Outline

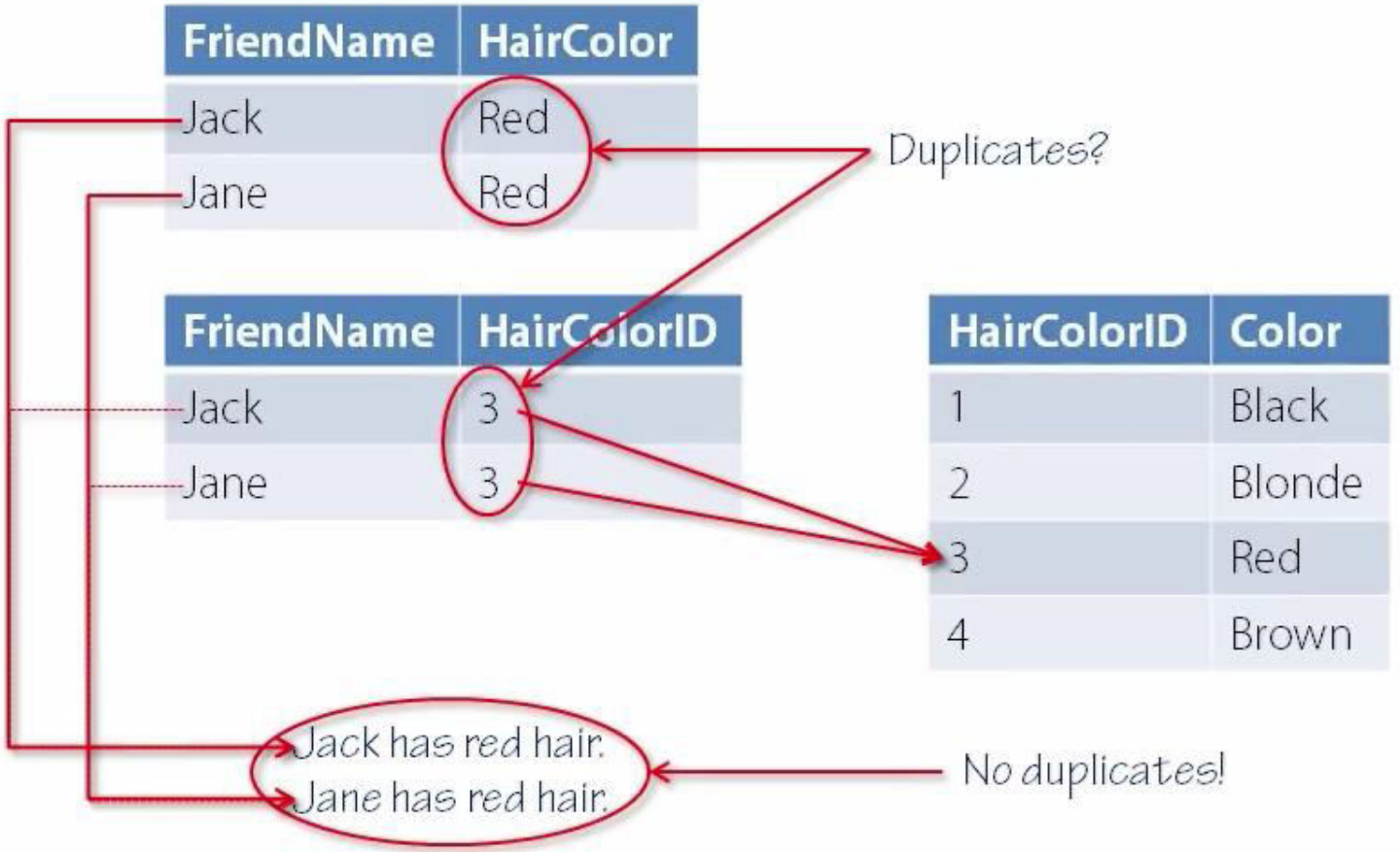
- **Normalization**
- **Functional dependencies**
- **Basic normal forms**
  - First Normal Form (1NF)
  - Second Normal Form (2NF)
  - Third Normal Form (3NF)
- **Finding functional dependencies**

Why normalize?

