

# Reprezentacja wiedzy: Grafy pojęć, byty mentalne i wnioskowanie o przekonaniach

Wojciech Jaworski

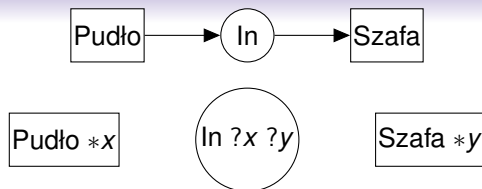
Instytut Informatyki  
Uniwersytet Warszawski

# Spis treści

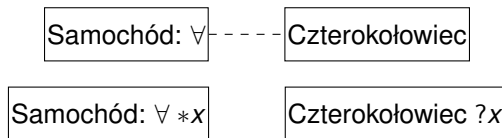
1 Grafy pojęć

2 Byty mentalne

3 Wnioskowanie o przekonaniach



- Symbole  $*x$  i  $*y$  to etykiety definiujące (*defining labels*). Wskazują miejsca, w których są wprowadzane zmienne.
- Symbole  $?x$  i  $?y$  są etykietami związanymi (*bound labels*), które wskazują użycie zmiennych.

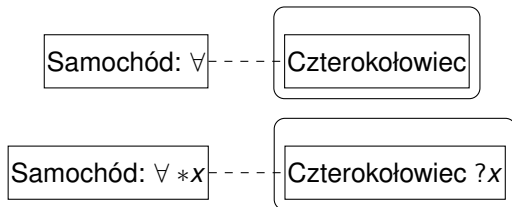


$(\forall x : \text{Samochód})\text{czterokołowiec}(x)$

# Negacja

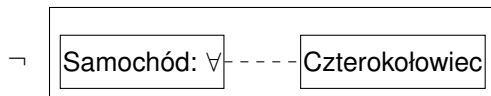
- Negację wprowadzamy umieszczając zanegowaną formułę w specjalnym kontekście:

$$(\forall x : \text{Samochód}) \neg \text{czterokołowiec}(x)$$



- Negację można też oznaczyć następująco:

$$\neg (\forall x : \text{Samochód}) \text{czterokołowiec}(x)$$

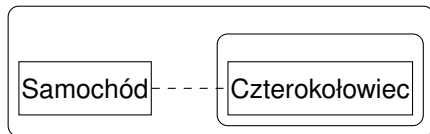


- Koniunkcja ( $\wedge$ ), negacja ( $\neg$ ) oraz kwantyfikator egzystencjalny ( $\exists$ ) pozwalają wyrazić pozostałe operatory logiczne.
- Zdanie

$(\forall x : \text{Samochód})\text{czterokołowiec}(x)$

jest równoważne

$\neg(\exists x : \text{Samochód})\neg\text{czterokołowiec}(x)$



- Grafy nie zawierające kwantyfikatora uniwersalnego mają jednoznaczną semantykę.

# Implikacja

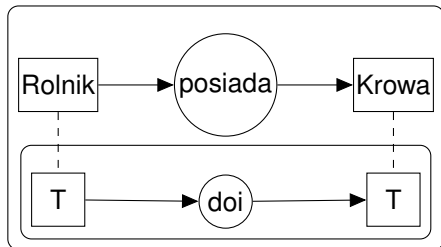
## Zdanie

*Jeśli rolnik ma krowę to ją doi.*

$$(\forall x : \text{Rolnik})(\forall y : \text{Krowa})(\text{posiada}(x, y) \implies \text{doi}(x, y))$$

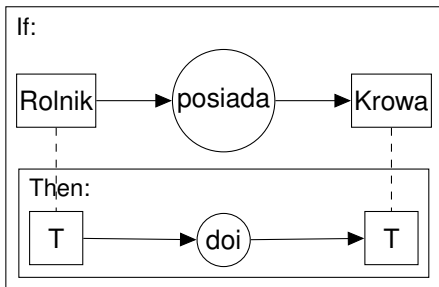
jest równoważne

$$\neg(\exists x : \text{Rolnik})(\exists y : \text{Krowa})(\text{posiada}(x, y) \wedge \neg \text{doi}(x, y))$$



# Implikacja

Implikację jawnie reprezentujemy wprowadzając kontekst If-Then:



Powyższa reprezentacja jest bliższa pierwotnemu zdaniu

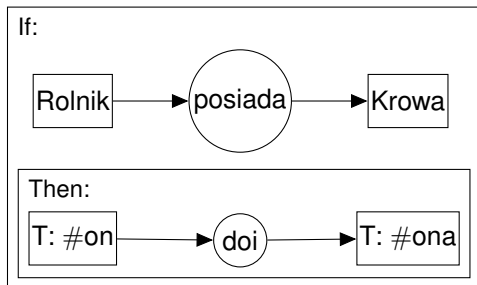
*Jeśli rolnik ma krowę to ją doi.*

niż formuła logiczna

$$(\forall x : \text{Rolnik})(\forall y : \text{Krowa})(\text{posiada}(x, y) \implies \text{doi}(x, y))$$

*Jeśli rolnik ma krowę to ją doi.*

- Analizując pierwotne zdanie uzyskamy najpierw formę pośrednią:



- A następnie będziemy szukać obiektów, na które wskazują wskaźniki #on i #ona.
- Jest to przykład rozwiązywania koreferencji (coreference resolution).



