MAP

A relation is a collection of ordered pairs. A function is a collection of ordered pairs with a restriction: the first item of every ordered pair in the collection is unique. Thus, a function is a relation of a special type.

Most relations in nature are not so restricted as functions, but much of science is devoted to functions. This fact perhaps reveals something about human beings. We are obsessed with uniqueness.

We use the words "function", "map", and "mapping" interchangeably, and prefer the shortest word "map". On occasions, we may replace the word "map" with the word "functional", "form", "transformation", "operator", or "operation". We will note such occasions when they arise.

Map

Map. A *map* is a collection of ordered pairs with a restriction: the first item of every ordered pair in the collection is unique. The second item of every ordered pair, however, need not be unique.

Examples. The collection of ordered pairs

$$\{(1,3),(2,3),(3,3),(4,1)\}$$

is a map, because the first items of every ordered pair is unique. In this map, not all second items in the ordered pairs are unique.

The collection of ordered pairs

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

is a relation, not a map, because "1" appears as the first item in two ordered pairs in the collection.

The set-building notation

$$\{(x,y) | x = y^2, y \text{ is a real number}\}$$

defines a collection of ordered pairs. This collection specifies a relation, but not a map. For example, both $(4,-2)$ and $(4,2)$ belong to the collection.

The collection of ordered pairs

$$\{(x,y) | x^2 = y, x \text{ is a real number}\}$$

specifies a map.

The items in each ordered pair need not be numbers. The grades of students in a class define a map:

$$G = \{(s,g) | \text{student } s \text{ receives grade } g\}.$$

Every student in the class receives a unique grade. That is, every student appears as the first item only in one ordered pair, but a grade may appear as the second item in no order pair, one ordered pair, or multiple ordered pairs.

The heights of these students—at a particular time—define another map:

$$H = \left\{ (s,h) \middle| \text{student } s \text{ is of height } h \right\}.$$

At a particular time, each student has a unique height.

The relation

$$\left\{ (x,y) \middle| \text{person } x \text{ was born in contury } y \right\}$$

is a map, because every person was born in a unique country.

The relation

$$\left\{ (x,y) \middle| \text{person } x \text{ has spent time in contury } y \right\}$$

is just a relation, not a map, because a person may have spent time in multiple countries.

The blood type is a map, but the blood relation is not a map.

A map is a biased relation. Recall that a relation is a collection of ordered pairs. In a relation, the first and second items in the ordered pairs play unbiased roles.

By contrast, in a map, the first and second items in the ordered pairs play biased roles. The bias comes from the restriction on the first items in the ordered pairs, but not on the second items in the order pairs.

A map is a set *M*, each element of which is an ordered pair, such that $x \neq a$ for any two ordered pairs (x,y) and (a,b) in *M*. If (x,y) and (x,z) belong to a map, then the two ordered pairs are identical, and $y = z$. The phrase "a multi-valued map" violates the definition, and is an abuse of language.

Notation of map. This bias permeates our language. If an ordered pair (x,y) belongs to the map *M*, we can of course write

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

This notation, however, does not reflect the bias, and is rarely used.

If an item *x* appears as the first item in an ordered pair in a map *M*, the item *x* will not appear as the first item in any other ordered pairs in *M*. This uniqueness lets us denote the second item of this unique ordered pair by $M(x)$. We write the map in a biased way:

$$x \mapsto M(x).$$

If an ordered pair (x,y) belongs to a map *M*, instead of writing $(x,y) \in M$, more commonly we write

$$y = M(x).$$

We call the first item *x* the independent variable, or input, or argument. We call the second item *y* the dependent variable, or output, or value. We say that *y* is a map of *x*, and that the map *M* sends item *x* to item *y*. In addition to the word "send", other words in common use include "map", "carry", "transform", "associate", and "relate".

Graph, domain, and range. Because maps are biased relations, the language of maps derives from that of relations, with modifications that reflect the bias.

We often specify a map by a property, such as "people and their countries of birth". This property leads to a collection of ordered pairs, such as

$$\left\{\left(\text{Amy, England}\right),...,\left(\text{Zac, China}\right)\right\}$$

We call the collection of ordered pairs the *graph* of the map. We let the letter *M* stand for all three things: the map, the property that specifies the map, and the graph of the map.

The first items in the ordered pairs in a map *M* define a set:

$$\left\{x \middle| \left(x,y\right) \in M\right\}.$$

We denote this set by *X*, and call it the *domain* of the map *M*.

The second items in the ordered pairs in *M* define another set:

$$\left\{y \middle| \left(x,y\right) \in M\right\}.$$

We denote this set by $M\left(X\right)$, and call it the *range* of the map *M*.

