Fall 2013

ER MODELING
## [CH 2: SECTIONS 2.1-2.4.2 AND CH3: SECTIONS 3.5.1 - 3.5.4]

Database Design

- Requirements Analysis
  - Data stored, operations, apps, …

- Conceptual Database Design
  - Model high-level description of the data, constraints, ER model

- Logical Database Design
  - Choose a DBMS and design a database schema

- Schema Refinement

- Physical Database Design

- Application and Security Design

ER Model Basics

- *Entity:* Distinguishable real-world object
  - Described by a set of *attributes,* Each attribute has a *domain*

- *Entity Set:* A collection of similar entities. E.g., all citizens.
  - All entities in an entity set have the same set of attributes. (Until we consider ISA hierarchies!)
  - *Key* : minimal set of attributes whose values uniquely identify an entity in an entity set
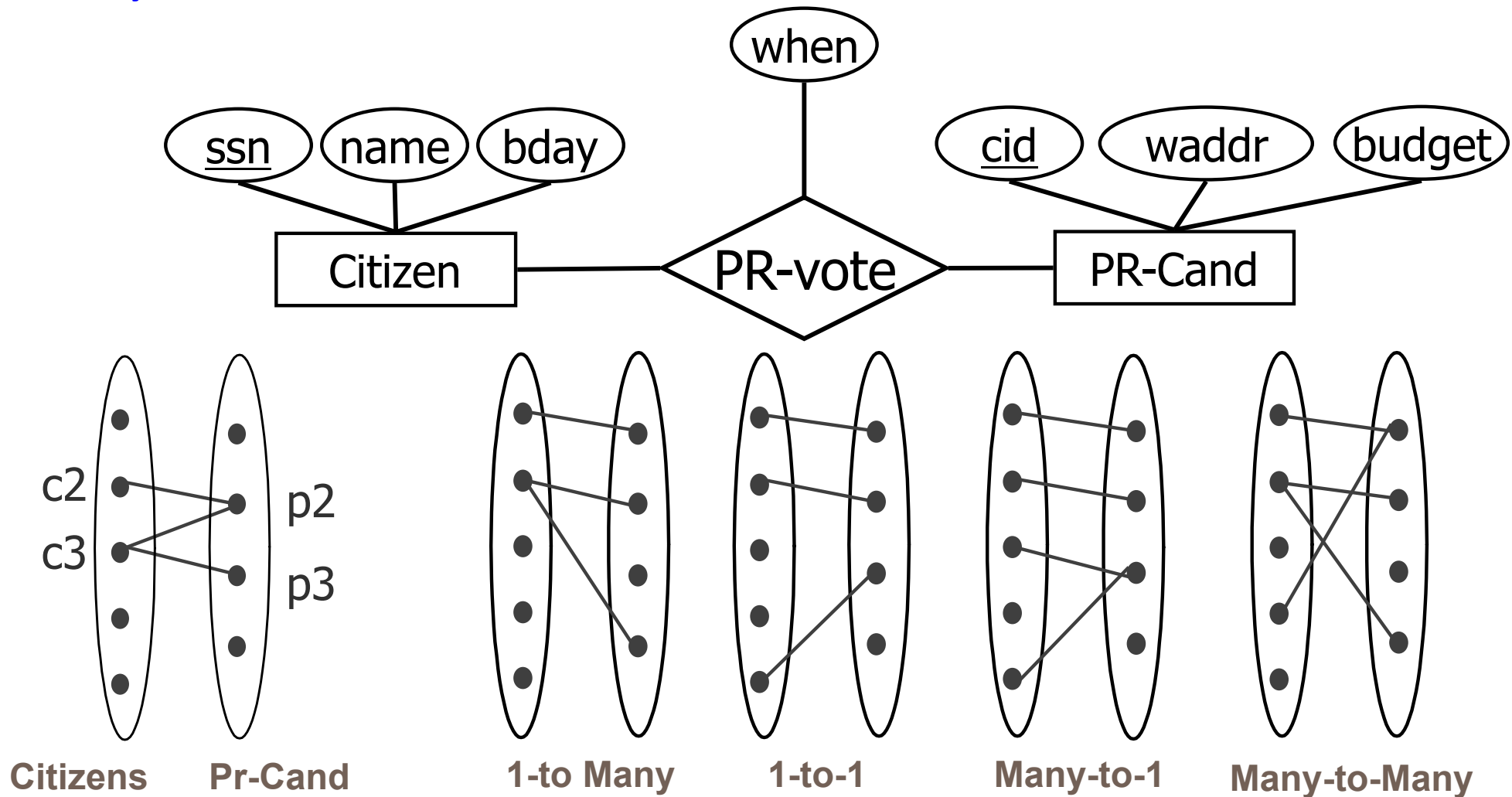   - Primary key
    - Candidate key

