QUE SIGNIFIE « NON » ?

Dans un premier temps, revenons à des
bases de faits, *sans ontologie*

THÉORIE DES BASES DE CONNAISSANCES
HMIN312M

Hypothèse du monde ouvert

OWA : open-world assumption

- On suppose une connaissance **incomplète** du monde

  → ce qui est dans la base de faits est vrai, mais pour le reste on ne sait pas

- Pour déterminer les réponses à une requête « de façon certaine », on considère toutes les extensions possibles de la base de faits

  → cela correspond à considérer **tous les modèles** de la base de faits

  donc la conséquence logique classique

- C'est l'hypothèse généralement faite dans les systèmes à base de connaissances (y compris dans le web sémantique)

  ex : arbre généalogique, liste des témoins d'un accident,
  liste des fans d'une star, ...

Hypothèse du monde clos

CWA : closed-world assumption

On suppose qu'on a une connaissance complète du monde

→ la base de faits contient tous les atomes vrais, les autres sont faux

Pour déterminer les réponses à une requête,
il suffit de considérer l'unique modèle associé à la base de faits
[c'est-à-dire le modèle isomorphe à la base de faits]

Rappel : une interprétation logique correspond à un monde « complètement connu »

- C'est l'hypothèse faite en bases de données

  Tout ce qui est stocké est vrai, le reste est faux
  ex : liste des utilisateurs enregistrés, plan du métro, ...

Cela fait-il une différence pour les requêtes conjonctives ?

Non

car, pour une BCQ Q,

tout modèle de *F* est un modèle de Q

ssi

le modèle isomorphe à *F* est un modèle de *Q*

Mais ajoutons à nos requêtes (et nos faits) des littéraux négatifs ...

Rappel : un **littéral** est un atome ou la négation d'un atome

Négation atomique dans les requêtes

- **CQ$^{\neg}$** : conjonction de littéraux quantifiée existentiellement

$$Q(x) = \exists y\, (p(x,y) \land \neg p(y,x))$$

Notation:
$Q = Q+ \cup Q-$

Comment calculer les réponses à une CQ$^{\neg}$?

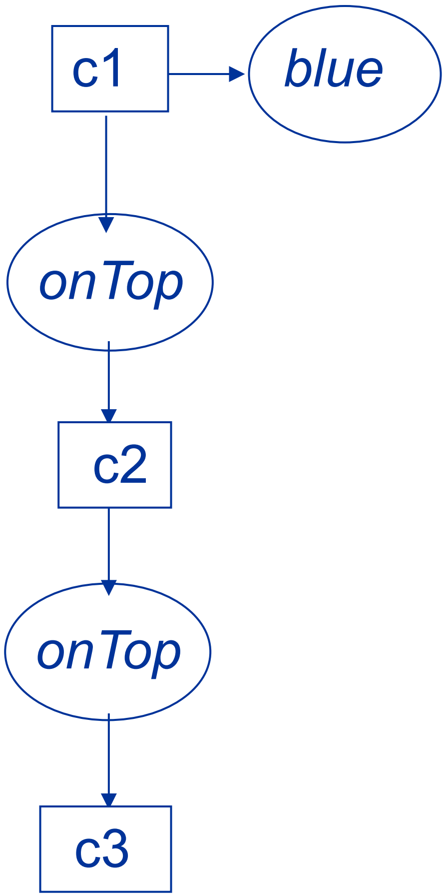
- en **monde ouvert** et avec une base de faits **positive** : ?
- en **monde clos** (et base de faits **forcément** positive) : ?

Prenons une CQ$^{\neg}$ booléenne Q :

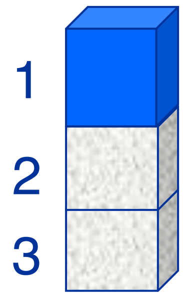
- CWA : il existe un homomorphisme h de Q+ dans F, tel que $h(Q-) \cap F = \emptyset$
  (que faire si certaines variables de Q n'apparaissent pas dans Q+?)
- OWA : si Q a au moins un littéral négatif, la réponse à Q est non, pour tout F

- si la base de faits est **positive**, la notion de CQ$^{\neg}$ n'a d'intérêt qu'avec l'**hypothèse du monde clos**
- on considère alors des **CQ$^{\neg}$ sûres** (safe) : toute variable de la requête apparait au moins dans un littéral positif