# Alternatywa

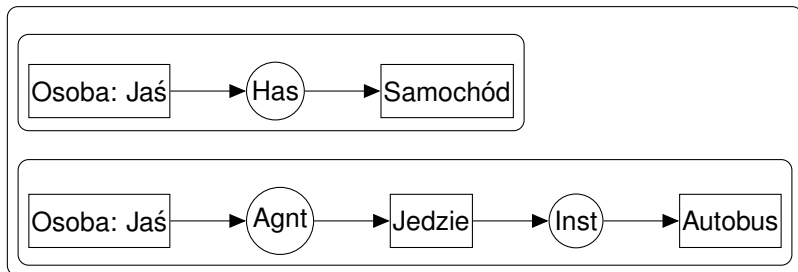
Korzystamy z tautologii

$$p \vee q \iff \neg(\neg p \wedge \neg q)$$

*Jaś ma samochód lub Jaś jedzie autobusem.*

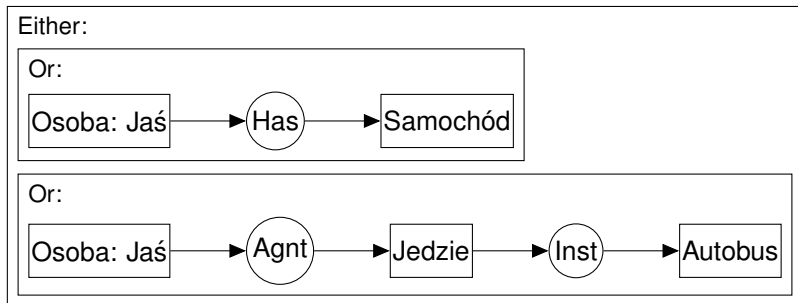
zamieniamy na

*Nie prawda, że Jaś nie ma samochodu i nie jedzie autobusem.*



# Alternatywa

Alternatywę reprezentujemy za pomocą kontekstu Either-Or:



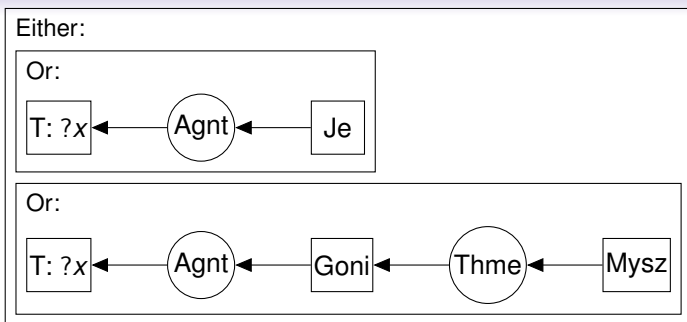
Z uwagi na to że

$$p_1 \vee p_2 \vee \dots \vee p_n \iff \neg(\neg p_1 \wedge \neg p_2 \wedge \dots \wedge \neg p_n)$$

w jednym kontekście Either może być dowolnie wiele kontekstów Or.

# Semantyka Either-Or

Kot:  $\forall *x$

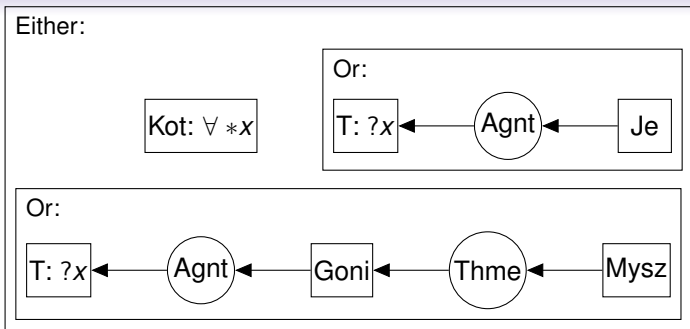


$$(\forall x : \text{Kot}) \neg (\neg ((\exists a : \text{Je}) \text{agnt}(a, x)) \wedge \neg ((\exists b : \text{Goni})(\exists y : \text{Mysz}) \text{agnt}(b, x) \wedge \text{thme}(b, y)))$$

czyli

$$(\forall x : \text{Kot}) (((\exists a : \text{Je}) \text{agnt}(a, x)) \vee ((\exists b : \text{Goni})(\exists y : \text{Mysz}) \text{agnt}(b, x) \wedge \text{thme}(b, y)))$$

# Semantyka Either-Or

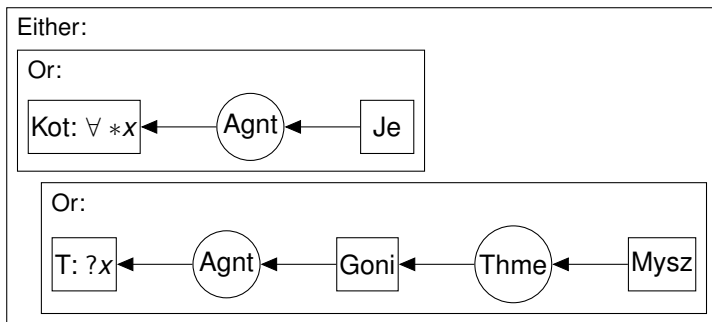


$$\neg(\forall x : \text{Kot})(\neg((\exists a : \text{Je})\text{agnt}(a, x)) \wedge \neg((\exists b : \text{Goni})(\exists y : \text{Mysz})\text{agnt}(b, x) \wedge \text{thme}(b, y)))$$

czyli

$$(\exists x : \text{Kot})(((\exists a : \text{Je})\text{agnt}(a, x)) \vee ((\exists b : \text{Goni})(\exists y : \text{Mysz})\text{agnt}(b, x) \wedge \text{thme}(b, y)))$$

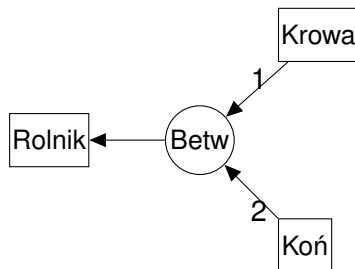
# Semantyka Either-Or



- Powyższy graf jest niepoprawny.
- Nie da się przetłumaczyć na formułę logiczną.
- $x$  jest użyte w kontekście rozłącznym z tym, w którym jest zdefiniowanie.