A Map from One Set to Another Set

Often, every ordered pair in a map *M* draws its first item from a set *X*, and draws its second item from another set *Y*. This observation motivates an alternative, but equivalent, definition of map.

A map from one set to another set. A *map M* from a set *X* to a set *Y* is a collection of ordered pairs, such that, for *every* element *x* in *X*, there exists a *unique* element *y* in *Y* to make $\left(x,y\right) \in M$. We call *X* the *domain* of the map, *Y* the *codomain* of the map, and the set of ordered pairs *M* the *graph* of the map.

The graph of the map is a subset of the Cartesian product of the domain and codomain:

$$M \subset X \times Y.$$

Thus, the map *M* is a total relation with respect to *X*, but may be a partial relation with respect to *Y*.

The domain of the map coincides with the set *X*, but the range of the map is a subset of the set *Y*.

To specify a map this way, we must specify three ingredients: a domain, a codomain, and a graph, such that every element in the domain is listed in one and only one ordered pair in the graph. For a map *M* from a set *X* to a set *Y*, we write

$$M : X \to Y,$$

or

$$X \xrightarrow{M} Y .$$

Specify a map by a property. We now look at the map "people and their countries of birth" in detail. This phase is a property, which specifies the map in the set-building notation:

$$M = \{(x,y) | \text{person } x \text{ was born in contury } y\}$$

This specification leaves us wondering who was born where, but does remind us to collect the data.

Specify a map by a table. Let X be a set of twelve people:

$$X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\} .$$

Let Y be a set of five countries:

$$Y = \{y_1, y_2, y_3, y_4, y_5\} .$$

The map M requires us to specify twelve ordered pairs—that is, we need to find out the country of birth for each of the twelve people. We find who was born where, and list the map M by a table:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
y_3	y_5	y_5	y_3	y_4	y_3	y_1	y_3	y_3	y_5	y_3	y_3

This representation takes the advantage that a map is a special kind of relation—every person was born in one and only one country. The table is hard on the eye if the domain is a large set, but the table is an effective way to store the data on a computer.

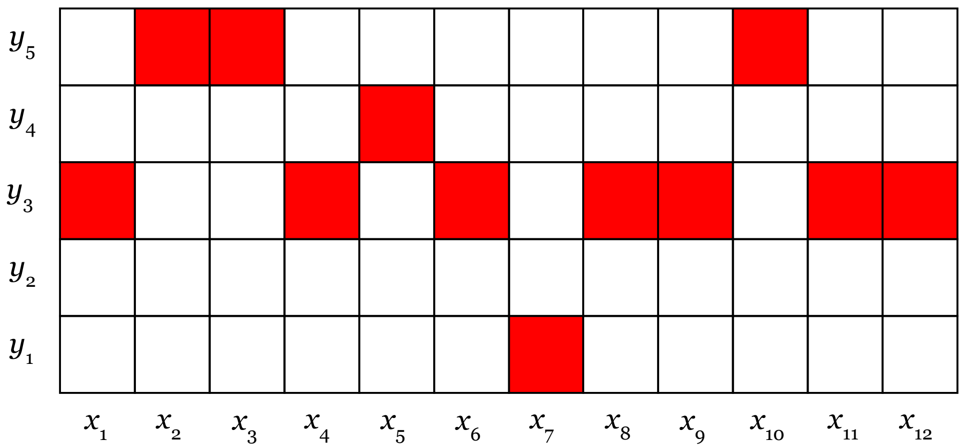
Because the set of people is an unordered set, any of 12! permutations of the set of people corresponds to a table. Thus, a single map can be represented by a total of 12! tables.

Specify a map by a graph. We can specify a map by a graph on a piece of paper the same way as we specify any kind of relation. We place the Cartesian product of the two sets as a table. We then mark all ordered pairs (x,y) that belong to the map M. We represent map from people to contour by a graph:

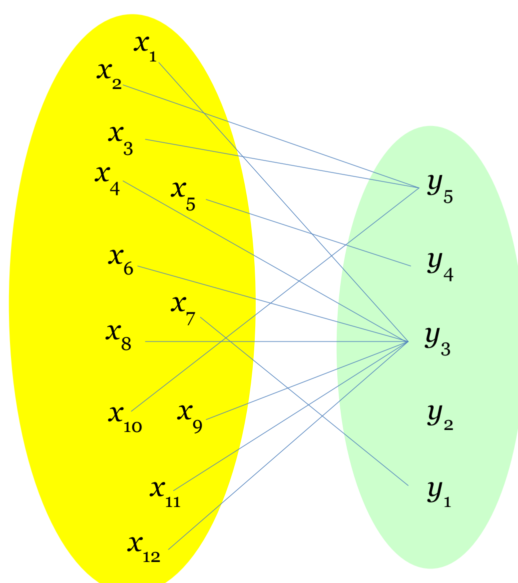
Every person was born in one and only one of the five countries. Multiple people were born in countries y_3 and y_5, one person was born in country y_1, another person was born in country and y_4, and nobody was born in country y_2. Every column has one and only one mark. Every row can have no mark, or one mark, or multiple marks.

