

## RELATION

### Binary Relation

**Relation.** A *binary relation* is a collection of ordered pairs. If there is no danger for confusion, we will simply call a binary relation a *relation*.

Thus, a relation is a set, each element of which is an ordered pair. If an ordered pair  $(x, y)$  belongs to a relation  $R$ , we write

$$(x, y) \in R.$$

We say that the items  $x$  and  $y$  stand in the relation  $R$ .

**Specify a relation by a collection of ordered pairs.** We specify a relation in the same way as we specify any set, by specifying the elements of the set.

Here is a collection of three ordered pairs:

$$\{(\text{red}, \text{blue}), (\text{conservative}, \text{liberal}), (\text{elephant}, \text{donkey})\}.$$

Do we call this collection a relation? Yes, any collection of any ordered pairs *defines* a relation. This set may or may not be a useful relation, but it surely fits the definition of a relation. We can, of course, try to discover the “meaning” of this relation, or the *property* that specifies this relation. But we do not have to do anything to call the set a relation.

**Specify a relation by a property.** This said, we do want to know the “meaning” of a relation. Like a useful set, a useful relation is often specified by a property, a rule that picks the ordered pairs in the relation.

Consider a collection of two ordered pairs:

$$\{(\text{Lisa}, \text{Eric}), (\text{Daniel}, \text{Michael})\}.$$

By talking to the people, we learn that this relation is specified by a property: each ordered pair is a pair of siblings.

We also learn that Lisa and Daniel stand in another relation: they are married. Among the four people, the relation of marriage has a single element: the ordered pair

$$(\text{Lisa}, \text{Daniel}).$$

Knowing these two relations, we can deduce another relation:

$$\{(\text{Lisa}, \text{Michael}), (\text{Daniel}, \text{Eric})\}.$$

The collection of ordered pairs specifies the siblings-in-law relation.

The people and their cities of birth also form a relation:

$$\{(\text{Lisa}, \text{Shanghai}), (\text{Daniel}, \text{Boston}), (\text{Eric}, \text{Detroit}), (\text{Michael}, \text{Santa Barbara})\}.$$

Here is yet another relation:

$$\{(\text{Lisa}, \text{fish}), (\text{Daniel}, \text{pork}), (\text{Eric}, \text{shrimp}), (\text{Michael}, \text{beef})\}.$$

We can guess the “meaning” of this relation: the four people have just ordered dishes in a restaurant.

**Specify a relation by the set-building notation.** Given a property that determines the membership of a relation, we can specify the relation using the set-building notation.

People and their countries of birth form a relation  $R$ , which can be defined using the set-building notation:

$$R = \{(x, y) \mid \text{person } x \text{ was born in country } y\}.$$

The property (phrase) “people and their countries of birth” does not tell us any people or any countries. Nor does it tell us who was born where. But the property is a call to build. We ask people where they were born, and build this set of ordered pairs.

The hyperlinks in webpages form a relation:

$$\{(x, y) \mid \text{webpage } x \text{ links to webpage } y\}.$$

This relation enables PageRank, an algorithm pioneered by Google to rank webpages.

A similar relation is

$$\{(x, y) \mid \text{paper } x \text{ cites paper } y\}.$$

This relation is significant to researchers.

**Unstructured set of ordered pairs.** In defining a relation, the order of the two items in each pair is significant, but the order in which all the pairs appear in the set is insignificant.

Consider the relation

$$L = \{(x, y) \mid \text{person } x \text{ likes person } y\}.$$

The fact “ $x$  likes  $y$ ” does not imply “ $y$  likes  $x$ ”. That is, the fact  $(x, y) \in L$  does not imply that  $(y, x) \in L$ .

For example, three people  $a$ ,  $b$ , and  $c$  can form a total of six ordered pairs:

$$(a, b), (a, c), (b, a), (b, c), (c, a), (c, b).$$

We find that only two pairs belong to the relation  $L$ :

$$L = \{(a, b), (c, b)\}.$$

The relation tells that  $a$  likes  $b$ , and  $c$  likes  $b$ . In particular, we find that

$$(b, a) \notin L, \quad (b, c) \notin L.$$

The relation  $L$  is the same if we change the order in which the two pairs appear in  $L$ :

$$\{(a, b), (c, b)\} = \{(c, b), (a, b)\}.$$

**Union and intersection.** A relation is a type of set. Given two relations, we can form their union and intersection.

Here are two examples:

$$\begin{aligned}\{(1,2), (a,b), (1,a)\} \cup \{(a,b), (2,b)\} &= \{(1,2), (a,b), (1,a), (2,b)\}, \\ \{(1,2), (a,b), (1,a)\} \cap \{(a,b), (2,b)\} &= \{(a,b)\}.\end{aligned}$$

**Subrelation.** A subset of a relation is called a *subrelation*. For example, the relation

$$\{(\text{red}, \text{blue}), (\text{conservative}, \text{liberal})\}$$

is a subrelation of the relation

$$\{(\text{red}, \text{blue}), (\text{conservative}, \text{liberal}), (\text{elephant}, \text{donkey})\}.$$

Compare two relations:

$$\begin{aligned}R &= \{(x, y) \mid \text{person } x \text{ was born in country } y\}, \\ Q &= \{(x, y) \mid \text{person } x \text{ has spent time in country } y\}.\end{aligned}$$

A person who was born in a country must have spent some time in the country. A person who has spent some time in a country may or may not be born in a country. Thus,  $R$  is a subset of  $Q$ . We write

$$R \subset Q.$$

We say that  $R$  is a subrelation of  $Q$ .

