Entity Relationship Modelling

P.J. McBrien

Imperial College London

Designing a Relational Database Schema

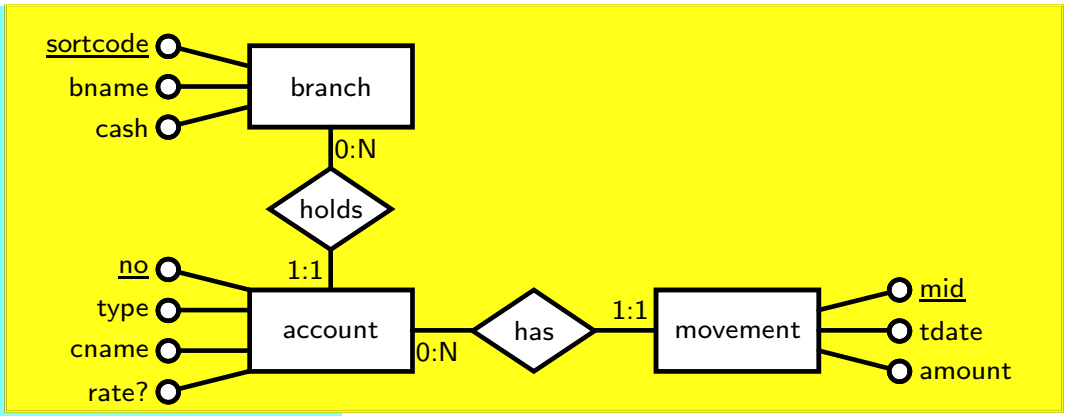
How do you design a relational database schema for a particular UoD?

1. Need some way to model the semantics of the UoD as a conceptual schema
   - ER (many variants exist)
   - UML class diagrams
2. Need to map the ER/UML schema into a relational schema
3. Need to ensure that the relational schema is a good design
   - Normalisation

Semantic Modelling: ER Schemas

```
CREATE TABLE branch
(       sortcode INTEGER NOT NULL,
        bname VARCHAR(20) NOT NULL,
        cash DECIMAL(10,2) NOT NULL,
        CONSTRAINT branch_pk PRIMARY KEY (sortcode)
)
CREATE TABLE account
(       no INTEGER NOT NULL,
        type CHAR(8) NOT NULL,
        cname VARCHAR(20) NOT NULL,
        rate DECIMAL(4,2) NULL,
        sortcode INTEGER NOT NULL,
        CONSTRAINT account_pk PRIMARY KEY (no),
        CONSTRAINT account_fk FOREIGN KEY (sortcode) REFERENCES branch
)

CREATE INDEX account_type ON account (type)
CREATE TABLE movement
(       mid INTEGER NOT NULL,
        no INTEGER NOT NULL,
        amount DECIMAL(10,2) NOT NULL,
        tdate DATETIME NOT NULL,
        CONSTRAINT movement_pk PRIMARY KEY (mid),
        CONSTRAINT movement_fk FOREIGN KEY (no) REFERENCES account
)
```

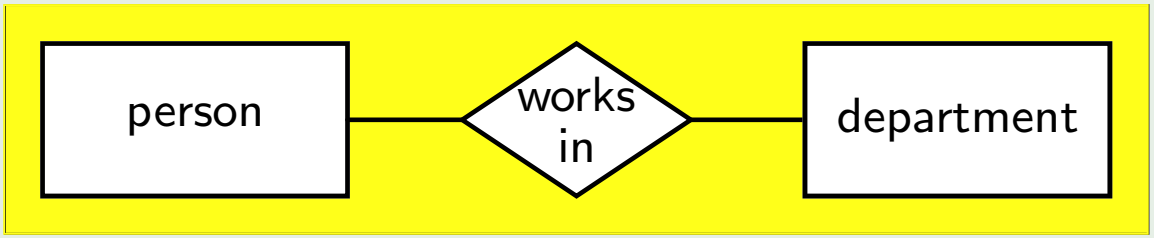


Core of ER: Entities and Relationships

- an entity E represents a set of objects in the UoD which conceptually are the same type of thing
  - **nouns** $\rightarrow$ entity set
  - proper nouns imply instances, which are not entity sets.
- a relationship R represents a set of tuples of objects in the UoD where each tuple is some type of conceptual association between entities E_1, E_2
  $R \subseteq \{\langle e_1, e_2 \rangle | e_1 \in E_1 \wedge e_2 \in E_2\}$
  - **verbs** $\rightarrow$ relationship
  - any objects appearing in a relationship must also appear in the entities the relationship connects

Identifying entities and relationships

In News Ltd, each person works in exactly one department; there are no restrictions on the number of persons a department may employ.