# Relacje wieloargumentowe

*Rolnik stoi pomiędzy krową a koniem.*



$$(\exists x : \text{Rolnik})(\exists y : \text{Krowa})(\exists z : \text{Koń})\text{betw}(y, z, x)$$

- Strzałka wychodząca do kółka wskazuje ostatni argument relacji.
- Pozostałe argumenty są reprezentowane przez strzałki wychodzące etykietowane liczbami wskazującymi ich kolejność.

# Spis treści

1 Grafy pojęć

2 Byty mentalne

3 Wnioskowanie o przekonaniach

# Przykład wiodący

- Rozpatrzmy robota szukającego wyjścia z labiryntu.
- Robot wykonuje akcje jakimi jest poruszanie się po labiryncie.
- Buduje model otaczającego go świata w miarę jak go odkrywa (uczy się labiryntu),
- Działa w celowo: to, że podejmuje akcje jest konsekwencją tego, że nie osiągnął jeszcze celu.
- Planuje swoje działania łącząc akcje w większe bloki, tworząc cele pośrednie, na przykład może chcieć opuścić ślepy zaułek.
- Przyjmijmy teraz, że labirynt jest fragmentem świata w grze sieciowej.
- Robot zaś jest programem grającym przez sieć w taką grę.



## *Robot buduje model otaczającego go świata*

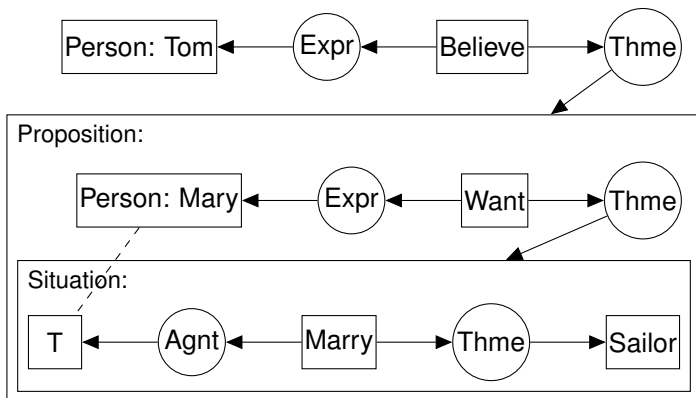
- Oznacza to, że robot ma jakąś strukturę danych, w której zbiera dostarczane mu informacje.
- Silnik gry dostarcza robotowi informacje o świecie, będące odpowiednikiem **wrażeń** rejestrowanych przez ludzi.
- Informacje te robot przetwarza do postaci pojęć przechowywalnych w jego strukturze danych, proces ten nazywamy **percepcją**.
- Następnie dodaje te informacje do struktury danych, **zapamiętuje** je.
- To jakie pojęcia są przechowywalne w modelu świata określa ontologia.

- Struktura danych modelująca świat może opierać się na założeniu, że ściany w labiryncie są odcinkami linii prostych i pamiętane są położenia końców tych odcinków w jakimś układzie współrzędnych.
- Strukturę tę możemy wyrazić za pomocą:
  - ▶ listy par punktów;
  - ▶ formuły logicznej.
  - ▶ grafu pojęć;
  - ▶ wypowiedzi w języku naturalnym;
- Struktura danych modelująca świat jest zbiorem **sądów** (Proposition) jakie robot ma o świecie.

# Nastawienia sędzeniowe

- **Nastawienia sędzeniowe** (Propositional Attitude) okrešlają postawę robota względem sądu.
- Sądy składające się na model świata robota są jego **przekonaniami** (Belief).
- Wyrażamy je za pomocą czasowników *wierzyć*, *być przekonanym*.
- Robot *działa celowo*, a więc posiada pewien sąd opisujący stan świata, który jest jego celem.
- **Pragnienie** (dążenie, Desire) to chęć, by świat był zgodny z sądem.
- Pragnienie opisujemy np. za pomocą czasownika *chcieć*.
- Np. *Robot chce znaleźć wyjście z labiryntu*.

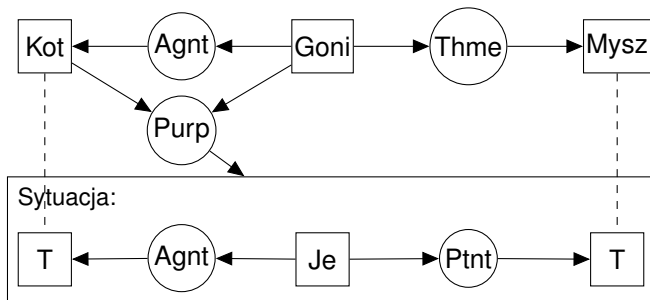
*Tom believes that Mary wants to marry a sailor.*



## Nastawienia sądeniowe: intencje

- Aby zrealizować swoje pragnienie robot wykonuje sekwencję akcji.
- Wykonując każdą z nich spodziewa się konkretnego efektu, zaistnienia określonej sytuacji i pragnie, by ta sytuacja zaistniała.
- Oznacza to, że robot ma **intencję** (Intention) realizacji celu wyrażonego przez ten sąd.
- Intencja to jest zamiar - ma wbudowany aspekt realizacji, łączy świat mentalny ze światem fizycznym.
- Intencję akcji wyrażamy jako okolicznik celu.
- Inne nazwy dla intencji: cel, zamiar, zamysł, aim, purpose.

