Data Modeling
CMPSCI 445

Fall 2008

Exercise

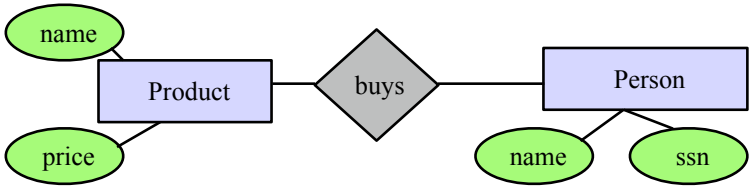
- Design a schema to store the data used by iTunes (or a similar application)
- Things to keep in mind:
  - Can all necessary information be represented?
  - Are properties represented in more than one place?
  - Are common operations efficient?

iTunes information

- Database includes: songs, artists, albums, album artwork, playlists
- Sample fields include: song name, artist, year, albumName, trackNum, genre, lastPlayedDate, playCount, albumArtwork, songRating, albumRating?, albumReleaseDate, artistDOB
- playLists: an ordered list of songs

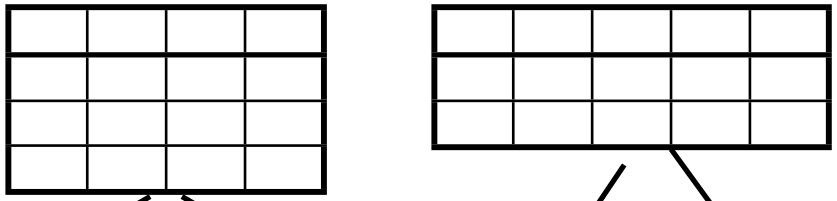
Relational Schema Design

Conceptual Design



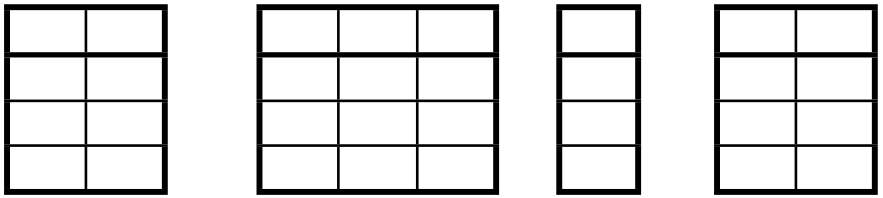
ER Model

Logical design



Relational Schema plus Integrity Constraints

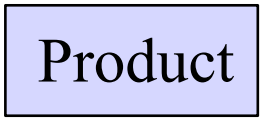
Schema Refinement



Normalized schema

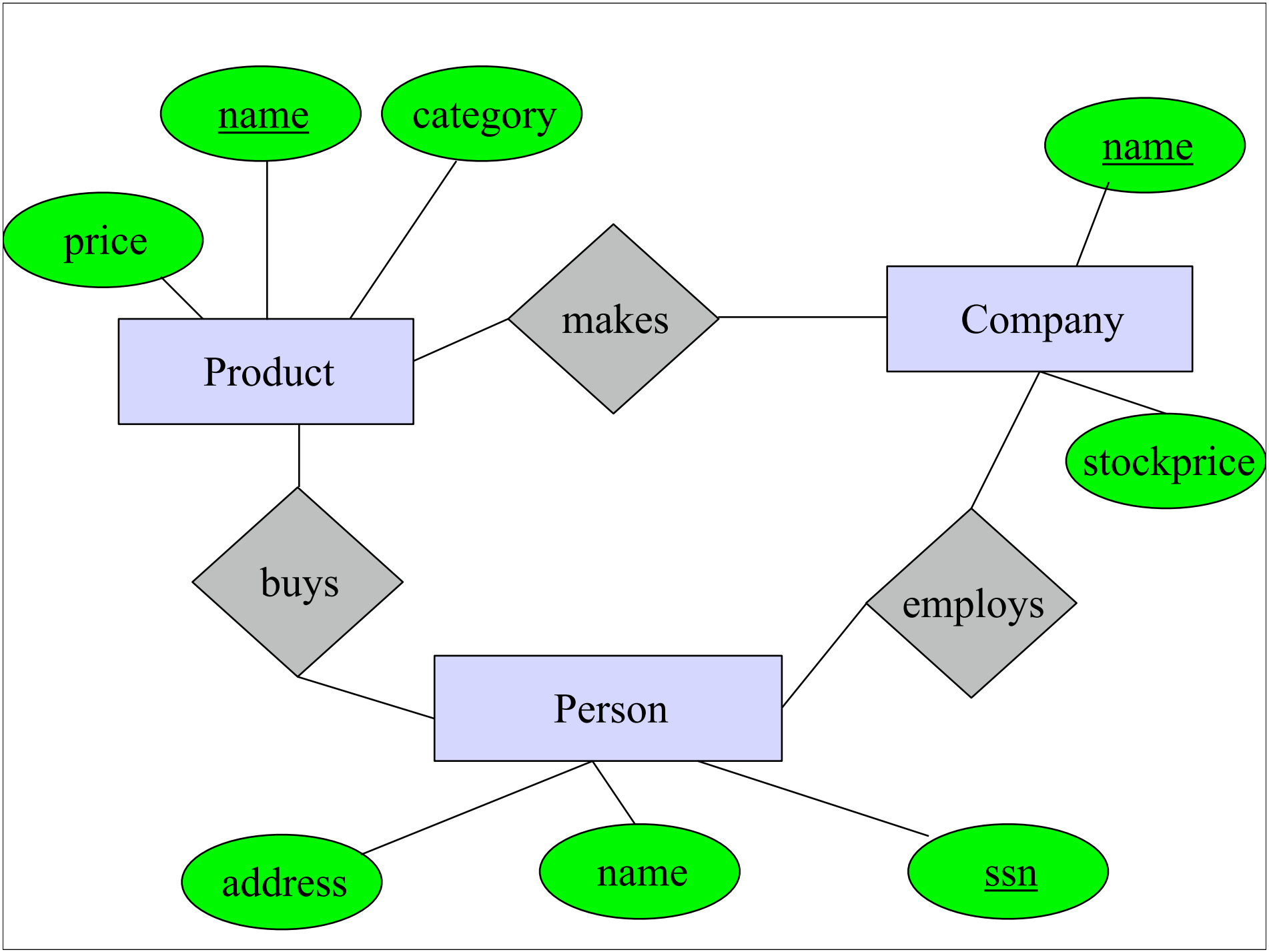
Entity / Relationship Diagrams

Entity sets



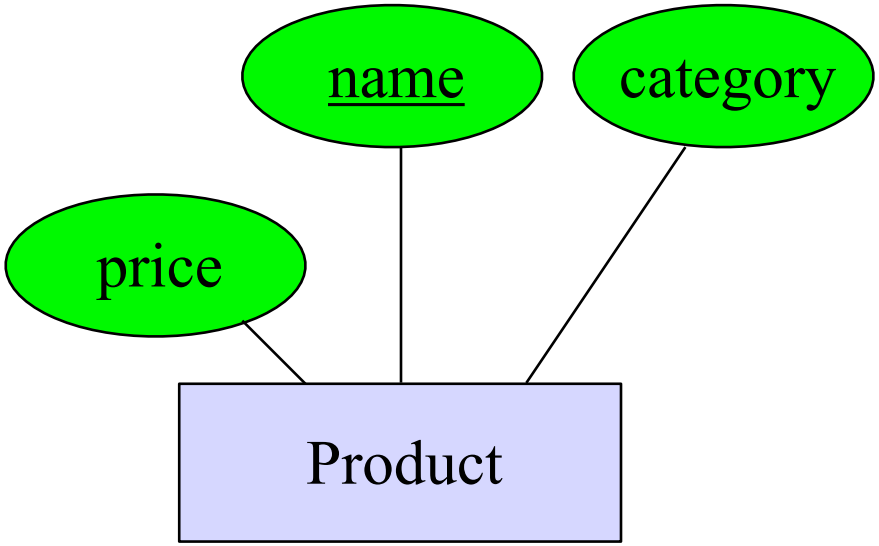
Attributes

Relationships


name
category
price
Product
makes
name
Company
stockprice
buys
employs
Person
address
name
ssn

