamy zawsze na myśli pewnego wybranego reprezentanta określonej z góry kategorii ludzi, a nie wszystkie istoty ludzkie w ogóle. Oczywistym powodem tej restrykcji jest wspomniana subiektywność, idąca tak daleko, iż niektórzy ludzie tworzą swoje systemy przekonań w sposób całkowicie irracjonalny i niemożliwy do przewidzenia nawet w najbanalniejszych sytuacjach. Można sobie na przykład z łatwością wyobrazić filozofa (dialektyka?), który jest przekonany, że *p* i równocześnie wierzy, że *nie-p*. Podobnie ma się sprawa z podstawowymi zasadami logicznymi, np. regułą odrywania *(modus ponens)*, przechodniością implikacji itp.; mam w swoim otoczeniu humanistkę, która wie wprawdzie, że filozofowie są humanistami, zaś niektórzy filozofowie to durnie, lecz zdanie *Niektórzy humaniści są durniami* jest dla niej wnioskiem zbyt ogólnym (!), niemożliwym do przyjęcia. Osoby takie tworzą silne lobby w wielu społeczeństwach, stowarzyszają się nawet w różne komitety i rządy, tworząc zbiorowość szalenie interesującą dla socjologów i psychiatrów; dla logika jednak nie są to osobnicy akurat najbardziej zajmujący. Będziemy tu rozważać wyłącznie przekonania ludzi całkowicie **racjonalnych** i całkowicie **świadomych** swoich własnych przekonań. Ludzie ci muszą wierzyć np. w prawa logiki: wyraża to reguła RB z następnego paragrafu; nie mogą akceptować zdań wzajemnie sprzecznych: jest to treścią postulatu Ax.4; muszą wyraźnie odróżniać zdania, o których prawdziwości są przekonani, od zdań, o których prawdziwości przekonani nie są: tę cechę wyrażają aksjomaty Ax.2 i Ax.3; wreszcie ich systemy przekonań powinny być zamknięte na regułę odrywania, tzn. jeśli ktoś wierzy, że p implikuje q i równocześnie wierzy, że p, to nie śmie poddawać w wątpliwość prawdziwości q (por. Ax.5). Niczego więcej (ani mniej) o racjonalnym systemie świadomych przekonań nie będziemy na razie zakładać.

Spośród wielu możliwych (*wie, wierzy, dopuszcza, odrzuca* itp.) zdecydowaliśmy się na wybór funktora *wierzy* jako podstawy konstruowanego systemu; o wyborze tym przesądziły co najmniej dwa istotne względy. Przede wszystkim *wierzy* jest, obok *wie*, jednym z najczęściej analizowanych spójników pragmatycznych. Po drugie, przy jego pomocy można zdefiniować wiele, przypuszczalnie nawet „większość", spośród pozostałych spójników wyrażających postawy zdaniowe, na przykład:

„x dopuszcza możliwość, że p" := „nieprawda, że x wierzy, że nie-p",
„x odrzuca p" := „x wierzy, że nie-p",
„x wie, że p" := „p i x wierzy, że p".

Fakt, iż rozważany przez nas osobnik jest przekonany (wierzy), że p, notujemy w postaci Bp; symbol B pochodzi od czasownika *believe* i został, zdaje się, po raz pierwszy użyty przez Carnapa.

System LB

Przez L oznaczamy język *formuł klasycznych*, zbudowanych ze zmiennych zdaniowych i spójników negacji ($\sim$), koniunkcji ($\wedge$) i alternatywy ($\vee$); implikację ($\rightarrow$) i równoważność ($\equiv$) definiuje się standardowo. Przez PC rozumiemy zbiór klasycznych *tautologii*, t oznacza dowolną ustaloną tautologię, np. $p \vee \sim p$, zaś f – dowolną ustaloną kontrtautologię, np. $p \sim p$. Przez *teorię klasyczną* rozumiemy dowolny zbiór klasycznych formuł zawierających PC i zmknięty na regułę odrywania MP.