We regard both the set of people and the set of countries as unordered sets. In drawing the graph of the map on a piece of paper, we have to place the elements in each set in some order. These orders are not inherent to the map.

We can place the set of people in any of the 12! permutations, and place the set of countries in any of the 5! permutations. Each map from the set of people to the set of countries can have a total of 12!5! graphs. We should guard ourselves against any visualization that may bias us.



Specify a map by linking arguments and values. Place elements in the domain in one bubble, and place elements in the codomain in another bubble. Draw lines to link the arguments and the values in the map. Every element in the domain has one and only one link. Every element in the codomain has no link, or one link, or multiple links.



Varieties of Maps

A general relation is a many-many relation. A map is a many-one relation. We further classify maps as follows.

Function. A map is also called a function. We often write a function as

$$y = f(x).$$

Given a set *X* of m elements, and a set *Y* of n elements, how many functions from *X* to *Y* are possible?

A function can send each element in *X* to any one of the n elements in *Y*. Thus, there exist a total of n^m functions from *X* to *Y*.

Injection. A map is injective (one-one) if it sends each element in the domain to a distinct element in the codomain. That is, for every two distinct elements a and b in *X*, $f(a)$ and $f(b)$ are distinct elements in *Y*. An injective map is called an *injection*.

Given a set *X* of m elements, and a set *Y* of n elements, how many injections from *X* to *Y* are possible?

If $m > n$, no injection is possible. If $m \le n$, a map can send the first element in *X* to any one of the n elements in *Y*, send the second element in *X* to any one of the remaining $n-1$ elements in *Y*,..., and send the last element in *X* to any one of the remaining $n-m+1$ elements in *Y*. Thus, the number of possible injections is

$$n(n-1)...(n-m+1).$$

Surjection. A map is surjective (onto) if it sends at least one element in the domain to every element in the codomain. That is, for every element y in *Y*, there is at least one element x in *X* to satisfy $y = f(x)$. A surjective map is called a *surjection*. The image of the domain under a surjection is the entire codomain.

Given a set *X* of m elements, and a set *Y* of n elements, how many surjections from *X* to *Y* are possible?

If $m < n$, no surjection is possible. ...

Bijection. A map is bijective (one-one and onto) if the following properties hold. For every element x in *X*, there exists a unique element y in *Y* such that $(x,y) \in f$. For every element y in *Y*, there exists a unique element x in *X* such that $(x,y) \in f$. A bijective map is called a *bijection*, or a *one-one correspondence*.

For a bijection to exist between a set *X* of n elements and a set *Y*, the latter must also have n elements. How many bijections are possible between the two sets? A map can send the first element in X to any one of the n elements in Y, send the second element in X to any one of the remaining $n-1$ elements in Y,..., and send the last element in X to the last element in Y. Thus, the number of possible bijection is $n!$.

A bijection is both injective and surjective. When the domain and codomain have the same number n of elements, either injection or surjection implies bijection.

	surjective	non-surjective
injective	bijection	injection only
non-injective	surjection only	function

Inversion. A bijection is an unbiased relation: the two sets play unbiased roles. We speak of a bijection *between* two sets.

Unfortunately, our notation for bijection is biased. We call f a bijection from X to Y, and write

$$y = f\left(x\right).$$

The map f sends an element x in X to an element y in Y.

If we wish to speak of the same bijection, but switch the roles of X and Y, we write

$$x = f^{-1}\left(y\right),$$

and call f^{-1} the inverse map. The inverse map sends an element y in Y to an element x in X.

Of course, a map is invertible if and only if the map is bijective.

Image

A map is a biased relation: the domain and codomain of a map play biased roles. This bias leads to the definition of two terms: image and preimage. Let us look at image first.