- Pictorially …
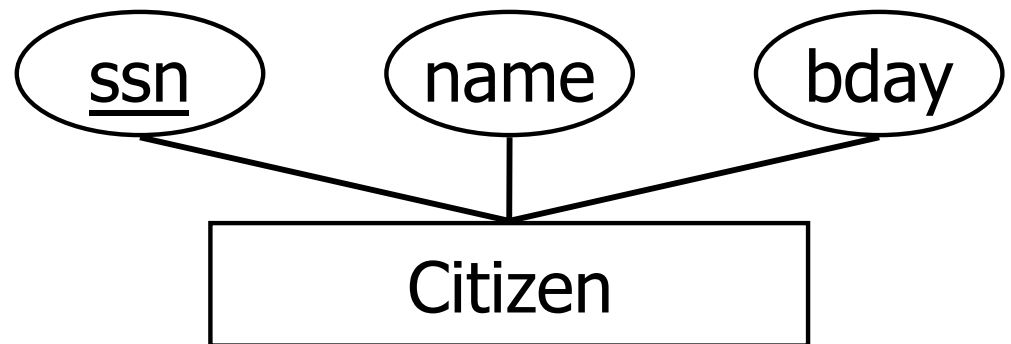
- *Relationship* : Association among two or more entities
- *Relationship Set* : Collection of similar relationships
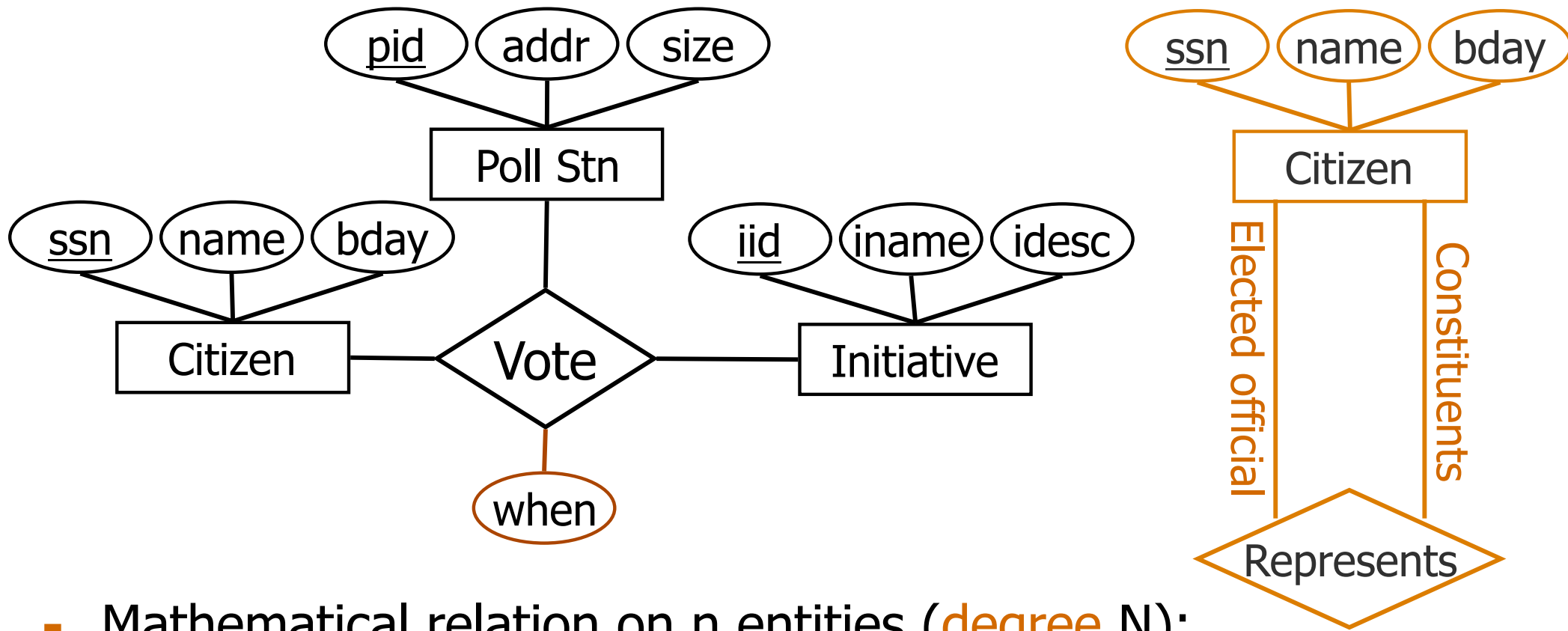
- Mathematical relation on n entities (degree N):