*Kot goni mysz, żeby ją zjeść.*



- Intencja wiąże agenta, akcję i sytuację.
- Reprezentujemy ją za pomocą trójargumentowej relacji **Purp**.

# Nastawienia sędzeniowe: wyrażenia

- Nastawienia sędzeniowe mogą być też wyrażone.
- Są to np.: obietnice, stwierdzenia, kłamstwa, czy deklaracje.
- **Wyrażenie** (Expression) jest sędem wyrażonym za pomocą jakiegoś medium.
- Wyrażenia składają się z symboli obiektów istniejących w świecie oraz symboli relacji zachodzących między tymi obiektami.
- Środkiem wyrazu może być język naturalny, język logiki, graf pojęć, schemat, itd.

## Przykład wiodący cd.

- Stan rzeczy w grze jest zapisany w pamięci silnika gry.
- Możemy go opisać za pomocą pudełka z typu sytuacja zawierającego graf pojęć.
- Robot posiada niepełny i niekoniecznie prawdziwy obraz tego stanu rzeczy.
- Jego obraz świata opiszemy za pomocą pudełka z typu sąd zawierającego graf pojęć reprezentujący zawartość jego struktury danych.
- Sąd ten będzie przedmiotem przekonań robota.



- Podmiot, byt, który żywi nastawienia sędzeniowe, działa celowo.
- Agent ma zamiary (intends) określone przez jego intencje.
- Agent nie musi być świadomy.
- Działać celowo może: człowiek, organizacja, społeczeństwo, program komputerowy.
- Na podstawie pragnień agent układa **plan**, który jest strukturą złożoną z akcji i innych planów, i wykonuje go.
- Jest również przekonany (believes) odnośnie rezultatów swoich działań.
- Nawet świadomi agenci (ludzie) mogą realizować pragnienia i intencje, których nie są świadomi.

- Świat mentalny opisujemy przez analogię do świata fizycznego.
- **Przedmioty mentalne**, stanowiące byty trwałe w czasie to sądy i nastawienia sądzeniowe.
- **Procesy mentalne** (myślenie, zapamiętywanie, wyobrażanie itp.) są operacjami na obiektach mentalnych.
- **Stany mentalne** to odpowiedniki fizycznych sytuacji.
- **Reprezentacje mentalne** znaki, które zastępują elementy świata zewnętrznego (np. surogaty).

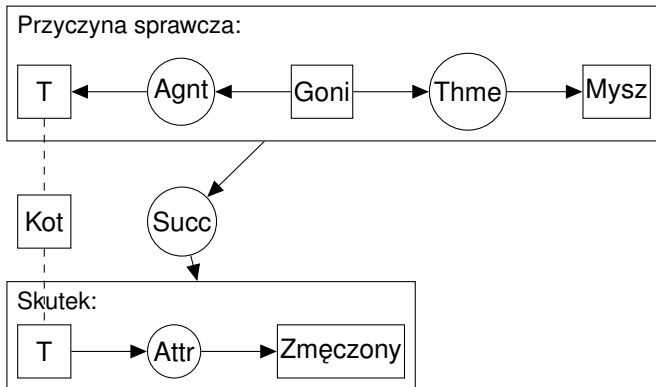
# Role epistemiczne

- **Role epistemiczne** (Epistemic Role) są to role odgrywane przez sądy i sytuacje, w procesach rozumowania, wnioskowania.
- Określają również prawdziwość — relację sądu ze światem.
- Sądy jednocześnie pełnią role epistemiczne i są przedmiotami nastawień sądzeniowych.
- **Obserwacje** (Observation) są to przekonania, które powstały w wyniku percepcji (przetwarzania doznań zmysłowych), mają odpowiadać jakiejś obserwowanej sytuacji.
- **Oczekiwanie** w sensie nadziei (antycypacja, Expectation) jest rolą sytuacji będącej przedmiotem intencji.

# Role epistemiczne

- **Fakt:** sąd odpowiadający zaistniałej sytuacji, np *Na ulicy leży denat.*
- **Przyczyna sprawcza** (causa efficiens; LKIF: Cause): Proces, który doprowadził do zaistnienia sytuacji, powstania rzeczy: *Został dźnięty nożem.*
- **Przyczyna celowa** (causa finalis; LKIF: Reason): Cel, któremu służyła zaistniała sytuacja, powstała rzecz (*padł ofiarą napadu rabunkowego*).
- Przyczyna celowa jest blisko związana z intencją i oczekiwaniem.
- W mowie potocznej przyczyna i powód (cause, reason) wydają się być synonimami.
- **Skutek:** drugi w czasie człon związku przyczynowo-skutkowego, efekt akcji.

*Kot gonit mysz, aż się zmęczył.*



# Spis treści

- 1 Grafy pojęć
- 2 Byty mentalne
- 3 Wnioskowanie o przekonaniach**

*Jaś nie umie całkować,  
ale jest przekonany, że umie całkować.*

- Dwa poziomy: obiektywny i poziom przekonań Jasia.
- Dwa światy: świat obiektywny i świat jakim widzi go Jaś.
- Jaś nie jest wszechwiedzący, więc jego przekonania nie wyznaczają dokładnie opisu świata.
- Co więcej, Jaś może się mylić: jego przekonania mogą być sprzeczne ze stanem faktycznym.