Image. Let f be a map from a set X to a set Y. Let A be a subset of X. The map f sends every element in A to a unique element in Y. The elements in Y associated with all elements in A form a set, which is called the *image* of A under f, and is written as $f\left(A\right)$. Thus,

$$f\left(A\right) = \left\{ y \middle| \left(x,y\right) \in f, x \in A \right\}.$$

Properties of image. For every subset A of X, $A \subset X$, the image of A under f is a unique subset of Y,

$$f\left(A\right) \subset Y.$$

For instance, in the example of twelve people born in five countries, we note that

$$f\left(\left\{x_1,x_4,x_7\right\}\right)=\left\{y_3,y_1\right\}.$$

The image of the entire domain, $f\left(X\right)$, is called the *range* of the map. The range is a subset of the codomain *Y*:

$$f\left(X\right)\subset Y.$$

We list these properties along with a few others:

$$f\left(A\right)\subset Y,$$
$$f\left(X\right)\subset Y,$$
$$f\left(\varnothing\right)=\varnothing,$$
$$A_1\subset A_2\Rightarrow f\left(A_1\right)\subset f\left(A_2\right),$$
$$f\left(A_1\cup A_2\right)=f\left(A_1\right)\cup f\left(A_2\right),$$
$$f\left(A_1\cap A_2\right)\subset f\left(A_1\right)\cap f\left(A_2\right).$$

Here A, A_1 and A_2 are subsets of *X*.

Map vs. image. For any map $f:X\to Y$, the image of every element *x* in *X* under *f* is always a unique element *y* in *Y*:

$$y=f\left(x\right).$$

This usage recovers the notation for the map that sends *x* to *y*. An individual element in *x* is a special subset of *X*, and an individual element in *y* is a special subset of *Y*. Thus, the expression "$f\left(A\right)\subset Y$" generalizes the expression "$f\left(x\right)=y$".

Preimage

Preimage. Let *f* be a map from a set *X* to a set *Y*. Let *B* be a subset of *Y*. The elements in *X* associated with all elements in *B* constitute a subset in the domain *X*. We call the subset the preimage of *B* under *f*, and write the subset as $f^{-1}\left(B\right)$. Thus,

$$f^{-1}\left(B\right)=\left\{x\middle|\left(x,y\right)\in f,y\in B\right\}.$$

Properties of preimage. For every subset *B* of *Y*, the image of *B* under *f* is a subset of *X*,

$$f^{-1}\left(B\right)\subset X.$$

The preimage of an individual element y in Y, written as $f^{-1}(y)$, is also a subset of X:

$$f^{-1}(y) \subset X$$

The preimage of the entire set Y is X:

$$f^{-1}(Y) = X .$$

We list these properties along with a few others:

$$f^{-1}(B) \subset X,$$
$$f^{-1}(Y) = X,$$
$$f^{-1}(\varnothing) = \varnothing,$$
$$B_1 \subset B_2 \Rightarrow f^{-1}(B_1) \subset f^{-1}(B_2),$$
$$f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2),$$
$$f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2).$$

Here B, B_1 and B_2 are subsets of X.

Inversion vs. preimage. We have used the notation $f^{-1}(y)$ to mean two distinct things: the inversion of a bijective map, and the preimage of an element y in Y under an arbitrary map.

For the example of twelve people born in five countries, we

$$f^{-1}(\{y_1\}) = \{x_7\}$$
$$f^{-1}(\{y_2\}) = \varnothing$$
$$f^{-1}(\{y_3\}) = \{x_1, x_4, x_6, x_8, x_9, x_{11}, x_{12}\}$$
$$f^{-1}(\{y_4\}) = \{x_5\}$$
$$f^{-1}(\{y_5\}) = \{x_2, x_3, x_{10}\}$$

For a bijective map $f: X \to Y$, the notation $f^{-1}(y)$ means the unique element x in X, such that f sends x to y in Y:

$$x = f^{-1}(y), \quad y = f(x).$$

The two expressions denote the same bijection.

For an arbitrary map $f: X \to Y$, the notation $f^{-1}(y)$ means the preimage of an element y in Y under f, and the preimage is a subset of X:

$$f^{-1}(y) \subset X .$$

For an arbitrary map, $f^{-1}(y)$ is a set, which may contain no element, one element, or multiple elements in *X*.

Preimage generalizes inversion. When a map is not bijective, we cannot speak of the inverse map, but can still speak of preimages. When the map is non-injective, the preimage of an element in *Y* may contain multiple elements in *X*. When the map is surjective, a preimage may be an empty set. When a map is bijective, the preimage of an individual element in *Y* is a unique element in *X*.