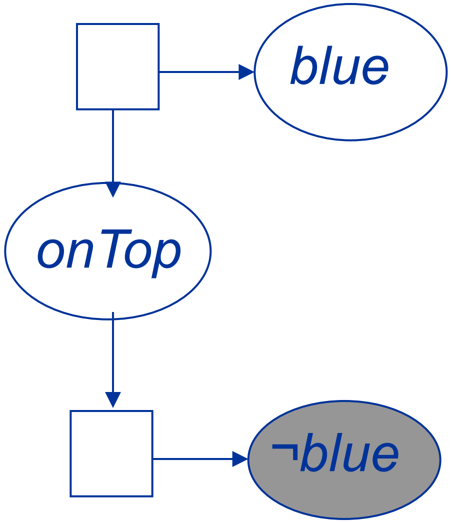
Closed World Assumption



F



Q



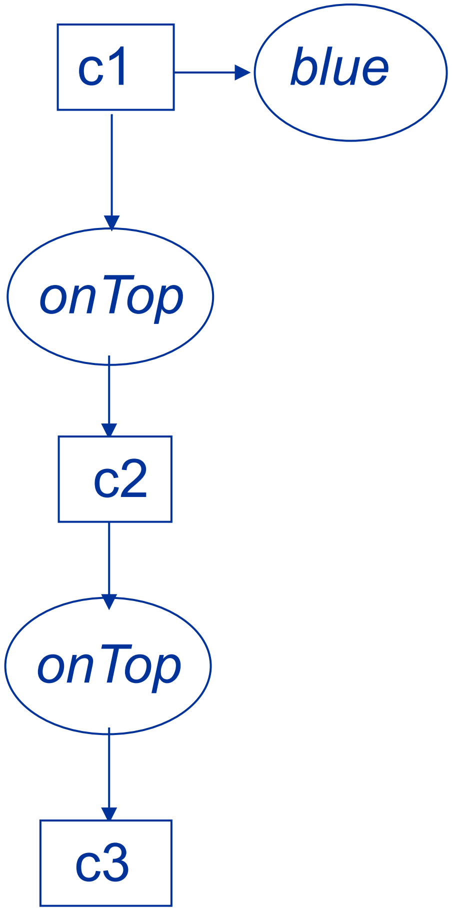
[Q as a Boolean CQ¬]
∃x ∃y (blue(x) ∧ ¬blue(y) ∧ onTop(x,y)) ?

yes { () }

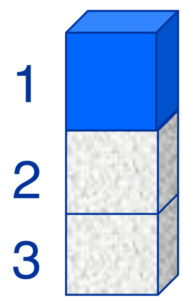
[Q with answer variable x]
∃y (blue(x) ∧ ¬blue(y) ∧ onTop(x,y))

{ (c1) }

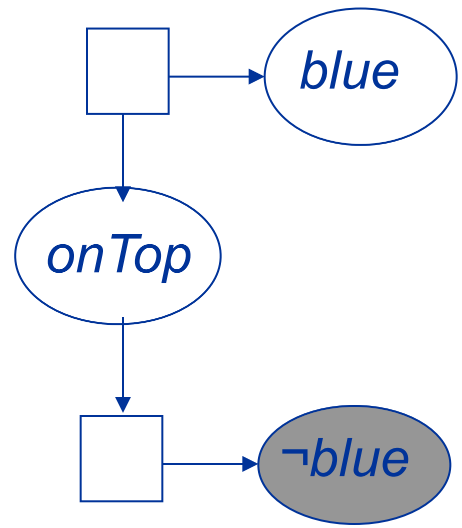
Open World Assumption



F



Q



[Q as a Boolean CQ¬]
$\exists x \, \exists y \, (\text{blue}(x) \wedge \neg\text{blue}(y) \wedge \text{onTop}(x,y))$?

no {}

[Q with answer variable x]
$\exists y \, (\text{blue}(x) \wedge \neg\text{blue}(y) \wedge \text{onTop}(x,y))$