Przez S oznaczamy zbiór wszystkich *formuł* zbudowanych ze zmiennym użyciem $\sim$, $\wedge$, $\vee$ oraz specyficznego jednoargumentowego spójnika B; formułę B czytamy: wierzę, że α. Przez Taut rozumiemy zbiór wszystkich podstawień klasycznych tautologii formułami języka S. Przyjmujemy, że aksjomatami są formuły o następującym kształcie ($\alpha, \beta \in S$):

Ax.1. α, gdzie $\alpha \in$ Taut
Ax.2. $B\alpha \equiv BB\alpha$
Ax.3. $\sim B\alpha \equiv B \sim B\alpha$
Ax.4. $B \sim \alpha \rightarrow \sim B\alpha$
Ax.5. $B(\alpha \rightarrow \beta) \rightarrow (B\alpha \rightarrow B\beta)$

oraz przyjmujemy następujące reguły wnioskowania:

MP: z α i $\alpha \rightarrow \beta$ można wyprowadzić β;
RB: z α można wyprowadzić $B\alpha$.

W świetle propagandowego paragrafu 1, powyższe postulaty są w większości zrozumiałe same przez się; pewnego komentarza wymaga być może jedynie aksjomat Ax.3 i reguła RB. Ax.3 wraz z poprzedzającym go Ax.2 stwierdza, że nasza wiara na temat własnych przekonań jest w istocie **wiedzą** : zarówno gdy wierzę w α (Ax.2), jak i gdy nie wierzę w α (Ax.3), czynię to świadomie, po rozważeniu argumentów. Podkreślmy jednak, na wszelki wypadek, że niewiara w sąd p nie obliguje mnie automatycznie do poglądu, że nie-p: jak się przekonamy dalej, formuła $\sim Bp \rightarrow B \sim p$ nie da się wyprowadzić z powyższych aksjomatów. Regułę RB należy interpretować następująco: jeśli α jest **prawem logiki**, to mam obowiązek wierzyć w α. Reguła ta nie służy do tworzenia dowolnych teorii i stosuje się ją jedynie do zbioru tez; jej przyjęcie nie oznacza, że mam obowiązek wierzenia we wszystkie zdania prawdziwe. Jak to się mówi, nie jest ona regułą **wtórną**, a jedynie **dopuszczalną**, podobnie jak w logice modalnej znana reguła Gödla: jeśli α jest tezą, to również α jest tezą.

Przez LB oznaczamy zbiór wszystkich formuł wyprowadzalnych z aksjomatów Ax.1 - Ax.5 przy pomocy reguł MP i RB. Zamiast pisać $\alpha \in$ LB, będziemy często pisać $\vdash_{LB} \alpha$ lub po prostu $\vdash \alpha$ i będziemy w takim wypadku mówić, że α jest *tezą* logiki LB. Przez *teorię* w logice LB rozumiemy dowolny zbiór formuł zawierający LB i zamknięty na (samą!) regułę odrywania MP. Oto kilka przykładów tez logiki LB:

T.1 $\vdash \sim \mathrm{B}(\alpha \wedge \sim \alpha)$

T.2 $\vdash \mathrm{B}(\alpha \equiv \beta) \to (\mathrm{B}\alpha \equiv \mathrm{B}\beta)$

T.3 $\vdash (\mathrm{B}\alpha \vee \mathrm{B}\beta) \to \mathrm{B}(\alpha \vee \beta)$

T.4 $\vdash \mathrm{B}(\alpha \wedge \beta) \equiv (\mathrm{B}\alpha \wedge \mathrm{B}\beta)$

T.5 $\vdash \mathrm{B}\alpha \to \sim \mathrm{B} \sim \alpha$

T.6 $\vdash \mathrm{B}(\alpha \vee \beta) \wedge \sim \mathrm{B}\alpha \to \sim \mathrm{B} \sim \beta$

T.7 $\vdash \mathrm{B}(\alpha \vee \beta) \wedge \mathrm{B} \sim \alpha \to \mathrm{B}\beta$