**Partition of relation.** Recall that a partition of a set  $S$  is a family of nonempty subsets of  $S$ , which are disjoint, and whose union is  $S$ . Given a relation, we can form its partitions.

Consider the relation

$$\{(x, y) \mid \text{person } x \text{ has spent time in country } y\}.$$

This relation can be partitioned into two subrelations:

$$\begin{aligned}&\{(x, y) \mid \text{person } x \text{ has spent time in an Asian country } y\}, \\ &\{(x, y) \mid \text{person } x \text{ has spent time in a non-Asian country } y\}.\end{aligned}$$

## Graph and Ranges

**Graph of a relation.** A collection of ordered pairs defines a relation  $R$ . Often we define a relation  $R$  by a property. We then call the actual collection of ordered pairs the *graph* of the relation  $R$ .

Thus,  $R$  stands for the relation, the property that defines the relation, and the graph of the relation.

**Examples.** Let us define a relation  $R$  by the property “people and their cities of birth”. We express this definition in the set-building notation:

$$R = \{(x, y) \mid \text{person } x \text{ was born in city } y\}.$$

The graph of the relation  $R$  is a collection of ordered pairs:

$$\{(Lisa, Shanghai), (Daniel, Boston), (Eric, Detroit), (Michael, Santa Barbara)\}.$$

The relation

$$\{(x, y) \mid \text{webpage } x \text{ links to webpage } y\}$$

consists of a huge number of ordered pairs. We are unable to list them on paper, but a search engine stores these ordered pairs on computers.

**Ranges of a relation.** Given a relation  $R$ , we can define two sets:

$$P = \{p \mid (p, q) \in R\},$$

$$Q = \{q \mid (p, q) \in R\}.$$

We call  $P$  and  $Q$  the first and second *ranges* of the relation  $R$ . We say that the relation  $R$  is *over* the two sets  $P$  and  $Q$ , and that the two sets  $P$  and  $Q$  are *under* the relation  $R$ .

**Examples.** The relation

$$\{(\text{red}, \text{blue}), (\text{conservative}, \text{liberal}), (\text{elephant}, \text{donkey})\}$$

has its two ranges:

$$\{\text{red}, \text{conservative}, \text{elephant}\},$$

$$\{\text{blue}, \text{liberal}, \text{donkey}\}.$$

The first range is a list of items commonly associated with the Republican Party of the United States. The second range is a list of items commonly associated with the Democratic Party of the United States.

The relation

$$\{(Lisa, Shanghai), (Daniel, Boston), (Eric, Detroit), (Michael, Santa Barbara)\}$$

has its two ranges:

$$\{Lisa, Daniel, Eric, Michael\},$$

$$\{Shanghai, Boston, Detroit, Santa Barbara\}.$$

The first range lists the four people, and the second range lists four cities.

The relation

$$\{(a, 1), (a, 2), (b, 3)\}$$

has its ranges:

$$\{a, b\},$$

$$\{1, 2, 3\}.$$

Note that item  $a$  appears in two ordered pairs in the relation, but should only appear once in the range.

**Graph and ranges.** The graph of the relation  $R$  is a subset of the Cartesian product of its ranges:

$$R \subset P \times Q.$$

If  $R$  is a nonempty set,  $P$  and  $Q$  are nonempty sets.

### Image

**Image.** Let  $R$  be a relation,  $P$  and  $Q$  be the ranges of  $R$ , and  $A$  be a subset of  $P$ . The *image* of  $A$  under  $R$  is the collection of the elements in  $Q$ , each of which stands in the relation  $R$  with at least one element in  $A$ .

Denote the image of  $A$  under  $R$  by  $\langle A, \square \rangle_R$ , and express the above definition in the set-building notation:

$$\langle A, \square \rangle_R = \{q \mid (p, q) \in R, p \in A\}.$$

The image of a subset  $A$  of  $P$  under the relation  $R$  is a subset of  $Q$ :

$$\langle A, \square \rangle_R \subset Q.$$

Similarly, the image of a subset  $B$  of  $Q$  under  $R$ , denoted as  $\langle \square, B \rangle_R$ , is the collection of the elements in  $P$ , each which stands in the relation  $R$  with at least one element in  $B$ . We can express this definition in the set-building notation:

$$\langle \square, B \rangle_R = \{p \mid (p, q) \in R, q \in B\}.$$

The image of a subset  $B$  of  $Q$  under the relation  $R$  is a subset of  $P$ :

$$\langle \square, B \rangle_R \subset P.$$

**Examples.** Consider the relation

$$R = \{(p, q) \mid \text{person } p \text{ has spent time in country } q\}.$$

The relation has its ranges: the set of all people ( $P$ ), and the set of all countries ( $Q$ ).

The phrase “people who have spent time in Asian countries” mentions two sets. The set of Asian countries, denoted as  $B$ , is a subset of  $Q$ . The set of people who have spent time in Asia is the image of  $B$  under the relation  $R$ . We denote this set by  $\langle \square, B \rangle_R$ .

