COMP-421
The Entity-Relationship Model

Database:
collection of
structured,
mission-critical
data of an
application domain

model
needs
to be
correct

Steps in Database Design

1) **Requirement Analysis**
   - Identify the data that needs to be stored
     - data requirements
   - Identify the operations that need to be executed on the data
     - functional requirements
2) **Conceptual Database Design**
   - **Use a semantic data model to develop semantic schema**
3) Map semantic schema to relational schema

Step 1 Requirement Analysis

- Similar to first phase of general software design
- focus: data & functional requirements
- Questions to ask:
  - What are the **entities** and **relationships** in the application?
  - What information about these entities and relationships should we store in the database?
  - What are the integrity constraints or business rules that hold?
  - What operations do we want to execute on the entities and relationships?
- Tools known from SE
  - data-flow diagrams, sequence diagrams…
- In this course:
  - half-formal, structured specification in plain English

eg
- who can see what data

Requirement Analysis: Minerva/Banner University Database

- The database captures crucial information about university objects and subjects such as students, employees, financial data, courses and much more.
- The data can be accessed via interfaces that allow
  - students access to registration, address changes, fee assessment, transcripts;
  - Instructors access to enrollment information, grade input
  - Principal researchers access to grant information, travel reimbursements, …

entities

who can do what

Requirement Analysis: Student Database

- The database contains information about *persons:*
  - Data:
    - each person has an identifying ID
    - each person has a PIN, name, permanent address + phone number, 0 or more emergency contacts
  - Functionality
    - all data can be viewed and changed
- A person can be a *student*. A student has
  - Data:
    - …
  - Functionality:
    - Register for course…
- A person can be an instructor.
  - Functionality:
    - Can teach a course
    - Can enter grades

Student Database (contd)

- A course
  - Data
    - Term: the term the course takes place
    - course id (special format)
    - CRN
    - type
    - Credits: how many credits; typically 1-4
    - title
    - days
    - time
    - capacity
    - enrollment
    - one or more *instructors*
    - *room*
    - …
  - Some properties are specific for one term; others always hold for this course

presented as table one minerva
not a table in DB

consider:
change data
for one year
but want to
keep the same
for past years

prof …

title, course code …

Student Database (contd)

- A course
  - Functionality:
    - Students can register
    - Course can be cancelled
    - Course can be assigned instructor

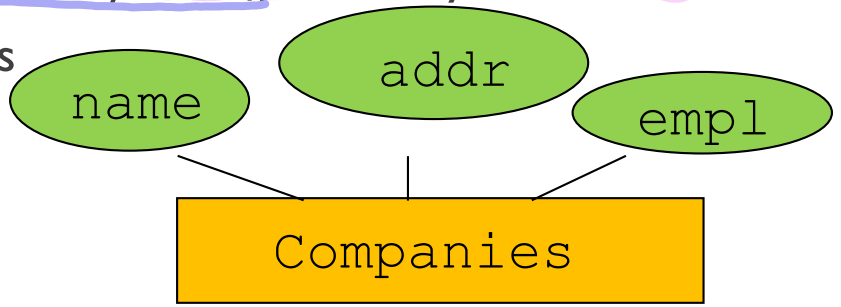
Semantic Data Model: Entity-Relationship Model (ER)

- The *E/R model* is a language that allows for a pictorially description of the data determined through the requirement analysis
  - Conceptually similar to UML class diagrams
- An *E/R schema* is a representation of the structure of the data of an application using the E/R model
- An ER schema should be understandable by non-computer scientists.
- The main concepts of the E/R model are entity sets that group entities and relationship sets that group relationships
- An ER schema can be translated into the relational schema in a quite straightforward way.
- Many dialects
  - We use the one from the book….

Entity

- **Entity**: Real-world object distinguishable from other objects. An entity is described using a set of **attributes.**
  - similar to an object/instance in OO
- **Entity Set:** A collection of similar entities, e.g., all companies registered in Quebec (similarity to a "class" in OO)
  - All entities in an entity set have the same attributes (until we consider ISA hierarchies later…). An attribute describes a property of the entity set.
  - An entity set must have a **key:** (underlined)
    - Minimum set of attributes whose values *uniquely identify* an entity in the set
    - What would be a good key for Companies
    - Often: artificial key — no meaning, just id

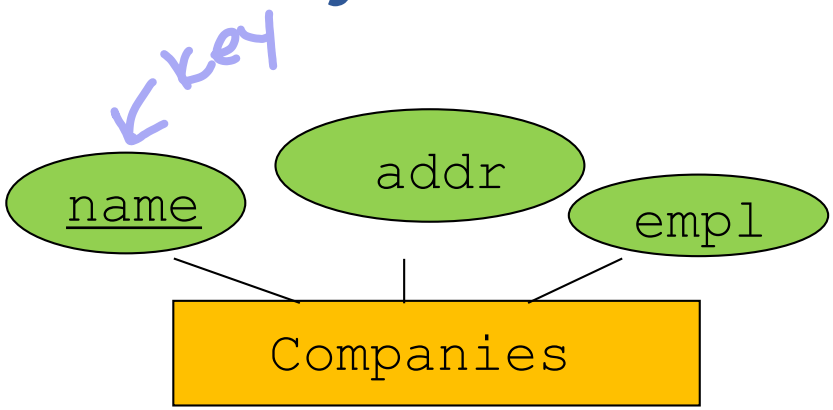
Something unique



Entity set = rectangle, Attribute = oval

Key: use attribute if possible and appropriate

Basic Entity Sets → Tables



entities → rows

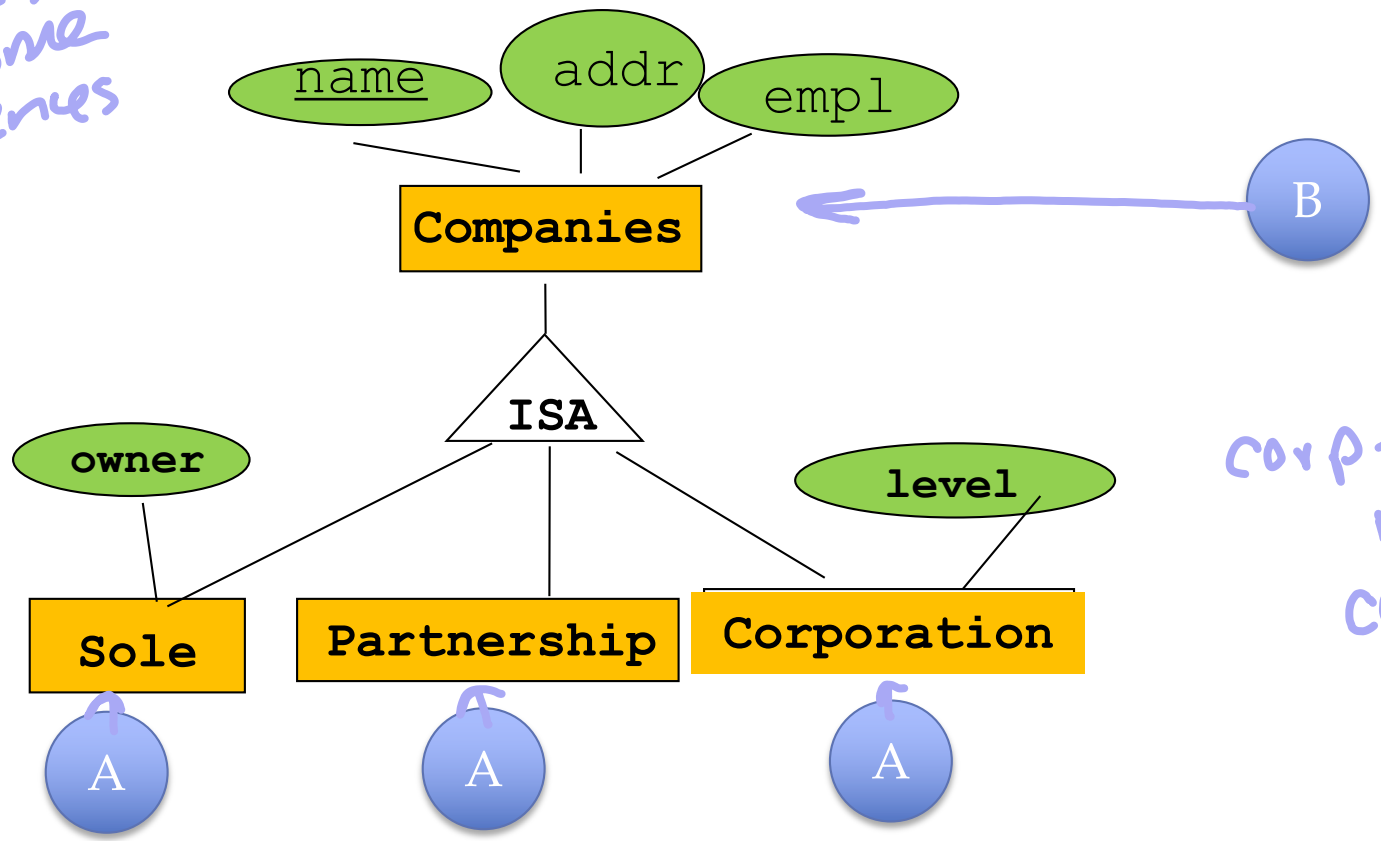
attributes → column

name	addr	empl
BiggestEngCompanyEver	Eng. Av., H3X…	25,000
BiggestConstCommpanyEver	Constr. St. H4E…	47,000
NoNameCompany	Whatever St., …	200
…		

ISA ("is a") Hierarchies: Subclasses

- Subclass = special case = fewer entities = more properties
- "A ISA B", then every A entity is also a B entity
  - Key only in B.