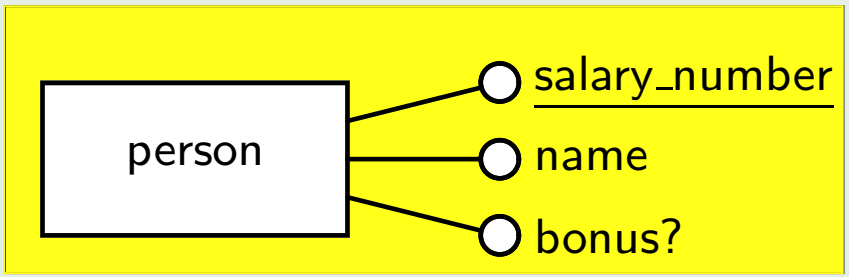


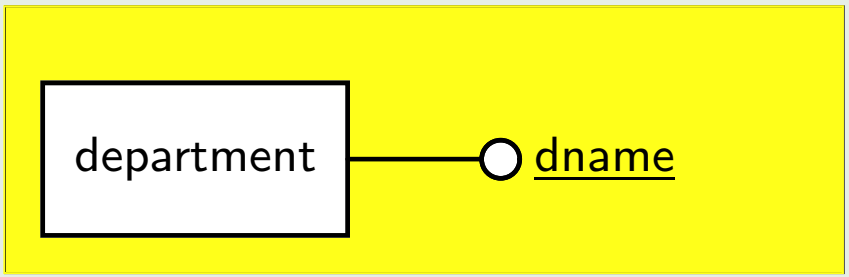
Core ER: Attributes of Entities

- An attribute $E.A$ is a function that maps from an entity set E to a value set V.
  1. $E.A \subseteq \{\langle e, v\rangle | e \in E \wedge v \in V\}$
  2. unique: $\langle e, v_1\rangle \in E.A \wedge \langle e, v_2\rangle \in E.A \rightarrow v_1 = v_2$
  3. mandatory: $E = \{e | \langle e, v\rangle \in E.A\}$
- In $\text{ER}^{\mathcal{O}}$: an **optional attribute** removes property (3)
- In $\text{ER}^{\mathcal{K}}$: certain attribute(s) $E.A_1 \ldots E.A_n$ of E are denoted **key attributes** such that
  $E = \{\langle v_1, \ldots, v_n\rangle | \langle e, v\rangle \in E.A_1 \wedge \ldots \wedge \langle e, v_n\rangle \in E.A_n\}$
- **adjective**, **adjective noun** $\rightarrow$ attribute

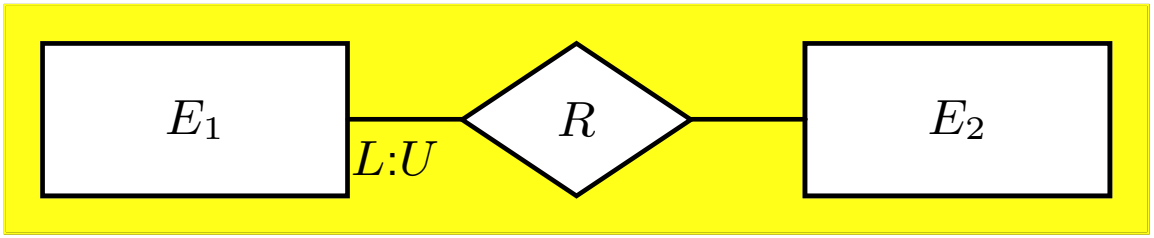
Identifying attributes

We record the name of each person working in the department; and identify them by their salary number. Optionally they might have a bonus figure recorded. Departments are identified by their name.