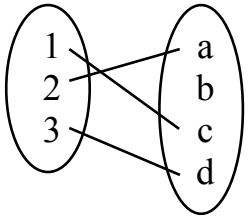
Keys in E/R Diagrams

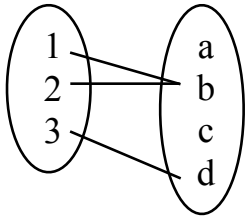
- Every entity set must have a key

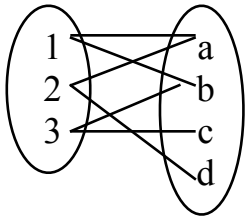


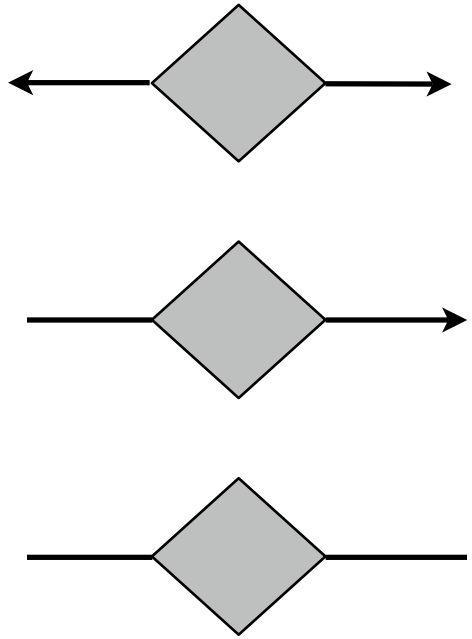
Multiplicity of E/R Relations

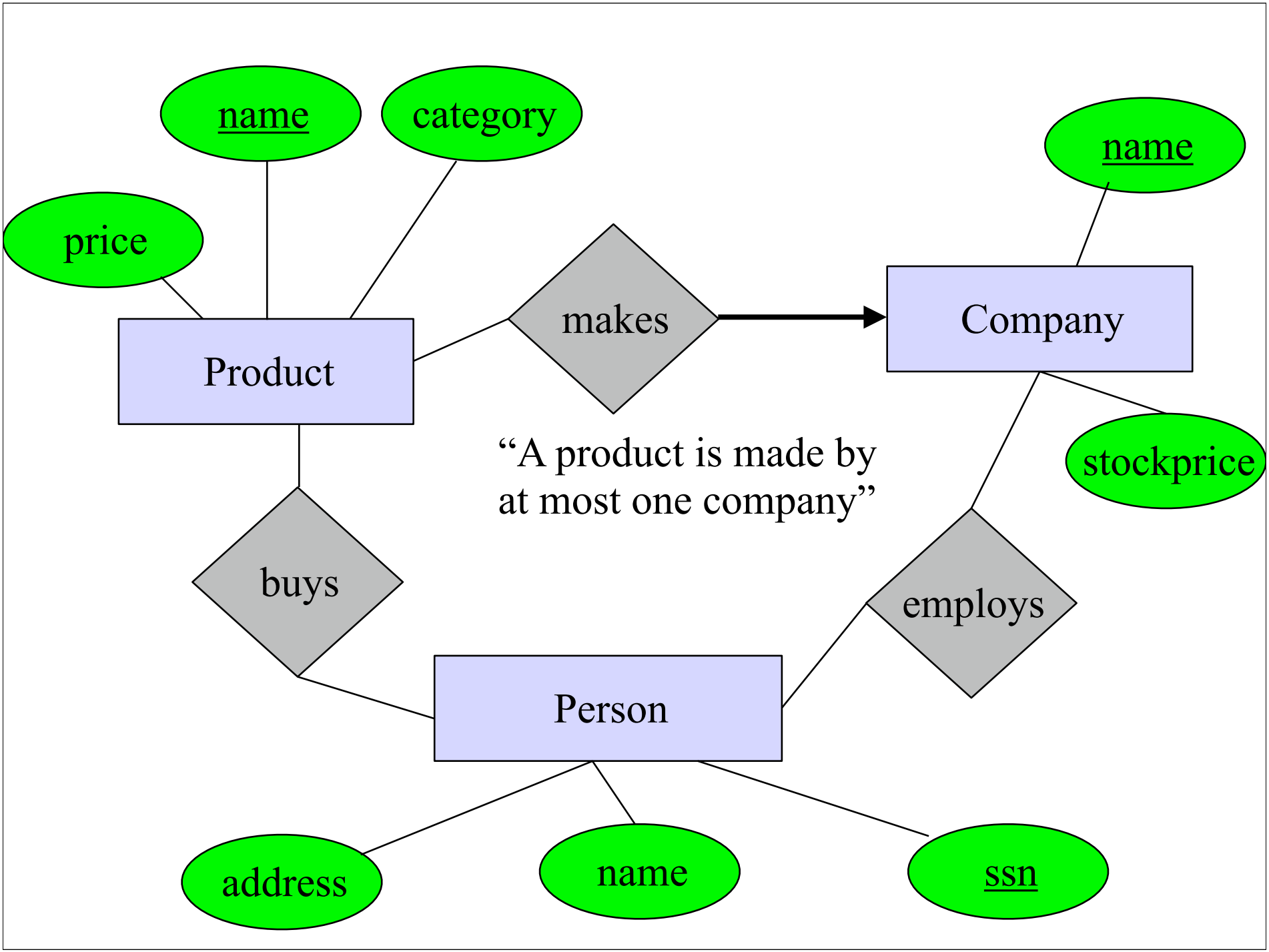
- one-one:
- many-one
- many-many



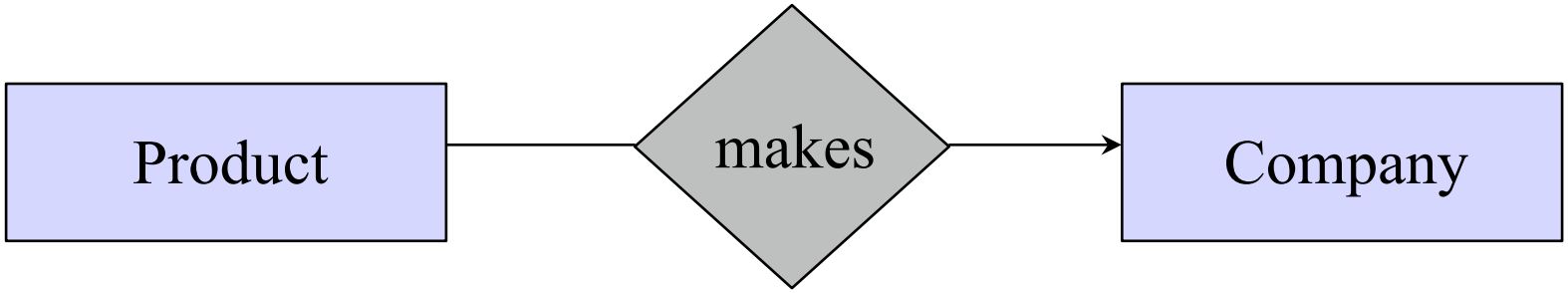




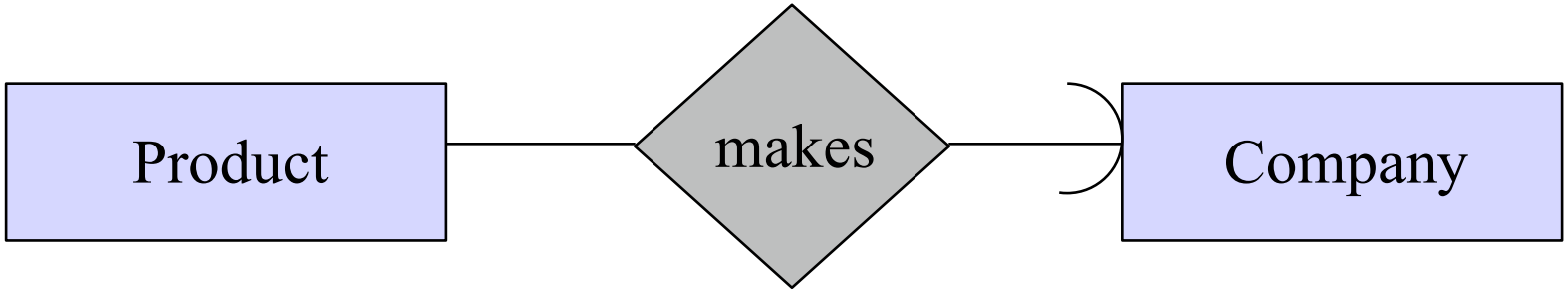



name
category
price
name
makes
Product
Company
"A product is made by
at most one company"
stockprice
buys
employs
Person
address
name
ssn