Set of common attributes but some differences



Corporation is a company

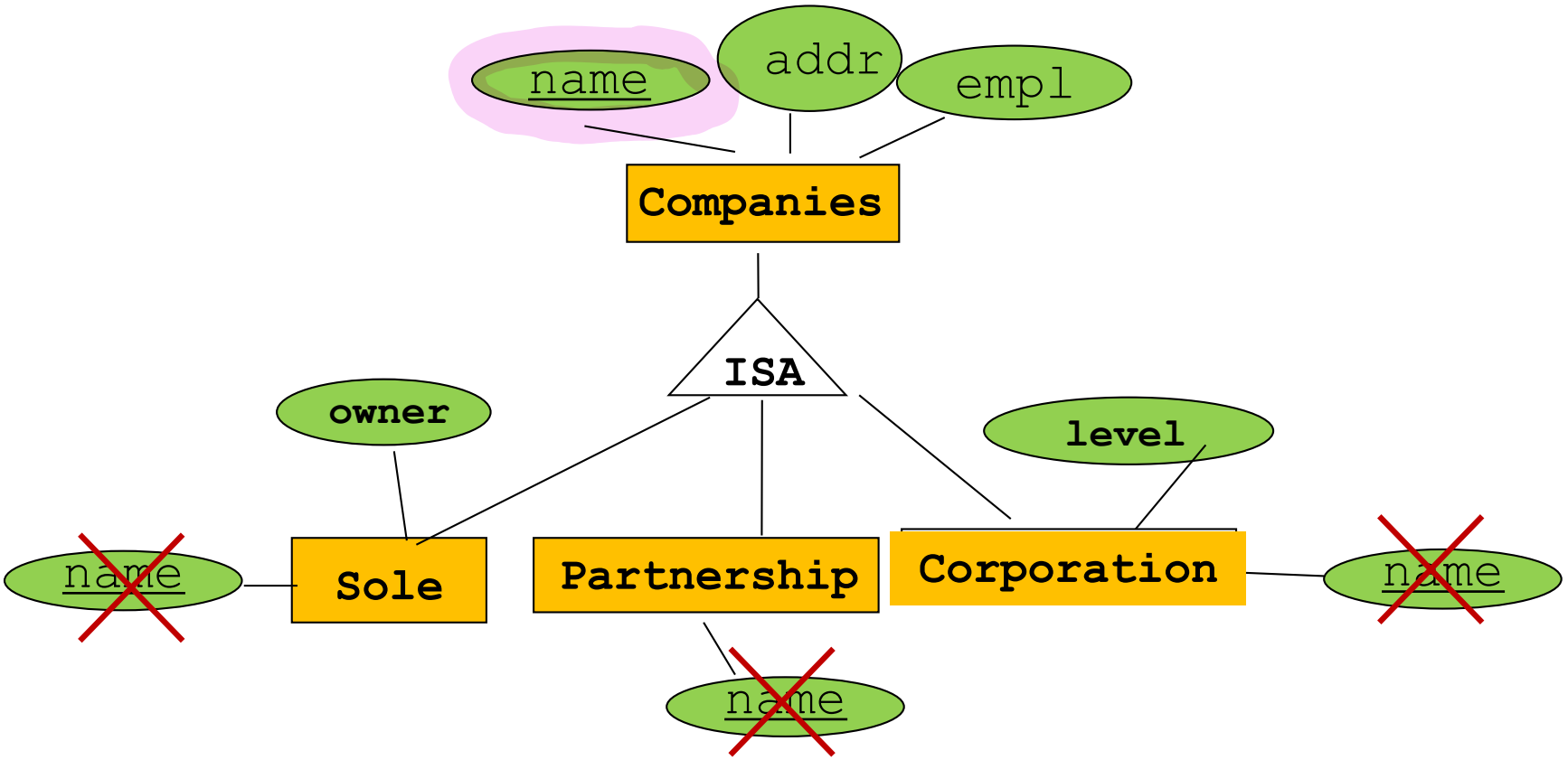
ISA ("is a") Hierarchies: Subclasses

- Reasons for using Subclasses:
  - Additional descriptive attributes specific for a subclass
  - Identification of a subclass that participates in a relationship (will see later)
- Where do the entities reside:
  - E/R-viewpoint:: An entity has a component in each entity set to which it logically belongs. Its attributes are the union of these entity sets.
  - Contrast OO-viewpoint: An object belongs to exactly one class. The subclass inherits the attributes of its superclass.

belongs to both entities

ISA ("is a") Hierarchies: Keys

- Only the highest entity sets has the key attribute!!



Q: can you have overlap of sub-entities
↳ overlapping hierarchy

entity & instance

ISA Hierarchies (Contd.)

- **Overlap Constraint**: Can an entity be in more than one subclass? (allowed/disallowed)
- **Covering Constraint**: Must every entity of the superclass be in one of the subclasses? (yes/no)

don't need a subclass w/ no info

- Developing Class Hierarchies
  - **Specialization**: Top-down approach
    - The superclass is there first
    - Then identify subsets of an entity set (the superclass) that share some distinguishing characteristic (special attributes / relationship).

Find unique

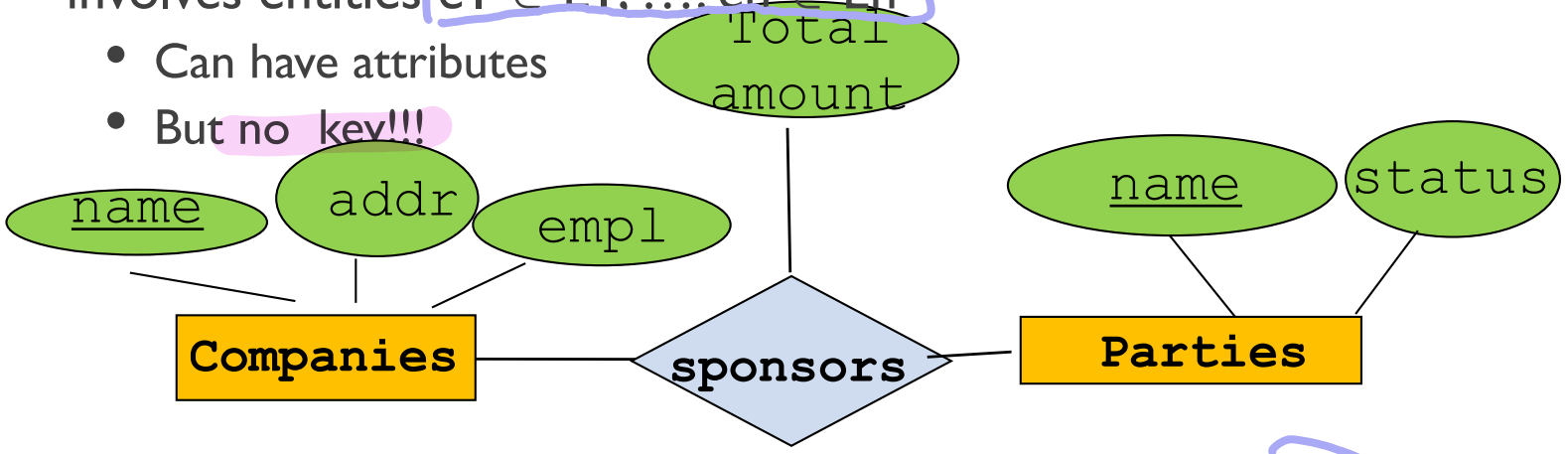
  - **Generalization**: Bottom-up approach
    - Several entity sets are generalized into a common entity set
    - common characteristics of a collection of entity sets are identified, a superclass entity set is built with these characteristics as attributes.

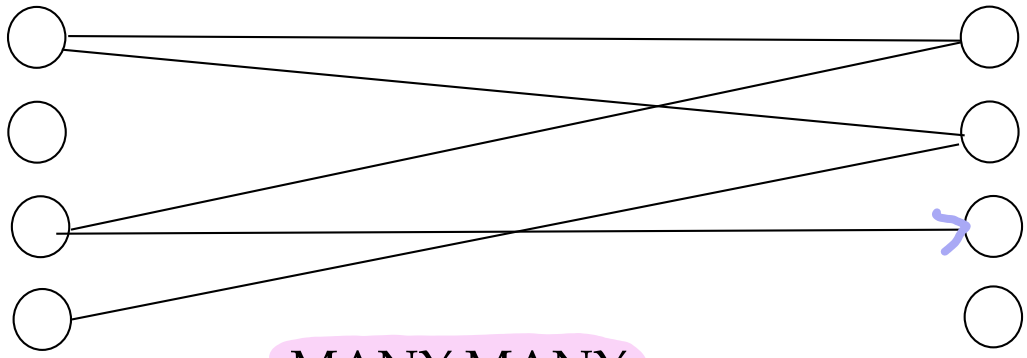
Find common

- Theoretically multiple inheritance possible, in practice not used

Relationship

- **Relationship:** Association among two or more entities. E.g. Company X has sponsored the Liberal Party (with a total amount of ….)
- **Relationship Set:** Collection of similar relationships
  - An n-ary relationship set R relates n entity sets E1…En; each relationship in R involves entities e1 ∈ E1, …, en ∈ En
    - Can have attributes
    - But no key!!!





0/1/+ connections

all in one sponser set

MANY-MANY

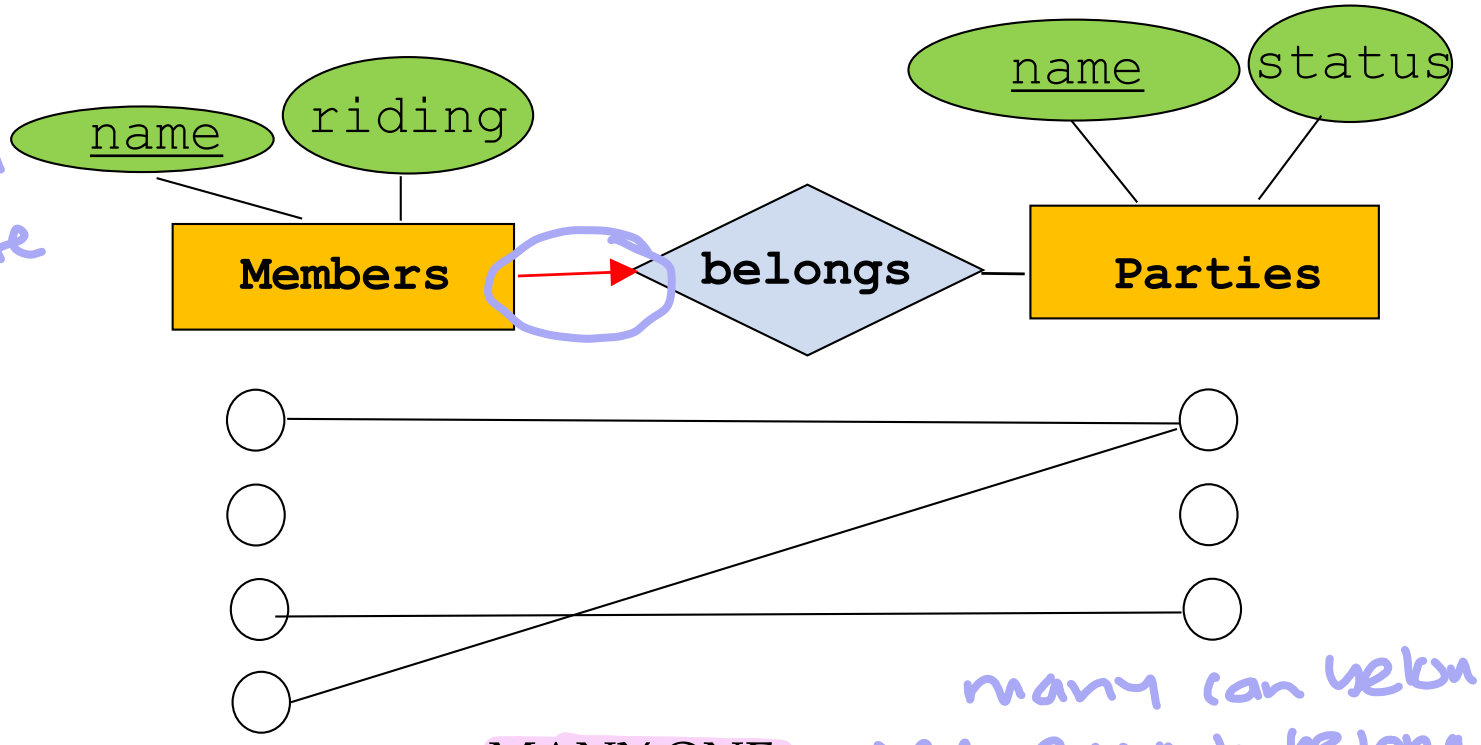
directed

Many-many

- Each entity of `Companies` can have relationships with many entities from `Parties`
  - Company *BiggestEngCompanyEver* sponsors *Liberals* and *PQ*
- Each entity of `Parties` can have relationships with many entities from `Companies`
  - The *Liberals* are sponsored by *BiggestEngCompanyEver* and *BiggestConstCompanyEver*
- Each relationship is uniquely defined by the primary keys of participating entities
  - *BiggestEngCompanyEver* cannot sponsor twice the *Liberals*
  - If we wanted to keep track of each sponsorship transaction, we would have to model this differently….