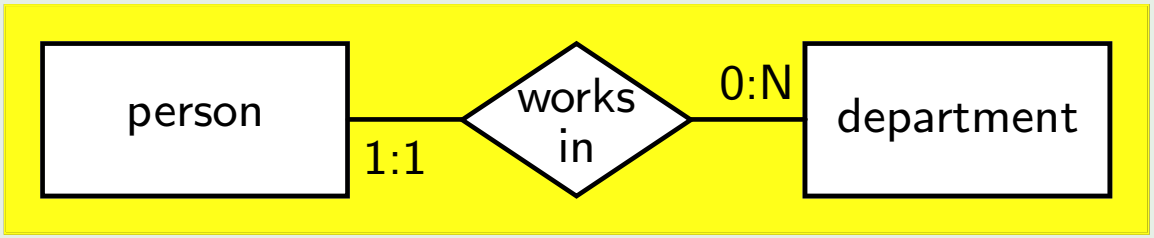
$\text{ER}^{\mathcal{L}}$: Look-Here Cardinality Constraints



- this course uses **look-here** cardinality constraints, where the constraints for an entity are written next to the entity
- an upper bound cardinality constraint U states that each instance of E_1 may appear at most U times in R
  - an upper bound of N indicates no limit
- additionally with $\text{ER}^{\mathcal{O}}$: a lower bound cardinality constraint L states that each instance of E_1 must appear at most L times in R

Adding look-here cardinality constraints in $\text{ER}^{\mathcal{LO}}$

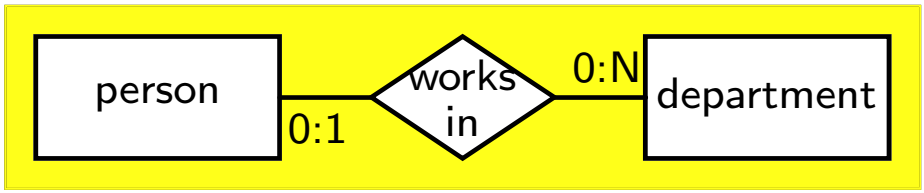
Each person works in exactly one department; there are no restrictions on the number of persons a department may employ.



Quiz 1: Extent of Relationships

person = ‘Peter’, ‘Jane’, ‘Mary’

dept = ‘CS’, ‘Maths’



Which is not a possible extent of works_in?

A

works_in={⟨‘Peter’,‘Maths’⟩, ⟨‘Peter’,‘CS’⟩, ⟨‘Mary’,‘Maths’⟩, ⟨‘Jane’,‘Maths’⟩}

B

works_in={⟨‘Peter’,‘Maths’⟩, ⟨‘Mary’,‘Maths’⟩, ⟨‘Jane’,‘Maths’⟩}

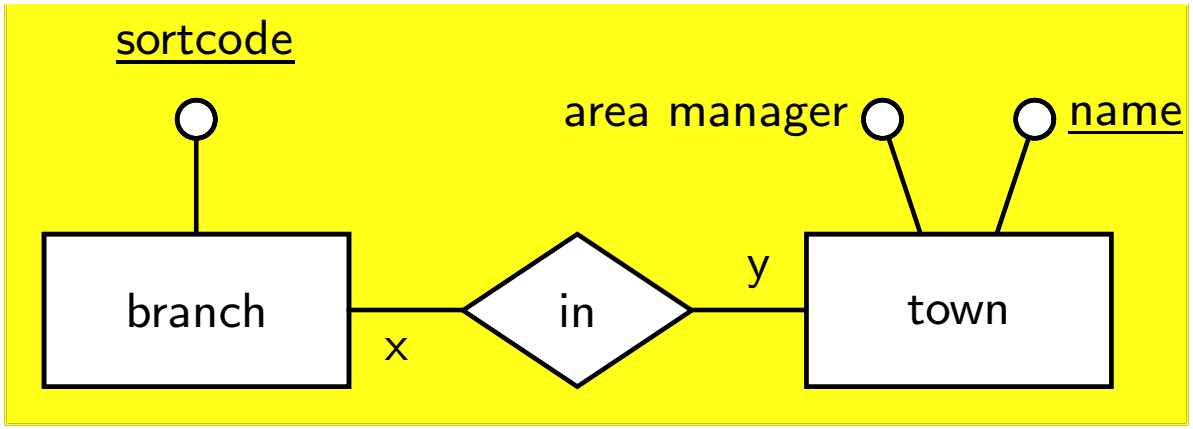
C

works_in={⟨‘Peter’,‘CS’⟩, ⟨‘Mary’,‘Maths’⟩, ⟨‘Jane’,‘Maths’⟩}

D

works_in={⟨‘Peter’,‘CS’⟩, ⟨‘Jane’,‘Maths’⟩}

Quiz 2: Cardinality Constraints on Relationships



Branches based in towns are all asigned to an area manager for that town; and area managers are only assigned to towns that have branches

What should be the cardinality constraints of in?

