

Believe It or Not – Adding belief annotations to databases

Wolfgang Gatterbauer, Magda Balazinska,
Nodira Khoussainova, and Dan Suciu

University of Washington

<http://db.cs.washington.edu/beliefDB/>

High-level overview

- DBMS: manage consistent data
- Applications need inconsistent DM
 - Scientific databases reason: disagreement !
 - Internet community databases
- Community DBMS: manage inconsistent views

- This work: **Belief databases**
 - manage data and curation
 - grounded in modal and default logic
 - implemented on top of relational model

Agenda

- Motivating example
- Logic foundations
- Relational implementation
- Discussion

Motivating application

- NatureMapping project (<http://depts.washington.edu/natmap/>)
 - volunteer contribute animal observations
 - one person curates the database

problem: does not scale!

Observations

<u>id</u>	uid	species	date	location	comment
2	Alice	Crow	06-14-08	Lake Placid	found feathers

Sightings (S)

<u>sid</u>	uid	species	date	location
s2	Alice	Crow	06-14-08	Lake Placid

Comments (C)

<u>cid</u>	comment	sid
c1	found feathers	s2

1. Distinct database instances



Alice

S

<u>sid</u>	uid	species	date	location
s2	Alice	Crow	06-14-08	Lake Placid



Bob

S

<u>sid</u>	uid	species	date	location
s2	Alice	Raven	06-14-08	Lake Placid

D1: Belief worlds: individually consistent,
mutually possibly inconsistent

1. Distinct database instances



Alice

S

<u>sid</u>	uid	species	date	location
s2	Alice	Crow	06-14-08	Lake Placid



Bob

S

<u>sid</u>	uid	species	date	location
s2	Alice	Raven	06-14-08	Lake Placid

BeliefSQL

I: Alice believes that she saw a crow.

```
insert into BELIEF 'Alice' Sightings
values ('s2','Alice','Crow','06-14-08','Lake Placid')
```

I: Bob believes that she actually saw a raven.


```
insert into BELIEF 'Bob' Sightings
values ('s2','Alice','Raven','06-14-08','Lake Placid')
```

Q: Who believes something different than Alice and what?

```
select U2.name, S1.species, S2.species
from Users as U,
BELIEF 'Alice' Sightings as S1,
BELIEF U.uid Sightings as S2,
where S1.sid = S2.sid
and S1.species <> S2.species
```

A: {'Bob', 'Crow', 'Raven'}

2. Open world assumption



Alice

S

<u>sid</u>	uid	species	date	location
s2	Alice	Crow	06-14-08	Lake Placid



Bob

S

<u>sid</u>	uid	species	date	location
s2	Alice	Raven	06-14-08	Lake Placid



Carol


S

<u>sid</u>	uid	species	date	location	
s2	Alice	Crow	06-14-08	Lake Placid	⊖
s2	Alice	Raven	06-14-08	Lake Placid	⊖

Adapted key constraints !

D2: Model incomplete knowledge with explicit negative beliefs

2. Open world assumption



Alice


s

sid	uid	specie
s2	Alice	Crow

I: Carol does not believe that Alice saw a crow nor a raven.

insert into BELIEF 'Carol' not Sightings
 values ('s2','Alice','Crow','06-14-08','Lake Placid')

insert into BELIEF 'Carol' not Sightings
 values ('s2','Alice','Raven','06-14-08','Lake Placid')



Bob

s

sid	uid	specie	date	location
s2	Alice	Raven	06-14-08	Lake Placid



Carol

s

sid	uid	species	date	location
s2	Alice	Crow	06-14-08	Lake Placid
s2	Alice	Raven	06-14-08	Lake Placid

⊖
⊖

2. Open world assumption



Alice

S

<u>sid</u>	uid	species	date
s2	Alice	Crow	06-



Bob

S

<u>sid</u>	uid	species	date
s2	Alice	Raven	06-



Carol

S

<u>sid</u>	uid	species	date	location
s2	Alice	Crow	06-14-08	Lake Placid
s2	Alice	Raven	06-14-08	Lake Placid




Q: Who disagrees with a sighting from '06-14-08' that Alice believes?

```
select U.name, S1.species
from Users as U,
BELIEF 'Alice' Sightings as S1,
BELIEF U.uid not Sightings as S2
where S1.sid = S2.sid
and S1.uid = S2.uid
and S1.species = S2.species
and S1.date = '06-14-08'
and S2.date = '06-14-08'
and S1.location = S2.location
```