{}

Monde ouvert: négation dans les **faits** et les **requêtes**

- **Base de faits**:
  - conjonction de **littéraux** instanciés (ground) : sans variables
  - extension possible : conjonction de littéraux fermée existentiellement
- Une base de faits peut être inconsistante (insatisfiable)
  ce qu'on peut vérifier facilement : une BF est inconsistante ssi
  elle contient deux littéraux opposés $p(e_1 \dots e_k)$ et $\neg p(e_1 \dots e_k)$

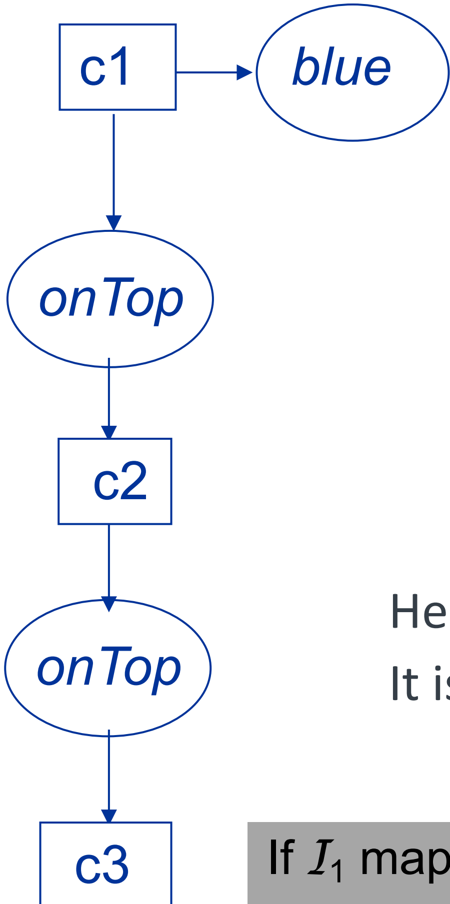
  [Si pas de littéraux opposés, le modèle isomorphe à (la partie positive de) la BF est un modèle de la BF]
- **$CQ^{\neg}$**
- Lorsque *F* est consistante, la notion de réponse est la même qu'avec les CQ

The answer to a $BCQ^{\neg}$ Q on F is *yes* if $F \models Q$

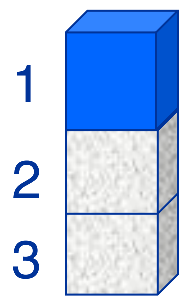
A tuple $(a_1, \dots, a_k)$ of *constants* is a (certain) answer to a $CQ^{\neg}$ $Q(x_1,\dots,x_k)$ on F if $F \models Q[a_1,\dots,a_k]$, where $Q[a_1,\dots,a_k]$ is obtained from $Q(x_1,\dots,x_k)$ by replacing each x_i by a_i.

mais le calcul devient plus complexe

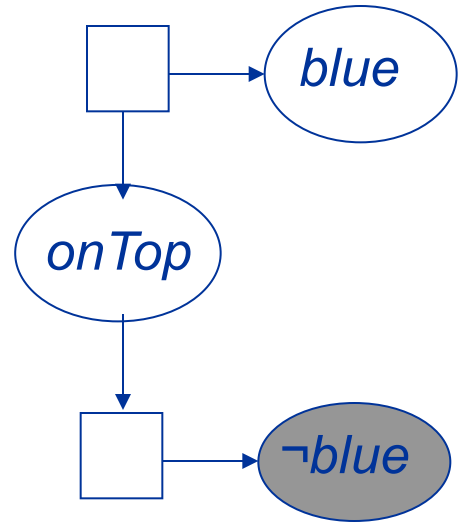
Open World Assumption (positive factbase)



F



Q



Here, the model isomorphic to *F* is a model of *Q*

It is a universal model but it is not sufficient anymore

Why ?