**Identities of images.** Let  $R$  be a relation, and  $P$  and  $Q$  be the ranges of  $R$ . Let  $A$ ,  $A_1$  and  $A_2$  be subsets of  $P$ . When there is no danger of confusion, we drop the subscript  $R$  and denote the image of  $A$  under  $R$  by  $\langle A, \square \rangle$ .

You can confirm the following identities of images of subsets in  $P$  under relation  $R$ :

$$\begin{aligned}
\langle A, \square \rangle &\subset Q, \\
\langle P, \square \rangle &= Q, \\
\langle \emptyset, \square \rangle &= \emptyset, \\
A_1 \subset A_2 &\Rightarrow \langle A_1, \square \rangle \subset \langle A_2, \square \rangle, \\
\langle A_1 \cup A_2, \square \rangle &= \langle A_1, \square \rangle \cup \langle A_2, \square \rangle, \\
\langle A_1 \cap A_2, \square \rangle &\subset \langle A_1, \square \rangle \cap \langle A_2, \square \rangle.
\end{aligned}$$

Similar identities work for the images of subsets of  $Q$  under the relation  $R$ .

### Marketing Books\*

**Books and buyers.** A bookseller knows who has bought which book. These data define a relation:

$$R = \{(p, q) \mid \text{person } p \text{ has bought book } q\}.$$

This relation is the principal information to help the bookseller market books. For this purpose, the bookseller regards multiple copies of the same book as one book.

**Graph and ranges.** Consider a simple case of  $R$  being a set of eight ordered pairs:

$$R = \{(p_1, q_2), (p_1, q_3), (p_2, q_5), (p_3, q_4), (p_3, q_6), (p_3, q_1), (p_4, q_5), (p_5, q_6)\}.$$

The eight ordered pairs constitute the graph of the relation  $R$ .

The first range of the relation  $R$  is the set of five people:

$$P = \{p_1, p_2, p_3, p_4, p_5\}.$$

The second range of the relation  $R$  is a set of six books:

$$Q = \{q_1, q_2, q_3, q_4, q_5, q_6\}.$$

Both  $p_2$  and  $p_4$  have bought  $q_5$ , but we only list  $q_5$  once in  $Q$ .

**Images.** Inspecting the relation  $R$ , we find that two people

$$\{p_1, p_3\}$$

have bought five books:

$$\{q_1, q_2, q_3, q_4, q_6\}.$$

We write

$$\langle \{p_1, p_3\}, \square \rangle = \{q_1, q_2, q_3, q_4, q_6\}.$$

Inspecting the relation  $R$ , we also find that

$$\langle \square, \{q_1, q_2, q_3, q_4, q_6\} \rangle = \{p_1, p_3, p_5\}.$$

Observe that  $\langle A, \square \rangle = B$  does *not* imply  $\langle \square, B \rangle = A$ , because both  $p_3$  and  $p_5$  have bought  $q_6$ .

**Marketing books.** Let us look at a marketing strategy. For a subset of books  $B$  of  $Q$ , the bookseller searches the relation  $R$  and finds every person who has bought at least one book in  $B$ . Denote this subset of people by  $RB$ . If a person  $p$  in  $RB$  has not bought a book  $q$  in  $B$ , the bookseller will recommend book  $q$  to person  $p$ . The recommendation may appear on the website of the bookseller as “customers who bought this book also bought these other books”.

### Relation, Function, and Bijection

A relation is a collection of ordered pairs, each of which has two items. An item in one pair may just appear in this pair, or may also appear in other pairs. The unique or repeated appearance of items classifies relations into three varieties.

In this section,  $R$  denotes a relation, and  $P$  and  $Q$  denote the ranges of  $R$ .

**Many-many relation.** A relation  $R$  is a *many-many relation* if at least one element  $p$  in  $P$  appears in more than one ordered pair in  $R$ , and at least one element  $q$  in  $Q$  appears in more than one ordered pair in  $R$ .

The relation

$$\{(x, y) \mid \text{person } x \text{ has spent time in country } y\}$$

is a *many-many* relation. A person may have spent time in many countries, and a country has hosted many people.

Here are more examples of many-many relations:

$$\begin{aligned} &\{(x, y) \mid \text{webpage } x \text{ links to webpage } y\}, \\ &\{(x, y) \mid \text{person } x \text{ has read article } y\}, \\ &\{(x, y) \mid \text{person } x \text{ likes person } y\}, \\ &\{(x, y) \mid \text{person } x \text{ likes dish } y\}. \end{aligned}$$

Most useful relations are many-many relations, but most textbooks hardly mention them.

**Many-one relation (function).** A relation  $R$  is a *many-one relation* if every element  $p$  in  $P$  appears in a unique ordered pair in  $R$ , but at least one element  $q$  in  $Q$  appears in more than one ordered pair in  $R$ .

Human beings must have instinctive obsession with uniqueness. Many-one relations have received so much attention that we give them special names. We call them maps or functions. We will talk a lot more about them later.

The relation

$$\{(x, y) | \text{person } x \text{ was born in country } y\}$$

is a *many-one* relation. Every person was born in a unique country, but every country is a birthplace of many people.

Here are more many-one relations:

$$\{(x, y) | \text{person } x \text{ was born in year } y\},$$

$$\{(x, y) | \text{person } x \text{ has weight } y \text{ now}\},$$

$$\{(x, y) | \text{country } x \text{ has population } y \text{ now}\}.$$

We may change the roles of  $P$  and  $Q$  and define one-many relation.