# Możliwe światy

- Zbiór  $\mathcal{K}$  zawierający byty zwane *możliwymi światami*.
- Wyróżniony świat  $w_0$  będący światem obiektywnym.
- *Relacja dostępności*  $R(u, v)$  wiążąca ze sobą możliwe światy.
- *Ewaluacja*  $\Phi(w, p)$  stwierdzająca, czy zdanie  $p$  jest prawdziwe w świecie  $w$ .
- Za pomocą zbioru możliwych światów opisujemy statyczną sytuację w świecie, w którym występuje agent mający przekonania o stanie tego świata.
- Predykat  $\text{believes}(a, p)$  oznacza, że agent  $a$  wierzy w prawdziwość zdania  $p$ .

$$\Phi(w, \text{believes}(a, p)) \iff (\forall v \in \mathcal{K})(\Phi(v, p) \iff R(w, v))$$

- Relacja  $R$  wyznacza wszystkie światy, które agent uważa za możliwe, dla zadanego świata aktualnego.



- Sytuacje są fragmentami światów wybranymi zgodnie z intencją obserwatora.
- Na przykład w świecie aktualnym *Jaś nie umie całkować*
- Możemy stworzyć sytuację  $s$ , która będzie obcięciem świata  $w_0$  do tego faktu.
- Sytuacja ta nie stanowi wycinka przestrzennego świata.
- Rozszerzymy relację  $\Phi$  tak by wiązała sytuacje z faktami.

$\Phi(s, \text{"Jaś nie umie całkować"})$

- Niektórzy uważają, że zbiór możliwych światów jest ideą o wątpliwej sensowności.
- Zamiast mówić o świecie możliwym, w który Jaś wierzy, możemy określić zbiór sądów definiujących ten świat i stwierdzić, wierzy on w prawdziwość tych sądów.
- Przekonania stają się wtedy sądami opakowanymi w pudełka (konteksty), które definiują ich rolę.
- Podobnie sytuacje możemy zinterpretować jako konteksty zawierające sądy o świecie.
- Pozwala to porzucić koncepcję możliwych światów na rzecz
  - ▶ jednego zmieniającego się świata, którym rządzą prawa logiki oraz
  - ▶ składających się ze zbiorów formuł logicznych przybliżonych opisów tego świata w umysłach agentów.

*Jaś nie umie całkować,  
ale jest przekonany, że umie całkować.  
Ponadto, Jaś wierzy, że jeśli ktoś umie całkować,  
to zaliczy kolokwium z całkowania*

- Niech  $U(x)$  oznacza, że  $x$  umie całkować,
- a  $Z(x)$  oznacza, że  $x$  zaliczył kolokwium z całkowania.

$$\Phi(s_1, \neg U(\text{Jaś}) \wedge \text{believes}(\text{Jaś}, U(\text{Jaś}) \wedge (\forall x)U(x) \Rightarrow Z(x)))$$

*Jaś dowiadyuje się z USOSa, że nie zaliczył kolokwium z  
całkowania.*

- Mamy tutaj fakt oraz obserwację:

$$\Phi(s_2, \neg U(\text{Jaś}) \wedge \neg Z(\text{Jaś}) \wedge \text{believes}(\text{Jaś}, U(\text{Jaś}) \wedge (\forall x)U(x) \Rightarrow Z(x)) \\ \wedge \text{observes}(\text{Jaś}, \neg Z(\text{Jaś})))$$

# Jaś rewiduje przekonania

$$\Phi(s_2, \neg U(\text{Jaś}) \wedge \neg Z(\text{Jaś}) \wedge \text{believes}(\text{Jaś}, U(\text{Jaś}) \wedge (\forall x)U(x) \Rightarrow Z(x)) \\ \wedge \text{observes}(\text{Jaś}, \neg Z(\text{Jaś})))$$

- Załóżmy, że Jaś jest racjonalnym agentem, czyli nie dopuszcza do tego by jego przekonania były sprzeczne.
- Przyjmijmy również, że Jaś potrafi wnioskować w logice tak, by zauważyć sprzeczność.
- W wyniku obserwacji Jaś rewiduje przekonania:

$$\Phi(s_3, \neg U(\text{Jaś}) \wedge \neg Z(\text{Jaś}) \wedge \text{believes}(\text{Jaś}, \neg Z(\text{Jaś}) \wedge (\forall x)U(x) \Rightarrow Z(x)))$$

- Agent dokonuje wniosków na podstawie swoich przekonań. Musi wierzyć w związek między swoimi umiejętnościami a zdaniem kolokwium, żeby wnioskować

# Rewizja (korekta) przekonań

- Przekonania są to sądy (zdania w logice) na podstawie których agent modeluje świat
- System przekonań jest to teoria powstała w oparciu o te sądy
- Zakładamy, że agent jest racjonalny, czyli że jego przekonania stanowią niesprzeczną teorię.
- Zakładamy też że agent posiada wszechwiedzę logiczną, czyli potrafi określić wszystkie logiczne konsekwencje swoich przekonań.
- Agent nabywa nową informację, która zmusza go do zmiany przekonań.
- W szczególności nowa informacja może być sprzeczna z jego dotychczasowymi przekonaniami.
- Agent musi w takiej sytuacji porzucić niektóre z dotychczasowych przekonań.
- Operację rewizji przekonań wyrażamy za pomocą operatora  $*$
- $K * \varphi$  dla zadanej teorii (zbioru sądów)  $K$  oraz nabywanego sądu  $\varphi$  generuje nowy zbiór sądów.