Key-constraints: one-many, many-one

- Belongs Relationship Set:
  - Members (of the National Assembly) and (Political) Parties
  - A member of the national assembly can belong to at most one party
  - One party can have several members in the National Assembly
  - The condition "each member belongs to only one party" is called a **key constraint** on the `belong` relationship set
    - depicted with arrow from `Members` to `Parties`



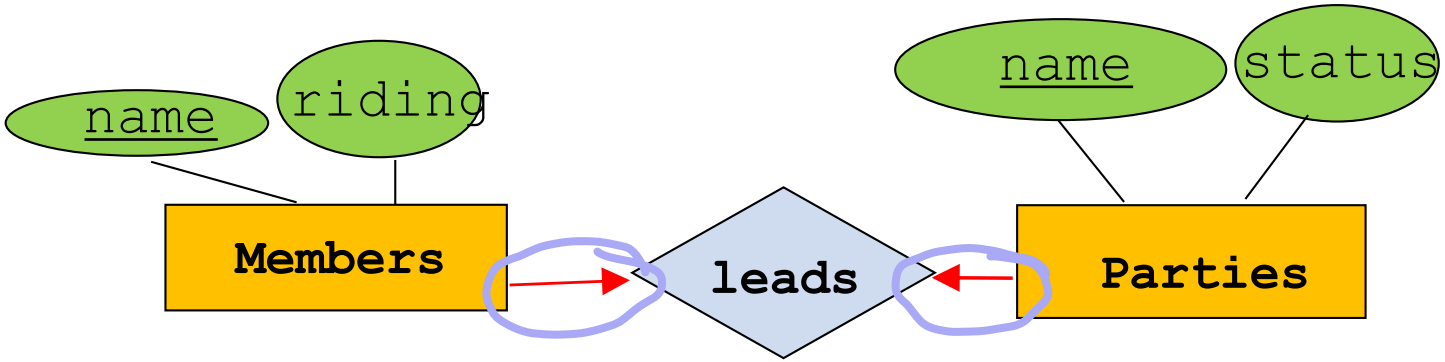
member does not have "party" attribute

0,1

many can belong to one
one can not belong to many

one-one

- Lead
  - Each member of the National Assembly can be leader of at most one party
  - Each party can have at most one leader
  - One-to-one relationship set between `Members` and `Parties`
  - Key constraints in both directions



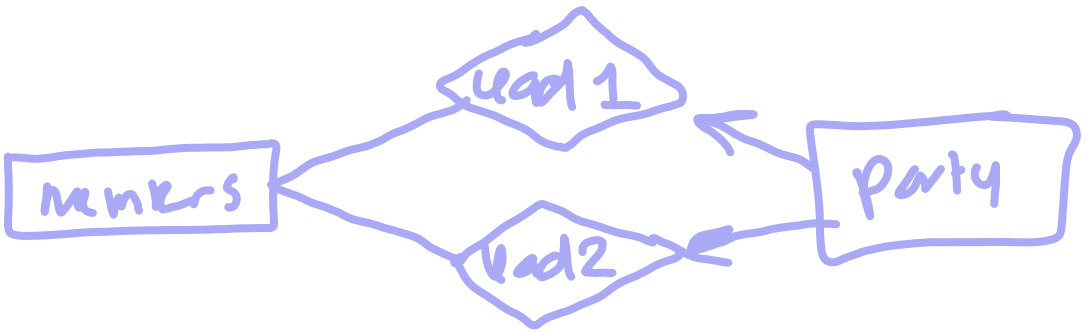
constraints = smaller schema but they need to be correct

ONE-ONE

Specific Cardinalities

- Some E/R languages let you specify a many-many with specific # of relations

- In our language:
  - option to have 2 relationship sets with key constraints
    - e.g. leads1 and leads2 from party to member

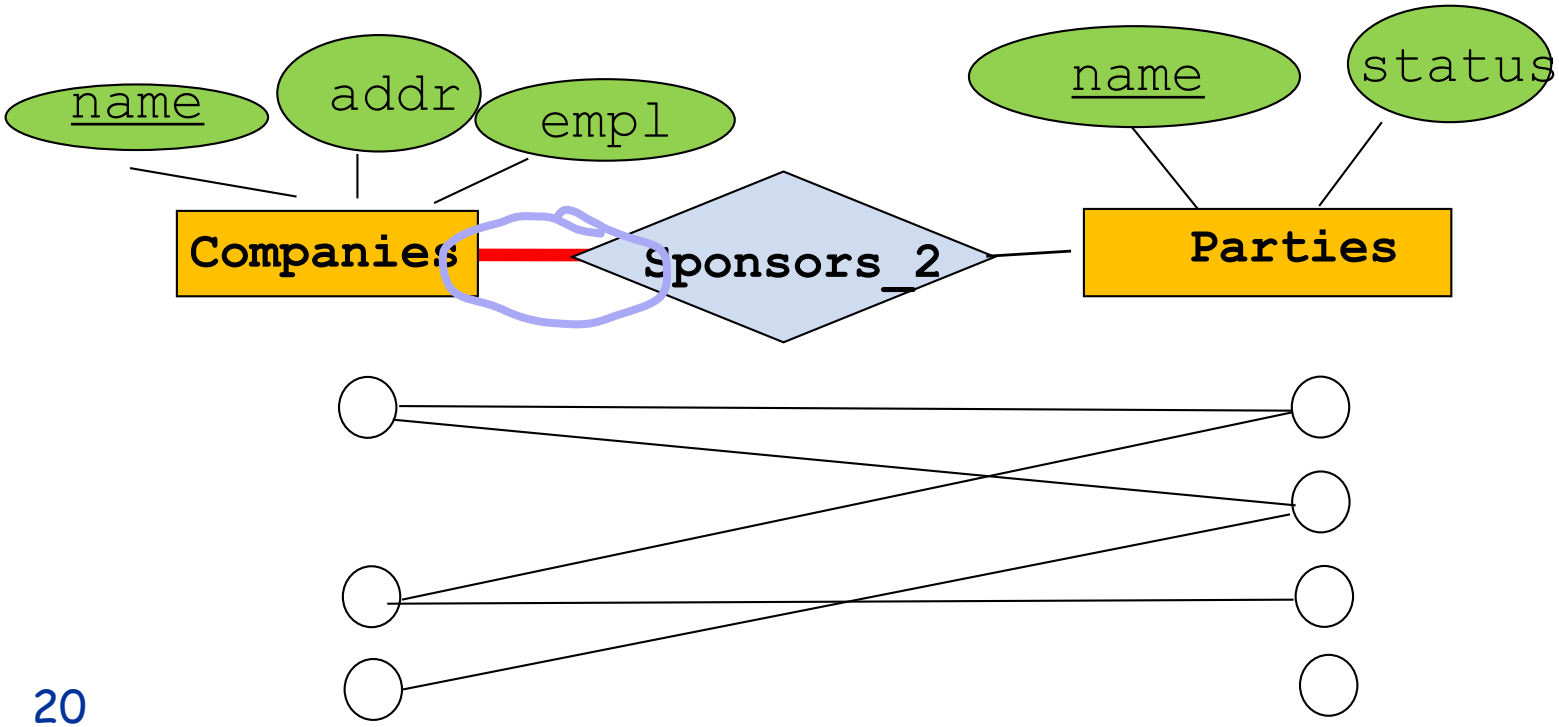


Note on Key Constraints

- The existence (or non-existence) of key constraints has a large influence on the final database design
  - Key constraints: less tables in the relational schema
  - No key constraints: more tables in the relational schema
- Before indicating key constraints, make sure that they really hold in all circumstances
  - Invariants of the application
  - Local example, Quebec Solidaire has two leaders.
  - in Germany, the Green party had for a long time always two leaders → key constraint of last slide does not hold
- If it later turns out that a key constraint does not hold, then your database design might have problems to capture all data

Participation Constraints

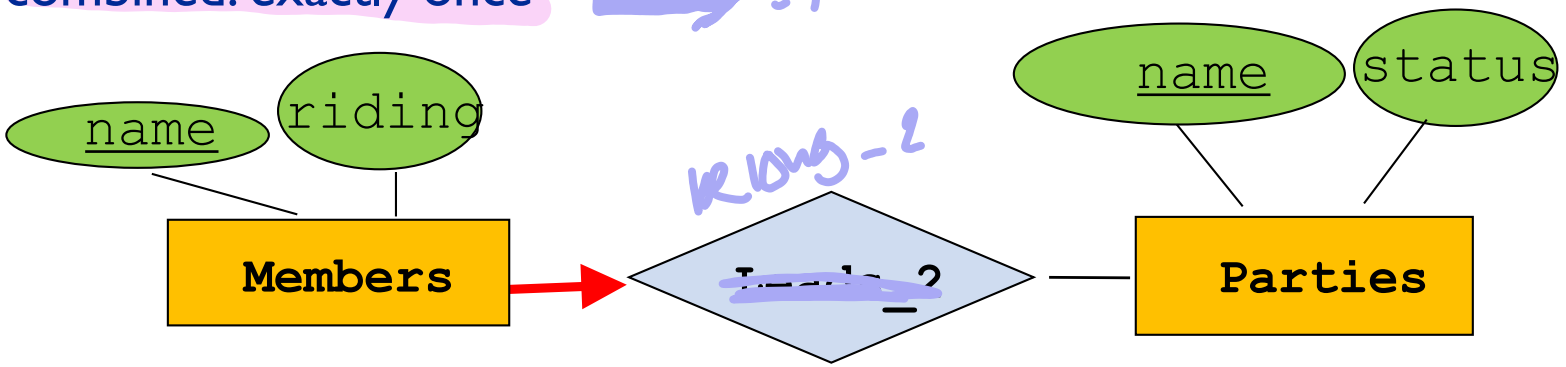
- Sponsor_2 participation:
  - Each company must sponsor at least one party
    - (we don't keep track of companies that don't provide sponsorship….)
  - the **participation constraint** requires that the participation of `Company` in `Sponsor` is *total* (vs. *partial*) (depicted with a **thick** line)