**One-one relation (bijection).** A relation  $R$  is a *one-one relation* if every element  $p$  in  $P$  appears in a unique ordered pair in  $R$ , and every element  $q$  in  $Q$  appears in a unique ordered pair in  $R$ .

One-one relations have also received so much attention that we give them special names. We call them bijections, or one-one correspondences. We will talk a lot more about them later.

Consider the relation

$$\{(x, y) | \text{person } x \text{ sits in seat } y\}.$$

We further stipulate that every person sits in one seat, and every seat has a person in it.

**Avoid needless bias.** We regard a function as special type of relation, and a bijection a special type of function.

In a general relation  $R$ , the two ranges  $P$  and  $Q$  play unbiased roles. Some writers would say, “a relation *from* a set  $P$  *to* a set  $Q$ ”. We avoid such biased language. We say “a relation *over* two sets  $P$  and  $Q$ ”.

Later we will introduce a relation over any number of sets. The unbiased language will work well, but the “from-to” language will break down.

In a function, the two sets  $P$  and  $Q$  do bias the roles of the two sets. This bias will then permeate our language of functions. For example, we say a function from set  $P$  to set  $Q$ .

In a bijection, once again the two sets  $P$  and  $Q$  play unbiased roles. We say a bijection between sets  $P$  and  $Q$ .

### Relation over Two Different Sets

Often, every ordered pair in a relation  $R$  draws its first item from a set  $X$ , and draws its second item from another set  $Y$ . This observation motivates a somewhat different definition of relation.

**Relation over two different sets.** A relation  $R$  is a subset of the Cartesian product of two sets  $X$  and  $Y$ :

$$R \subset X \times Y.$$



We say that  $R$  is a relation *over* the two sets  $X$  and  $Y$ , and that  $X$  and  $Y$  are the two sets *under* the relation  $R$ . The sets  $X$  and  $Y$  are called the *domains* of the relation  $R$ , and the set of ordered pairs  $R$  is called the *graph* of the relation. We say that elements  $x$  and  $y$  stand in the relation  $R$  if

$$x \in X, y \in Y, \text{ and } (x, y) \in R.$$

To specify a relation, we need to specify three ingredients: a set  $X$ , a set  $Y$ , and a subset  $R$  of  $X \times Y$ .

To remind us of the *three* ingredients, we may also define a relation as a triple  $(X, Y, R)$ , where  $X$  is a set,  $Y$  is another set, and  $R$  is a subset of the Cartesian product  $X \times Y$ .

**Examples.** Let  $X$  be a set of students, and  $Y$  be a set of courses. The selection of courses by the students is a relation over the two sets:

$$\{(x, y) \mid \text{student } x \text{ takes course } y\}.$$

Let  $X$  be the set of west-east coordinates, and  $Y$  be the set of south-north coordinates. A path is a relation over the two sets:

$$\{(x, y) \mid \text{points on the path}\}.$$

Coordinates of a set of buildings also form a relation over  $X$  and  $Y$ :

$$\{(x, y) \mid \text{coordinates of a building}\}.$$

Each point on the surface of the Earth corresponds to an ordered pair of longitude and altitude. The territory of a nation is a set of points. That is, the territory of the nation is a relation over two sets: the set of longitudes and the set of altitudes.

Let  $S$  be a set of train stations, and  $T$  be a set of time. The schedule of a train is a relation over  $S$  and  $T$ :

$$\{(s, t) \mid \text{the train reaches stop } s \text{ at time } t\}.$$

**How many relations can we create over two sets?** A set  $X$  has  $m$  elements, and a set  $Y$  has  $n$  elements. The two sets form a collection of  $mn$  ordered pairs. Each subset of the collection of ordered defines a relation. Thus, we can create a total of  $2^{mn}$  relations. All relations over two sets  $X$  and  $Y$  constitute the power set of the Cartesian product  $X \times Y$ .

**Range vs. domain.** Let  $R$  be a relation over two sets  $X$  and  $Y$ . Let  $P$  be a range of the relation:

$$P = \{x \mid (x, y) \in R\}.$$

In general, not every element in  $X$  stands in the relation  $R$ . Consequently, the range  $P$  is a subset of  $X$ :

$$P \subset X.$$

We call  $P$  the *range* of the relation  $R$  in the domain  $X$ . We can similarly define the range of the relation  $R$  in the domain  $Y$ .

When several people go to a restaurant, the event forms a relation:

$$R = \{(x, y) \mid \text{person } x \text{ has dish } y\}.$$

The relation  $R$  is over two sets: the set  $X$  of the people in the group, and the set  $Y$  of the dishes on the menu. Typically, the range of the relation in the domain  $X$  is the same as  $X$ : every person in the group eats. Also typically, the range of the relation  $R$  in the domain  $Y$  is a subset of  $Y$ : not every dish on the menu is ordered by a person in the group.

**Totality.** Let  $R$  be a relation over two sets  $X$  and  $Y$ . The relation  $R$  is called *total* with respect to  $X$  if every element in  $X$  stands in the relation, namely, if the range of the relation with respect to  $X$  equals  $X$ :

$$P = X.$$

The relation  $R$  is called *partial* with respect to  $X$  if  $P \neq X$ . Similar definitions apply with respect to  $Y$ .