# Notacja

- $L$  zbiór wszystkich zdań języka, w naszym przypadku logiki pierwszego rzędu.
- Niech  $\Gamma$  będzie zbiorem zdań z  $L$ .
- $\Gamma \vdash \varphi$  oznacza, że  $\varphi$  możemy udowodnić, jeśli założymy że zdania z  $\Gamma$  są prawdziwe.
- Przez  $Cn(\Gamma)$ , oznaczmy zbiór wszystkich logicznych konsekwencji  $\Gamma$ , czyli

$$Cn(\Gamma) = \{\psi \in L : \Gamma \vdash \psi\}$$

- Zbiór wszystkich teorii nad  $L$  oznaczamy jako  $\mathbb{K}_L$
- Dla teorii  $K$  definiujemy

$$K + \Gamma = Cn(K \cup \Gamma)$$

- Dla zdania  $\varphi$  notacja  $K + \varphi$  jest skrótem od  $K + \{\varphi\}$

# Postulaty rewizji przekonań (Alchourron, Gardenfors, Makinson)

Funkcja rewizji przekonań  $*$  :  $\mathbb{K}_L \times L \rightarrow \mathbb{K}_L$  spełnia następujące aksjomaty:

( $K * 1$ )  $K * \varphi$  jest teorią w  $L$

( $K * 2$ )  $\varphi \in K * \varphi$

- Postulat ( $K * 1$ ) mówi, że agent, pozostaje logicznie wszechwiedzący, gdy zrewiduje przekonania.
- Postulat ( $K * 2$ ) mówi, że nową informację  $\varphi$  należy zawsze zawrzeć w nowym zbiorze przekonań.
- ( $K * 2$ ) przywiązuje olbrzymią wagę do wiarygodności  $\varphi$ .
- Nowa informacja jest postrzegana jako tak wiarygodna, że ma pierwszeństwo przed wszystkimi sprzecznymi z nią przekonaniem, niezależnie od tego, co to za przekonania.

## Postulaty rewizji przekonań cd.

$$(K * 3) \quad K * \varphi \subseteq K + \varphi$$

$$(K * 4) \quad \text{Jeżeli } \neg\varphi \notin K, \text{ to } K + \varphi \subseteq K * \varphi$$

- $(K * 3)$  i  $(K * 4)$  razem stanowią, że jeśli nowa informacja  $\varphi$ , nie ma powodu, by usuwać jakiegokolwiek dotychczasowe przekonanie.
- W takiej sytuacji nowy system przekonań  $K * \varphi$  będzie zawierał całe  $K$ , nową informację  $\varphi$  i wszystko co jest logiczną konsekwencją  $K$  i  $\varphi$ .
- Zasadniczo  $(K * 3)$  i  $(K * 4)$  wyrażają koncepcję minimalnej zmiany w przypadku, gdy nowa informacja jest niesprzeczna z dotychczasowymi przekonaniem.



## Postulaty rewizji przekonań cd.

(K \* 5) Jeżeli  $\varphi$  jest niesprzeczna, to  $K * \varphi$  jest niesprzeczna

(K \* 6) Jeżeli  $\vdash \varphi \Leftrightarrow \psi$ , to  $K * \varphi = K * \psi$

- (K \* 5) stanowi, że agent powinien dążyć do niesprzeczności za wszelką cenę.
- Jedynie w przypadku kiedy nowo nabyte przekonanie jest sprzeczne, agent, z uwagi na postulat (K \* 2) musi przyjąć sprzeczny zbiór przekonań.
- (K \* 6) jest postulatem nieistotności składni.
- Mówi, że składnia sądu (informacji) nie ma wpływa na efekt rewizji; liczy się jedynie jego znaczenie.
- Równoważne logicznie zdania  $\varphi$  i  $\psi$  zmieniają teorię  $K$  w taki sam sposób.

## Postulaty rewizji przekonań cd.

$$(K * 7) \quad K * (\varphi \wedge \psi) \subseteq (K * \varphi) + \psi$$

$$(K * 8) \quad \text{Jeżeli } \neg\psi \notin K * \varphi, \text{ to } (K * \varphi) + \psi \subseteq K * (\varphi \wedge \psi)$$

- Postulaty  $(K * 7)$  i  $(K * 8)$  mówią, że dla dowolnych zdań  $\varphi$  i  $\psi$ ,
  - ▶ jeśli rewidując początkowy zbiór przekonań  $K$  za pomocą  $\varphi$  agent uzyska zbiór przekonań  $K * \varphi$  zgodny z  $\psi$ ,
  - ▶ to, żeby wygenerować  $K * (\varphi \wedge \psi)$  wystarczy rozszerzyć  $K * \varphi$  za pomocą  $\psi$ .

symbolicznie:  $K * (\varphi \wedge \psi) = (K * \varphi) + \psi$ .

- Racjonalne uzasadnienie jest następujące:  $K * \varphi$  jest najmniejszą zmianą  $K$  pozwalającą włączyć  $\varphi$ , nie da się zatem uzyskać  $K * (\varphi \wedge \psi)$  z  $K$  za pomocą "mniejszej zmiany".
- Z drugiej strony jeśli  $\psi$  jest niesprzeczne z  $K * \varphi$ , rewizja przekonań może się ograniczyć do dodania  $\psi$  do  $K * \varphi$  i ich domknięcia.