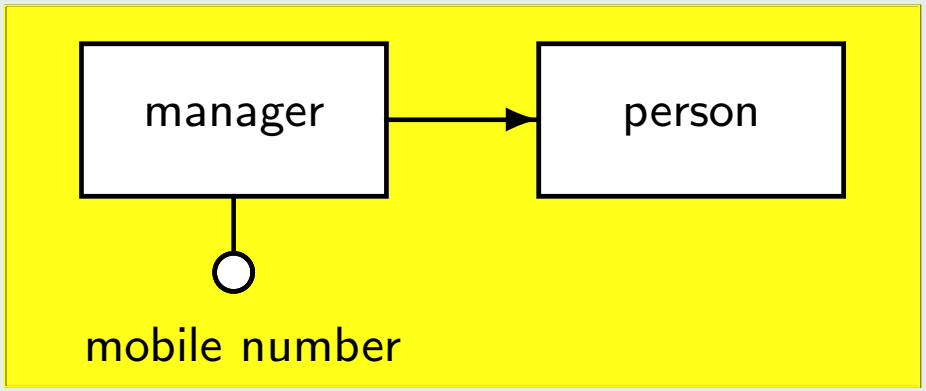
A	B	C	D
x = 1:1, y = 0:N	x = 0:1, y = 0:N	x = 0:N, y = 1:N	x = 0:1, y = 1:N

ER$^{\mathcal{S}}$: Subset/isa hierarchies

- In ER$^{\mathcal{S}}$: if it is found that the instances of one entity E_s are a subset of a another entity E, we may add a **subset** constraint.
  $E_s \subseteq E$
- **specialisation of nouns** $\rightarrow$ subset

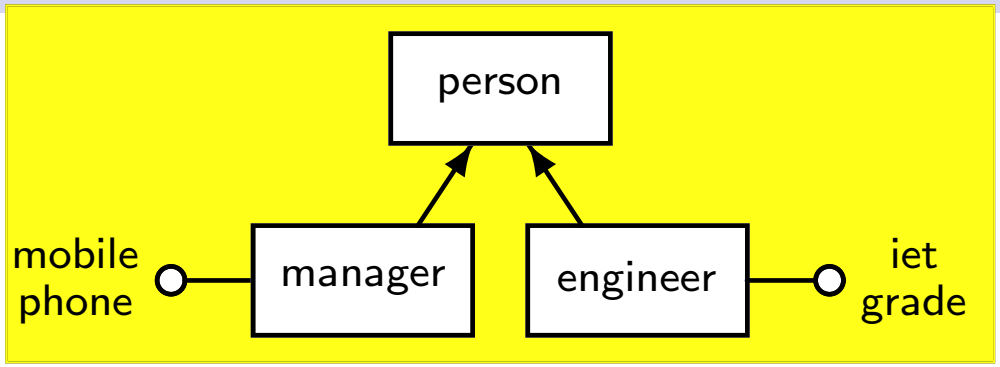
Identifying subsets with ER$^{\mathcal{S}}$

Some employees are ranked as managers, and receive a mobile phone.



Quiz 3: Extent of subset and superset entities

manager = {'Jane', 'Mary'}



Which is not a possible extent?