$$\{(e_1, e_2, ..., e_n) \mid e_1 \in E_1, e_2 \in E_2, ..., e_n \in E_n\}$$

- Entity sets:
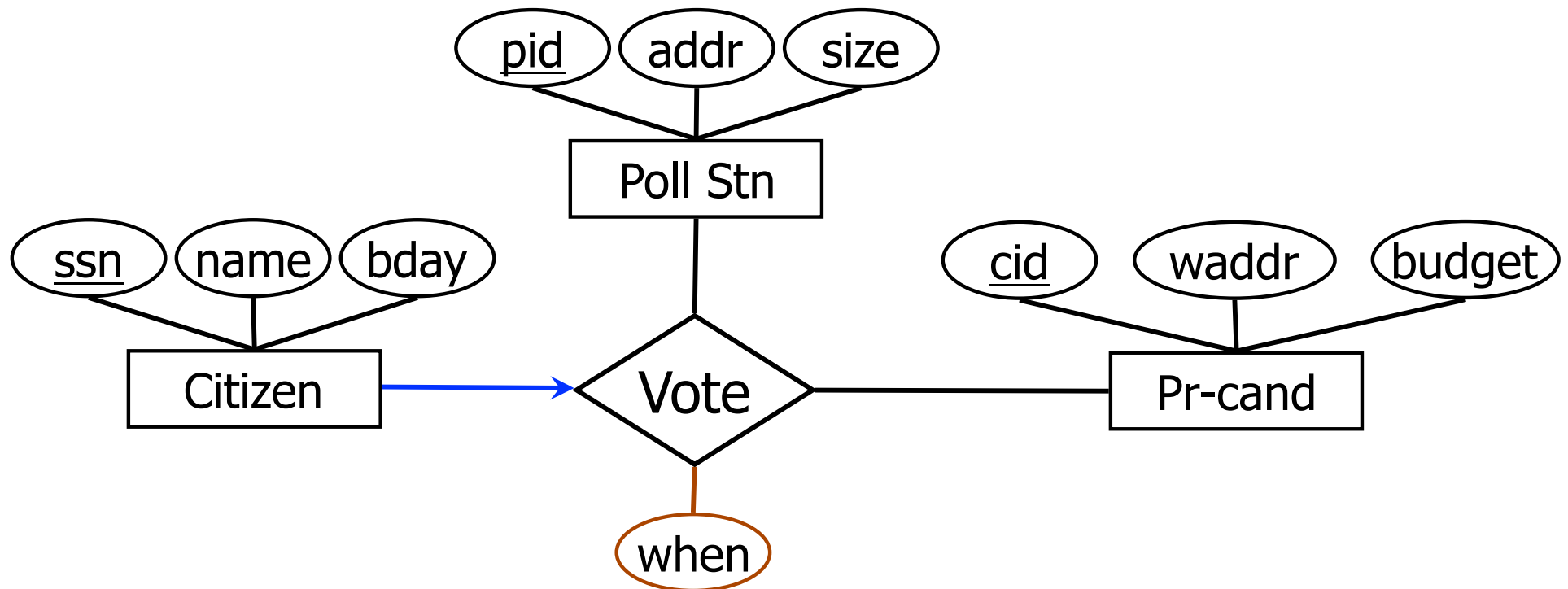 - Can participate in > 1 relationship sets in different "roles".

Key Constraints

*Key Constraint* : Each citizen votes at most once

Key Constraints: Generalize

Each voter votes at most once (for one candidate) and
at a single location

Participation Constraints

- Key Constraint: A citizen has a single vote

- Q: Must every citizen vote?

  - This is a *participation constraint* : Every citizen must participate (*total* vs. *partial*).

Every party must have exactly one candidate?
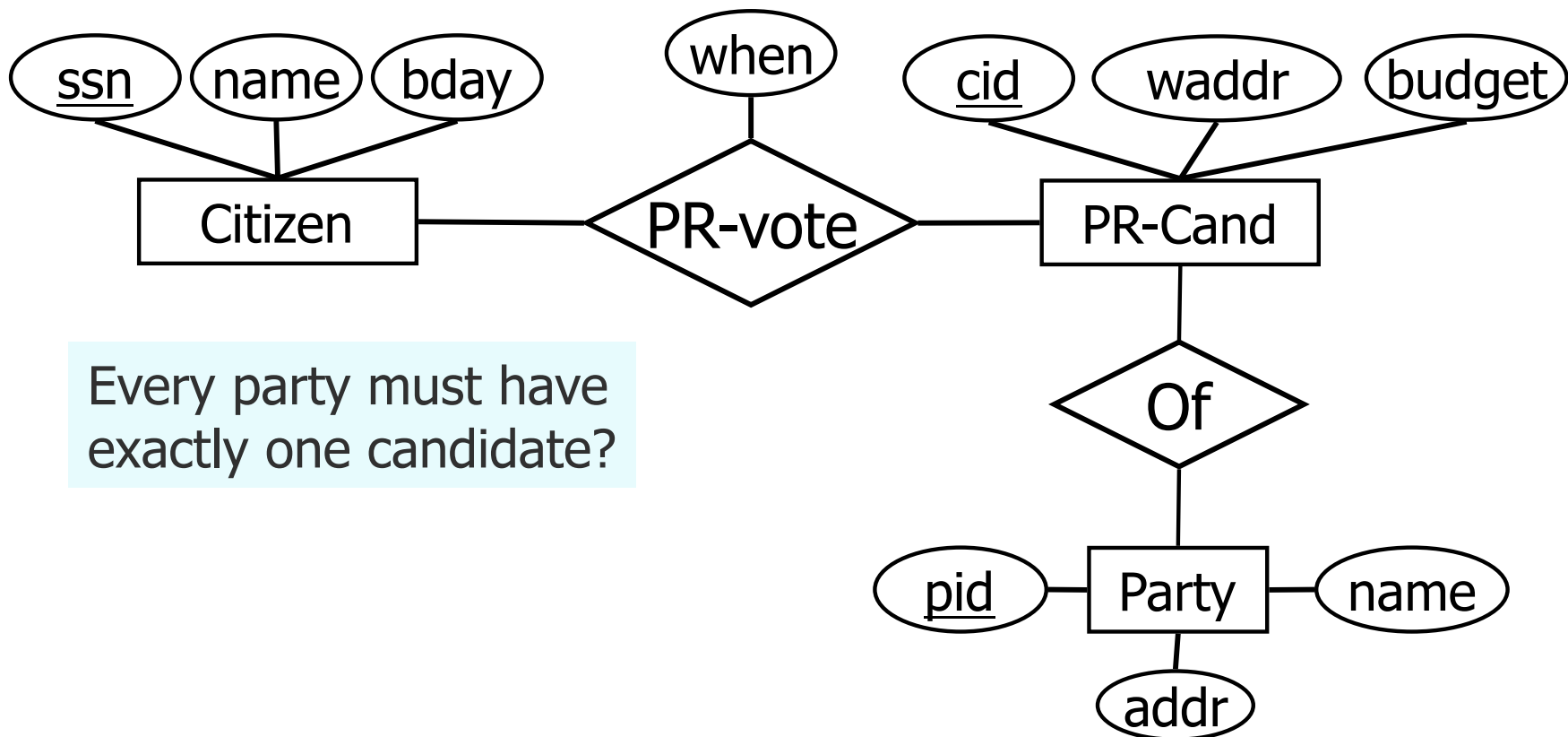
Other Constructs in the ER model