## Jaś rewiduje przekonania

Przekonania Jasia przed ogłoszeniem wyników kolokwium opisuje następująca teoria

$$K = \text{Cn}(U(\text{Jaś}), (\forall x)U(x) \Rightarrow Z(x))$$

Po zapoznaniu się z wynikami Jaś dowiadyuje się że ma miejsce fakt

$$\varphi = \neg Z(\text{Jaś})$$

W związku z czym Jaś dokonuje rewizji przekonań uzyskując teorię  $K * \varphi$ . Zgodnie z aksjomatem  $K * 2$  mamy  $\varphi \in K * \varphi$ . Zgodnie z aksjomatem  $K * 5$   $K * \varphi$  jest niesprzeczne. Teoria

$$K' = \text{Cn}((\forall x)U(x) \Rightarrow Z(x), \neg Z(\text{Jaś}))$$

spełnia aksjomaty. Teoria

$$K'' = \text{Cn}(U(\text{Jaś}), \neg Z(\text{Jaś}))$$

również spełnia aksjomaty.

# Jaś rewiduje przekonania

- Teoria

$$K' = Cn((\forall x)U(x) \Rightarrow Z(x), \neg Z(\text{Jaś}))$$

opisuje rewizję przekonań polegającą na tym, że Jaś uświadomił sobie, że nie umie całkować.

- W przypadku teorii

$$K'' = Cn(U(\text{Jaś}), \neg Z(\text{Jaś}))$$

Jaś zwątpił w związek pomiędzy swoimi umiejętnościami a oceną.

- Którą z powyższych teorii Jaś powinien uzyskać po rewizji przekonań?
- Warto jeszcze zauważyć, że w wyniku rewizji przekonań w powyższym przypadku Jaś nie powinien uznać prawdziwości zdania  $\psi$  oznaczającego *W pokoju jest słoń*.

# Kontrakcja (porzucanie) przekonania (Alchourron, Gardenfors, Makinson)

- Racjonalne usunięcie przekonania ze zbioru przekonania.
- Następuje, gdy agent traci wiarę w przekonanie i rezygnuje z niego.
- Nie wystarczy usunąć przekonania z teorii, trzeba usunąć również wszystkie fakty, które pozwalają je wyprowadzić.

## Postulaty kontrakcji przekonań

Funkcja kontrakcji przekonań  $\div : \mathbb{K}_L \times L \rightarrow \mathbb{K}_L$  spełnia następujące postulaty:

$(K \div 1)$       $K \div \varphi$  jest teorią w  $L$

$(K \div 2)$       $K \div \varphi \subseteq K$

$(K \div 3)$      Jeżeli  $\varphi \notin K$ , to  $K \div \varphi = K$

$(K \div 4)$      Jeżeli  $\not\vdash \varphi$ , to  $\varphi \notin K \div \varphi$

$(K \div 5)$      Jeżeli  $\varphi \in K$ , to  $K \subseteq (K \div \varphi) + \varphi$

- Postulat  $(K \div 3)$  mówi, że jeśli  $\varphi$  nie należy do systemu przekonań nie trzeba go zmieniać.
- $(K \div 4)$  stanowi, że  $\varphi$  jest usuwane z systemu przekonań, chyba że jest tautologią.
- Odpowiada to postulatowi  $(K * 2)$ .
- Na mocy  $(K \div 5)$  i  $(K * 2)$  kontrakcja, a następnie dodanie  $\varphi$  pozwala nam uzyskać pierwotną teorię.
- $(K \div 5)$  realizuje postulat minimalnej zmiany.

# Postulaty kontrakcji przekonań

$(K \div 6)$  Jeżeli  $\vdash \varphi \Leftrightarrow \psi$ , to  $K \div \varphi = K \div \psi$

$(K \div 7)$   $(K \div \varphi) \cap (K \div \psi) \subseteq K \div (\varphi \wedge \psi)$

$(K \div 8)$  Jeżeli  $\varphi \notin K \div (\varphi \wedge \psi)$ , to  $K \div (\varphi \wedge \psi) \subseteq K \div \varphi$

- $(K \div 6)$  jest postulatem nieistotności składni.
- Ostatnie dwa postulaty wiążą sekwencyjną kontrakcję dwu zdań i kontrakcją ich koniunkcji.
- Aby porzucić  $\varphi \wedge \psi$  musimy porzucić albo  $\varphi$ , albo  $\psi$ , albo oba z nich.
- Przekonanie, które nie jest porzucane w wyniku kontrakcji  $\varphi$ , ani kontrakcji  $\psi$  nie jest z nimi związane, nie jest więc związane również z  $\varphi \wedge \psi$ .
- Zatem zgodnie z zasadą minimalnej zmiany nie powinno ono zostać porzucone ( $K \div 7$ ).
- $\varphi$  wynika z  $\varphi \wedge \psi$ , zatem kontrakcja  $\varphi \wedge \psi$  powinna wywołać większy efekt niż kontrakcja  $\varphi$  ( $K \div 8$ ).

# Związek kontrakcji z rewizją

Tożsamość Leviego :

$$K * \varphi = (K \div \neg\varphi) + \varphi$$

## Theorem

*Każda funkcja  $\div : \mathbb{K}_L \times L \rightarrow \mathbb{K}_L$  spełniająca postulaty  $(K \div 1)$ – $(K \div 8)$  generuje za pomocą tożsamości Leviego funkcję  $*$  spełniającą postulaty  $(K * 1)$ – $(K * 8)$*

Tożsamość Harpera:

$$K \div \varphi = (K * \neg\varphi) \cap K$$

## Theorem

*Każda funkcja  $*$  :  $\mathbb{K}_L \times L \rightarrow \mathbb{K}_L$  spełniająca postulaty  $(K * 1)$ – $(K * 8)$  generuje za pomocą tożsamości Leviego funkcję  $\div$  spełniającą postulaty  $(K \div 1)$ – $(K \div 8)$*



# Problem słońia

- Korzystamy z identyczności Leviego by opisać za pomocą kontrakcji rewizję przekonań u Jasia.
- Najpierw usuwamy przekonanie

$$\varphi = \neg Z(\text{Jaś})$$

uzyskując zgodnie z postulatami kontrakcji (K 2) dwie możliwe teorie

$$Cn((\forall x)U(x) \Rightarrow Z(x))$$

$$Cn(U(\text{Jaś}))$$

- Następnie dodajemy  $\varphi$ .
- Jak widać mamy w ten sposób zapewnione to, że Jaś nie zacznie wierzyć, że w pokoju jest słoń.