The relation  $R$  is called a total relation over  $X$  and  $Y$  if  $P = X$  and  $Q = Y$ . That is, every element in  $X$  and every element in  $Y$  stand in the relation  $R$  over  $X$  and  $Y$ . A total relation is also called a *surjection*.

Every person was born in a country, and every country is a birthplace for some people. Thus, the relation

$$\{(x, y) \mid \text{person } x \text{ was born in country } y\}$$

is a total relation over the set of all people and the set of all countries.

Some people never buy any book, and some books are never sold. Thus, the relation

$$\{(x, y) \mid \text{person } x \text{ has bought book } y\}$$

is a partial relation over the set of all people and the set of all books.

When several people go to a restaurant, every person orders a dish from the menu, but not every dish on the menu is ordered by a person in the group. Thus, the relation

$$\{(x, y) \mid \text{person } x \text{ has dish } y\}$$

is total with respect to the set of people in the group, but is partial with respect to the set of dishes on the menu.

### Ways to Specify a Relation

Now let us look at one relation in detail. Let  $X$  be a set of researchers, and  $Y$  be a set of instruments. We wish to know which researcher knows to use which instrument.

**Specify a relation by a property.** We make our wish clear by defining a relation  $R$  over the two sets  $X$  and  $Y$ :

$$R = \{(x, y) \mid \text{researcher } x \text{ knows how to use instrument } y\}.$$

This definition specifies the set  $R$  by a property. The property, however, is so obscure that it does not tell us which researcher knows to use which instrument. Nor does it tell us who the researchers are, or what the instruments are.

**Specify a relation by listing the elements of the three sets.** We specify the three ingredients of the relation. Let  $X$  be a set of five researchers:

$$X = \{x_1, x_2, x_3, x_4, x_5\}.$$

Let  $Y$  be a set of twelve instruments:

$$Y = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}\}.$$

After talking to the researchers, we find who knows how to use what:

$$R = \{(x_1, y_2), (x_1, y_3), (x_2, y_5), (x_3, y_4), (x_3, y_6), (x_3, y_{11}), (x_4, y_{10}), (x_5, y_6)\}.$$

For example, the ordered pair  $(x_2, y_5)$  means that researcher  $x_2$  knows how to use instrument  $y_5$ . The relation  $R$  has a total of eight ordered pairs—that is,  $R$  is an eight-element set. Of course, we can list the eight ordered pairs in any order.

A list of a large number of ordered pairs is hard on the eye, but is an excellent way to store the relation on a computer.

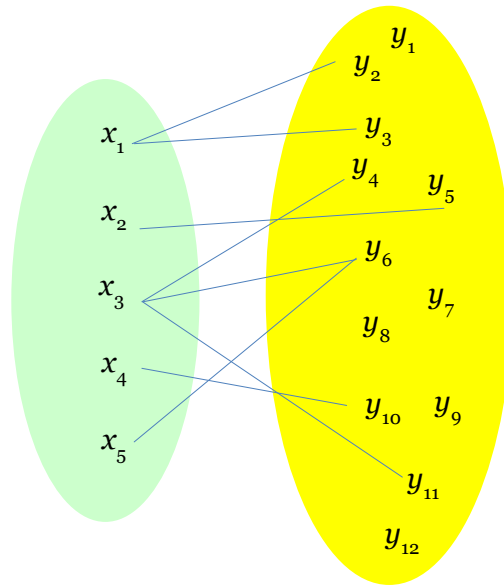
**Specify a relation by a graph.** In this method, we specify the three ingredients of the relation as follows. The Cartesian product  $X \times Y$  is a  $5 \times 12$  table, in which we mark the ordered pairs that belong to the relation  $R$ . This table is easy on the eye, but is only effective for sets of relatively small numbers of elements

	$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_6$	$y_7$	$y_8$	$y_9$	$y_{10}$	$y_{11}$	$y_{12}$
$x_1$												
$x_2$												
$x_3$												
$x_4$												
$x_5$												

Neither the order of the researchers nor the order of the instruments matters to this relation. However, in graphing the relation, we have to place the elements in each set in some order. We can place the set of researchers in any of

the  $5!$  permutations, and place the set of instruments in any of the  $12!$  permutations. We have placed researchers in columns, and instruments in rows. We can also place them the other way. Consequently, each relation over a five-element set and a twelve-element set can have a total of  $2 \times 5! \times 12!$  graphs. We should guard ourselves against any visualization that may bias us.

**Specify a relation by lines linking ordered pairs.** In this method, we specify the three ingredients of the relation as follows. Place elements in the set  $X$  in one bubble, and place elements in the set  $Y$  in another bubble. Draw lines to link the ordered pairs that belong to the relation  $R$ .



### Relation over a Set and Itself

**Relation on a set.** As a special type of relation, any subset of  $S \times S$  defines a relation  $R$  over the set  $S$  and itself. We say that  $R$  is a relation *on* the set  $S$ .