- **Non-atomic values**
  - Complex code required
  - Performance impact
- **Redundancy**
  - Same fact stored multiple times
  - Storage space wasted
  - Performance impact
  - Possibility of conflicting data
  - Derived facts: special case of redundancy

Redundancy: Misconceptions

- Repeating a value is not redundant



Redundancy: Misconceptions

- Repeating a value is not redundant

FriendName	HairColor
Jack	Red
Jane	Red

No duplicates!

FriendName	HairColorID
Jack	3
Jane	3

HairColorID	Color
1	Black
2	Blonde
3	Red
4	Brown

Redundancy: Misconceptions

- **Not all redundancy is bad**
  - Redundancy can help performance
  - Derived data may be impossible to derive again later
  - Derived data may be too expensive to derive every time

- **Uncontrolled redundancy *IS* bad!**
  - Mark duplicated data as such
  - Mark derived data as such
  - Prevent inconsistent data

Why normalize?

- **Non-atomic values**
  - Complex code required
  - Performance impact
- **Redundancy**
  - Same fact stored multiple times
  - Storage space wasted
  - Performance impact
  - Possibility of conflicting data
  - Derived facts: special case of redundancy
- **Modification anomalies**
  - Design causes modifications to have unwanted side effects

Modification anomalies

Tournament	Player Name	Player Phone
2012 Christmas Tournament	Dave	**801-555-0124**
2013 Midsummer Tournament	Dave	801-555-0123
2013 Midsummer Tournament	Joanna	801-555-9007
…	…	…

Modification anomalies

Tournament	Player Name	Player Phone
2012 Christmas Tournament	Dave	**801-555-0124**
2013 Midsummer Tournament	Dave	**801-555-0124**
…	…	…

How to normalize?

- **Steps**
  - First Normal Form (1NF)
  - Second Normal Form (2NF)
  - Third Normal Form (3NF)
  - Elementary Key Normal Form (EKNF)
  - Boyce-Codd Normal Form (BCNF)
  - Fourth Normal Form (4NF)
  - Fifth Normal Form (5NF)
  - Domain/Key Normal Form (DKNF)
  - Sixth Normal Form (6NF)
- **Normal forms apply to table**
- **Normal form of database = lowest normal form of all its tables**