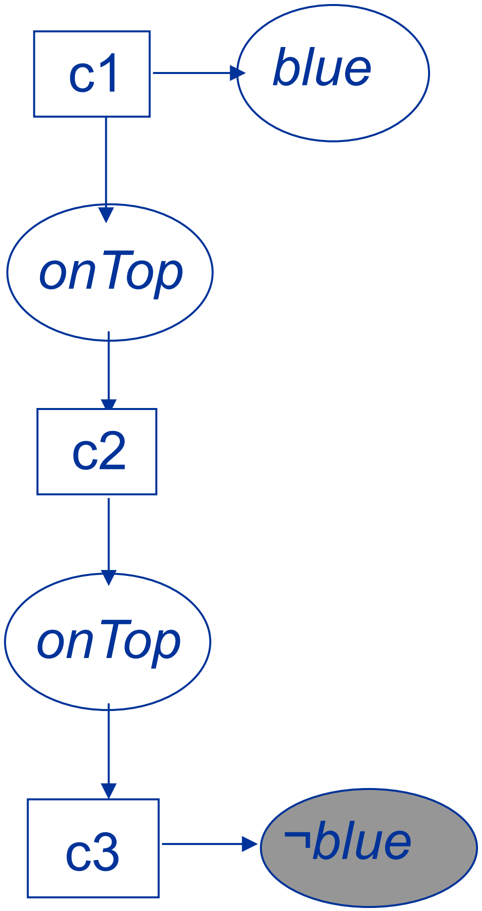
If $\mathcal{I}_1$ maps by homomorphism to $\mathcal{I}_2$ then, for any closed *f* in FOL(∃,∧),

$\mathcal{I}_1$ model of *f* $\Rightarrow$ $\mathcal{I}_2$ model of *f*

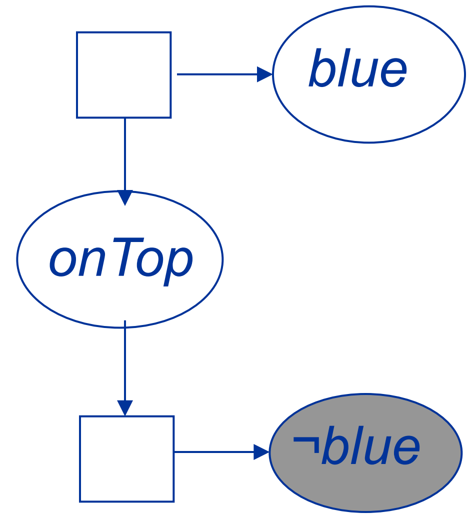
This property becomes false if *f* is a BCQ$^{\neg}$

Open World Assumption

F



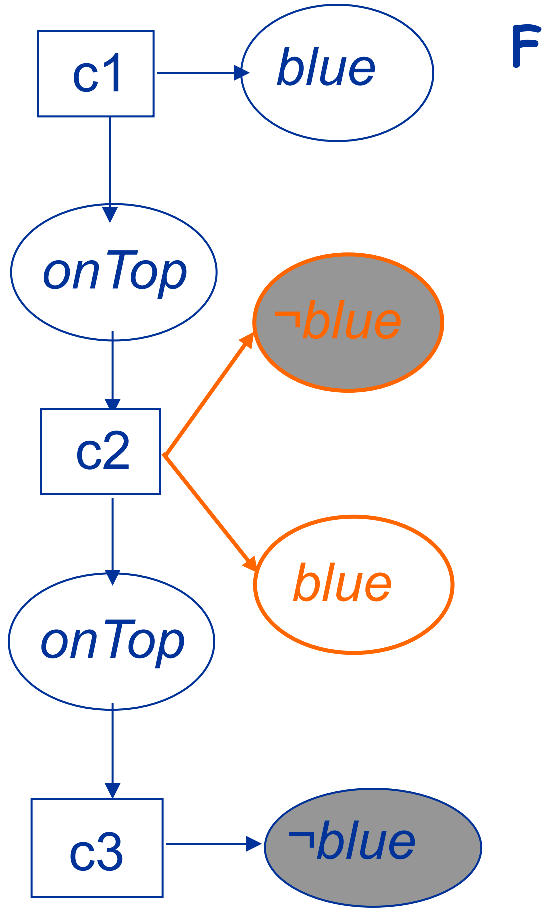
Q

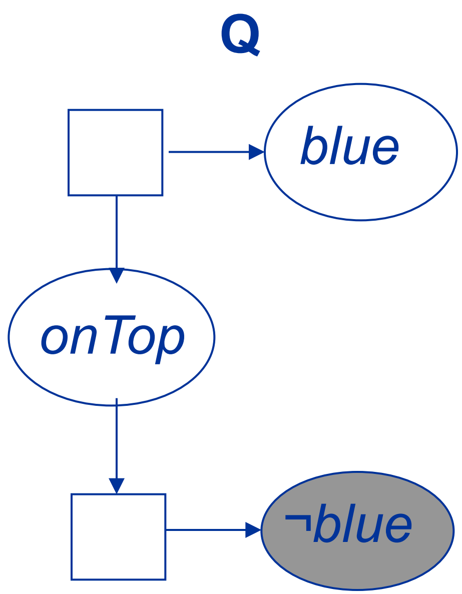


[Q as a Boolean CQ¬]
$\exists x\, \exists y\, (blue(x) \wedge \neg blue(y) \wedge onTop(x,y))$?