Referential Integrity Constraints

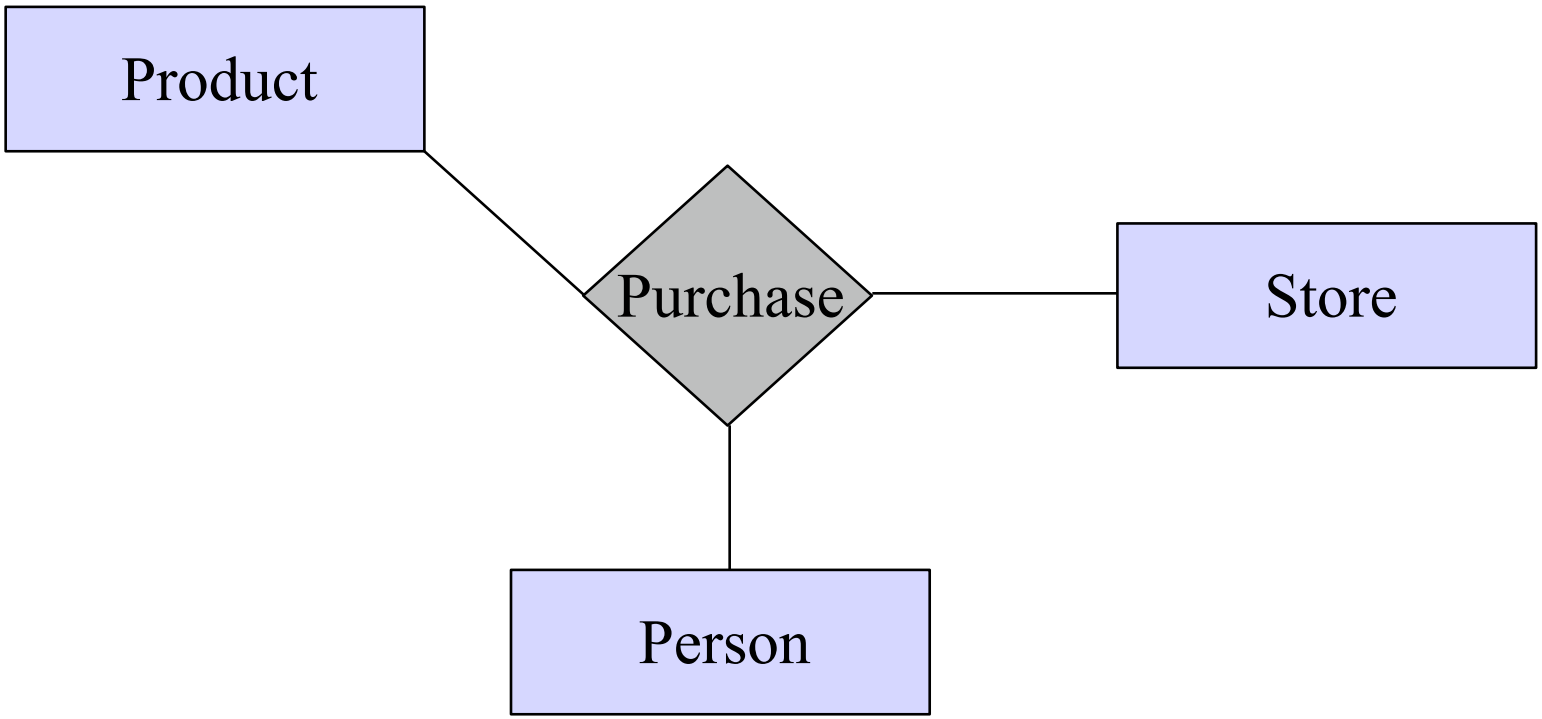


Each product made by at most one company.
(Some products made by no company)



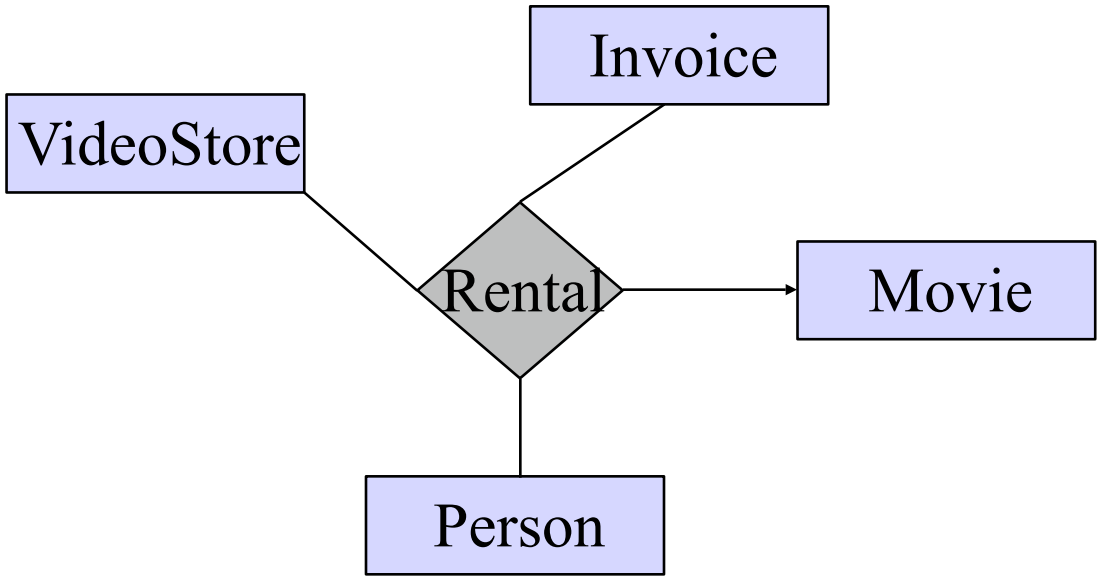
Each product made by *exactly* one company.

Multi-way Relationships



Arrows in Multiway Relationships

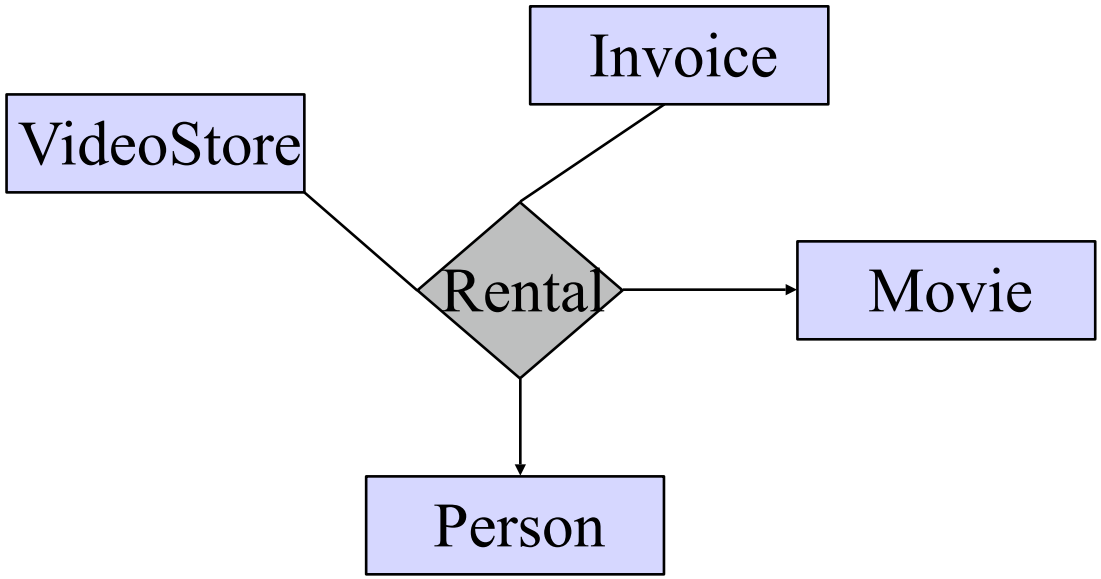
Q: what does the arrow mean ?