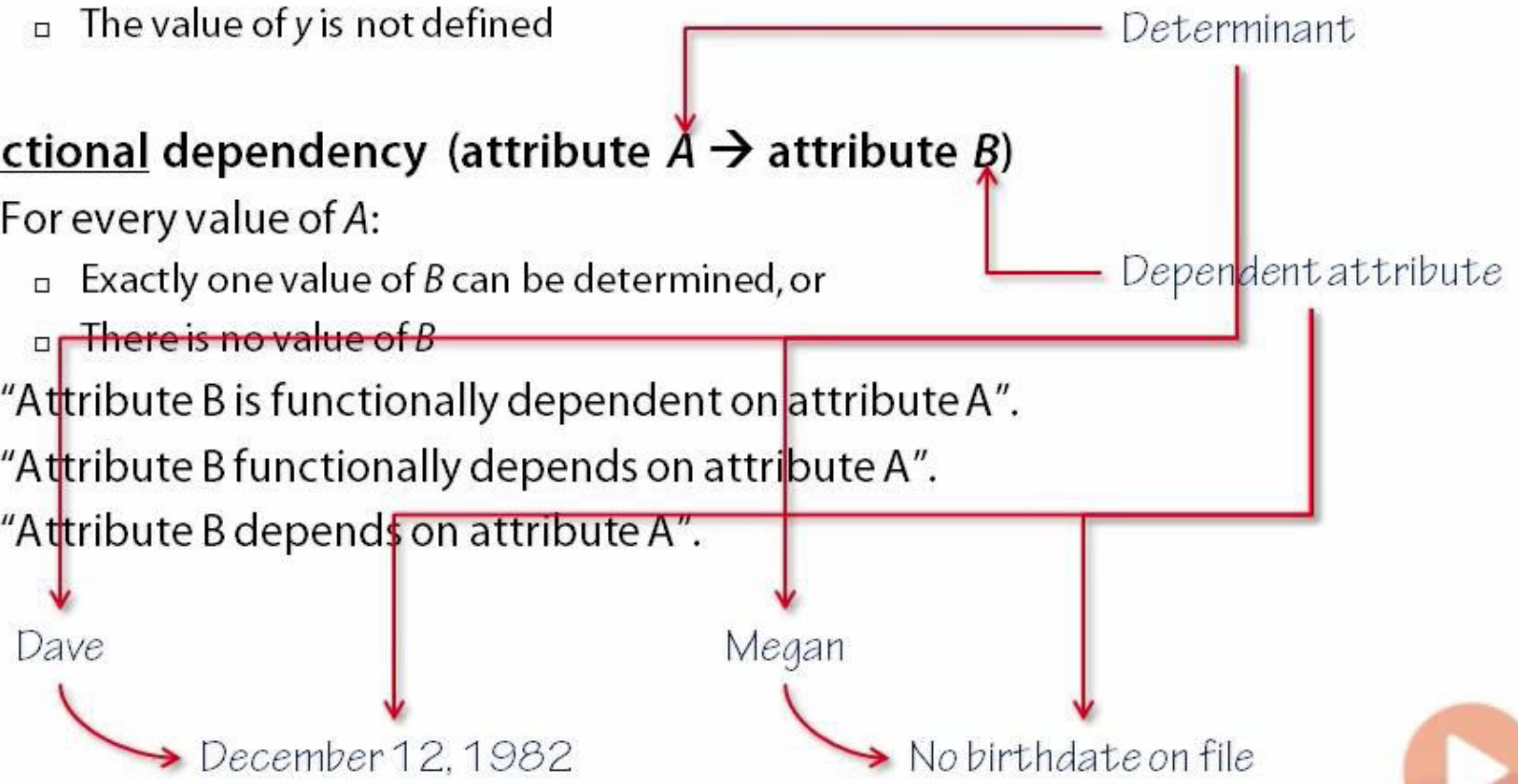
When to normalize?

- **Most common**
  - Convert Entity Relationship model to relational tables
  - Normalize relational tables
  - **Disadvantage:** Changes must be ported back to ER model

- **Alternative**
  - Normalize Entity Relationship model
  - Convert normalized ER model to relational tables
  - **Disadvantage:** Normalization is a bit more complicated
    - Normalize "every object that will eventually become a table"
    - For IDEF1X:
      - every entity type
      - every many-to-many relationship

Functional dependencies

- **Mathematical function** ($f(x) \rightarrow y$)
  - For every value of *x*:
    - Exactly one value of *y* can be computed, or
    - The value of *y* is not defined
- **Functional dependency (attribute *A* → attribute *B*)**
  - For every value of *A*:
    - Exactly one value of *B* can be determined, or
    - There is no value of *B*
  - "Attribute B is functionally dependent on attribute A".
  - "Attribute B functionally depends on attribute A".
  - "Attribute B depends on attribute A".



Properties of functional dependencies

- **Can be mutual**
  - Most are not!
- **Can be on a combination of two or more attributes**
- **Depend on "Universe of Discourse"**
  - Beware when making assumptions!
- **If *X* depends on *Y*, it also depends on each superset of *Y***
  - Dependency on two or more attributes can sometimes be reduced!
  - *Full* dependency: Functional dependency that cannot be reduced
- **Every attribute depends on itself (and on each superset of itself)**
  - *Trivial* dependency

Functional dependencies and normalization

- **Normalization uses functional dependencies that are:**
  - Non-trivial
  - Full
- **How to find all functional dependencies?**
  - *Most* are obvious
  - But what about the rest?
- **Guaranteed method**
  - Combines finding functional dependencies with normalization
  - Tedious and time-consuming; use only when needed

Functional dependencies and derived attributes

- Derived attributes may show up as functionally dependent
- These dependencies are different from "normal" dependencies

* Tax Amount *derives* from Gross Amount and Tax Rate