- Weak Entities: an entity that is dependent on another entity

- Hierarchies: Model "IS A" hierarchies.

- Aggregation: Make a relationship act like an entity when participating in another relationship.

See the text book if you are interested in these advanced constructs
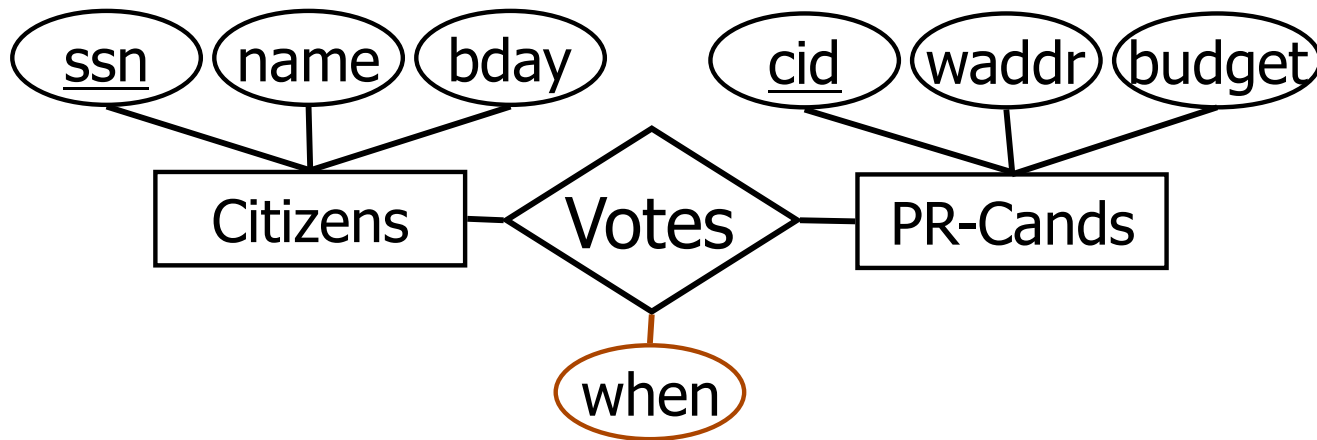
Next Step

- Need to map the ER diagram to SQL DDL statements

Logical DB Design: ER to Relational

ssn name bday
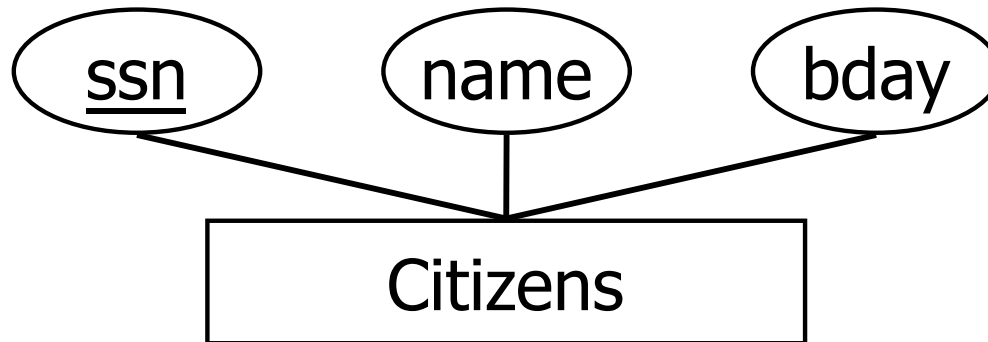
Citizens

```
CREATE TABLE Citizens
      (ssn   CHAR(11),
       name  CHAR(20),
       bday  DATE,
       PRIMARY KEY (ssn))
```