A: if I know the store, person, invoice, I know the movie too

Arrows in Multiway Relationships

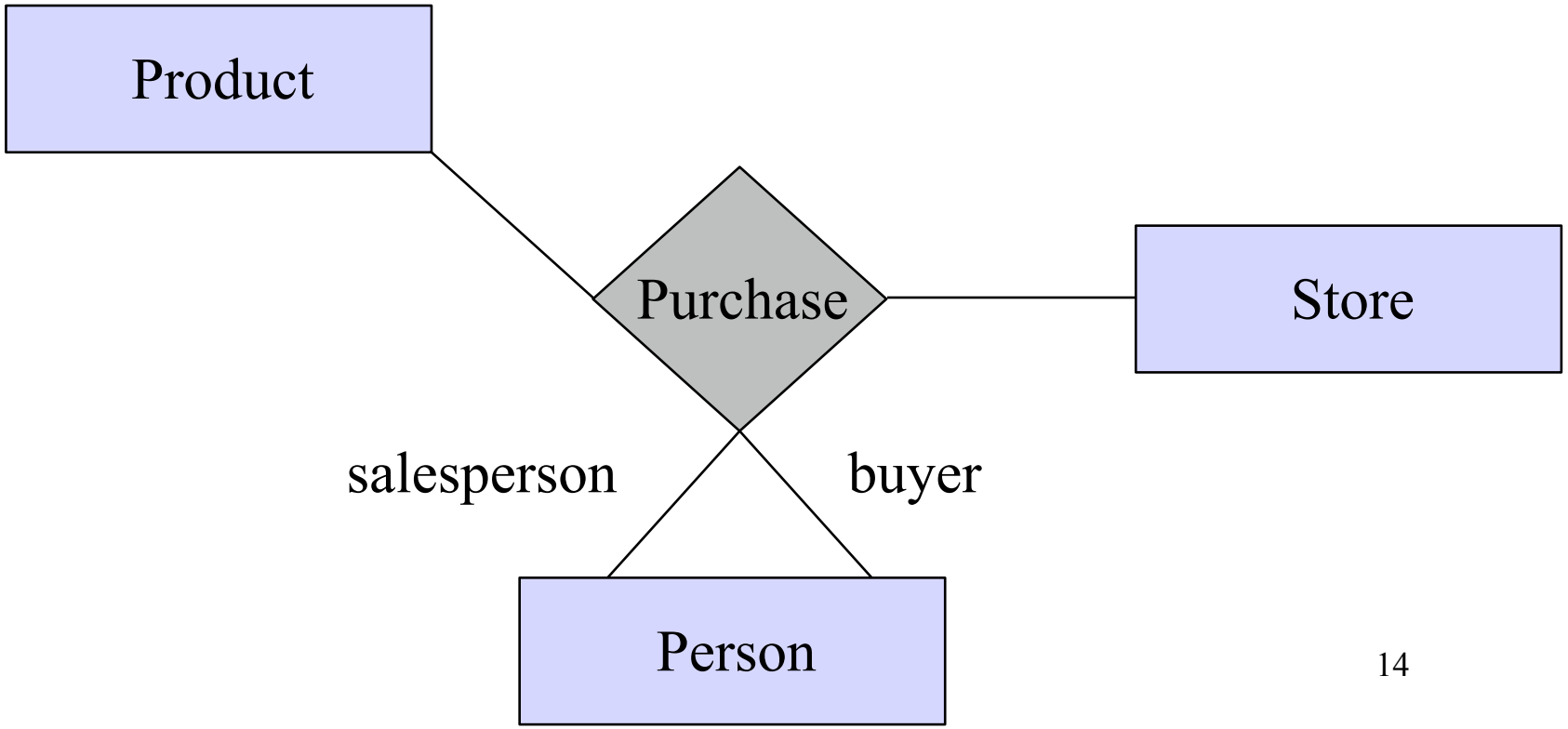
Q: what do these arrow mean ?



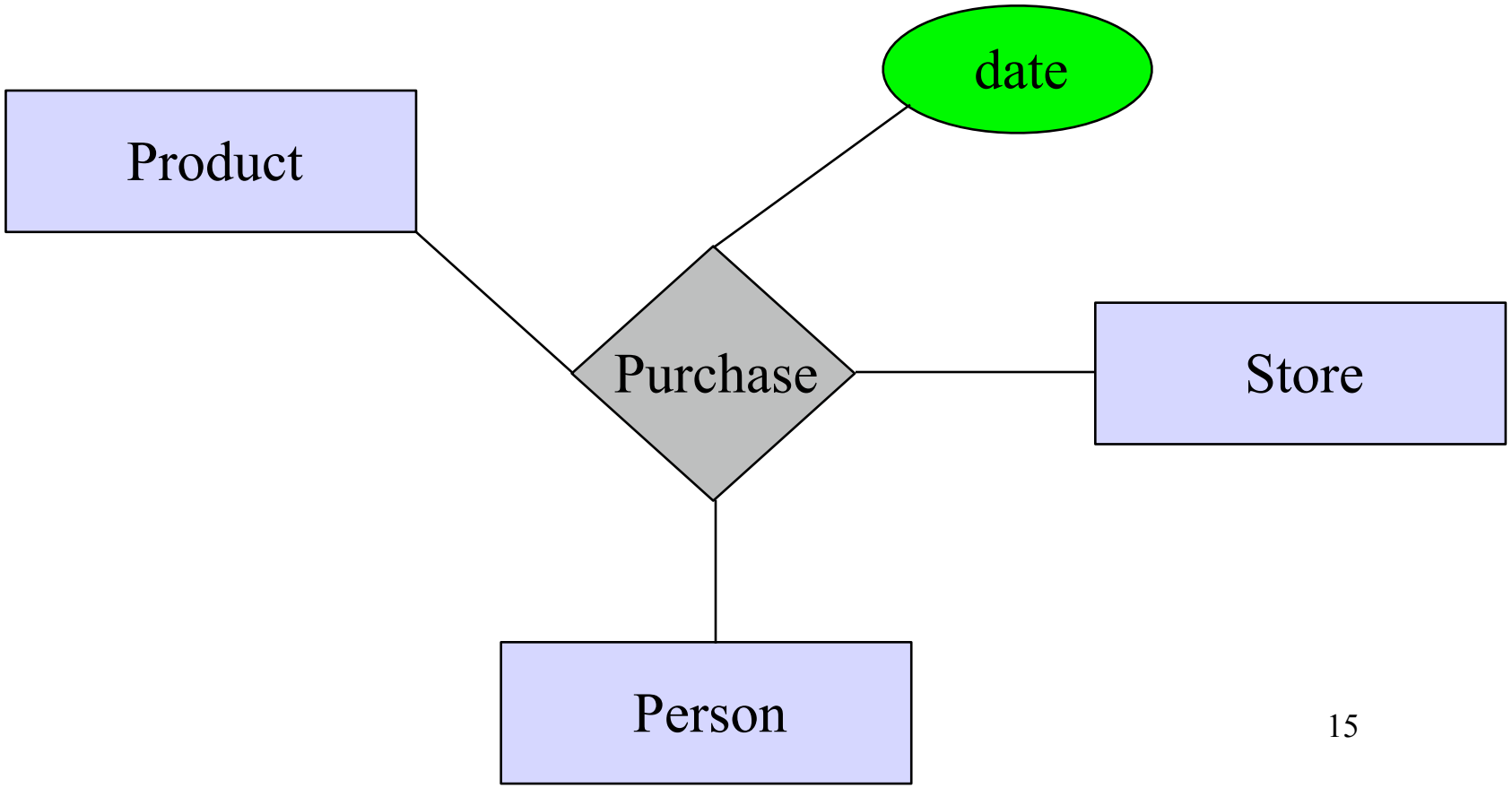
A: store, person, invoice determines movie and store, invoice, movie determines person