A: {('Bob', 'Crow'), ('Carol', 'Crow')}

3. Higher-order beliefs



Alice

S

<u>sid</u>	uid	species	date	location
s2	Alice	Crow	06-14-08	Lake Placid




Bob

S

<u>sid</u>	uid	species	date	location
s2	Alice	Raven	06-14-08	Lake Placid

C

<u>cid</u>	comment	sid
c1	plain black feathers	s2




Bob Alice

C

<u>cid</u>	comment	sid
c1	purple-black feathers	s2

D3: Beliefs about other user's beliefs: allow discussion between users

3. Higher-order beliefs




Alice

S	<u>sid</u>	uid	species
	s2	Alice	Crow

I: According to Bob, Alice believes that the feathers of the sighted animal were plain black.


insert into **BELIEF 'Bob'** **BELIEF 'Alice'** Comments values ('c1', 'plain black feathers', 's2')



Bob

S	<u>sid</u>	uid	species		
	s2	Alice	Raven	06-14-08	Lake Placid

C	<u>cid</u>	comment	sid
	c1	plain black feathers	s2



Bob Alice

C	<u>cid</u>	comment	sid
	c1	purple-black feathers	s2

3. Higher-order beliefs



Alice

S

<u>sid</u>	uid	species
s2	Alice	Crow

Q: Which comments does Alice believe according to Bob, which he does not believe himself?

```

select  C1.cid, C1.comment
from    BELIEF 'Bob' BELIEF 'Alice' Comments as C1,
        BELIEF 'Bob' not Comments as C2
where   C1.cid = C2.cid
and     C1.comment = C2.comment
and     C1.sid = C2.sid
    
```



Bob

S

<u>sid</u>	uid	species
s2	Alice	Raven

C

<u>cid</u>	comment
c1	plain black feathers

A: {{('c1','plain-black feathers')}}



Bob

Alice

C

<u>cid</u>	comment	sid
c1	purple-black feathers	s2

3. Higher-order beliefs



Alice

S

sid	uid	species
s2	Alice	Crow

Q: Which comments does Alice believe according to somebody, which (s)he does not believe themselves?

```

select  U.name, C1.sid, C1.comment
from    Users as U,
        BELIEF U.uid BELIEF 'Alice' Comments as C1,
        BELIEF U.uid not Comments as C2
where   C1.cid = C2.cid
and     C1.comment = C2.comment
and     C1.sid = C2.sid
    
```

A: {'Bob', 'c1', 'plain-black feathers'}



Bob

S

sid	uid	species
s2	Alice	Raven

C

cid	comment
c1	plain black feathers



Bob



Alice

C

cid	comment	sid
c1	purple-black feathers	s2



4. Message board assumption



Alice

S

<u>sid</u>	uid	species	date	location
s2	Alice	Crow	06-14-08	Lake Placid



Bob

S

<u>sid</u>	uid	species	date	location
s2	Alice	Raven	06-14-08	Lake Placid

C

<u>cid</u>	comment	sid
c1	plain black feathers	s2



Bob

Alice

S


<u>sid</u>	uid	species	date	location
s2	Alice	Crow	06-14-08	Lake Placid

C

<u>cid</u>	comment	sid
c1	purple-black feathers	s2


D4: Default assumption: models a trusting attitude & avoids repeated inserts

4. Message board assumption



Alice


S	<u>sid</u>	uid	species
	s2	Alice	Crow



Bob

S	<u>sid</u>	uid	species
	s2	Alice	Raven

C	<u>cid</u>	comment
	c1	plain black feathers



Bob Alice

S	<u>sid</u>	uid	species		
	s2	Alice	Crow	06-14-08	Lake Placid

C	<u>cid</u>	comment	sid
	c1	purple-black feathers	s2

Q: Which animal sightings does Alice believe according to Bob, which he does not believe himself?

```

select  S1.sid, S1.species
from    BELIEF 'Bob' BELIEF 'Alice' Sightings as S1,
        BELIEF 'Bob' not Sightings as S2
where   S1.sid = S2.sid
and     S1.uid = S2.uid
and     S1.species = S2.species
and     S1.date = S2.date
and     S1.location = S2.location
    
```