# Wsparcie Epistemiczne (Epistemic Entrenchment)

- Agent przypisuje wartości epistemiczne poszczególnym przekonaniom w swoim systemie.
- Na przykład ogólne prawo  $(\forall x)U(x) \Rightarrow Z(x)$  może być dla Jasia ważniejsze niż fakt  $U(\text{Jaś})$ .
- Zatem Jaś musząc wybrać pomiędzy nimi zachowa  $(\forall x)U(x) \Rightarrow Z(x)$ .
- Wsparcie epistemiczne przekonania określa jak bardzo jest ono odporne na zmianę.
- Im większe jest wsparcie epistemiczne tym mniej prawdopodobne jest usunięcie przekonania w wyniku kontrakcji.

# Aksjomaty wsparcia epistemicznego

Wsparcie epistemiczne jest relacją  $\leq$  na  $L$  spełniającą następujące warunki:

(EE1)      Jeśli  $\varphi \leq \psi$  i  $\psi \leq \chi$  to  $\varphi \leq \chi$

(EE2)      Jeśli  $\varphi \vdash \psi$  to  $\varphi \leq \psi$

(EE3)       $\varphi \leq \varphi \wedge \psi$  lub  $\psi \leq \varphi \wedge \psi$

- Aksjomat (EE1) stanowi o przechodniości  $\leq$ .
- (EE2) mówi, że przekonania logicznie silniejsze mają mniejsze wsparcie.
- Jeśli zachodzi  $\varphi \vdash \psi$  możemy zrezygnować z  $\varphi$  i zachować  $\psi$ , ale nie na odwrót.
- Z uwagi na to, że  $K$  jest teorią, nie możemy usunąć  $\varphi \wedge \psi$  nie usuwając jednocześnie  $\varphi$  lub  $\psi$ , co daje (EE3)
- Z (EE1)–(EE3) wynika, że  $\leq$  jest relacją totalną.

(EE4) Dla niesprzecznego  $K$  zachodzi:

$\varphi \notin K$  wtw. dla każdego  $\psi \in L$  zachodzi  $\varphi \leq \psi$

(EE5) Jeśli dla każdego  $\psi \in L$  zachodzi  $\psi \leq \varphi$  to  $\vdash \varphi$

- (EE4) stanowi, że fakty, w które agent nie wierzy mają minimalne wsparcie.
- Z kolei tautologie mają maksymalne wsparcie (EE5).

# Subiektywność wsparcia epistemicznego

- Dla ustalonego systemu przekonań istnieje zwykle wiele różnych relacji spełniających aksjomaty (EE1)–(EE5)
- Odpowiada to subiektywnemu charakterowi wsparcia epistemologicznego.
- Co ciekawe, intuicyjnie jesteśmy skłonni wiązać wybór relacji wsparcia z racjonalnością.
- To, że nie zaliczywszy kolokwium Jaś uzna, że nie umie całkować, jest z jego strony bardziej racjonalne niż zwątpienie w system oceniania.

# Związek wsparcia epistemicznego z kontrakcją

- Kontrakcję możemy zdefiniować za pomocą wsparcia epistemicznego w następujący sposób:

$$(C-) \quad \psi \in K \div \varphi \text{ wtw. } \psi \in K \text{ oraz } \varphi < \varphi \vee \psi \text{ lub } \vdash \varphi$$

## Theorem

- *Niech  $K$  będzie teorią.*
- *Jeśli  $\leq$  jest relacją spełniającą (EE1)–(EE5) to funkcja zdefiniowana za pomocą (C-) jest funkcją kontrakcji.*
- *Odwrotnie, jeśli  $\div$  jest funkcją kontrakcji to istnieje relacja spełniająca (EE1)–(EE5) oraz (C-).*

# Jaś rewiduje przekonania korzystając ze wsparcia i kontrakcji

$$\varphi = Z(\text{Jaś}), \quad \psi = U(\text{Jaś}), \quad \chi = (\forall x)U(x) \Rightarrow Z(x)$$

$$K = \text{Cn}(\psi, \chi) = \text{Cn}(\psi, \chi, \varphi, \varphi \vee \psi, \varphi \vee \chi)$$

Zakładamy, że  $\psi < \chi$ , co wyznacza nam wsparcie epistemiczne:

$$\varphi \leq \psi \leq \varphi \vee \psi < \chi \leq \varphi \vee \chi$$

Chcemy policzyć  $K * \neg\varphi$ . Zgodnie z tożsamością Leviego zachodzi

$$K * \neg\varphi = (K \div \varphi) + \neg\varphi$$

Wiemy, że  $K \div \varphi \subseteq K$ . Na mocy (C-) otrzymujemy

Jeśli  $\varphi < \varphi \vee \psi$  to  $\psi \in K \div \varphi$

Jeśli  $\varphi < \varphi \vee \chi$  to  $\chi \in K \div \varphi$