A

person={'Peter','Jane','Mary'} engineer={'Jane','Mary'}

B

person={'Peter','Jane','Mary','John'} engineer={}

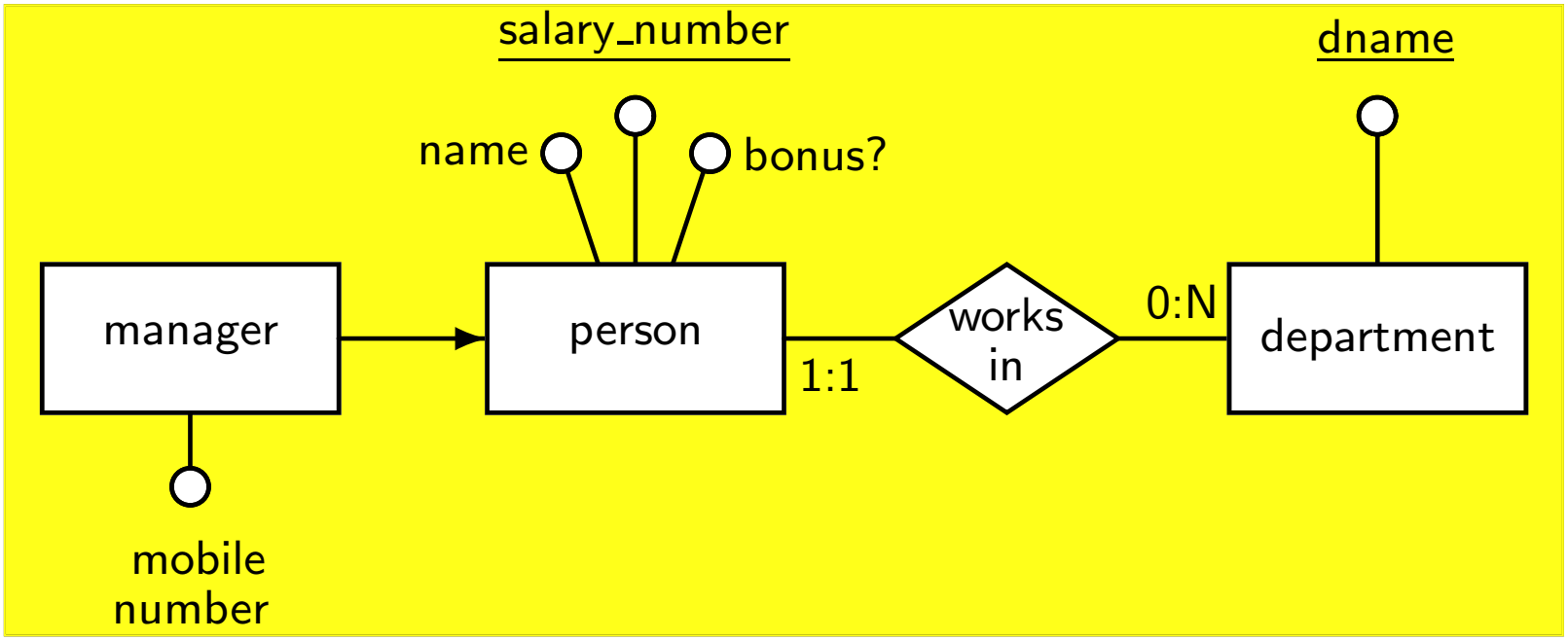
C

person={'Peter','Jane','Mary'} engineer={'John'}

D

person={'Peter','Jane','Mary','John'} engineer={'Peter','John'}

Combining Fragments



Worksheet: ER Modelling

Draw an ER$^{\mathcal{KLOS}}$ schema to describe the following domain

> *The payroll system for* BIG Inc *records the salaries, status, joining date, name, and payroll number for all of the corporation's 30,000 employees. Each employee works for one division, and each division has an account number for paying its staff. We identify divisions by their name, and record the address where the division's HQ is located.*
> *For employees sent abroad by* BIG Inc, *we record the address, country and telephone number of the foreign tax office that will handle the employee. It is assumed that each country has one central tax office that we have to deal with. All other employees have their tax affairs dealt with by the Inland Revenue.*