Roles in Relationships

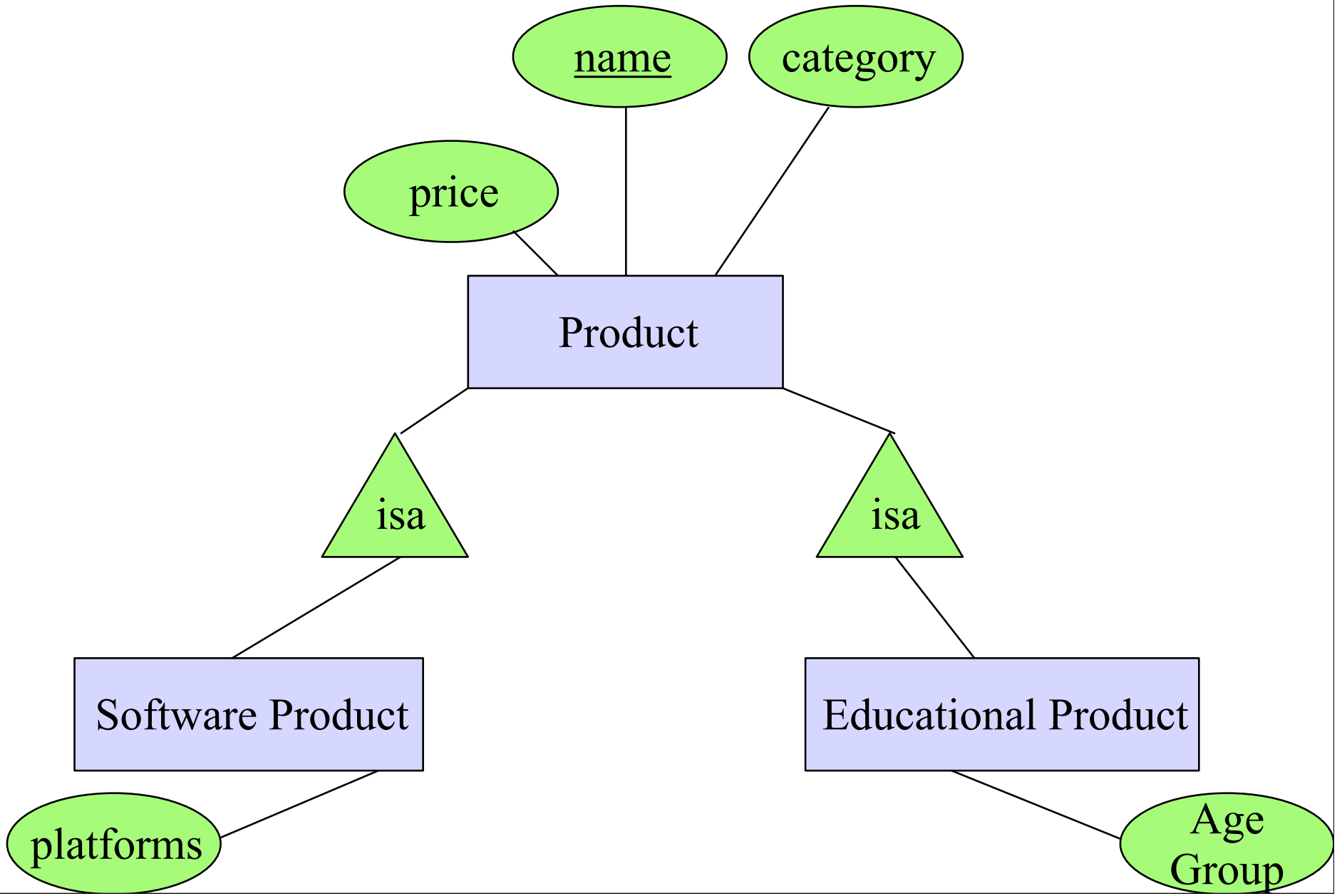
What if we need an entity set twice in one relationship?



Attributes on Relationships

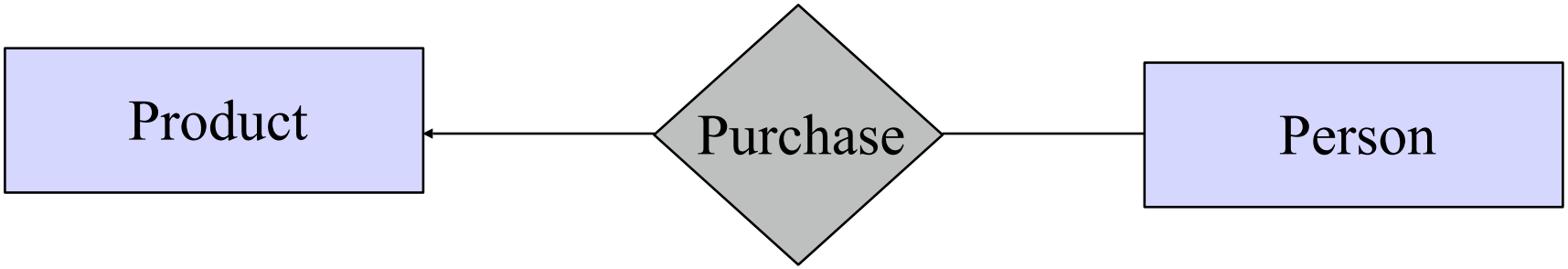


Subclasses

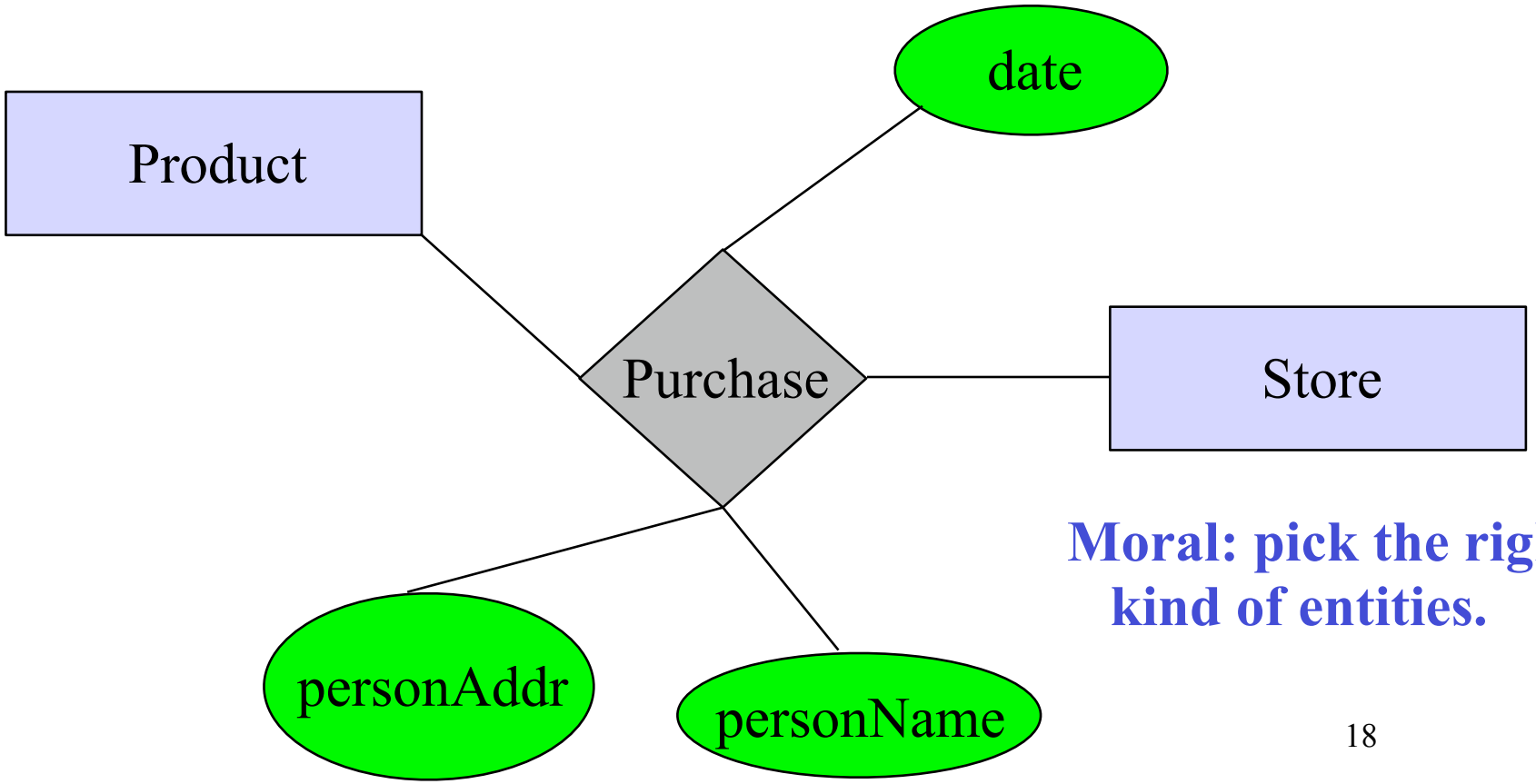

name
category
price
Product
isa
isa
Software Product
Educational Product
platforms
Age Group

Design Principles

What's wrong?