T.8 $\vdash \mathrm{B}(\alpha \to \beta) \to (\mathrm{B} \sim \beta \to \mathrm{B} \sim \alpha)$

T.9 $\vdash \sim \mathrm{B}(\alpha \to \beta) \to \sim \mathrm{B} \sim \alpha \wedge \sim \mathrm{B}\beta$

T.10 $\vdash \mathrm{B}(\mathrm{B}\alpha \vee \mathrm{B}\beta) \to \mathrm{B}(\alpha \vee \beta)$

T.11 $\vdash \mathrm{B}(\alpha \to \beta) \to \mathrm{B}(\mathrm{B}\alpha \to \beta)$

T.12 $\vdash \mathrm{B}(\mathrm{B}\alpha \to \alpha)$

Jeśli M jest zbiorem formuł (języka S), to przez T(M) oznaczamy zbiór tych formuł klasycznych, które po poprzedzeniu spójnikiem B należą do M. Łatwo wykazać, że jeśli M jest niesprzeczną LB-teorią, to T(M) jest niesprzeczną teorią klasyczną. Formalna definicja zbioru T(M) jest następująca:

$$\mathrm{T(M)} := \{\alpha \in \mathrm{L} : \mathrm{B}\alpha \in \mathrm{M}\}$$

Interpretacje

Niech T będzie dowolną teorią klasyczną, zaś α – dowolną formułą języka S. Określimy specjalną funkcję I_T, która formule α przyporządkowuje jej *interpretację w teorii* T, czyli pewną formułę klasyczną. Funkcja I_T „przegląda" formułę α, niczego nie zmieniając, tak długo, aż natknie się w α na podformułę o budowie $\mathrm{B}\beta$, gdzie $\beta \in$ L. Teraz funkcja I_T zastępuje wszystkie takie napotkane formuły $\mathrm{B}\beta$ przez t lub f, w zależności od tego, czy β było twierdzeniem teorii T, czy też nie. Jeśli w formule powstającej po wykonaniu tej operacji pozostał jeszcze jakiś spójnik B, to procedurę się powtarza, aż do jego zupełnego wyeliminowania. Ścisła indukcyjna definicja funkcji I_T jest następująca:

1° $I_T = p$, gdy p jest zmienną
2° $I_T \sim \gamma = \sim I_T\gamma$
3° $I_T(\gamma \wedge \delta) = I_T\gamma \wedge I_T\delta$
4° $I_T(\gamma \vee \delta) = I_T\gamma \vee I_T\delta$
5° $I_T B\gamma = \begin{cases} t & gdy \quad I_T\gamma \in T \\ f & gdy \quad I_T\gamma \notin T \end{cases}$

TWIERDZENIE 1. *Formuła α jest tezą LB (tzn. $\vdash \alpha$) wtedy i tylko wtedy, gdy dla każdej niesprzecznej teorii T, $I_T\alpha$ jest klasyczną tautologią (tzn. $I_T\alpha \in PC$).*

Niech $<A, -, \cap, \cup>$ będzie algebrą Boole'a. Symbolami 0 i 1 oznaczamy, odpowiednio, najmniejszy i największy jej element. Niech dalej **U** $= <A, -, \cap, \cup, *>$ będzie algebrą powstającą przez dołączenie dowolnego jednoargumentowego działania * w A. Mówimy, że formuła $\alpha \in S$ jest *prawdziwa w algebrze* **U**, symbolicznie $\underline{\mathbf{U}} \models \alpha$, gdy dla każdego *wartościowania* h zmiennych języka elementami z A mamy: $h\alpha = 1$. Podzbiór F zbioru A nazywamy *filtrem*, jeśli $1 \notin F$ oraz dla wszelkich $a, b \in A$, $a \cap b \in F$ wtw $a \in F$ i $b \in F$; filtr F jest *właściwy*, gdy $0 \in F$; jest *ultrafiltrem*, gdy jest maksymalnym filtrem właściwym. Największym filtrem jest więc cały zbiór A (jest to jedyny filtr niewłaściwy), najmniejszym filtrem jest singleton 1. Układ $\underline{\mathbf{U}} = <A, -, \cap, \cup, *, F>$ nazywamy *algebrą filtrową*, gdy $<A, -, \cap, \cup>$ jest algebrą Boole'a, F jest w niej filtrem, zaś * jest tzw. funkcją charakterystyczną zbioru F, tzn. dla dowolnego a A spełniony jest warunek:

$$^*a = \begin{cases} 1 & gdy \quad a \in F \\ 0 & gdy \quad a \notin F \end{cases}$$

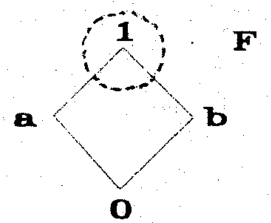
Rodzinę wszystkich algebr filtrowych oznaczamy FA. Jeśli K jest dowolną klasą algebr filtrowych, tzn. $K \supseteq FA$, to mówimy, że formuła jest *prawdziwa w klasie* K, symbolicznie $K \models \alpha$, gdy dla dowolnej algebry $\underline{\mathbf{U}} \in K$, $\underline{\mathbf{U}} \models \alpha$. Symbolem PFA oznaczamy klasę tych algebr filtrowych $<A, -, \cap, \cup, *, F>$, w których F jest filtrem właściwym.

TWIERDZENIE 2. $\vdash \alpha$ *wtw* $PFA \models \alpha$

Dowody powyższych twierdzeń są raczej zawiłe i nie podejmę się tutaj przedstawiania choćby ich szkicu. Wspomnę tylko, że istotną rolę odgrywa tu następujący lemat: jeśli M jest teorią maksymalną w logice LB, to $\alpha \in M$ wtw $I_{(M)}\alpha \in M$. Zresztą Twierdzenie 2 jest przede wszystkim narzędziem, rodzajem prostej w obsłudze maszynki do sprawdzania czy dana formuła jest twierdzeniem LB; wewnętrzna budowa narzędzia i technologia jego produkcji nie musi akurat interesować konsumenta. Zademonstrujmy więc raczej sposób posługiwania się twierdzeniem o pełności (czyli właśnie Twierdzeniem 2 na przykładzie trzech formuł: $\sim$Bp $\rightarrow$ B $\sim$ p; Bp $\rightarrow$ p; B(Bp $\rightarrow$ p).

A. $\nvdash$ $\sim$Bp $\rightarrow$ B $\sim$ p.

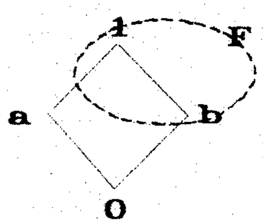
Ponieważ działanie * przyjmuje wyłącznie wartości 0 i 1, więc wystarczy wskazać taką algebrę z filtrem F, w której dla pewnego elementu a, ani a nie należy do F (wówczas -*a = 1), ani -a nie należy do F (wówczas *-a = 0). Łatwo wskazać najmniejszą taką algebrę: jest to algebra czteroelementowa z filtrem jednostkowym:



Kładziemy więc hp = a i otrzymujemy h($\sim$ Bp $\rightarrow$ B$\sim$ p) = 0, zatem na mocy Tw. 2 badana formuła nie jest tezą logiki LB.

B. $\nvdash$ Bp $\rightarrow$ p.

Wystarczy wskazać algebrę filtrową, w której pewien element b różny od 1 należy do filtru. Wtedy bowiem *b = 1, więc dla hp = b otrzymamy h(Bp $\rightarrow$ p) = b $\neq$ 1. Algebrą taką jest np.