A: {{'s2', 'Crow'}}

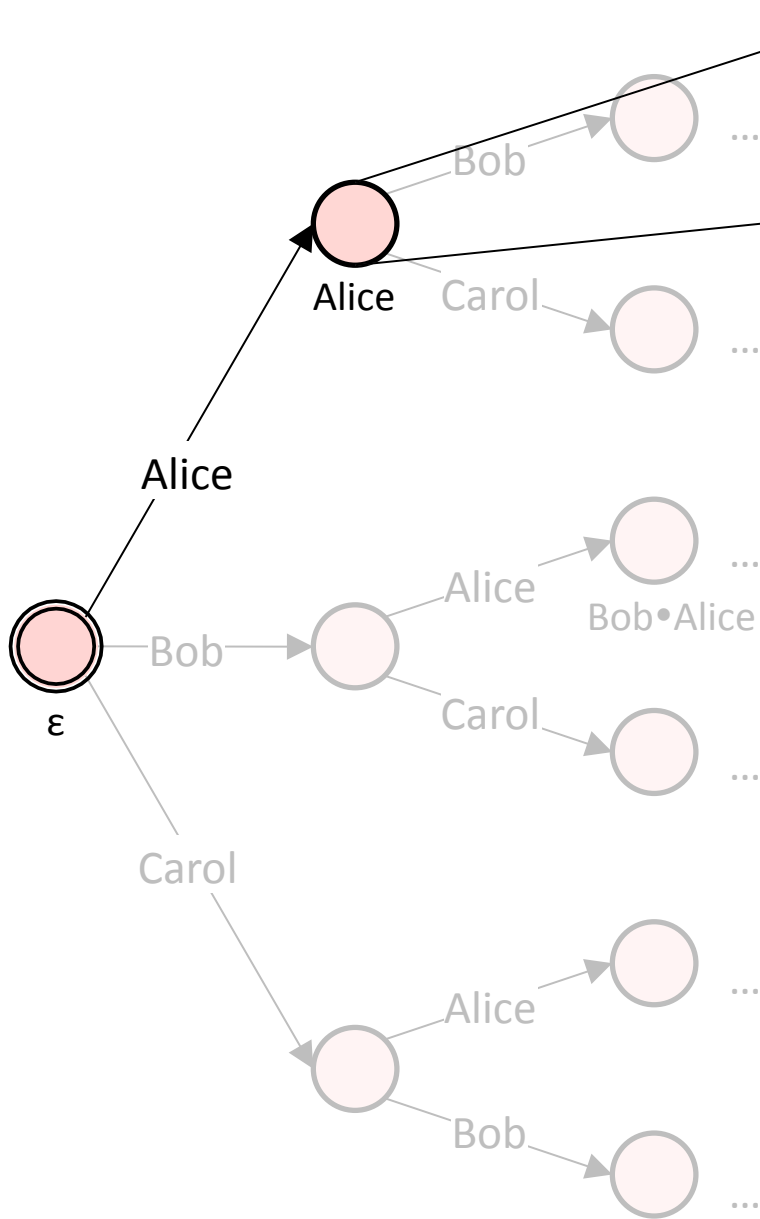
What we have seen so far

- 4 Design decisions for Belief Database model
 - Distinct belief worlds
 - Open world assumption (OWA)
 - Higher-order beliefs
 - Message board assumption
- **BeliefSQL**
 - SQL + 'BELIEF' (+ 'not')

Agenda

- Motivating example
- **Logic foundations**
- Relational implementation
- Discussion

Logic foundations: Belief statements



S

sid	uid	species	...
s2	Alice	Crow	...

insert into **BELIEF 'Alice'** S
values ('s2', 'Alice', 'Crow',...)