Worksheet: ER Modelling

Draw an ER$^{\mathcal{KLOS}}$ schema to describe the following domain

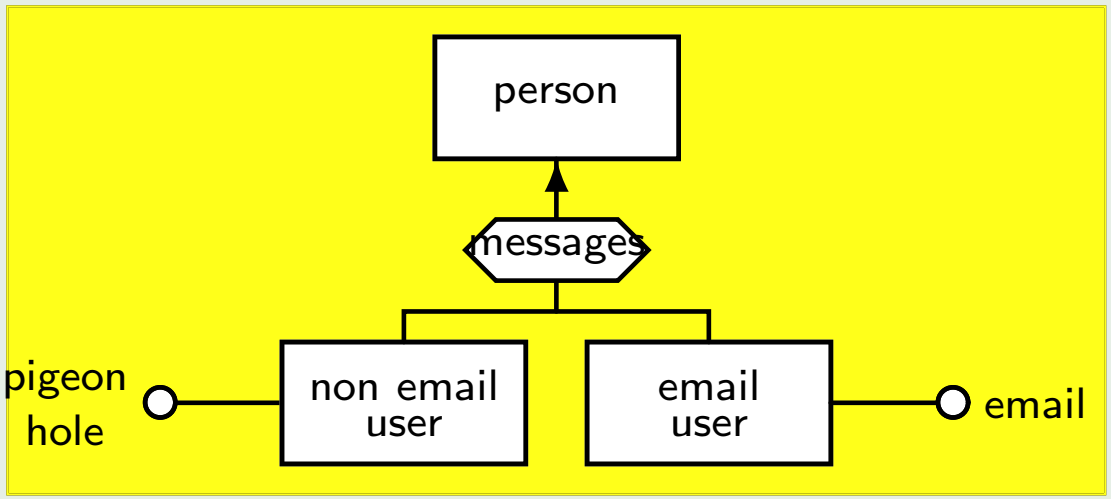
> *The payroll system for* BIG Inc *records the salaries, status, joining date, name, and payroll number for all of the corporation's 30,000 employees. Each employee works for one division, and each division has an account number for paying its staff. We identify divisions by their name, and record the address where the division's HQ is located.*
> *For employees sent abroad by* BIG Inc, *we record the address, country and telephone number of the foreign tax office that will handle the employee. It is assumed that each country has one central tax office that we have to deal with. All other employees have their tax affairs dealt with by the Inland Revenue.*

$ER^{\mathcal{D}}$: Disjointness and Generalisation Hierarchies

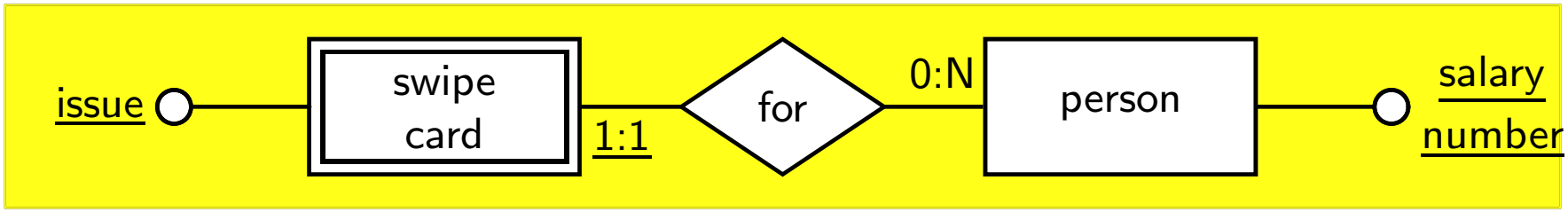
- In $ER^{\mathcal{D}}$: the disjointness of entities $E_1 \ldots E_n$ may be specified, enforcing that $\forall x, y.x \neq y \rightarrow E_x \cap E_y = \emptyset$
- The notion of **generalisation hierarchies** combines the use of disjointness and subset.
- **disjoint specialisation of nouns** $\rightarrow$ generalisation

Identifying generalisation hierarchies in $ER^{\mathcal{SD}}$

Employees may also be divided, according to how they like to receive messages, into email users and non-email users. The former must have a email address recorded, the later must have a pigeon hole number recorded.