[Q with answer variable x]
$\exists y\, (blue(x) \wedge \neg blue(y) \wedge onTop(x,y))$

Where the « law of the excluded middle » comes in ...





[Q as a Boolean CQ¬]
∃x ∃y (blue(x) ∧ ¬blue(y) ∧ onTop(x,y)) ?

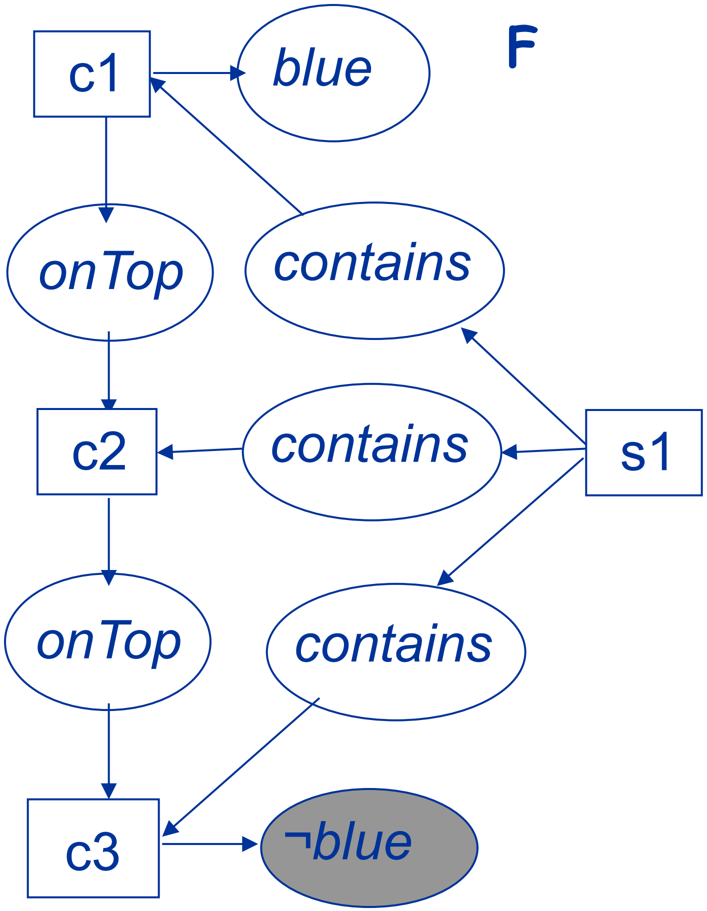
yes

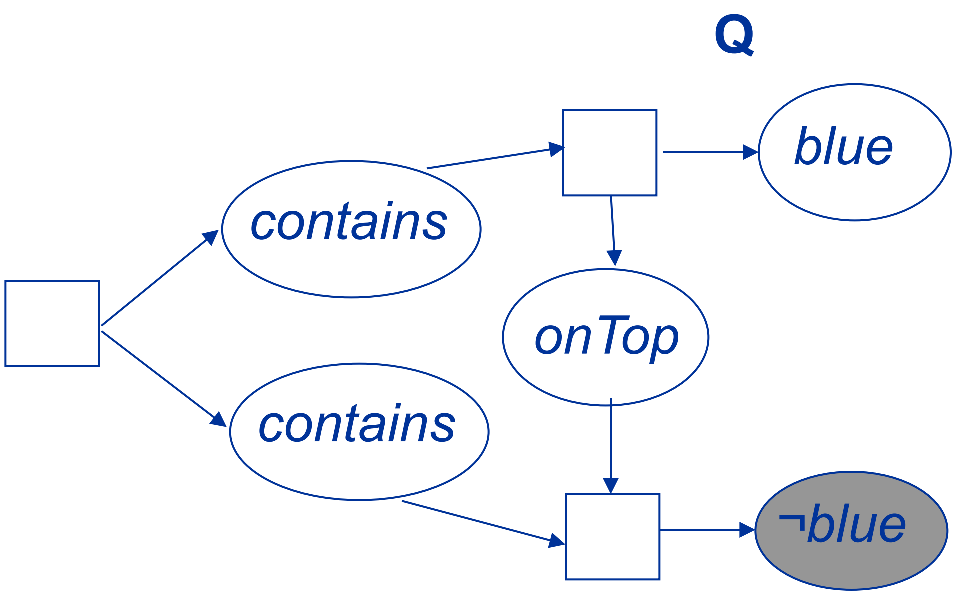
[Q with answer variable x]
∃y (blue(x) ∧ ¬blue(y) ∧ onTop(x,y))