Relationship Sets to Tables

Relationship set -> Table

Attributes:
- Participating entity set primary keys
  - Foreign key
  - Superkey
- Descriptive attributes
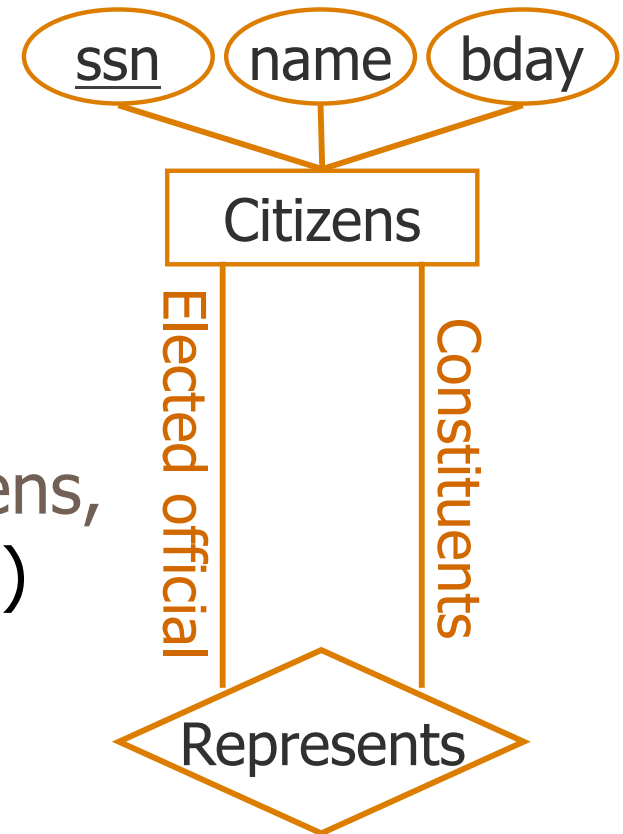
```
CREATE TABLE Votes(
       ssn      CHAR(11),
       cid      INTEGER,
       when  DATE,
PRIMARY KEY (ssn, cid),
FOREIGN KEY (ssn) REFERENCES Citizens,
FOREIGN KEY (cid) REFERENCES PR-Cands)
```

Can ssn have a null value?

Can generalize to n-ary relationships

Relationship Sets to Tables

CREATE TABLE Represents(
    elected_ssn CHAR(11),
    cons_ssn CHAR(11),
PRIMARY KEY (elected_ssn, cons_ssn),
FOREIGN KEY (elected_ssn) REFERENCES Citizens,
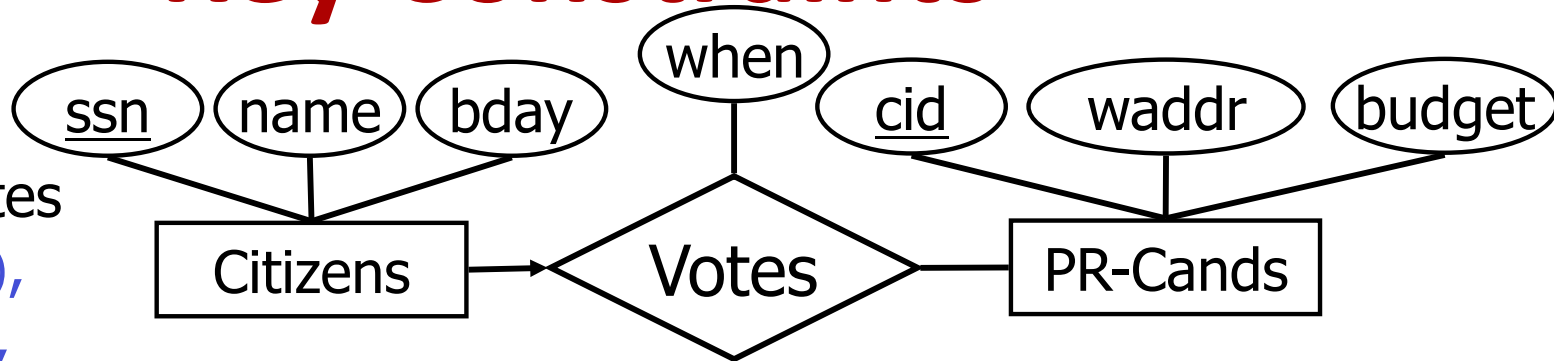FOREIGN KEY (cons_ssn) REFERENCES Citizens)