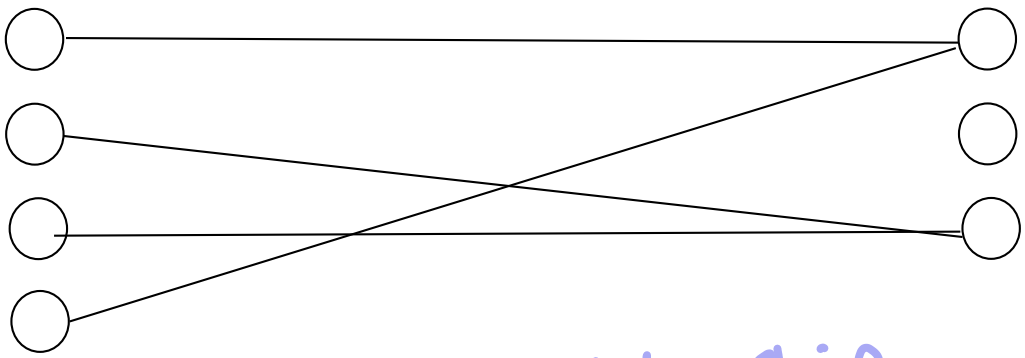


combine w/ key constraints

Participation Constraints

- belongs_2
  - Each member must belong to exactly one party
  - key constraint: at most once (at most one party) → ≤1
  - participation constraint: at least one → ≥1
  - combined: exactly once → =1



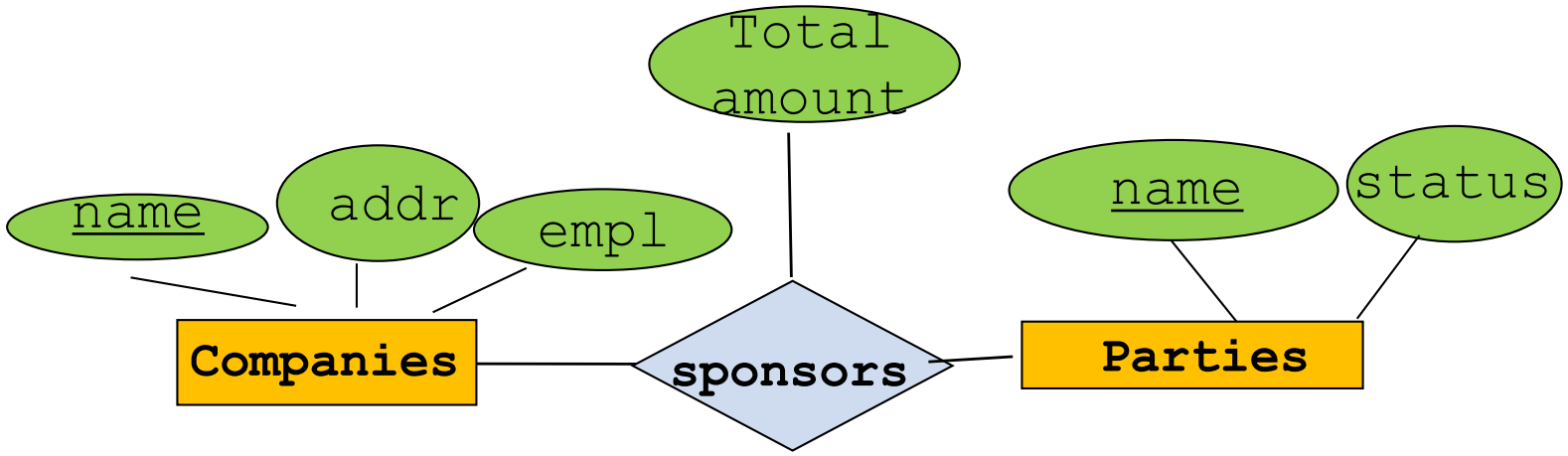


for exactly 2, need 2 relationships with thick arrow

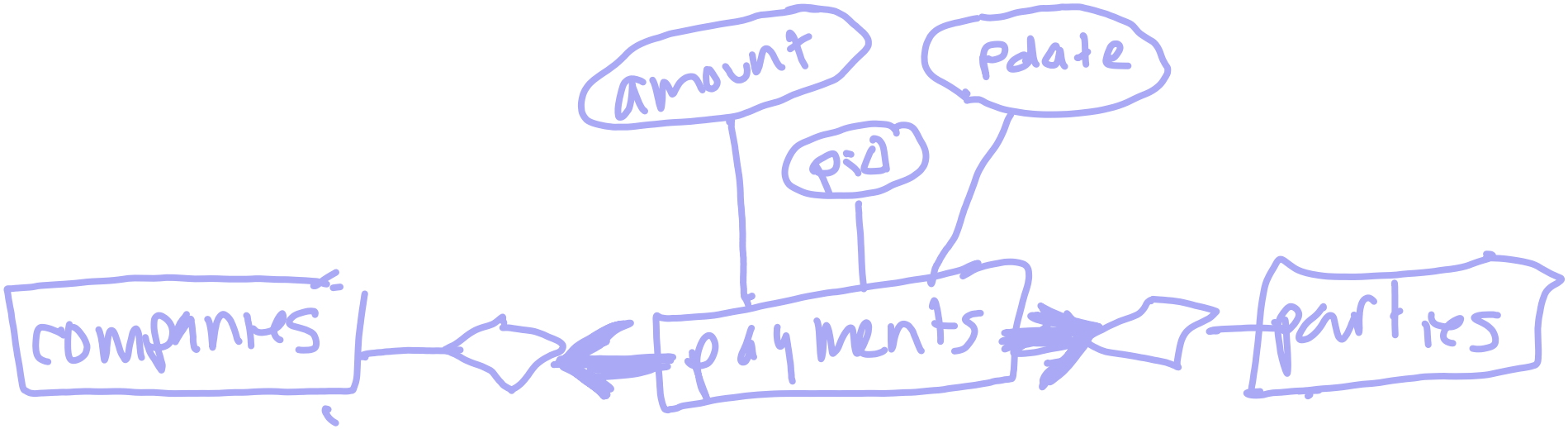
thin line – might not have a relationship
thick line → must have a relationship

Model Task

- A company can make several sponsorship payments to a party. Each payment has a date and an amount.
- Redesign below to allow for above situation



make payments an extra entity set

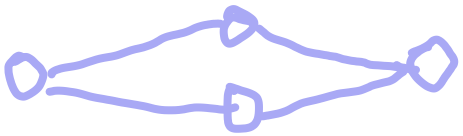


payments is an entity set containing each payment by company A to party B

payment can only have one payer and one receiver

companies can give many payments

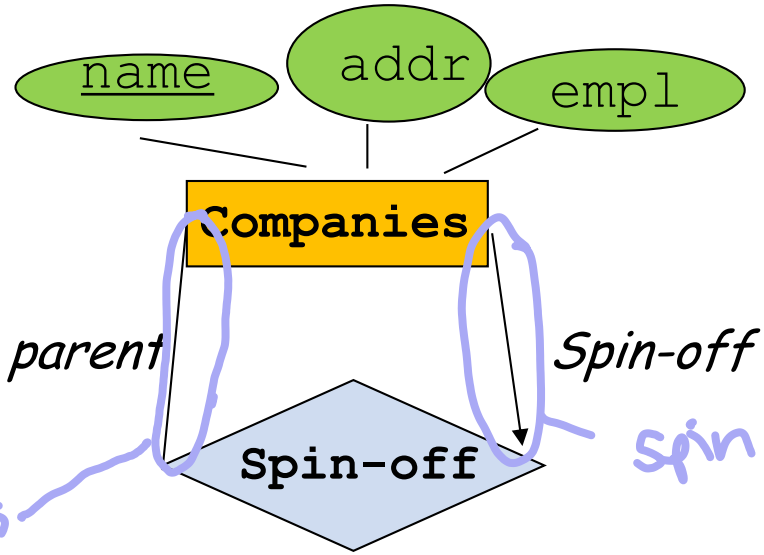
parties can receive many payments



entity set → relationship set

Relationships contd.

Relationship between companies



Ex: Spin-off

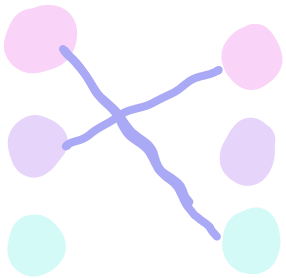
Alphabet ====== Google

Spin-off has 1 parent

Companies have multiple spin offs

Role indicator:

Parent company / spin-off

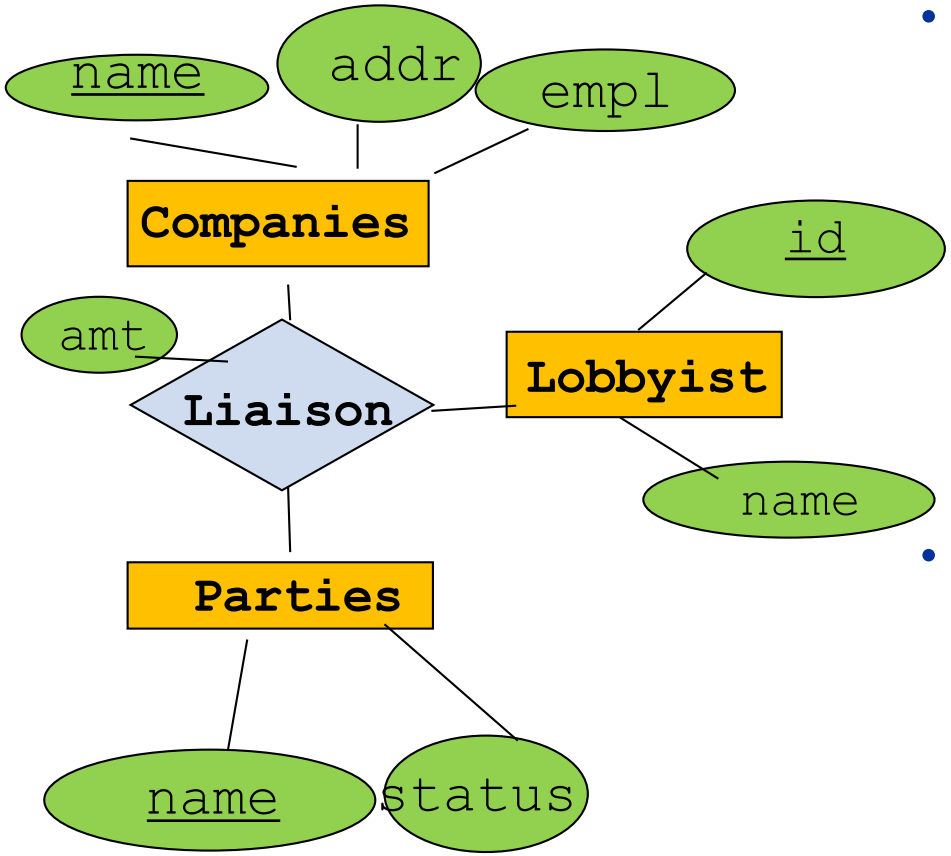


Parent

Spin-off

Purple parent of pink
pink parent of blue

Ternary Relationship



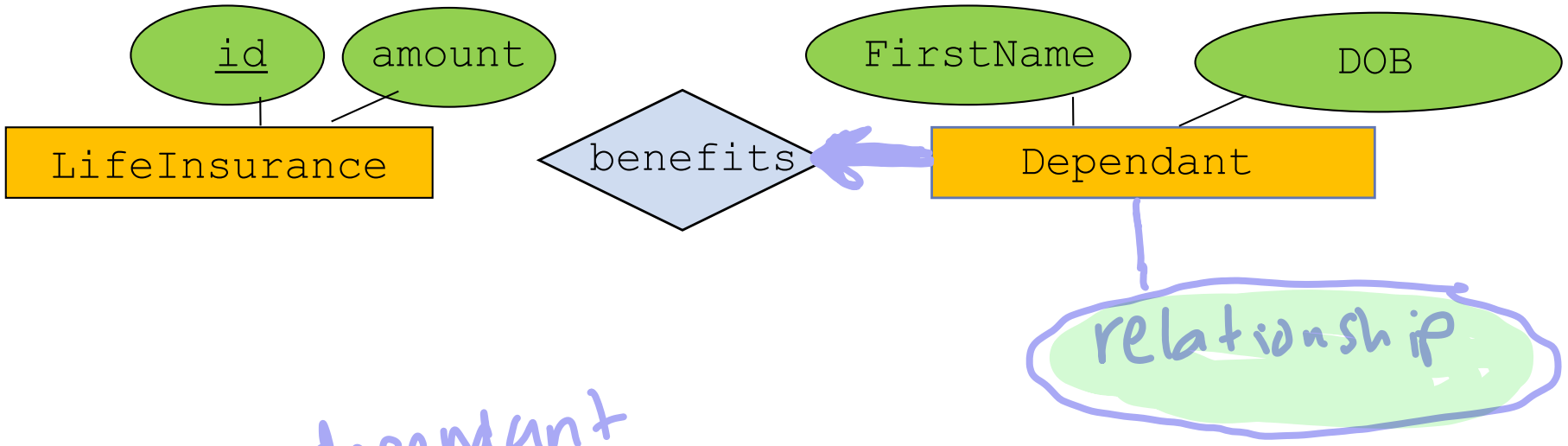
Examples

- Lobbyist Miller with id 007 builds the liaison for
  - BiggestEngFirmEver and the Liberal Party;
  - BiggestConsFirmEver and the Liberals
  - BiggestConsFirmEver and the CAQ
- Lobbyist Max with id 008 builds the liaison for
  - BiggestEngFirmEver and the Liberal Party;

- Note that given relationship set is many-many-many

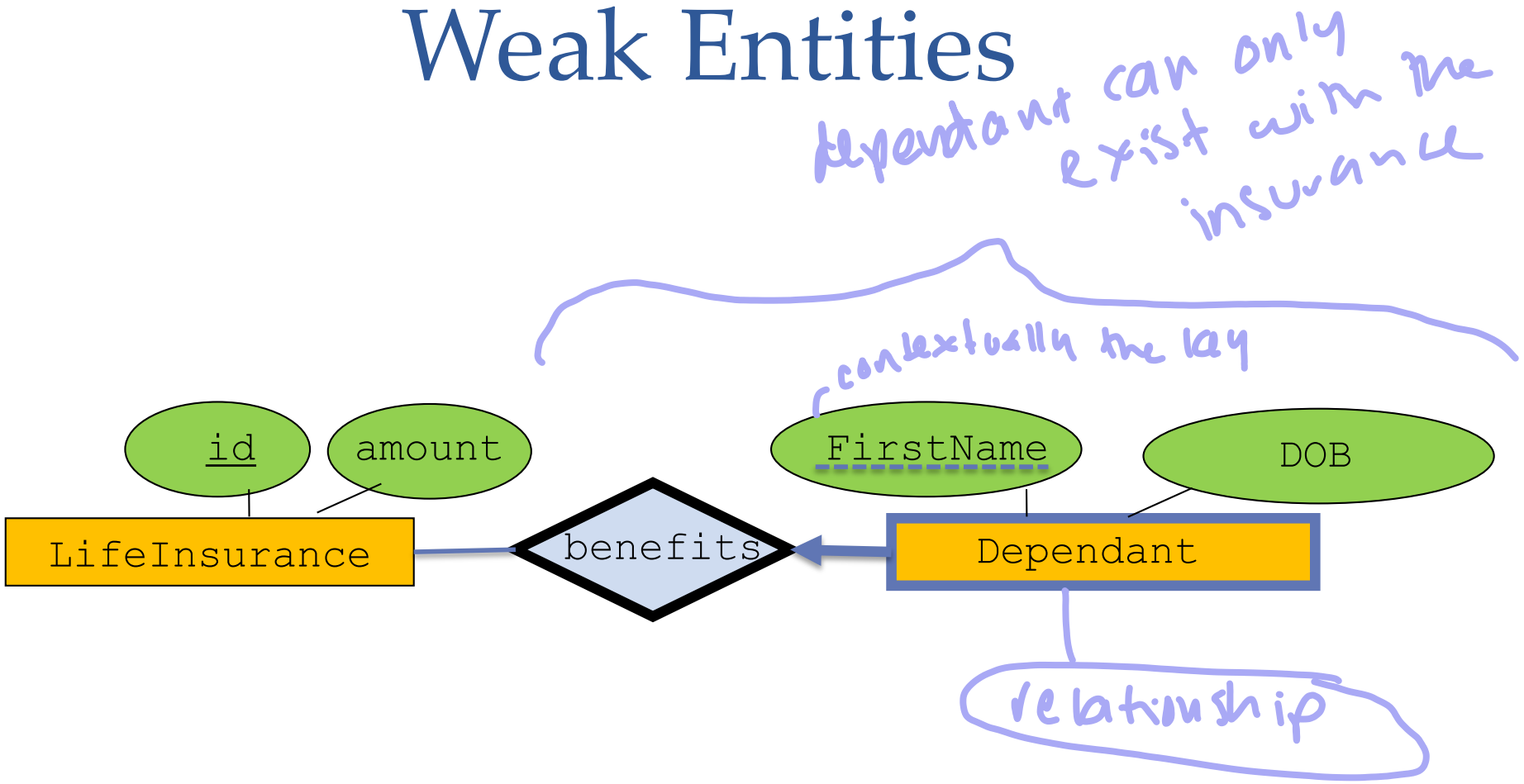
Requirements Quiz

Weak Entities



dependant has exactly one health insurance

Weak Entities

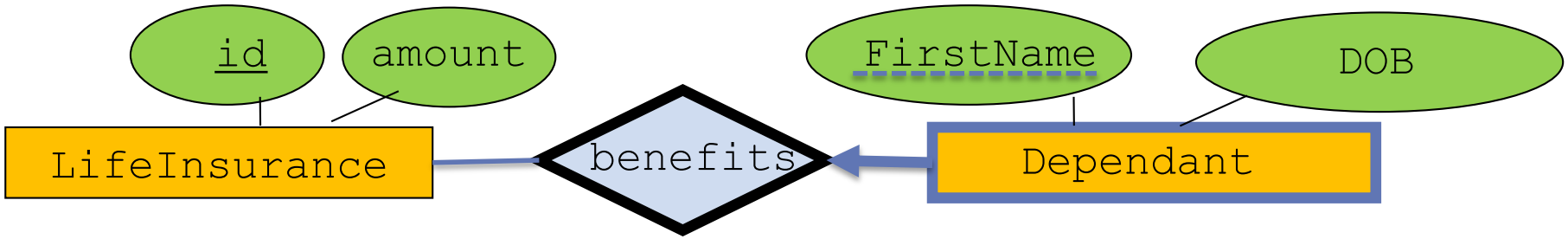


Weak Entities

- A **weak** entity "needs help" from another entity to be uniquely identifiable
- That is, a weak entity can be identified uniquely only by considering the primary key of another (owner) entity.
  - The key in weak entity set is the union of the key of the owner entity set and a set of its own attributes (underlined with dashes)
  - Owner entity set and weak entity set must participate in a *supporting one-to-many* relationship set (one owner, many weak entities).
  - Weak entity set must have total participation in this *identifying* set.

Example: Dependants covered

- First names are not unique
- But the first names among siblings are unique
  - Combination of life insurance id and first name are unique
- E/R:
  - Weak entity set in **bold**
  - Relationship set to supporting entity set with key and participating constraint (**bold and arrowed**)
  - Relationship set in **bold**
  - Partial key in weak entity set with **dashed line**



Don't overdo Weak Entities

- Many things appear to depend on entities of other entity sets and need their keys to identify.
- In reality:
  - Examples
    - Car and Owners
      - Make License plate key
    - Online users and their purchases
      - Generate unique purchase order id

if an entity set has important info for all the data set, maybe it needs to not be weak

When to use weak entities

- If the partial key within the weak entity set has some "meaning"
  - A name, an order (1st, 2nd, …)
- If there is no global authority to create global identifiers

what do you need to combine to make it unique

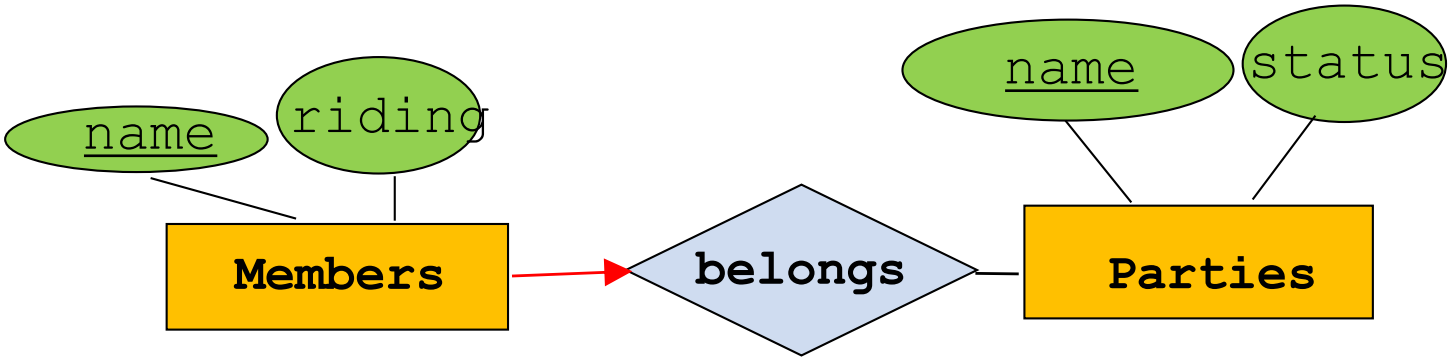
Conceptual Design with ER

- Design Principles:
  - Keep it simple
    - Don't use an entity set when an attribute is sufficient
  - Avoid redundancies (or not)
  - Capture as many constraints as possible
- Design Choices
  - Entity or attribute?
  - Attributes and relationships
  - Identifying relationships: Binary or ternary?