C. $\vdash B(Bp \rightarrow p)$.

Przypuśćmy, że powyższą formułę można obalić w pewnej algebrze z filtrem F, tzn. $hB(Bp \rightarrow p) \neq 1$ dla pewnego h. Ponieważ $hB(Bp \rightarrow p) = {}^*h(Bp \rightarrow p) \in \{0,1\}$, więc $hB(Bp \rightarrow p) = 0$, czyli $h(Bp \rightarrow p) \in F$. Wówczas jednak $hp \notin F$, więc $hBp = {}^*hp = 0$; stąd $h(Bp \rightarrow p) = 1 \in F$ - sprzeczność. Badana formuła musi więc być prawdziwa w każdej algebrze z PFA i na mocy Tw.2 jest tezą.

Mutacje

A. System LCB. System LCB, będący rozszerzeniem LB, stanowi opis przekonań tych osób, które w absolutnie każdej sprawie mają wyrobiony, niezachwiany pogląd (choć, oczywiście, niekoniecznie pogląd obiektywnie trafny). Osoby takie określa się mianem *besserwisser.* Do Ax.1 - Ax.5 dołączamy nowy aksjomat

Ax.6. $B\,\alpha \vee B \sim \alpha$

i LCB określamy jako zbiór formuł wyprowadzalnych z Ax.1. - Ax.6 przy pomocy reguł MP i RB. Oto trzy przykłady tez systemu LCB; w istocie każda z formuł T.13 - T.15 jest równoważna Ax.6 i może zostać równie dobrze przyjęta jako aksjomat:

T.13 $\vdash_{LCB} \sim B\,\alpha \rightarrow B \sim \alpha$

T.14. $\vdash_{LCB} B\,(\,(\alpha \vee \beta) \equiv (B\,\alpha \vee B\,\beta)$

T.15. $\vdash_{LCB} B\,(\,(\alpha \rightarrow \beta) \equiv (B\,\alpha \rightarrow B\,\beta)$

Przez UFA oznaczamy klasę tych algebr filtrowych $\langle A, -, \cap, \cup, {}^*, F\rangle$, w których F jest ultrafiltrem.

TWIERDZENIE 3. $\vdash_{LCB}$ *wtw* UFA $\models \alpha$.

B. System LIB. Logika LB wyklucza możliwość przekonań sprzecznych - efekt ten daje aksjomat Ax.4. Można jednak rozważyć wariant tolerujący sprzeczne przekonania - logika taka byłaby nawet bardziej realistyczna. Powstaje ona przez odrzucenie Ax.4: LIB definiujemy jako system formuł wyprowadzalnych z Ax.1, Ax.2, Ax.3, Ax.5 przy użyciu MP i RB.

TWIERDZENIE 4. $\vdash_{LIB} \alpha$ *wtw* FA $\models \alpha$.

C. System S5. Różnica statusu między wiarą a wiedzą polega na istnieniu wyraźnego związku między wiedzą a obiekywną prawdziwością:

ze zdania *x wie, że p* wynika, że zdanie *p* jest prawdziwe. Prawdziwość *p* jest wręcz presuponowana przez zdanie *x wie, że p*. Zatem dołączenie nowego aksjomatu

Ax.7. $B\alpha \rightarrow \alpha$

do Ax.1 - Ax.5 zmienia w istotny sposób treść spójnika B; wyrażenie Bp należałoby w tak rozszerzonej logice czytać raczej *wiem, że p*, niż *wierzę, że p*. Tak się składa, że system oparty na aksjomatach Ax.1 - Ax.5 i Ax.7 jest równoznaczny systemowi S5 Lewisa, co dostarcza interesującej i bardzo naturalnej interpretacji tej starej i szacownej logiki. Łatwo zauważyć, że aksjomat Ax. 7 jest prawdziwy wyłącznie w tych algebrach filtrowych $<A, -, \cap, \cup, *, F>$, w których F jest filtrem jednostkowym, tzn. w których $F = \{1\}$. Oznaczamy klasę tych algebr przez JFA. W ten sposób otrzymujemy dla S5 następujące, udowodnione przez Wajsberga w latach trzydziestych, twierdzenie o pełności:

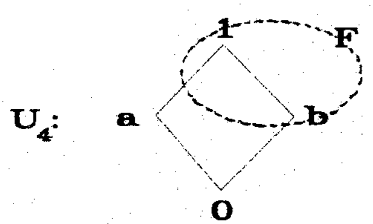
TWIERDZENIE 5. (Wajsberg, 1933). $\vdash_{S5}$ *wtw* IFA $\models \alpha$.