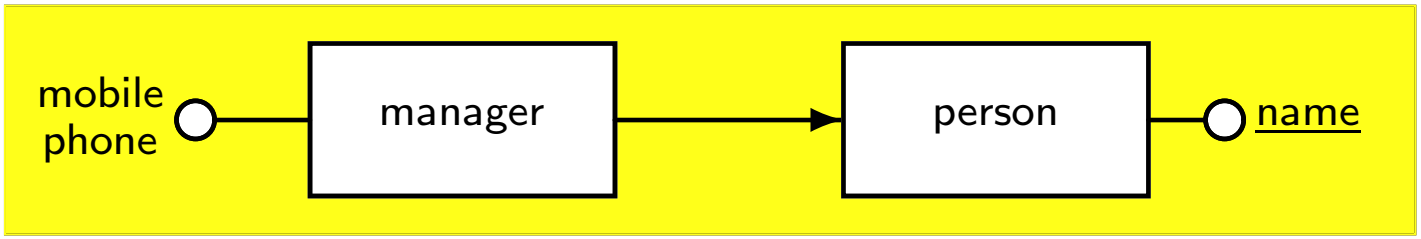


$ER^{\mathcal{W}}$: Weak entities



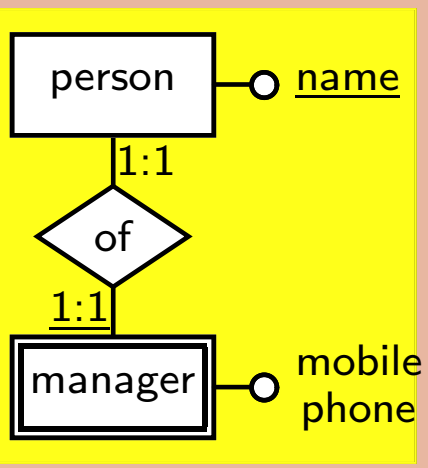
- If we allow the partipation of an entity in a relationship to be part of the entity key, we have a **weak entity**

Quiz 4: Subsets and weak entities

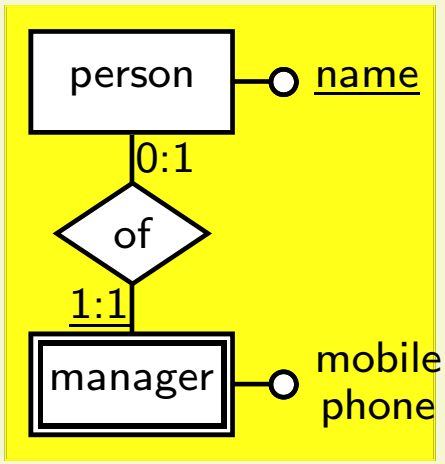


Which of the following is equivalent to the schema above?

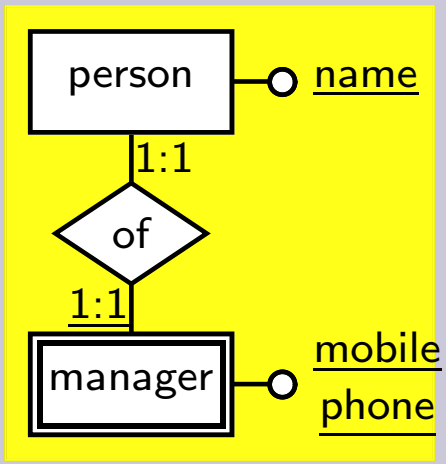
A



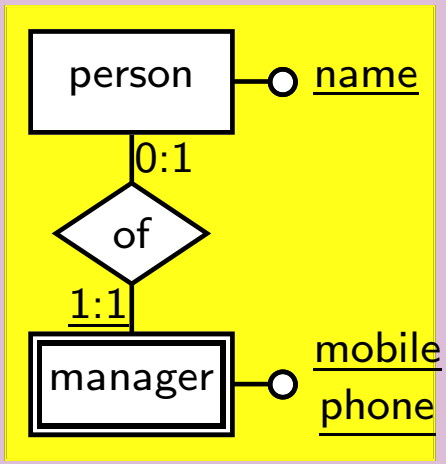
B



C



D

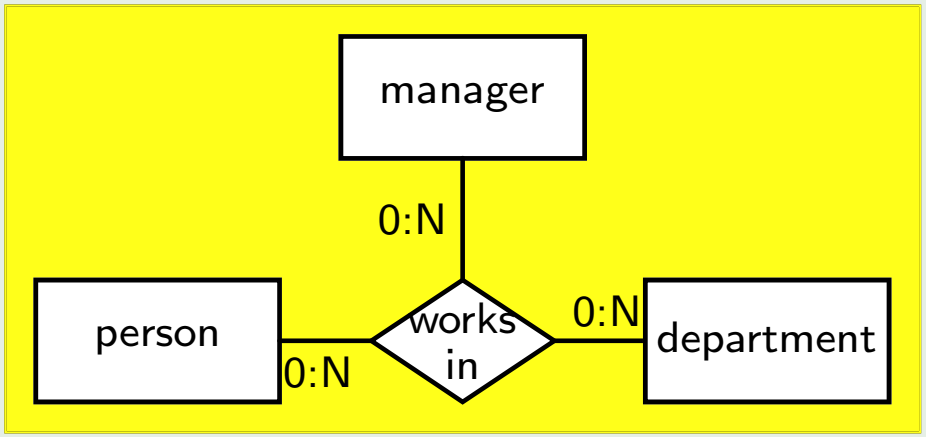


ER$^{\mathcal{H}}$: Allowing an n-ary relationship

- In graph theory, an edge connecting more that two nodes is called a **hyper-edge**.
- In ER$^{\mathcal{H}}$: allow n-ary relationships between entities, rather than just binary
- An n-ary relationship is equivalent to a weak entity with n binary relationships