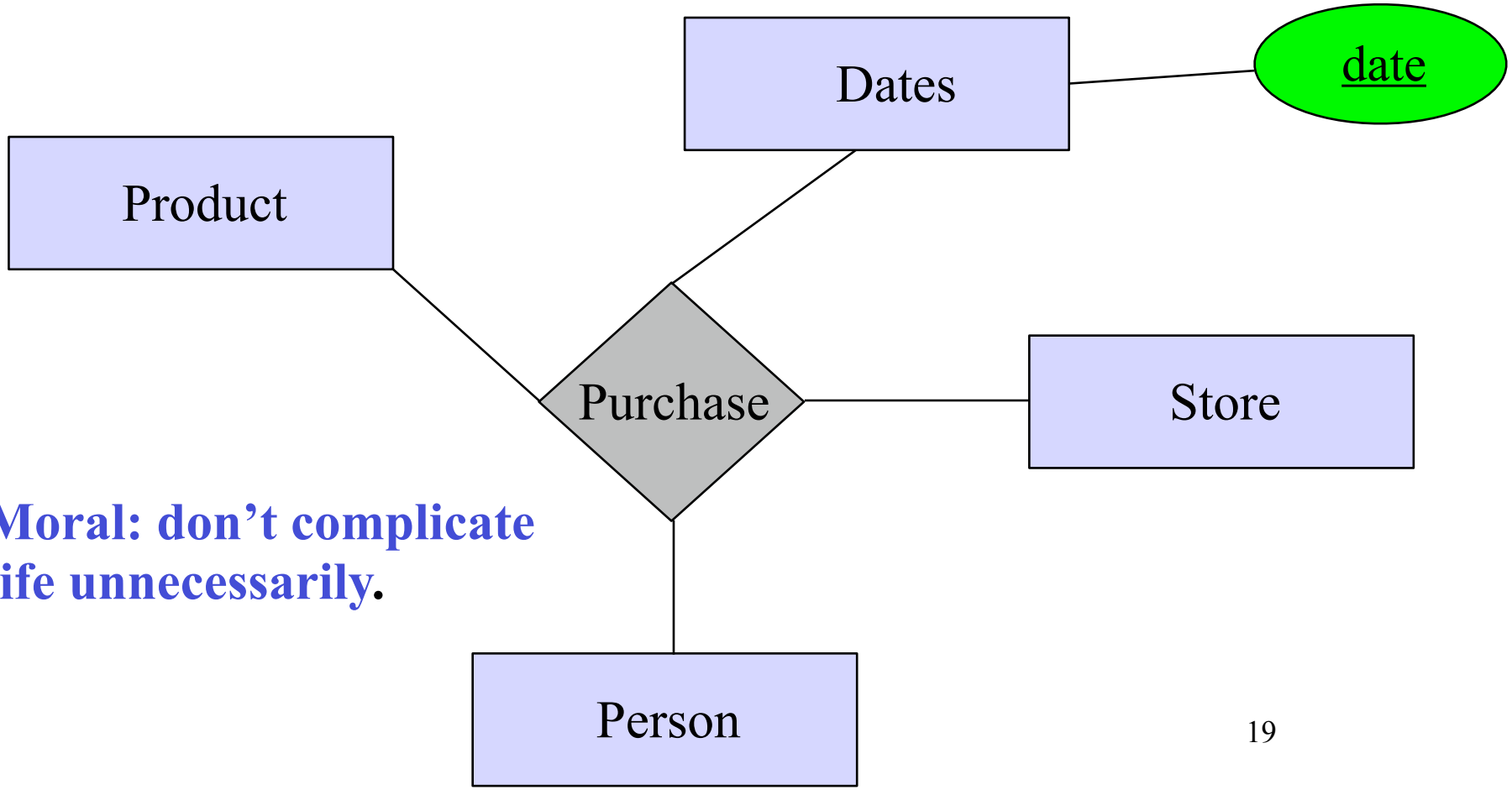


Design Principles: What's Wrong?



Moral: pick the right kind of entities.

Design Principles: What's Wrong?

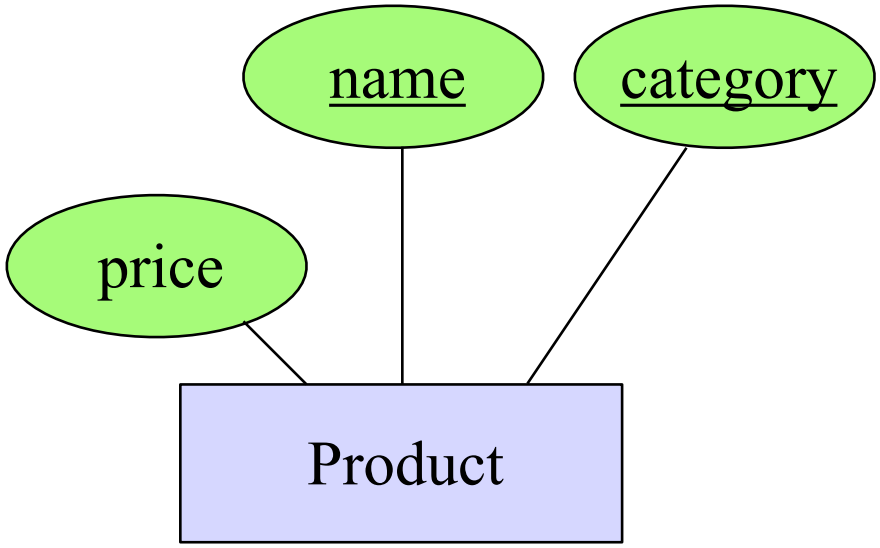


Moral: don't complicate life unnecessarily.

From E/R Diagrams to a Relational Schema

- Entity set → relation
- Relationship → relation

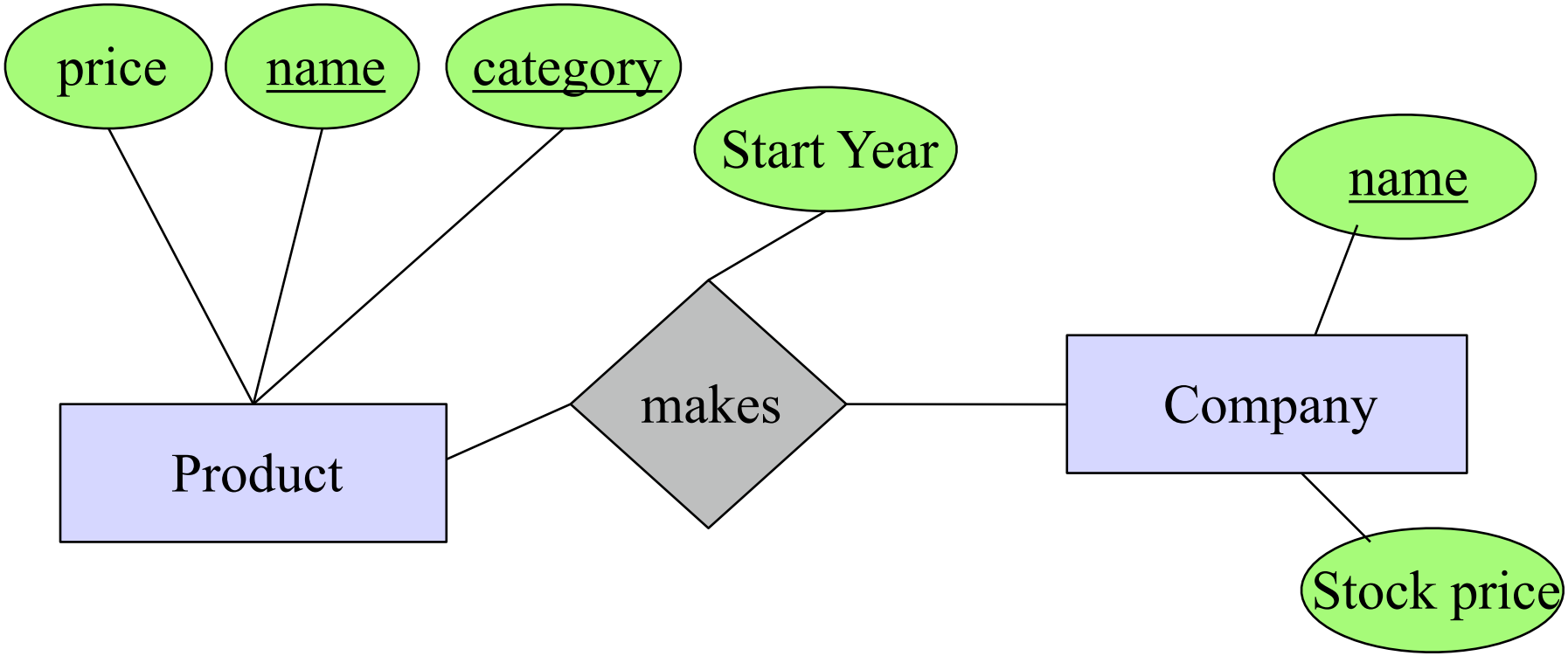
Entity Set to Relation



Product(name, category, price)