**Examples.** We can define many relations between people. As Chinese say, so far as people are concerned, it is all about relations. Let  $S$  be the set of people in the world. Blood relation is a relation on  $S$ :

$$\{(a,b) \mid \text{person } a \text{ and person } b \text{ are related by blood}\}.$$

Friendship on Facebook is a relation on  $S$ :

$$\{(a,b) \mid \text{person } a \text{ is friend of person } b\}.$$

Co-authorship is another relation on  $S$ :

$$\{(a,b) \mid \text{person } a \text{ and person } b \text{ have co-authored at least one paper}\}.$$

Alumni is yet another relation on  $S$ :

$$\{(a,b) \mid \text{persons } a \text{ and } b \text{ were graduated from the same university}\}.$$

Hyperlinks form a relation on the set of all websites:

$$\{(x,y) \mid \text{webpage } x \text{ links to webpage } y\}$$

**A comparison.** The relation

$$\{(x,y) \mid \text{person } x \text{ was born in country } y\}$$

is over two different sets. The sentence “Person  $x$  was born in country  $y$ ” means the same as “Country  $y$  was the birthplace of person  $x$ ”.

The relation

$$\{(x,y) \mid \text{person } x \text{ likes person } y\}$$

is over a set and itself. The sentence “ $x$  likes  $y$ ” is different from “ $y$  likes  $x$ ”.

## Equivalence

**Equivalence relation on a set.** An *equivalence relation* on a set  $S$  is a relation  $E$  with the following properties.

(E1) Reflexivity: For any element  $a$  in  $S$ ,  $(a,a) \in E$ .

(E2) Symmetry: For any two elements  $a$  and  $b$  in  $S$ , if  $(a,b) \in E$ , then  $(b,a) \in E$ .

(E3) Transitivity: For any three elements  $a$ ,  $b$ , and  $c$  in  $S$ , if  $(a,b) \in E$  and  $(b,c) \in E$ , then  $(a,c) \in E$ .

The reflexivity stipulates that every element in  $S$  stands in the relation  $E$ . Consequently, any equivalence relation  $E$  on a set  $S$  is a total relation. We say that  $a$  is equivalent to  $b$  with respect to  $E$  if

$$a \in S, b \in S, \text{ and } (a,b) \in E.$$

If the intended equivalence relation is clear from the context, we write

$$a \sim b.$$

**Examples.** Let  $S$  be the set of all people. Here are three equivalence relations on  $S$ :

$$R_1 = \{(a,b) \in S^2 \mid a \text{ and } b \text{ were born in the same century}\},$$

$$R_2 = \{(a,b) \in S^2 \mid a \text{ and } b \text{ have the same birthday}\},$$

$$R_3 = \{(a,b) \in S^2 \mid a \text{ and } b \text{ have the same blood type}\}.$$

The three equivalence relations are distinct relations on the same set  $S$ . They relate people differently.

The relation

$$R_4 = \{(a, b) \in S^2 \mid a \text{ loves } b\}$$

is not an equivalence relation. This relation is often painfully non-reflexive and non-symmetric, and is seldom transitive.

The relation

$$R_5 = \{(a, b) \in S^2 \mid a \text{ and } b \text{ are friends on Facebook}\}$$

is not an equivalence relation. The relation is symmetric, but not transitive. Is the relation reflexive?

The blood relation on  $S$  is not an equivalence relation. This relation is symmetric, but not transitive. For example, a son is blood-related to both his dad and mom, but the dad and mom are typically not blood-related. Is the relation reflexive?

Of all triangles, the triangles of the same shape (i.e., similar triangles) form an equivalence relation.

**Equivalent class.** In an equivalent relation on the set of all people, being equivalent people does not mean being the same person. Given an element  $x$  in a set  $S$ , all elements in  $S$  equivalent to  $x$  with respect to an equivalence relation  $E$  form a subset of  $S$

$$\{a \mid a \sim x\}.$$

We call this subset an *equivalent class* with respect to the equivalence relation  $E$ . We denote this equivalent class by  $[x]_E$ . We may drop the subscript  $E$  if the intended equivalence relation is clear from the context. The word “class” should remind us that an equivalent class is a set.

An equivalence relation on the set of all people is “people born in the same country”. In this equivalence relation, “People born in China” form an equivalent class. We can pick any particular person born in China as a representative, for example, Confucius, and denote the equivalent class by  $[\text{Confucius}]$ . Of course, we can also denote the same equivalent class by  $[\text{Mao}]$ .

An equivalence relation on the set of all forms of life is “being in the same species”. Each species is an equivalent class. Thus, we can also designate the equivalent class “human beings” by  $[\text{Confucius}]$ .

**Equivalence and partition.** Each equivalence relation on a set  $S$  generates a partition of  $S$ . Conversely, each partition of  $S$  defines an equivalence relation on  $S$ . Thus, equivalence relations on  $S$  one-one correspond to partitions of  $S$ . Each part in a partition of  $S$  is an equivalent class with respect to an equivalence relation on  $S$ .

For example, “people born in the same country” is an equivalence relation on a set of people  $S$ . We start with a person  $a$  in the set  $S$ , find all people in  $S$  born in the same country as  $a$ , and denote this subset of people by  $[a]$ . Then we find a person  $b$  among the remaining people in  $S$ , find all people in  $S$  born in the same country as  $b$ , and denote this subset of people by  $[b]$ . We can run this procedure until we find everyone in  $S$ . Thus, the equivalence relation “people born in the same country” partitions the set of people  $S$  in a family of equivalent classes,  $[a]$ ,  $[b]$ , ... Any two equivalent classes are disjoint, and the union of the equivalent classes is  $S$ . Given a set and an equivalence relation, this partition is unique.

As another example, “people having the same birthday” is also equivalence relation on the set of people  $S$ . This equivalence relation corresponds to a partition of  $S$ .