Avoid Redundancies

don't store information more than once

Correct Design

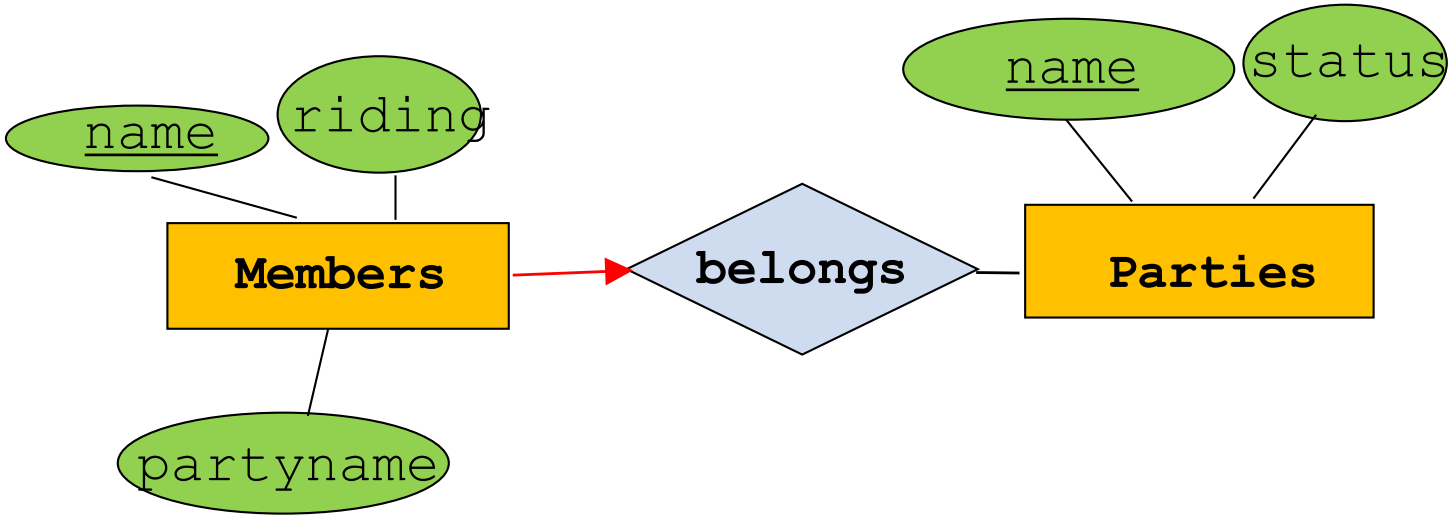


Each political party with its status is captured once as an entity in Political Parties

Avoid Redundancies

don't store information more than once

Bad Design

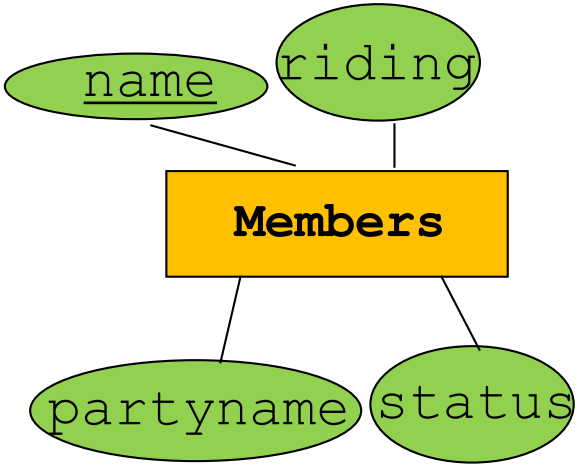


This design captures the name of the party twice. Once as attribute and once as entity

Avoid Redundancies

don't store information more than once

Bad Design



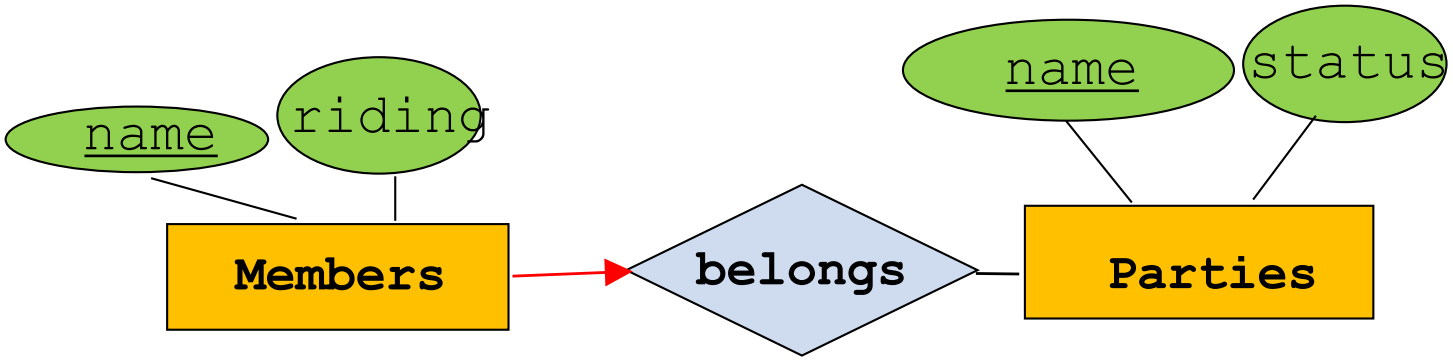
This design indicates the status of a party for each member of the party, and looses the information for a party that temporarily does not have members.

Entity Set vs. Attribute

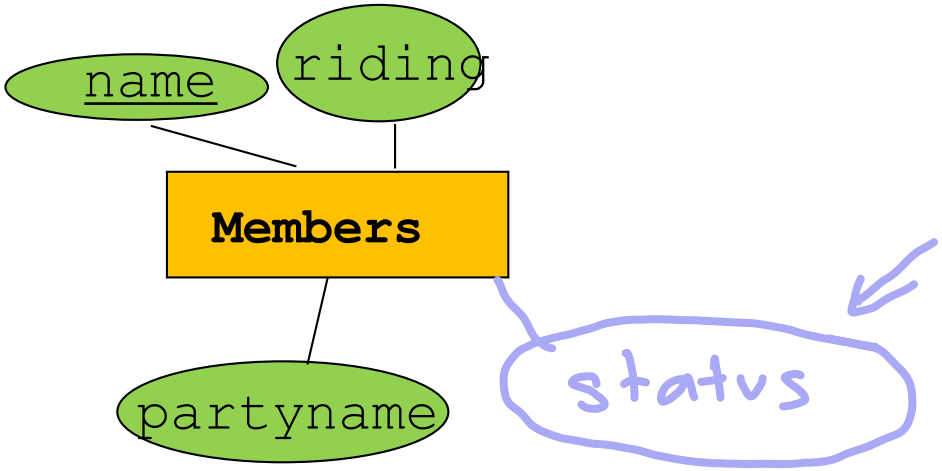
- A "thing" should satisfy at least one of the following to be represented as an entity set
  1. It is more than the name of something; at least one non-key attribute
  2. It is the many in a many-many or many-one relationship set
     - Example: A product can belong to many categories

Entity Set vs. Attribute

1. It is more than the name of something; at least one non-key attribute
   - if Party has additional attribute status ⇒ entity set



   - if No extra information about the party except the name (and a member belongs to at most one party) ⇒ attribute

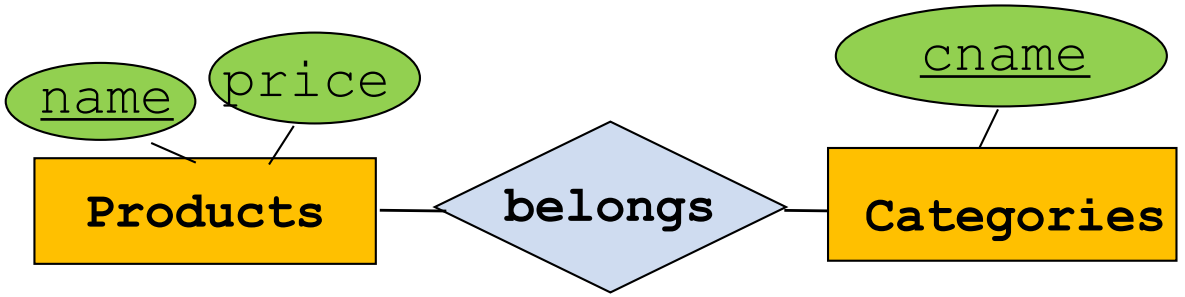


wrong! redundant ↳ repeating status of party for each member

Entity Set vs. Attribute

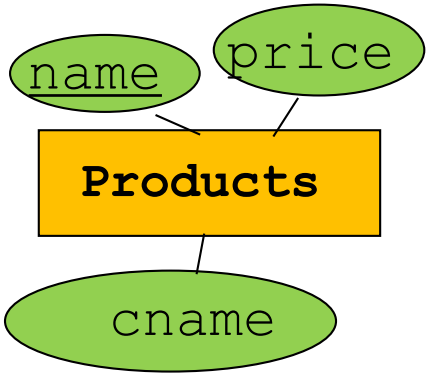
2. It is the many in a many-many or many-one relationship set

- Example: A product can belong to many categories



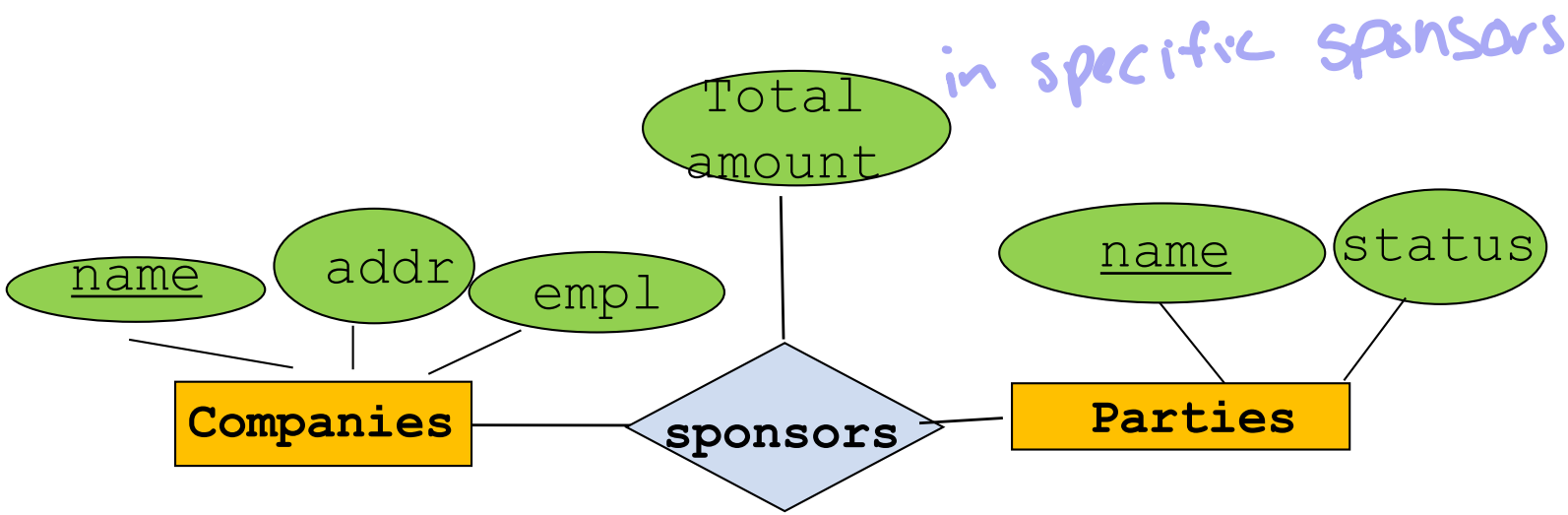
- A product belongs to only one category (and category is simply a name)