$i: \square_{\text{Alice}} S^+(\text{'s2', 'Alice', 'Crow', ...})$

modal operator
& belief path (w)

relational tuple (t)
sign (s)

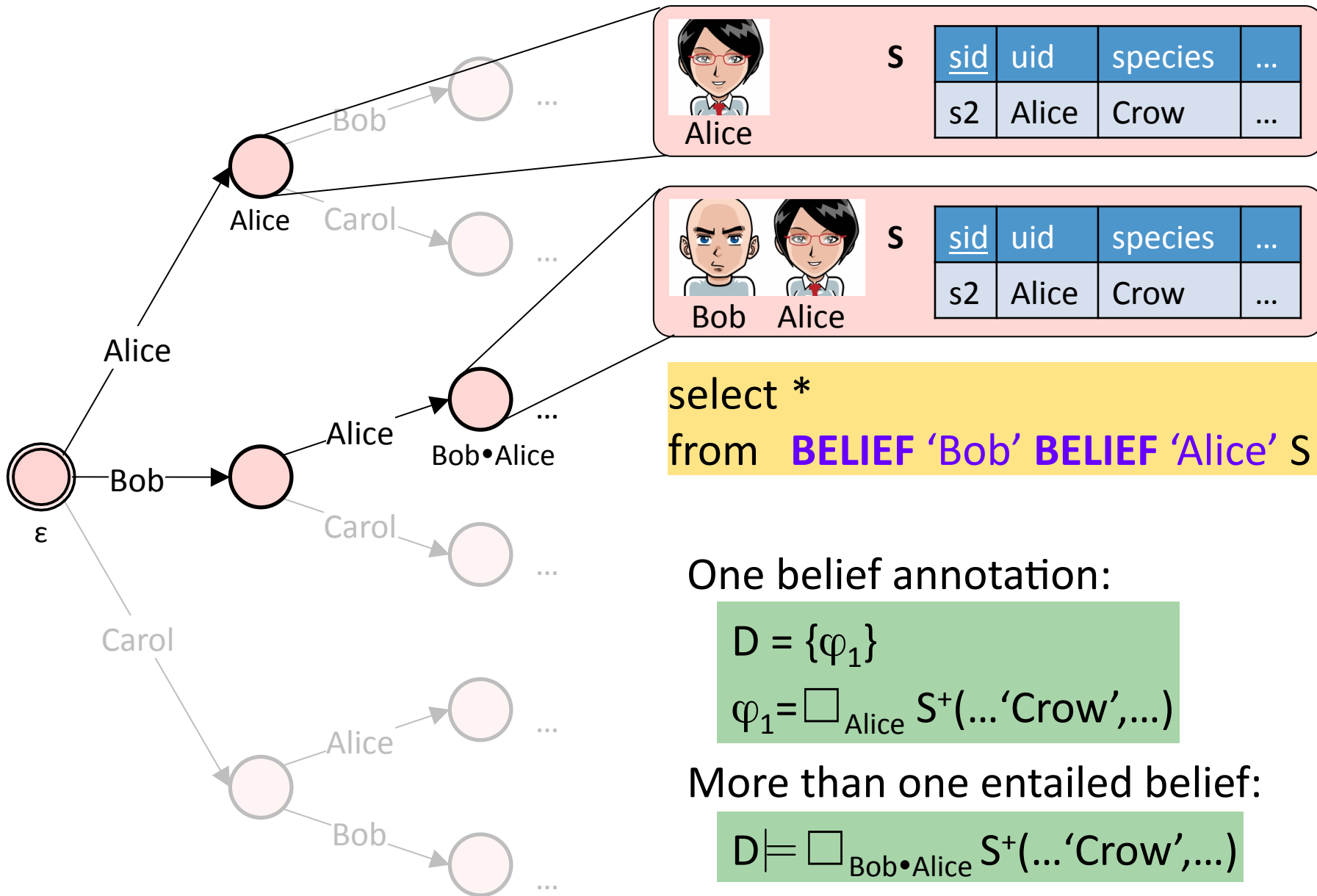
belief statement

$$\varphi = \square_w t^s$$

“annotation of
ground tuple”