Identifying an n-ary relationship

A person may work in multiple departments, and for each department that person works in, will be assigned a manager

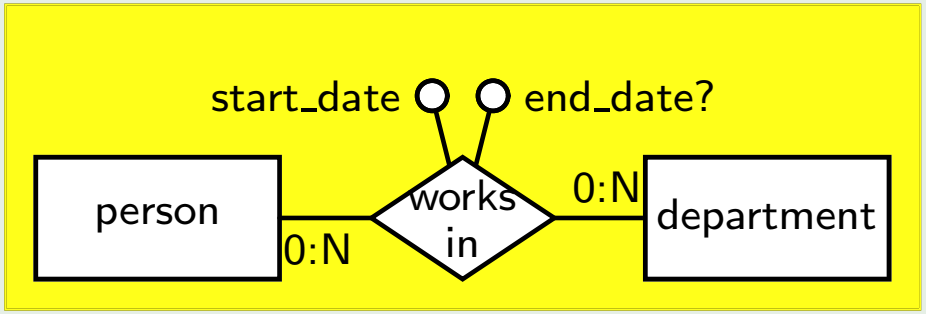


ER$^{\mathcal{A}}$: Allowing attributes on relationship

- Use when there are values to be associated with the relationship between entities

Identifying an attribute of a relationship

We record the **start_date** *when a person joined a department, and when the person leaves, record the* **end_date** *they left the department. We keep a history of all departments the person worked in.*

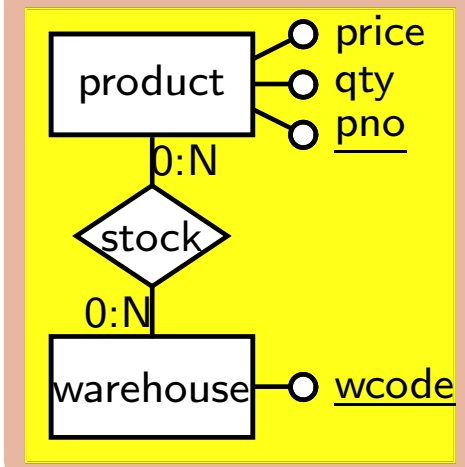


Quiz 5: Appropriate use of attributes on relationships

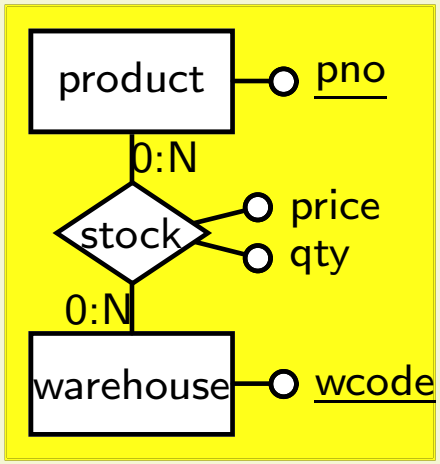
In the stock control system, we identify products by the **pno**, *and keep our stock in a number of warehouses identified by* **wcode**. *We record single* **price** *of each product, and the quantity* **qty** *of product we keep in each warehouse.*

Which of the following best models the above domain?

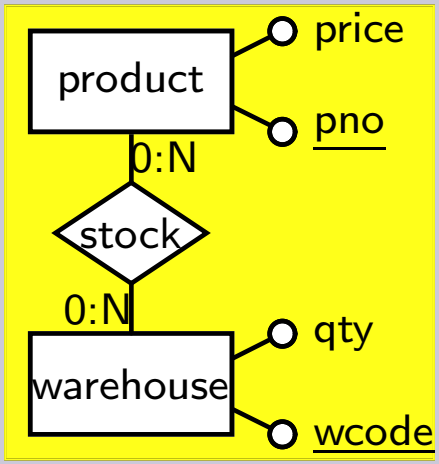
A



B



C



D