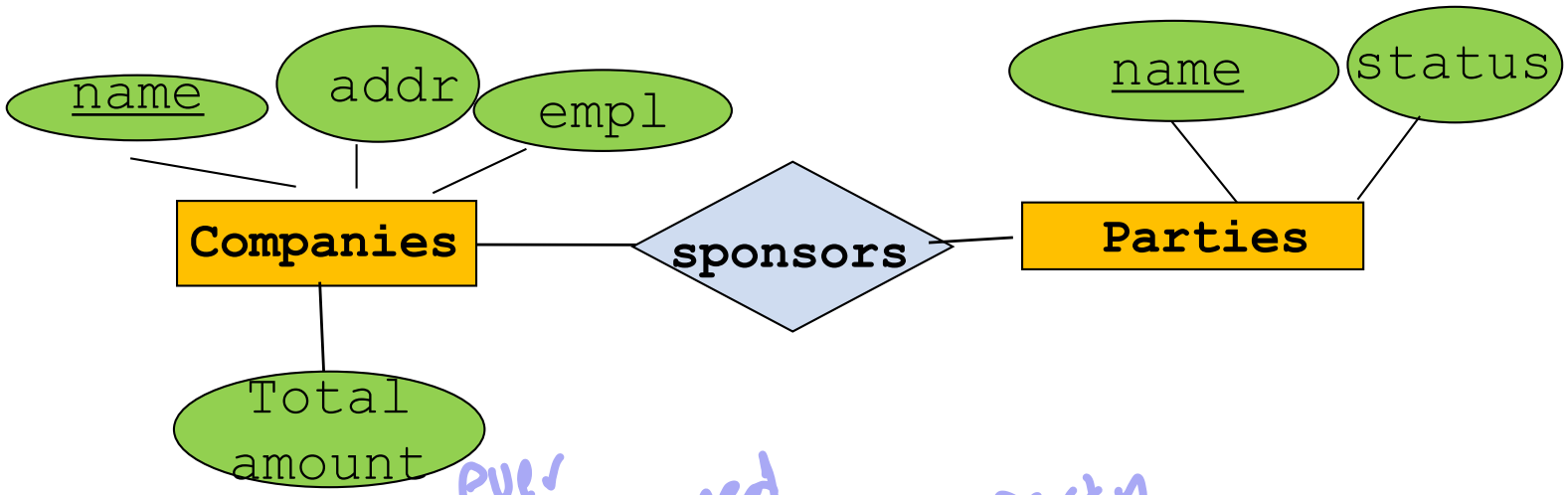
category has no special attribute



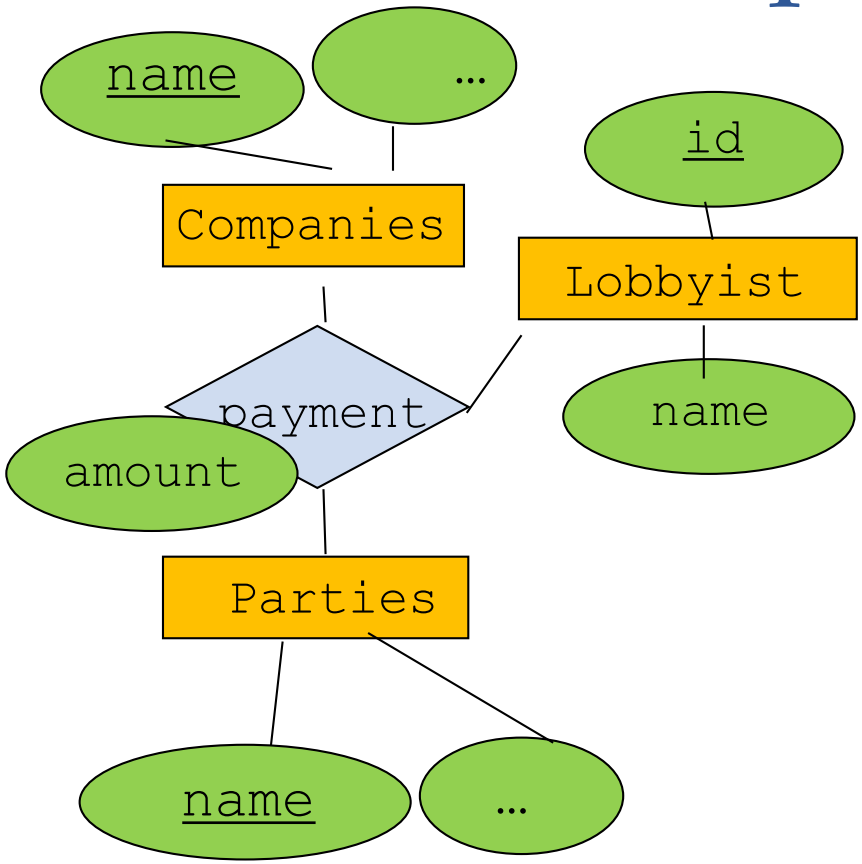
can't have a multivalued attribute

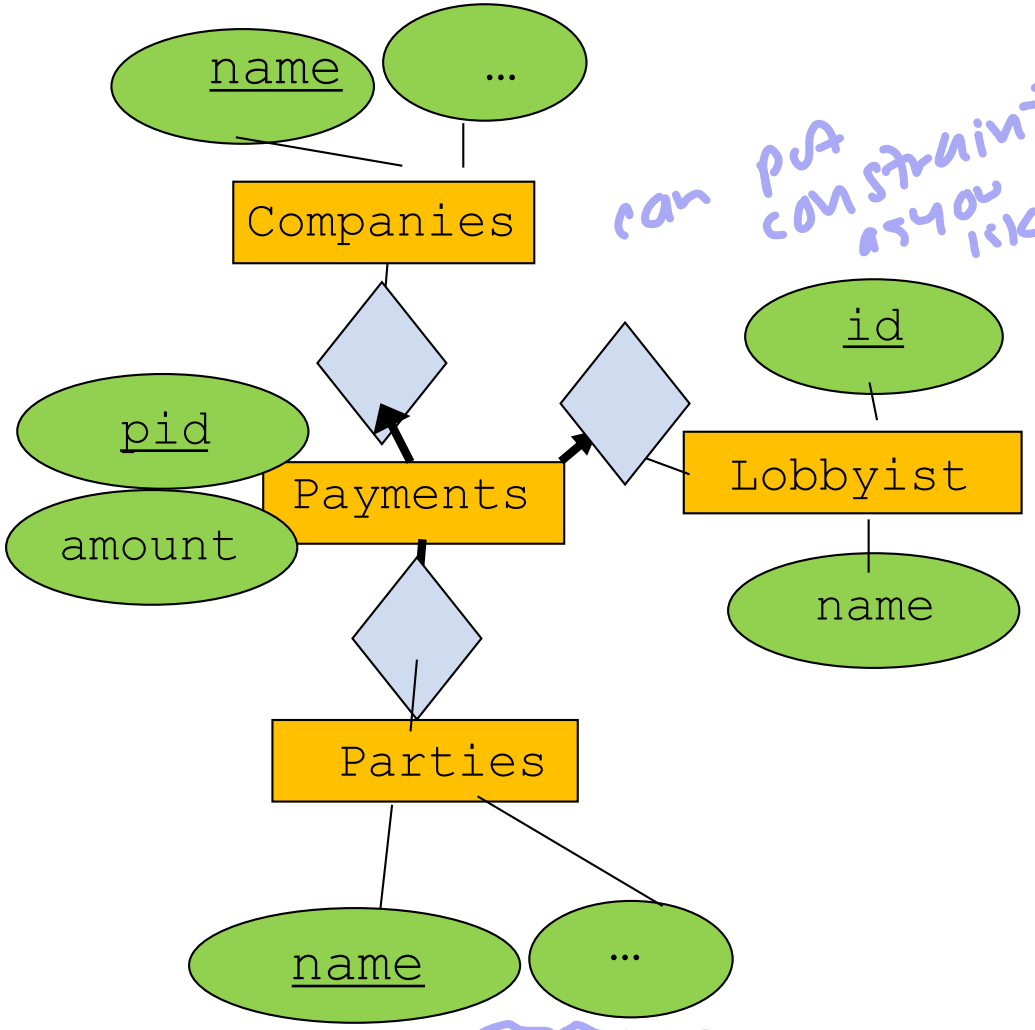
Relationships and Attributes





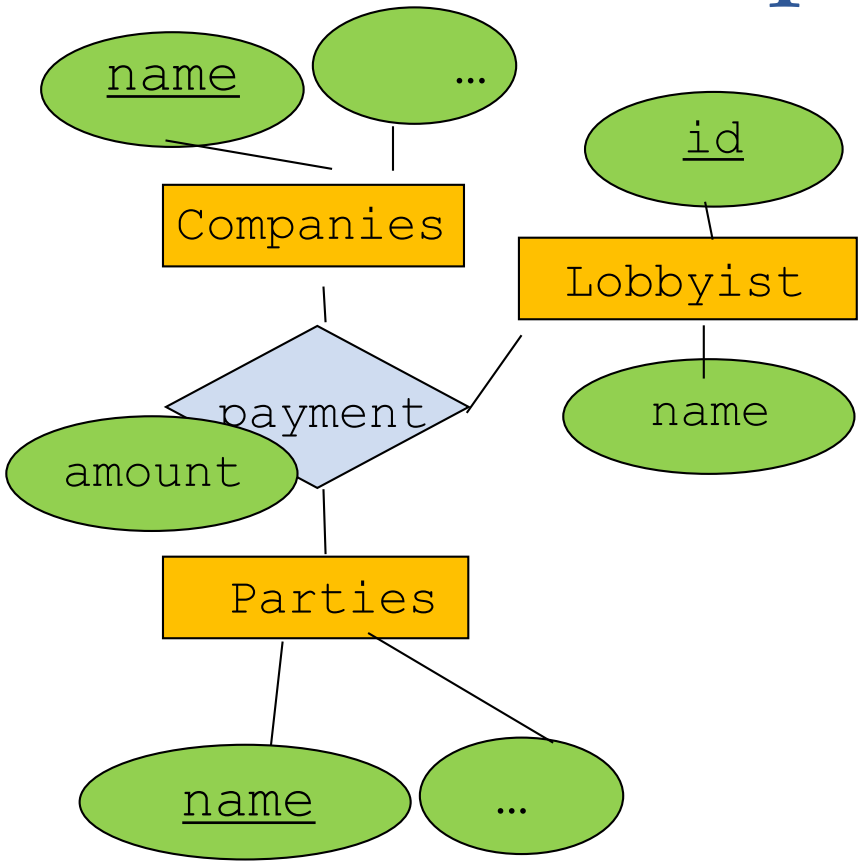
Relationships vs. Entity Sets

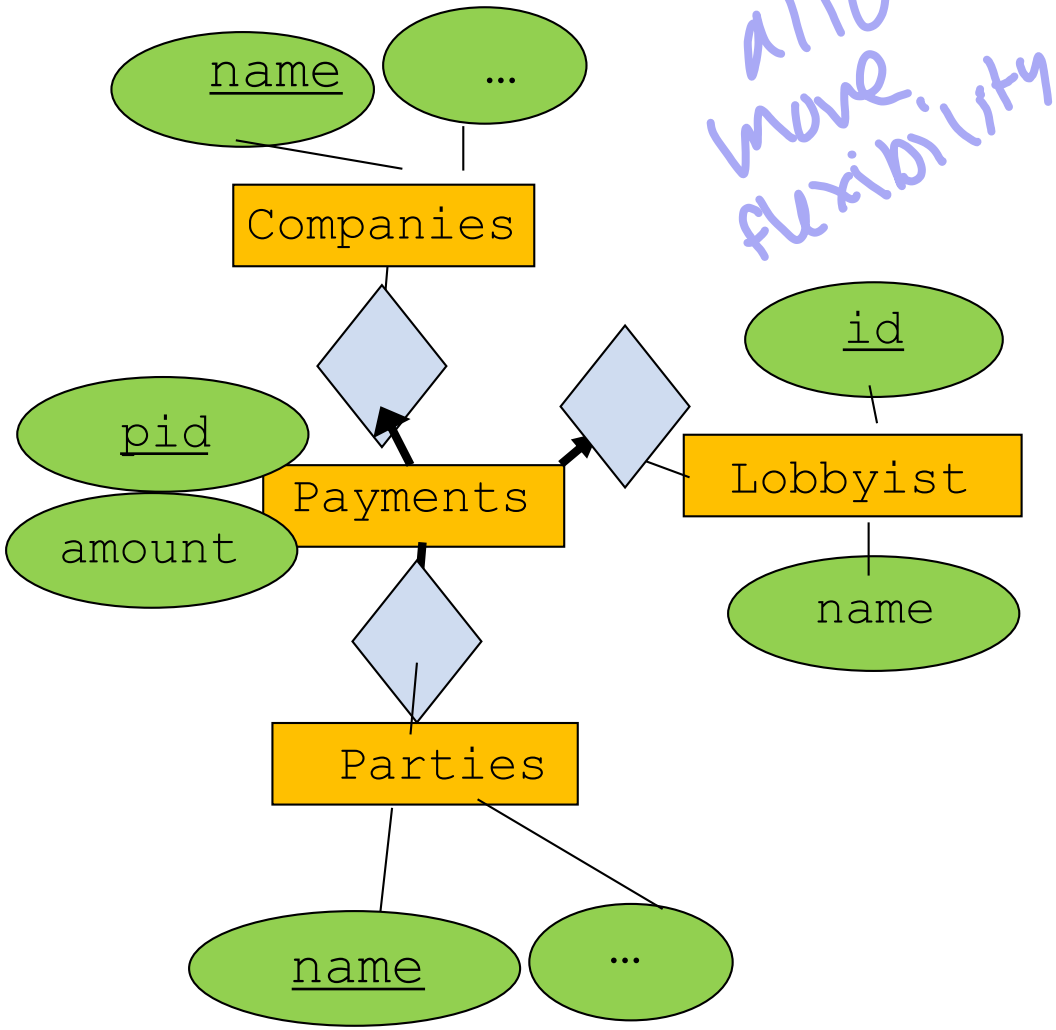




Ternary (payment as relationship set) vs. 3 binary relationship sets (payment as entity set)

Relationships vs. Entity Sets



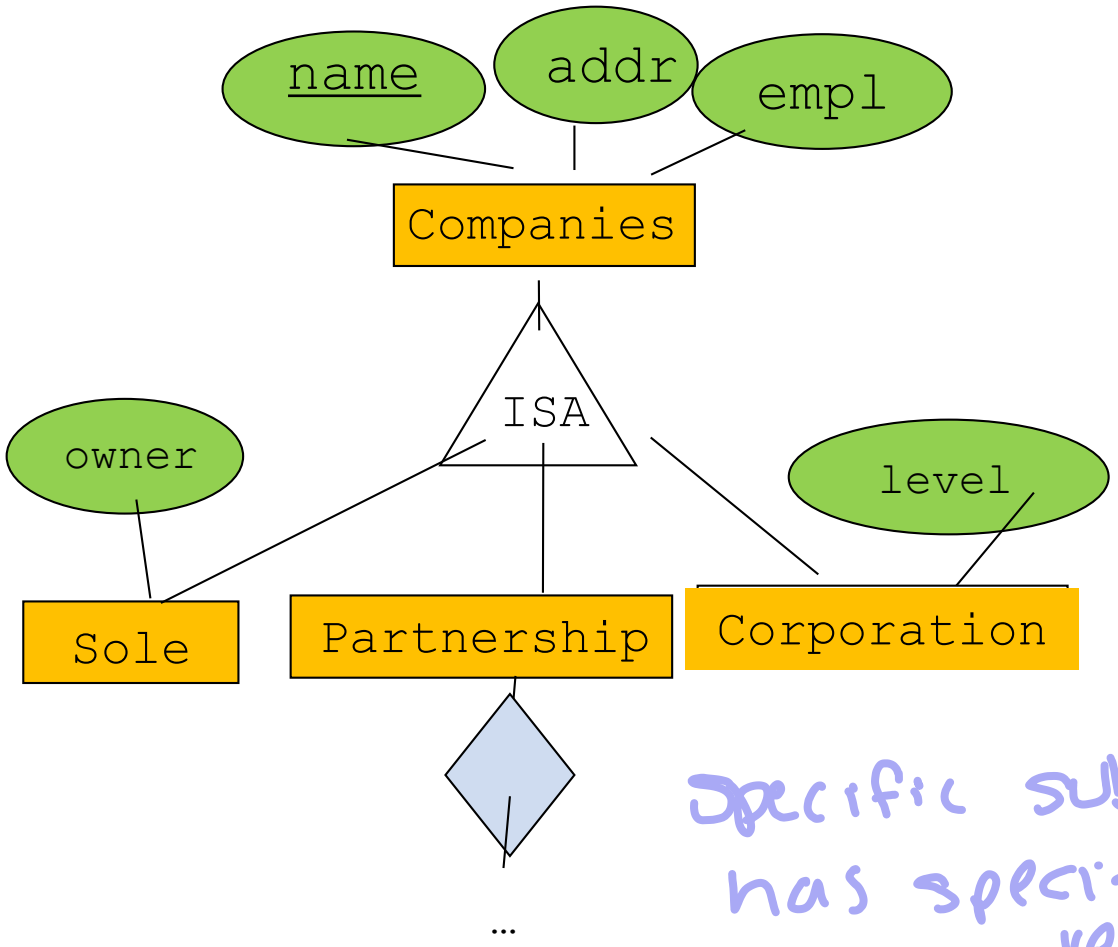


allows more flexibility

payment as entity set allows several payments of the same company to the same party by the same connection

ISA ("is a") Hierarchies