Belief database $D = \{\varphi_1, \dots, \varphi_n\}$

Logic foundations: Entailment



Logic foundations: Message board assumption

Message board assumption

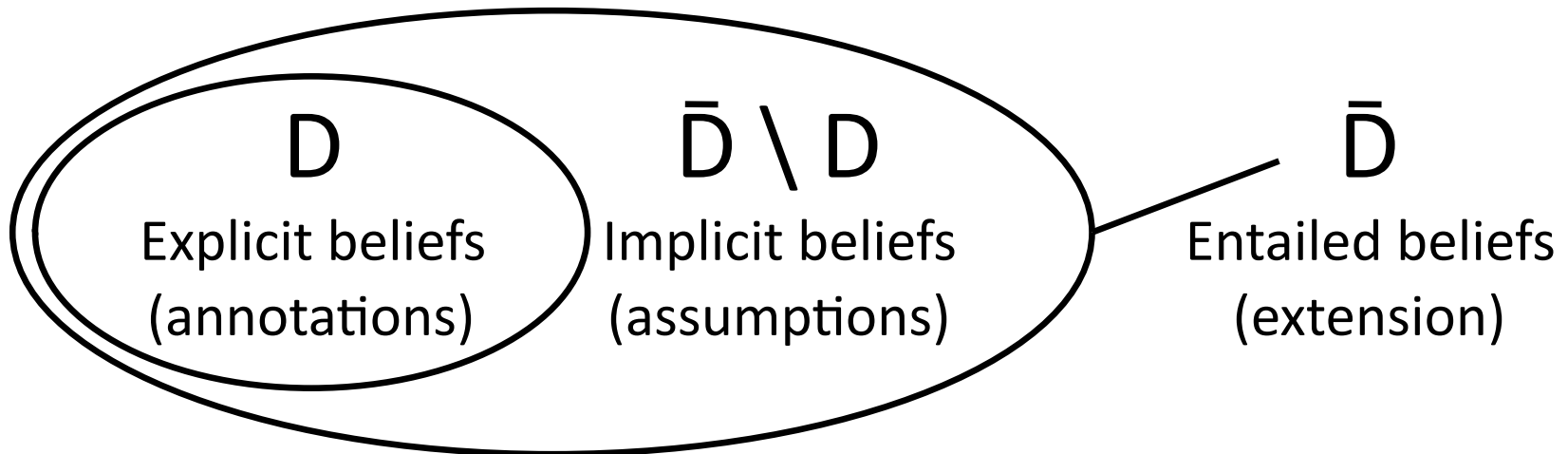
If $D \models \Box_w t^s$
and $\Box_{u \bullet w} t^s$ consistent with D
then $D \models \Box_{u \bullet w} t^s$



Default logic

$$\frac{\varphi : \Box_u \varphi}{\Box_u \varphi}$$

non-monotonic reasoning !



belief database “contains” more than the explicit belief annotations !

“Semi-formal” problem statement

INPUT

Belief statements

$$i_1: \varphi_1$$

$$i_2: \varphi_2$$

...

$$i_n: \varphi_n$$

Adapted key constraints

Message board assumption

$$\frac{\varphi : \Box_u \varphi}{\Box_u \varphi}$$

OUTPUT

$$D \models \varphi ?$$

$$D \models \Box_{w_1 \dots w_d} R^+(x_1, \dots, x_l) ?$$

$$q(\bar{x}) :- \Box_{\bar{w}} R_i^+(\bar{x}_i)$$

Belief Conjunctive Queries (BCQ)

$$q(\bar{x}) :- \Box_{\bar{w}_1} R_1^{s_1}(\bar{x}_1), \dots, \Box_{\bar{w}_g} R_g^{s_g}(\bar{x}_g)$$

Agenda

- Motivating example
- Logic foundations
- **Relational implementation**
- Discussion