{} there is no (certain) answer

In *any* model of *F*,
blue(c2) is either true or false

Where the « law of the excluded middle » comes in ...





[Q with answer variable x]

$\exists y\ \exists z\ (\ contains(x,y) \wedge contains(x,z) \wedge$

$\quad blue(y) \wedge \neg blue(z) \wedge onTop(y,z)\ )$

Answer ?

{ (s1) }

Egalité et différence (¬=)

- **Interprétation de l'égalité (et de la différence)**

  Une interprétation *I* satisfait $(t_1 = t_2)$ si $t_1^I = t_2^I$
  Une interprétation I satisfait $\neg(t_1 = t_2)$ si t_1^I et t_2^I sont différents
- **Egalité** : inutile dans les faits et les requêtes (on substitue)
- **UNA** (Unique Name Assumption) : pour toutes constantes distinctes *a* et *b*,
  on a implicitement $\neg(a = b)$

- **Exercice**. On décide d'enrichir les requêtes **conjonctives** avec la **différence** ; la base de faits ne comporte **pas de variables** (ground) et on fait l'**UNA**

  ex : $Q_1 = \exists x\, \exists y\, \exists z\, (r(x,z) \wedge r(y,z) \wedge \neg =(x,y))$
  $Q_2 = \exists u\, \exists v\, (r(u,v) \wedge \neg =(u,v))$

  1. l'interrogation change-t-elle, selon que l'on fait l'hypothèse du monde ouvert ou du monde clos ?
  2. comment calculer les réponses à une requête ?
  3. pour les BCQ on avait : $Q_1 \sqsubseteq Q_2$ ssi Q_2 s'envoie par homomorphisme dans Q_1. A-t-on une propriété similaire pour des CQs avec différence ? [Indication : voir les requêtes de l'exemple, penser au tiers-exclu]

Si on ajoute des règles ? OWA

- Déjà en logique des propositions, le mécanisme de marche avant n'est plus complet par rapport à la conséquence logique

  Même avec des règles positives :

  R1 : A → B

  R2 : A → C

  R3 : B ∧ C → D

  Fait : ¬D [Question : a-t-on ¬A ?]

- En logique du premier ordre, avec la négation OWA classique :
  la notion même d'application de règle devient problématique
  Ex: « si p(x,y) et q(z) et ¬p(y,z) alors r(x,y,z) »

- Exemple de méthode complète (pour toute la logique du 1er ordre) :
  méthode de résolution
  (nécessite de mettre les formules sous forme clausale)

Si on ajoute des règles ? CWA

- Bases de données déductives (Datalog)
  Faits : stockés dans la BD (relations dites extensionnelles)
  Requêtes incluant des règles qui permettent de déduire des faits
  non stockés dans la BD (relations dites intensionnelles)
- CWA : négation seulement dans le corps des règles
- La négation selon CWA rend la conséquence non monotone :
  en augmentant la base de données (base de faits), certains faits ne sont plus conséquences de la base
- Problème : ce qui est conséquence (par exemple l'ensemble des réponses à une requête donnée par une liste de règles) dépend de l'ordre d'application des règles → quelle sémantique déclarative donner à une telle base ?
- Cas où la sémantique peut être définie simplement :
  Datalog semi-positif (pas de problème d'ordre)
  Datalog stratifiable (toutes les stratifications donnent le même résultat)
- Cas général : Answer Set Programming