- Subclasses
  - Have extra attributes that only belong to the subclass OR
  - Have a relationship set that only holds for the subclass



Specific subgroup has specific relationship

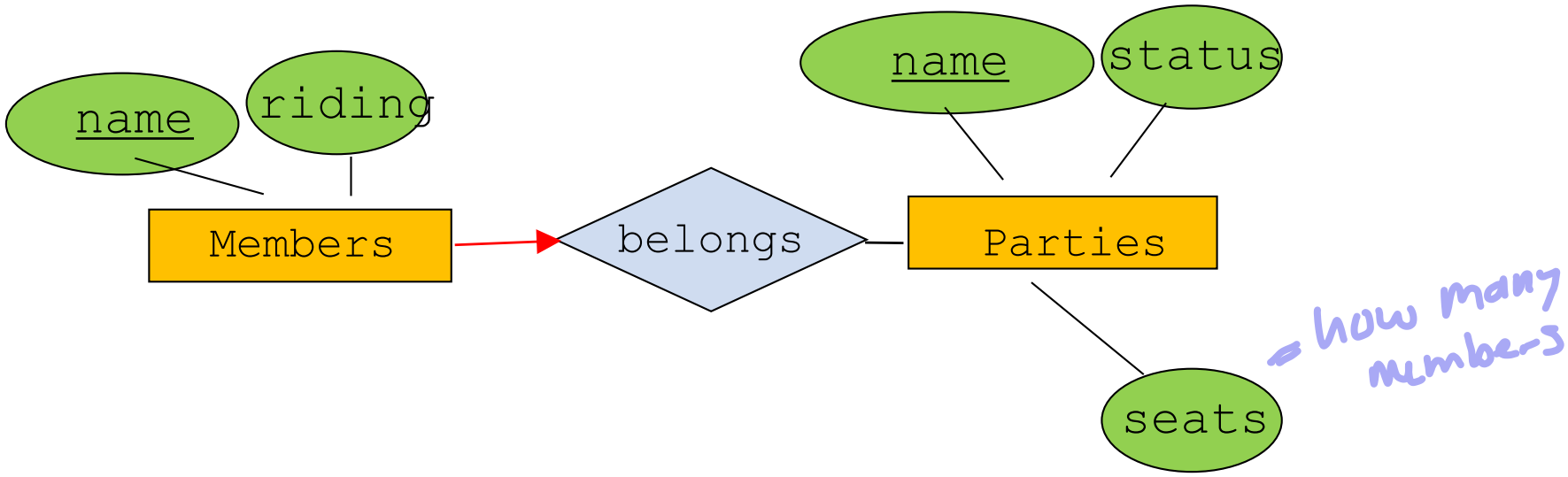
Super. class

- has a relationship that holds for all entities

- importance of unified key across all subclasses

- many queries that are interested in attributes that are shared

Redundant Attributes



- Attributes `seats` carries redundant information
- Why? – could query number of seats by counting number of relationships the party has in the set "belongs"
- Keep it nevertheless??
  - ↳ queried often
  - ↳ be aware, ensure consistency