Canonical Kripke structure

Belief statements*

$i_1: s1_1^+$

$i_2: \square_{\text{Bob}} s1_1^-$

$i_3: \square_{\text{Bob}} s1_2^-$

$i_4: \square_{\text{Alice}} s2_1^+$

$i_5: \square_{\text{Alice}} c1_1^+$

$i_6: \square_{\text{Bob}} s2_2^+$

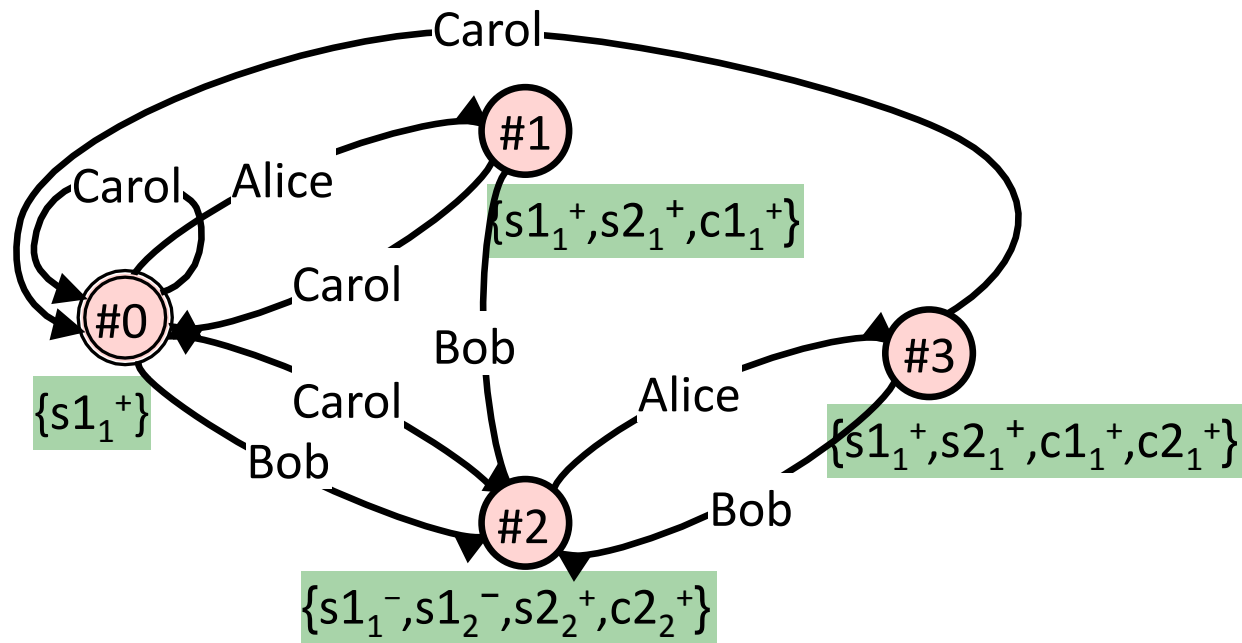
$i_7: \square_{\text{Bob} \cdot \text{Alice}} c2_1^+$

$i_8: \square_{\text{Bob}} c2_2^+$

Message board assumption

$\frac{\varphi : \square_i \varphi}{\square_i \varphi}$

$\square_i \varphi$



* Running example from the paper

Relational representation

Sightings_INTERNAL

tid	sid	uid	species	date	location
s1.1	s1	Carol	Bald eagle	06-14-08	Lake Forest
s1.2	s1	Carol	Fish eagle	06-14-08	Lake Forest
s2.1	s2	Alice	Crow	06-14-08	Lake Placid
s2.2	s2	Alice	Raven	06-14-08	Lake Placid

Comments_INTERNAL

tid	cid	comment	sid
c1.1	c1	found feathers	s2
c2.1	c2	plain black feathers	s2
c2.2	c2	purple-black feathers	s2

Sightings_V

wid	tid	sid	s	e
#0	s1.1	s1	+	y
#1	s1.1	s1	+	n
#1	s2.1	s2	+	y
#2	s1.1	s1	-	y
#2	s1.2	s1	-	y
#2	s2.2	s2	+	y
#3	s1.1	s1	+	n
#3	s2.1	s2	+	n

E

wid1	uid	wid2
#0	Alice	#1
#0	Bob	#2
#0	Carol	#0
#1	Bob	#2
#1	Carol	#0
#2	Alice	#3
#2	Carol	#0
#3	Bob	#2
#3	Carol	#0

Comments_V

wid	tid	cid	s	e
#1	c1.1	c1	+	y
#2	c2.2	c2	+	y
#3	c1.1	c1	+	n
#3	c2.1	c2	+	y

D

wid	d
#0	0
#1	1
#2	1
#3	2

S

wid1	wid2
#1	#0
#2	#0
#3	#1

Example Translation of a Belief CQ (BCQ)

Q: Who disagrees with a sighting from '06-14-08' that Alice believes?