ssn   name   bday

Citizens

Elected official

Constituents

Represents

Key Constraints: Review

Key Constraint : Each citizen votes at most once
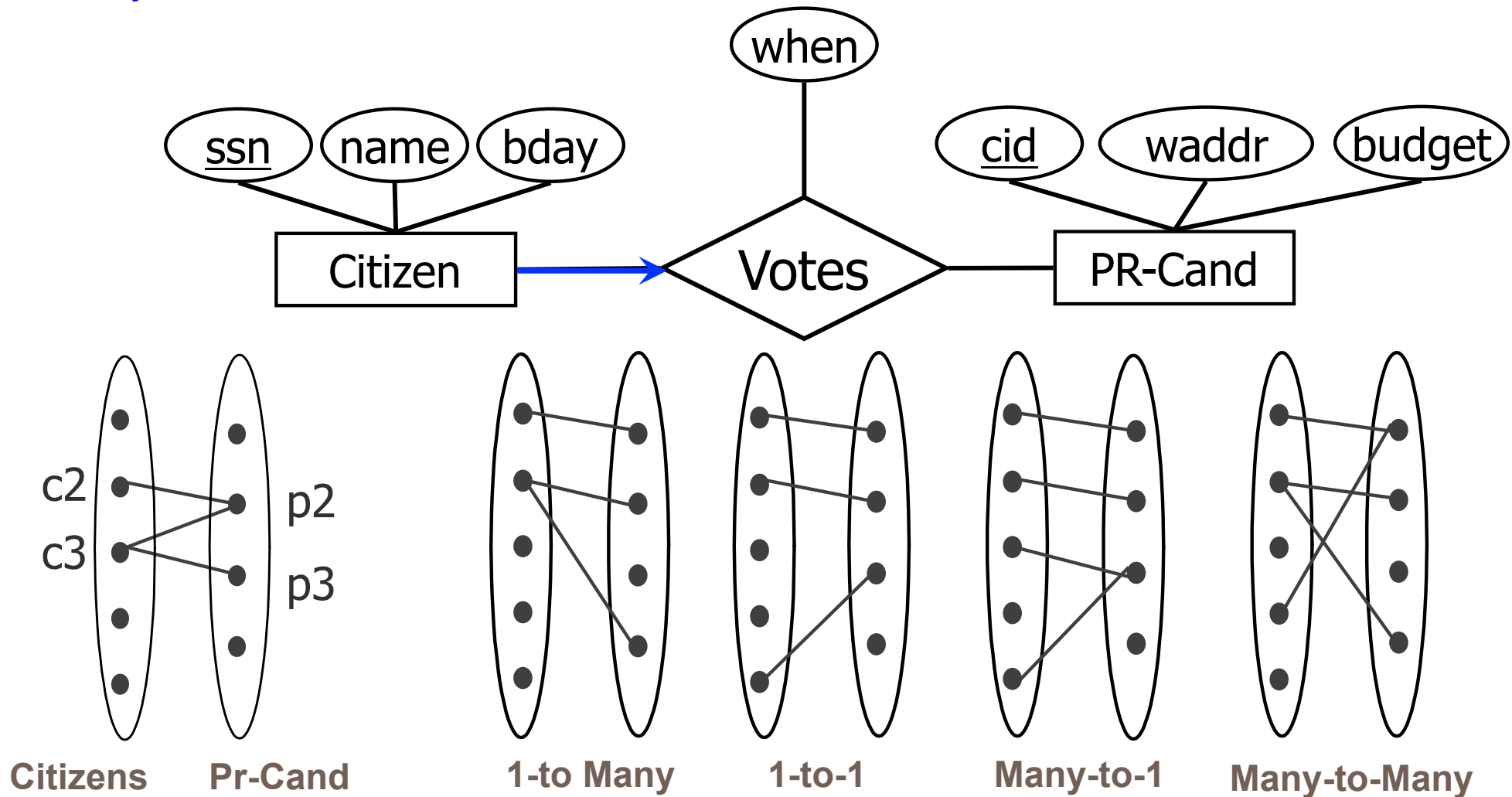
Key Constraints

CREATE TABLE Votes

(ssn   CHAR(11),
   cid   INTEGER,
   when   DATE,
   PRIMARY KEY (ssn),
      FOREIGN KEY (ssn) REFERENCES Citizens,
      FOREIGN KEY (cid) REFERENCES PR-Cands)

⇐ Approach 1: Three tables

CREATE TABLE Citizen_Votes (
   ssn   CHAR(11), name   CHAR(20),
   bday   DATE,   when DATE,
   cid   INTEGER,
PRIMARY KEY (ssn),
FOREIGN KEY (cid) REFERENCES PR-Cands)