name	category	price
gizmo	gadgets	$19.99

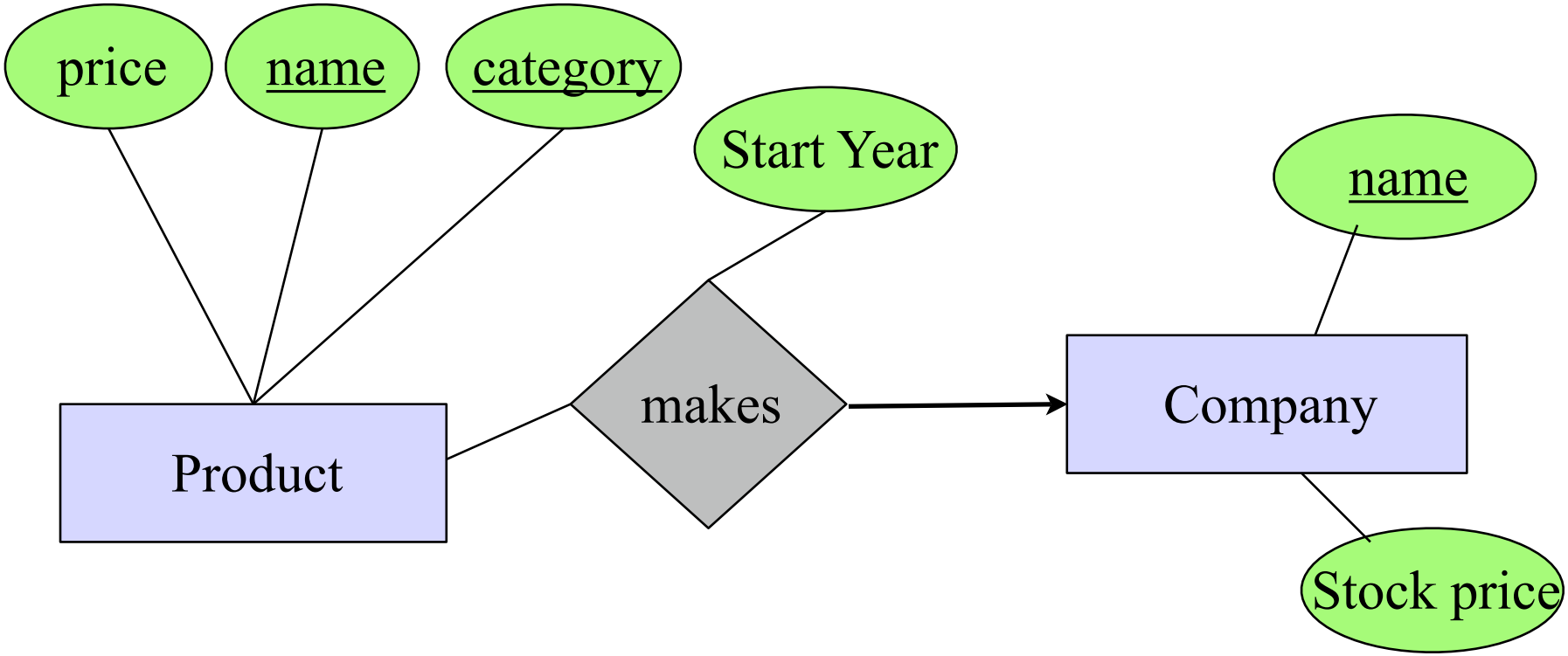
Relationships to Relations



Makes(product-name, product-category, company-name, year)

(watch out for attribute name conflicts)