BeliefSQL

```
select  U.name, S1.species
from    Users as U,
        BELIEF 'Alice' Sightings as S1,
        BELIEF U.uid not Sightings as S2
where   S1.sid = S2.sid
and     S1.uid = S2.uid
and     S1.species = S2.species
and     S1.date = '06-14-08'
and     S2.date = '06-14-08'
and     S1.location = S2.location
```

```
q(x,y) :- _Alice S+(u,v,y,'06-14-08',z),
          _x S-(u,v,y,'06-14-08',z)
```

Translation into SQL

```
select  E1.uid as uid1, V.tid, V.sid, R.uid, R.species, R.date, R.location, V.s
into    T2
from    E as E1, Sightings_V as V, Sightings_STAR as R
where   E1.wid1=0
and     V.wid=E1.wid2
and     V.tid=R.tid
and     E1.uid='1';

select  E1.uid as uid1, V.tid, V.sid, R.uid, R.species, R.date, R.location, V.s
into    T1
from    E as E1, Sightings_V as V, Sightings_STAR as R
where   E1.wid1=0
and     V.wid=E1.wid2
and     V.tid=R.tid;

select  T1.uid1, T2.species
from    T1 as T1, T2 as T2
where   T1.sid=T2.sid
and     ((T1.s=0 and T1.uid=T2.uid and T1.species=T2.species
and T1.date='6-14-08' and T1.location=T2.location) or
(T1.s=1 and (T1.uid<>T2.uid or T1.species<>T2.species
or T1.date<>'6-14-08' or T1.location<>T2.location)))

and     T2.s=1
and     T2.date='6-14-08';

drop   table T2;
drop   table T1;
```

Agenda

- Motivating example
- Logic foundations
- Relational implementation
- Discussion

Experiments

Size

Relative overhead $\rho := \frac{|R^*|}{n}$ $\rho = O(m^{d_{\max}})$

m ... #users
 d_{\max} ... maximum
depth of belief
annotation

In theory: e.g. 100 users, max. depth 2
 $\rho \rightarrow 10,000$

Experiments: $\rho \rightarrow 21 - 1,009$

Size not limitation of semantics, but of relational implementation!

Time

Depends on type of query (3 types in paper)

Q1: ~0.1 s

Experiments on 10,000 annotations ($\rho = 22.4$):

Q2: ~0.4 s

Q3: ~4.5 s

Considerable speed-up to come!

Inspirations and related work (excerpt)

- Annotations
 - Buneman et al. [ICDT 2001 / ICDT 2007]
 - Bhagwat et al. [VLDBJ 2005], Geerts et al. [ICDE 2006]
 - Srivastava & Velegrakis [SIGMOD 2007]
- Modal logic
 - Fagin et al. [1995]
 - Calvanese et al. [IS 2008]
 - Nguyen [LJ-IGPL 2008]
- Uncertain / incomplete information
 - Sarma et al. [ICDE 2006]
 - Green & Tannen [IEEE Data Eng. 2006]
 - Dalvi & Suciu [PODS 2007]
- Inconsistency / key violations
 - Arenas et al. [PODS 1999]
 - Fuxman et al. [SIGMOD 2005]
- Peer-to-peer computing / collaborative data sharing
 - Bernstein et al. [WebDB 2002]
 - Ives et al. [SIGMOD record 2008]

Conclusions

- Proposed BELIEF databases
 - Goal: manage, curate inconsistent data
- Implementation
 - Logical foundations
 - Relational translation
- Current work
 - making it compact and fast

BACKUP

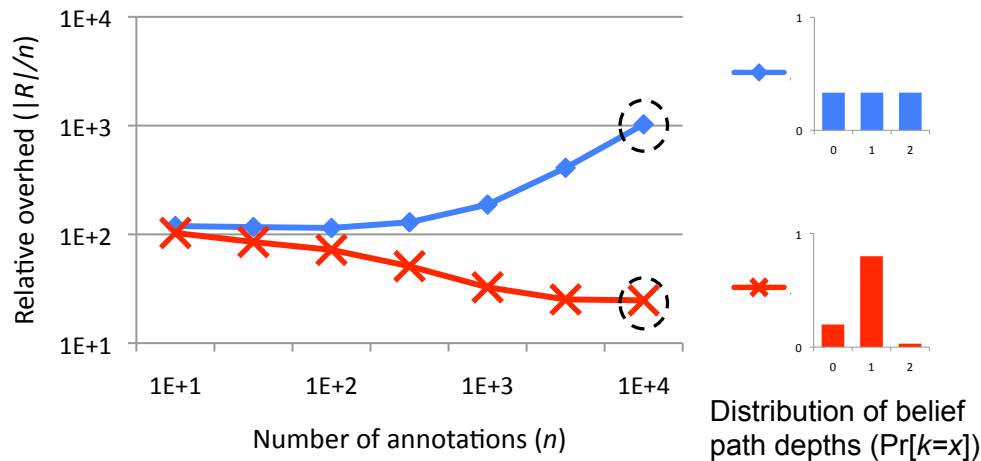
Relative overhead of relational representation