Approach 2: Two tables!
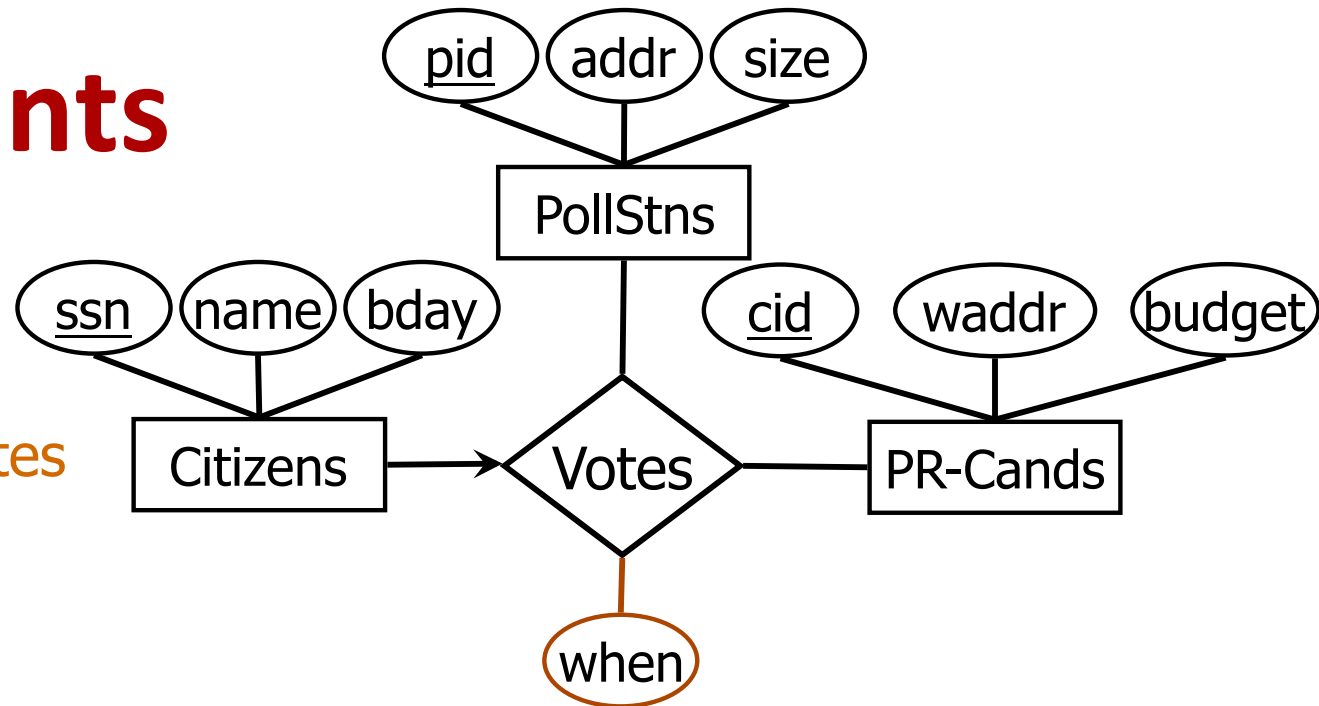 ▪ Fold into Citizens.
Q: Can cid be null?
Q: What if many citizens don't vote
Q: Which approach is better?

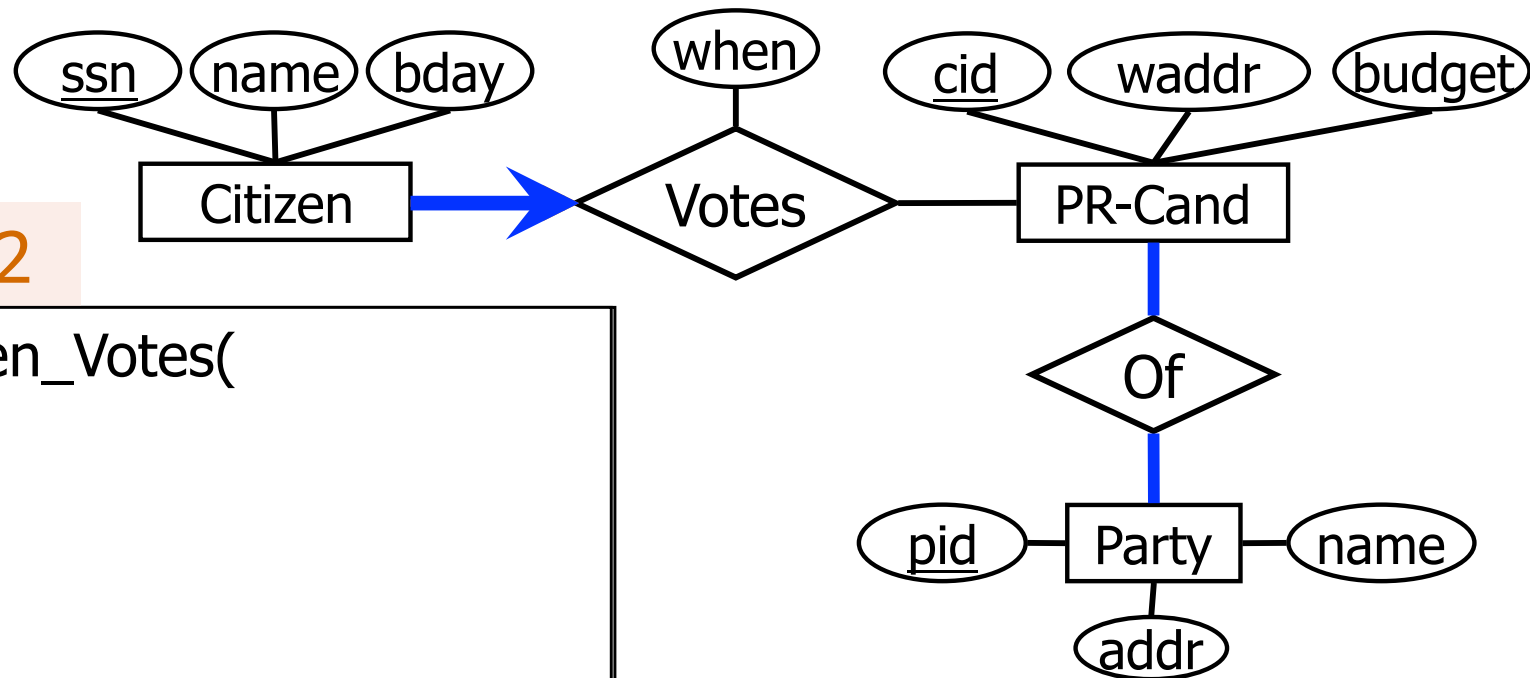Can generalize to n-ary relationships

Key Constraints

CREATE TABLE Citizens_Votes
(ssn   CHAR(11),
    name   CHAR(20),
    bday DATE,
    when   DATE,
    pid   INTEGER,
    cid   INTEGER,
PRIMARY KEY (ssn),
FOREIGN KEY (cid) REFERENCES PR-Cands,
FOREIGN KEY (pid) REFERENCES PollStns)