**How many equivalence relations can we form on a set?** The total number of partitions of an  $n$ -element set is the Bell number  $B_n$ , which is also the total number of equivalence relations on the  $n$ -element set.

## Equality

**Equality relation.** A very special type of equivalence relation is the relation “being equal”, or “being the same”, or “being identical”. If things  $a$  and  $b$  are identical, we write

$$a = b.$$

If things  $a$  and  $b$  are not identical, we say that they are distinct and write

$$a \neq b.$$

Recall that every equivalence relation on a set  $S$  corresponds to one and only one partition of  $S$ . The equality relation on  $S$  corresponds to a very special partition of  $S$ : each part in the partition contains a single element in  $S$ .

**Equality vs. equivalence.** “All men are created equal,” said Thomas Jefferson in the Declaration of Independence. The gender bias aside, he clearly did not mean that all men are identical. We interpret this insight as “All people belong to an equivalent class”.

Jefferson also answered the question, “equivalent with respect to what?” He wrote, “We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty and the pursuit of Happiness.”

When we speak of an equivalent class, we have in the back of our minds a larger set, of which the equivalent class is a subset. What might be the larger set, of which “all people” is a subset? An obvious larger set is “all living things”. But this thought has brought us away from the original context of the Declaration of Independence.

## Order

**Order on a set.** An *order* on a set  $S$  is a relation  $R$  with the following properties.

(O1) Totality: For any two distinct elements  $a$  and  $b$  in  $S$ , one and only one of the two statements,  $(a,b) \in R$  and  $(b,a) \in R$ , is true.

(O2) Transitivity: For any three elements  $a$ ,  $b$ , and  $c$  in  $S$ , if  $(a,b) \in R$  and  $(b,c) \in R$ , then  $(a,c) \in R$ .

We can define many orders on a set. If the intended order is clear from the context, we write

$$a < b.$$

The statement is read “ $a$  is less than  $b$ ”, or “ $a$  precedes  $b$ ”

The notation  $a \leq b$  means “ $a$  is either equal to, or less than,  $b$ ”. The notation  $b > a$  means  $a < b$ .

A set on which an order is defined is called an *ordered set*. A set on which no order is defined is called an *unordered set*.

**Examples.** All words in English form an ordered set according to the alphabetical order.

The set of integers is an ordered set, so is the set of rational numbers, and so is the set of real numbers. The set of complex numbers is an unordered set.

Many things in nature form ordered sets. Moments in time form an ordered set. Similarly, length, mass, energy, and temperature each forms an ordered set. Colors form an ordered set by frequency. We do not, however, have a “natural order” on the set of all smells.

**Order and permutation.** We can represent elements in a set  $S$  by distinct beads, and put them on a string in a sequence. Each sequence defines an order on  $S$ . Each sequence is called a **permutation**.

**How many orders can we define on an  $n$ -element set?** In specifying a set, the order of its elements does not matter. However, when displaying the elements of the set in a row, we place the elements in an ordered sequence.

We can choose the first element in a sequence among the  $n$  elements. We can choose the second element in the sequence among the  $n-1$  remaining elements. In this way we choose the third, the fourth, .., and the  $n$ th elements in the sequence. Thus, we can define a total of  $n!$  sequences (i.e., orders) on the  $n$ -element set.

**To order or not to order.** Given any set of  $n$  things, we can line them in  $n!$  permutations. Each permutation defines an order on the set. However, most sets do not have any “natural” or “preferred” order.



For example, the set of all nations in the world does not have any “natural order”. We can, of course, force an order on this set. For example, English speakers often list nations by the alphabetical order of their English names. This order makes no sense to non-English speakers.

To force an order on a set is to impose a structure that is not inherent to the set. If the imposition confuses our narrative, it is just a bad use of language. Alphabetical order is convenient, but incidental. We should never let this incidental order bias us, and should simply regard the set of nations as an unordered set.

**Avoid useless order.** I grew up on a campus of a university in China. The campus had many apartment buildings, but they were here and there, with no natural order. The university nonetheless labeled all the apartment buildings by a single sequence of numbers. The labels were placed on the walls of the buildings. The numbers seemed to turn the buildings into an ordered set. This order might be useful on an administrative spreadsheet, but confused visitors. It was hard to find a building by its number. I lived in Building 10, which was next to Building 4.

**Order does not obey arithmetic rules.** We often label elements in an ordered set using numbers. Using numbers to label elements does not authorize us to apply arithmetic rules.

For example, we usually label houses in a street by numbers. It is meaningless to add the addresses of two houses. House number 2 and house number 7 do not add up to become house number 9. The difference between the first house and the second house is not the same as the difference between the fifth house and the sixth house.

**Order-preserving coarsening of a partition.** Elements in an ordered set  $G$  can be represented by a string of beads. We can cut the string into segments. The beads on each segment form a subset of  $G$ , and preserve the order on  $G$ . All the segments together form a partition of  $G$ .

When we use an order-preserving partition of  $G$  to rank a set  $S$ , the rank consists of fewer equivalent classes, and is called a coarser rank.

For example, a teacher marks the papers of students by numbers from 0 to 100. She then divides the interval from 0 to 100 into five subintervals, and labels them as A, B, C, D, and E. The five subintervals form a partition of the interval from 0 to 100.