Bound for *relative overhead* $\frac{|\mathcal{R}^*|}{n} = \mathcal{O}(m^{d_{\max}})$

Measured relative overhead $\frac{|\mathcal{R}^*|}{n}$ for $n = 10,000$ annotations, $m = 100$ users, uniform or Zipf user participation, and 3 distributions of annotation depth:

$\Pr[d = \{0, 1, 2\}]$	uniform	Zipf
$[0.3, 0.3, 0.3]$	1,009	130
$[0.8, 0.19, 0.01]$	162	68
$[0.199, 0.8, 0.001]$	26	21

Measured relative overhead $\frac{|\mathcal{R}^*|}{n}$ for $m = 100$ users, uniform user participation, and 2 distributions of annotation depth:



Query types and execution times

1. *Query for content*: “What does Alice believe?” $d \in \{0, \dots, 4\}$:

$$q_{1,d}(x, y) : - \Box_w S^+(x, -, y, -, -), \text{ with } |w| \in \{0, \dots, 4\}$$

2. *Query for conflicts*: “Which animal sightings does Bob believe that Alice believes, which he does not believe himself?”

$$q_2(x, y) : - \Box_{2.1} S^+(x, z, y, u, v), \Box_2 S^-(x, z, y, u, v)$$

3. *Query for users*: Who disagrees with any of Alice’s beliefs of sightings at Lake Placid?”

$$q_3(x) : - \Box_x S^-(y, z, u, v, 'a'), \Box_1 S^+(y, z, u, v, 'a')$$

Execution times and size of result sets for example queries executed over a belief database with 10,000 annotations and relative overhead 22.4.

	$q_{1,0}$	$q_{1,1}$	$q_{1,2}$	$q_{1,3}$	$q_{1,4}$	q_2	q_3
E(Time) [msec]	105	145	146	152	144	436	4473
σ (Time) [msec]	120	168	153	162	162	186	661
Result size	1626	2816	2253	2061	1931	196	99

Belief Conjunctive Queries (BCQ)

Conjunctive Queries (CQ) in Datalog form:

$$q(\bar{x}) :- R_1(\bar{x}_1), \dots, R_g(\bar{x}_g)$$

Belief Conjunctive Queries (BCQ) in "Modal Datalog" form:

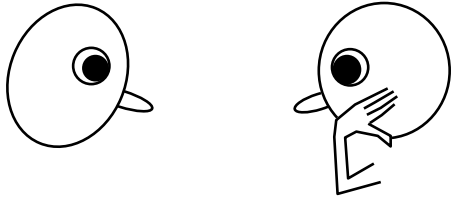
$$q(\bar{x}) :- \Box_{\bar{w}_1} R_1^{s_1}(\bar{x}_1), \dots, \Box_{\bar{w}_g} R_g^{s_g}(\bar{x}_g)$$

q_1 : “*Who disagrees with any sighting from '06-14-08' that Alice believes?*”

$$q_1(x, y) :- \Box_{\text{Alice}} S^+(u, v, y, \text{'06-14-08'}, z), \Box_x S^-(u, v, y, \text{'06-14-08'}, z)$$

$$q_1(D) = \{(\text{'Bob'}, \text{'bald eagle'}), (\text{'Bob'}, \text{'crow'})\}$$

Revisiting the semantics / the user



↓ (3) ?

-> Structured discourse

↓ (2) BeliefSQL

Conflicts in belief worlds:
OWA, keys, ML, DA

↓ (1) SQL

Standard relational model

BELIEF 'Alice' (...,'eagle',...)

-> 'Alice'ASSERTS (...,'eagle',...)

BELIEF 'Bob' BELIEF 'Alice'
(...,'black feathers',...)

-> 'Bob'SUGGESTS that the ASSUMPTION
(...,'black feathers',...) has led 'Alice' to her
original observation