(Total) Participation Constraints

Using Approach 2

```
CREATE TABLE  Citizen_Votes(

   ssn     CHAR(11),
   name    CHAR(20),
   bday    DATE,
   when    DATE,
   cid     INTEGER NOT NULL,
PRIMARY KEY  (ssn),
FOREIGN KEY  (cid) REFERENCES PR_Cands,
     ON DELETE NO ACTION)
```
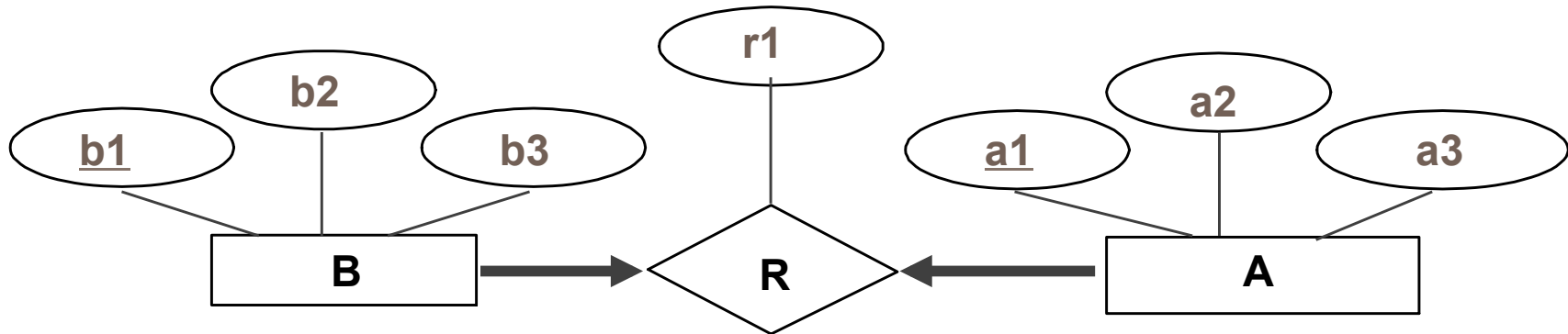
Approach 1?
- **Make ssn and cid not null**
- **ssn Fkey in Citizens pointing to Votes**

- **Participation constraint on OF**
  - Use Table constraints and assertions: Expressive but expensive!

Mapping Participating Constraints


```
CREATE TABLE  RAB(
        r1 Integer,
        a1 Integer,
        a2 Integer,
        a3 Integer,
        b1 Integer,
        b2 Integer,
        b3 Integer …)
```

Key constraints?

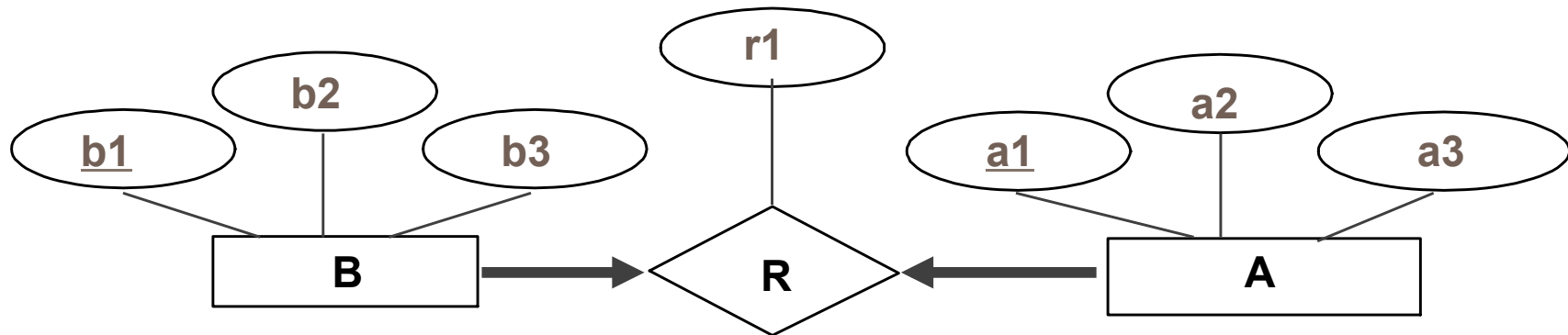# Mapping Participating Constraints


```
CREATE TABLE  RAB(
        r1 Integer,
        a1 Integer,
        a2 Integer,
        a3 Integer,
        b1 Integer NOT NULL,
        b2 Integer,
        b3 Integer,
        UNIQUE (b1), PRIMARY KEY (a1))
```

More modeling

- There are well-known techniques to map the advanced ER techniques (e.g. aggregation). See the book if you are interested.

- There are more than one ways to draw the ER diagram. A popular model is the "Crow's Foot Notation."

- Often the diagrams are generated using tools, and initial mappings are produced in automated ways.