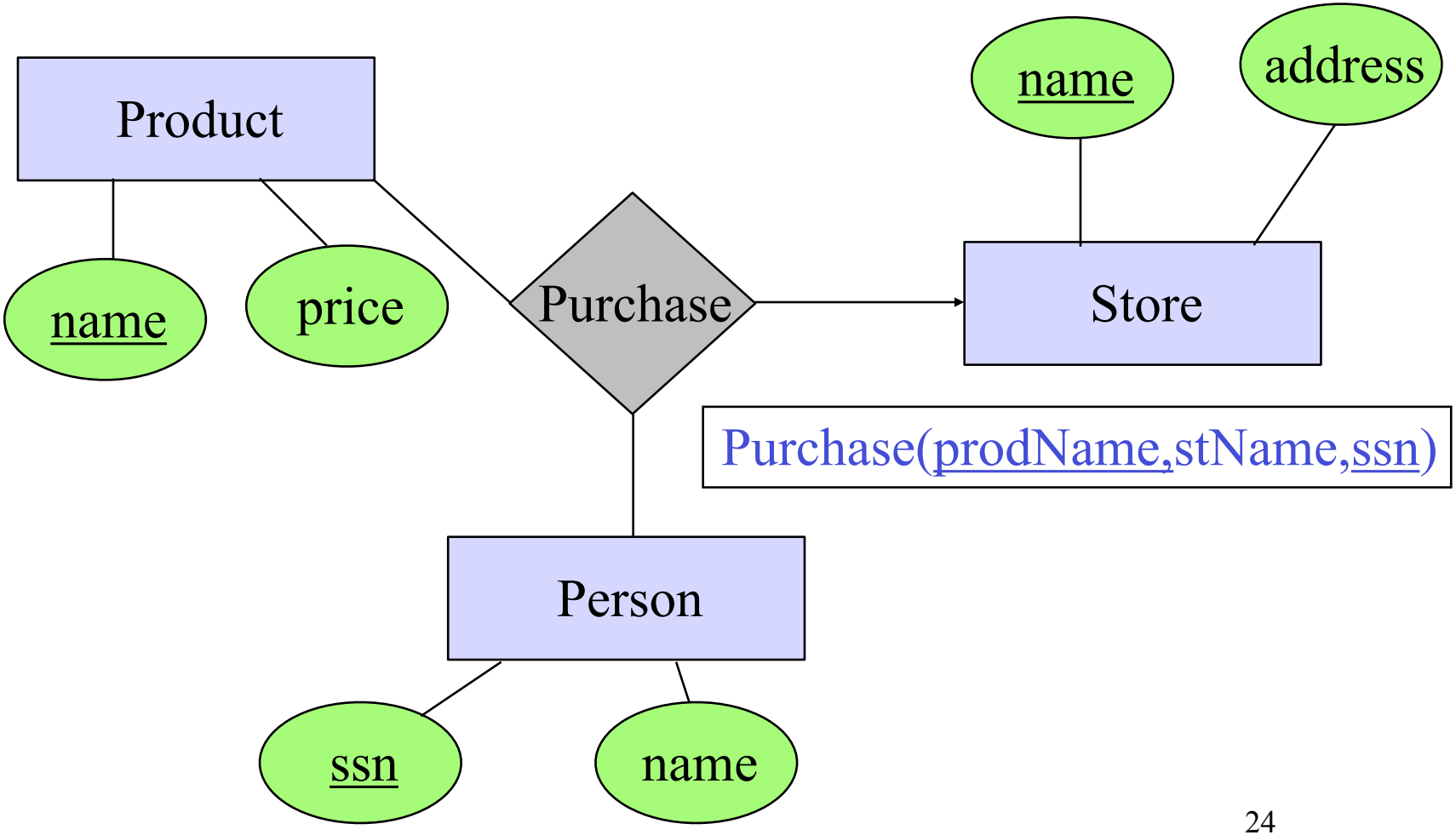
Relationships to Relations

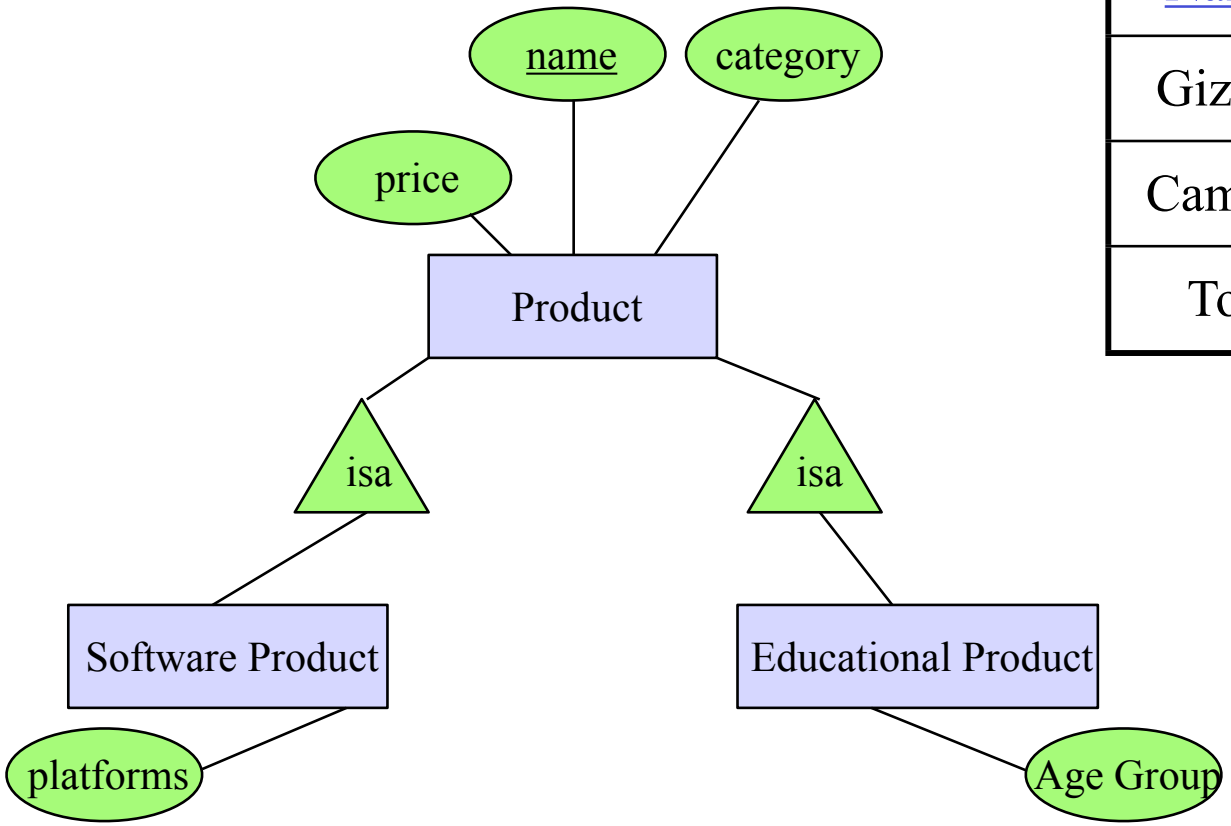


No need for **Makes**. Modify **Product**:

Product(name, category, price, startYear, companyName)

Multi-way Relationships to Relations





Product

Name	Price	Category
Gizmo	99	gadget
Camera	49	photo
Toy	39	gadget

Sw.Product

Name	platforms
Gizmo	unix

Ed.Product

Name	Age Group
Gizmo	todler
Toy	retired

Normalization

Evils of Redundancy

- *When a database schema is poorly designed we get anomalies.*
- *Redundancy* is at the root of several problems associated with relational schemas:

Redundant storage: data is repeated

Update anomalies: need to change in several places

Insertion anomalies: may not be able to add data we want to

Deletion anomalies: may lose data when we don't want to

Anomalies

Hourly_emps(ssn, name, lot, rating, hourly_wages, hours_worked)

Suppose hourly wages is determined by rating.

rating → hourly_wages

ssn	name	lot	rating	hourly_wages	hours_worked
29	brutus	48	8	10	40
85	art	22	8	10	30
95	bob	35	5	7	30
96	frodo	35	5	7	32
22	dustin	35	8	10	40

- Redundant storage: association between rating 8 and hourly wages 10 repeated 3 times.
- Update anomalies: hourly_wages updated in first tuple but not second
- Insertion anomalies: must know hourly_wage for rating value
- Deletion anomalies: delete all tuples with certain rating value, lost assoc.

Can null values fix problems?

- Not really.
- Insertion anomaly:
  - What if we know rating and hourly_wages for some rating, but there is no employee with that rating?
  - No. ssn can't be null.
- Deletion anomaly:
  - If last employee with some rating and hourly_wages value is deleted, replace with nulls?
  - No. ssn can't be null.

Schema Refinement

- Integrity constraints, in particular *functional dependencies*, can be used to identify schemas with such problems and to suggest refinements.
- Main refinement technique: *decomposition* (replacing ABCD with, say, AB and BCD, or ACD and ABD).
- Decomposition should be used judiciously:
  - Is there reason to decompose a relation?
  - What problems (if any) does the decomposition cause?

Data Anomalies

Persons may have several phones:

Name	SSN	PhoneNumber	City
Fred	123-45-6789	206-555-1234	Seattle
Fred	123-45-6789	206-555-6543	Seattle
Joe	987-65-4321	908-555-2121	Westfield

Anomalies:

- Redundancy = repeat data
- Update anomalies = Fred moves to “Bellevue”
- Deletion anomalies = Joe deletes his phone number:
  what is his city ?

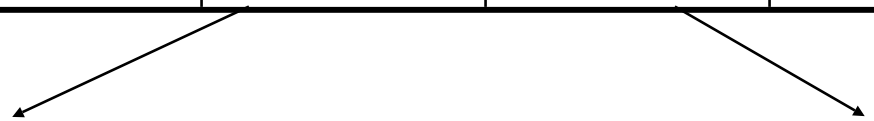
SSN → Name, City

but not SSN → PhoneNumber

Relation Decomposition

Break the relation into two:

Name	SSN	PhoneNumber	City
Fred	123-45-6789	206-555-1234	Seattle
Fred	123-45-6789	206-555-6543	Seattle
Joe	987-65-4321	908-555-2121	Westfield



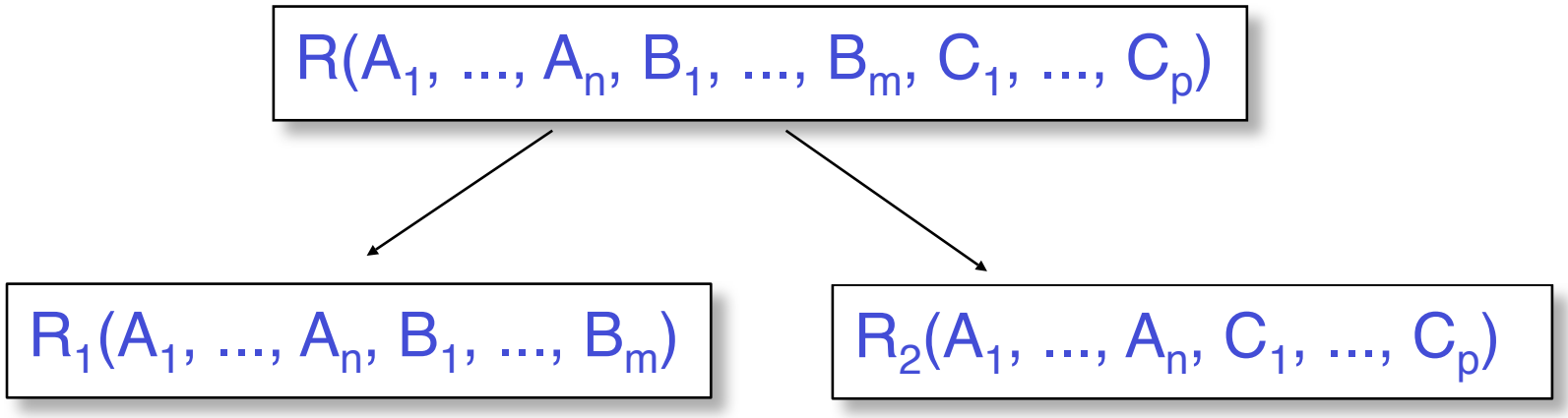
Name	SSN	City
Fred	123-45-6789	Seattle
Joe	987-65-4321	Westfield

SSN	PhoneNumber
123-45-6789	206-555-1234
123-45-6789	206-555-6543
987-65-4321	908-555-2121

Anomalies have gone:

- No more repeated data
- Easy to move Fred to “Bellevue” (how ?)
- Easy to delete all Joe’s phone number (how ?)

Decompositions in General



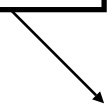
R_1 = projection of R on A_1, ..., A_n, B_1, ..., B_m

R_2 = projection of R on A_1, ..., A_n, C_1, ..., C_p

Lossless Decomposition

- Sometimes it is correct:

Name	Price	Category
Gizmo	19.99	Gadget
OneClick	24.99	Camera
Gizmo	19.99	Camera



Name	Price
Gizmo	19.99
OneClick	24.99

Name	Category
Gizmo	Gadget
OneClick	Camera
Gizmo	Camera

Lossy Decomposition

- Sometimes it is not:

Name	Price	Category
Gizmo	19.99	Gadget
OneClick	24.99	Camera
Gizmo	19.99	Camera

What's wrong ??

Name	Category
Gizmo	Gadget
OneClick	Camera
Gizmo	Camera

Price	Category
19.99	Gadget
24.99	Camera
19.99	Camera