Uzupełnienia

A. Rozstrzygalność. Można wykazać, że we wszystkich przedstawionych wyżej twierdzeniach o pełności (Twierdzenia 2 - 5) wystarczy się ograniczyć do algebr skończonych. Na mocy znanego twierdzenia Harropa otrzymujemy stąd następujący wniosek:

TWIERDZENIE 6. *Systemy* LB, LCB i LIB *i oczywiście* S5 *są rozstrzygalne.*

B. Moce modeli. W przypadku logik LB, LIB i S5 nie można mocy ich modeli ograniczyć żadną liczbą skończoną, tzn. dla dowolnej liczby n istnieje formuła nie będąca tezą S5 (a zatem także LB i LIB), która jednak jest prawdziwa w każdej algebrze z FA (a zatem także z PFA i JFA) mającej mniej niż n elementów. Inaczej przedstawia się sprawa z systemem LCB. Oznaczamy przez U_4 rozważaną już wyżej (w przykładzie B w punkcie „Interpretacje") algebrę o diagramie



TWIERDZENIE 7. $\vdash_{LCB} \alpha$ *wtw* $\underline{\mathrm{U}}_4 \models \alpha$.

C. Interpretacje klasyczne. Dla systemów LCB i LIB prawdziwe są następujące modyfikacje Twierdzenia 1 o interpretacjach formuł logiki przekonań w języku formuł klasycznych:

TWIERDZENIE 8.

a. $\vdash_{LCB} \alpha$ *wtw dla każdej maksymalnej teorii klasycznej* T, $\mathrm{I}_T\ \alpha \in$ PC;

b. $\vdash_{LIB} \alpha$ *wtw dla każdej teorii klasycznej* T, $\mathrm{I}_T\ \alpha \in$ PC.

D. Separacja zmiennych. System LB ma pewną ciekawą własność syntaktyczną, obrazującą fakt subiektywności, czy też arbitralności przekonań. W tezach tego systemu można mianowicie odseparować zmienne objęte funktorem B od zmiennych nie objętych tym funktorem, i uzyskana tą drogą formuła pozostanie nadal tezą logiki LB. Ujmijmy rzecz formalnie: zakładając, że zmiennymi języka S są $p_1, p_2, p_3, \ldots$, dla dowolnej liczby k definiujemy podstawienie e_k wzorem $e_k p_i := p_{k+i}$. Dla $\alpha \in$ S definiujemy z kolei formułę $w_k\ \alpha$ wzorami:

1° $w_k p_i = p_i$ dla i = 1, 2, 3,...
2° $w_k \sim \beta = \sim w_k \beta$
3° $w_k (\beta \wedge \gamma) = w_k \beta \wedge w_k \gamma$
4° $w_k (\beta \vee \gamma) = w_k \beta \vee w_k \gamma$
5° $w_k \mathrm{B}\beta = e_k \mathrm{B}\beta$.

Niech $n\alpha$ będzie największym spośród indeksów zmiennych występujących w α. Kładziemy

$$w\alpha := w_{n\alpha}\alpha.$$

Oto trzy przykłady działania operacji w : $w(\mathrm{B}(p_1 \vee p_2)) = \mathrm{B}(p_3 \vee p_4)$; $w(\mathrm{B}p_3 \rightarrow p_3) = \mathrm{B}p_6 \rightarrow p_3$; $w(p_1 \vee \mathrm{B}p_1 \rightarrow p_3 \wedge \sim \mathrm{B}p_2) = p_1 \vee \mathrm{B}p_4 \rightarrow p_3 \wedge \sim \mathrm{B}p_5$.

TWIERDZENIE 9. $\vdash_{LB} \alpha$ *wtw* $\vdash_{LB} w\alpha$