**Find an ordered subset in an unordered set.** The set of all people  $S$  does not have any “natural order”. Even the relation “being a male decedent” does not provide an order on the set of all people, because the relation does not satisfy the property of totality. We will regard  $S$  as an unordered set.

Many subsets of  $S$ , however, are ordered sets under the relation “being male decedent”. To construct such an ordered subset, we can start with a male,

and look for his father, grandfather, great-grandfather... This chain of males forms an ordered set. Each male in the world belongs to at least one such a chain.

**Order on a family of sets.** A family of sets is also a set, on which we can define orders. Consider a list of sets with the following property. Each set is a subset of the set after it, except for the last set in the list. This property defines an order on the list of sets. For example, various number systems form an ordered set:

$$\{N, Z, Q, R, C\}.$$

### Pinyin\*

**Romanization of Chinese.** The written Chinese words have long been the same in all parts of China. The same written Chinese word, however, sounds differently in various dialects. People speaking different dialects often cannot speak to each other. To unify the language in China, Children in schools are taught the Standard Chinese, also known as Mandarin.

In 1958, the Chinese government published the Pinyin system, a phonetic representation of the Standard Chinese in the Latin alphabet. For example, the word “China” means “中国”, which is spelled in Pinyin as “Zhong Guo”. The word “Pinyin” itself is the phonetic representation of two Chinese characters “拼音”, which means “spelled sound”.

Pinyin is used to teach Chinese. Pinyin is also used to spell Chinese names in the Latin alphabet. For example, the capital of China, “北京”, is spelled as “Beijing”. Pinyin also serves as a method to enter Chinese characters to computers.

**Pinyin and alphabetical order.** English-speakers usually list a set of people in the alphabetical order of their names. The alphabetical order implies no order with respect to any other relation on the set, and is merely a way to list the set of people. As we discussed before, in most situations it is wise to regard a set of people as an unordered set. The use of names turns an unordered set (people) to an ordered set (English words). The practice is so prevalent that we usually forget how ingenious and helpful it is. The name of a person is (nearly) permanent, but can be used to list any set of people.

This practice, of course, makes no sense in China. The Chinese-speakers may list a set of people in the order of the numbers of strokes in the Chinese characters of their names. Counting the numbers of strikes in Chinese characters is tedious. In most situations, we may as well regard the Chinese characters as an unordered set.

The Pinyin system turns an unordered set (Chinese characters) to an ordered set (words written in the Latin alphabet). Similar systems of romanization now enable us to list people in the alphabetical order of their names, regardless of their countries of origin.

## Finitary Relation

A binary relation is a set of ordered pairs. We now generalize this definition to a set of  $n$ -tuples.

**Finitary relation.** A set of  $n$ -tuples is called an  $n$ -ary relation. The integer  $n$  is called the *arity* of the relation.

Thus, a set of individual elements is called a *unary relation*, a set of ordered pairs is called a *binary relation*, and a set of triples is called a *ternary relation*.

If a triple  $(x, y, z)$  belongs to a ternary relation  $R$ , we write

$$(x, y, z) \in R.$$

We say that the items  $x$ ,  $y$ , and  $z$  stand in the ternary relation  $R$ .

**Graph, ranges, and images.** We can generalize the definitions of graph, ranges, and images.

The collection of  $n$ -tuples in an  $n$ -ary relation  $R$  is called the graph of the relation.

The relation  $R$  has  $n$  ranges, defined by

$$P_1 = \{p_1 \mid (p_1, \dots, p_n) \in R\}.$$

$$\dots\dots\dots P_n = \{p_n \mid (p_1, \dots, p_n) \in R\}.$$

Let  $A_i \subset P_i$  and  $A_j \subset P_j$ . The image of  $A_i$  and  $A_j$  under  $R$  is defined by the following expression:

$$\begin{aligned} \langle \square, \dots, \square, A_i, \square, \dots, \square, A_j, \square, \dots, \square \rangle_R = \\ \left\{ (p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_{j-1}, p_{j+1}, \dots, p_n) \mid (p_1, \dots, p_n) \in R, p_i \in A_i, p_j \in A_j \right\} \end{aligned}$$

**Relation over any number of sets.** A relation  $R$  over sets  $S_1, \dots, S_n$  is a subset of the Cartesian product of these sets:

$$R \subset S_1 \times \dots \times S_n.$$

We say that the sets  $S_1, \dots, S_n$  are under the relation  $R$ . The sets  $S_1, \dots, S_n$  are called the domains of the relation, and the set of tuples  $R$  is called the graph of the relation. We say that elements  $s_1, \dots, s_n$  stand in the relation  $R$  if

$$s_1 \in S_1, \dots, s_n \in S_n, \text{ and } (s_1, \dots, s_n) \in R.$$

In this definition,  $R$  stands for both the relation, the property that specifies the relation, and the graph of the relation.

In a relation, some domains can be the same set.

**Examples.** Immigration defines a ternary relation:

$$\{(p, x, y) \mid \text{persons } p \text{ immigrates from country } x \text{ to country } y\}.$$

This relation is over the set of all people  $P$ , the set of all countries  $X$ , and the set of all countries  $X$  again.

Let  $P$  be a set of people. Here are two ternary relations on  $P$ :

$$\{(a, b, c) \mid \text{persons } a \text{ and } b \text{ are parents of person } c\},$$

$$\{(a, b, c) \mid \text{person } a \text{ knows that person } b \text{ loves person } c\}.$$