Partition and preimage. Let *X* be a set of *m* elements:

$$X = \{x_1,...,x_m\}.$$

Let *Y* be a set of *n* elements:

$$Y = \{y_1,...,y_n\}.$$

For any map $f : X \to Y$, the preimage of an element y_i in *Y* is a subset of *X*:

$$f^{-1}(y_i) \subset X.$$

Because *f* sends every element in *X* to a *unique* element in *Y*, the preimages of any two distinct elements y_i and y_j in *Y* are disjoint:

$$f^{-1}(y_i) \cap f^{-1}(y_j) = \varnothing.$$

Because *f* sends *every* element in *X* to a unique element in *Y*, each element belongs to preimage of one element in *Y*. That is, the union of all preimages of individual elements in *Y* is *X*:

$$f^{-1}(y_1) \cup ... \cup f^{-1}(y_n) = X.$$

If *f* sends no element in *X* to an element y_k in *Y*, the preimage of this element is the empty set:

$$f^{-1}(y_k) = \varnothing.$$

In summary, for any map $f : X \to Y$, the preimages of individual elements in *Y* under *f* are disjoint subsets of *X*, and the union of these subsets is *X*. We reach a significant concussion: all nonempty preimages of individual elements in *Y* under *f* form a partition of *X*.

Partition of domain and partition of codomain. We can generalize the above observation. For every partition $B_1,...,B_n$ of *Y*, their preimages $f^{-1}(B_1)...,f^{-1}(B_n)$ form a partition of *X*.

Conversely, given two sets *X* and *Y*, for every partition $A_1,...,A_n$ of *X* and every partition $B_1,...,B_n$ of *Y*, there exists a map *f* such that

$$A_1 = f^{-1}(B_1)...,A_n = f^{-1}(B_n).$$

Probability*

Model of randomness. We have described a model of randomness using the language of sets. A trial of an experiment gives one and only one of multiple outcomes. Each outcome is called a sample, and the collection of all outcomes is called the sample space. A subset of a sample space is called an event. The collection of all events is called the event space.

Probability of an event. Consider an experiment of sample space $\left\{s_1,...,s_n\right\}$. Any model of probability must obey three rules.

First, for every event $\left\{s_i,...,s_k\right\}$, there exists a unique nonnegative real number, called the probability of the event $\left\{s_i,...,s_k\right\}$, and denoted by $P\left\{s_i,...,s_k\right\}$. For example, for an event $\left\{s_2,s_7,s_{10}\right\}$, we write its probability as

$$P\left\{s_2,s_7,s_{10}\right\}.$$

The notation is read "the probability of event $\left\{s_2,s_7,s_{10}\right\}$".

Second, the probability of the sample space is unity:

$$P\left\{s_1,...,s_n\right\}=1.$$

Third, the probability of any event is the sum of the probabilities of all samples in the event, namely,

$$P\left\{s_i,...,s_k\right\}=P\left\{s_i\right\}+...+P\left\{s_k\right\}.$$

Distribution of probability. In particular, we write the probabilities of individual samples as

$$P\left\{s_1\right\},...,P\left\{s_n\right\}.$$

This list of n numbers is called the *distribution of probability* of the experiment.

Once we know the probabilities of all the samples—that is, we know the distribution of probability, the third rule calculates the probability of any event.

Probability is a map from the event space to real numbers. In defining the probabilities of events, we have identified a map P, namely,

$$P:E\to\left[0,1\right].$$

The domain of the map P is the set of all events, E. The codomain of the map P is the real numbers in the interval $\left[0,1\right]$.

The distribution of probability is also a map, the domain of which is the sample space, and the codomain of which is the real numbers in the interval $[0,1]$.

Given an experiment, how do we obtain a distribution of probability? For example, a die is said to be fair if a roll results in every face with equal probability. For the fair die, the probability to get any particular face is 1/6.

As another example, a cheater fabricates a die, which is lighter on face *a*, so that a rolling is more probable to result in this face. She rolls the die for *N* times, where *N* is a large number. She finds that the die rolls *N*/3 times to face *a*, *N*/10 times to face *f*, and some equal times to the other four faces. Thus, the probability to get face *a* is 1/3, and the probability to get face *f* is 1/10. Let the probability for getting *b*, c, *d*, or *e* be *p*. The sum of the probabilities is unity:

$$\frac{1}{3}+p+p+p+p+\frac{1}{10}=1 .$$

Solving for *p*, we find that the probability to obtain face *b* is 17/120.

When a die is rolled, what is the probability to obtain a face in the subset $\{b,d,f\}$? For a fair die, the probability to obtain a face in the subset $\{b,d,f\}$ is calculated by the equation

$$P\{b,d,f\}=\frac{1}{6}+\frac{1}{6}+\frac{1}{6} .$$

For the cheater's die, the probability to obtain a face in the subset $\{b,d,f\}$ is calculated by the equation

$$P\{b,d,f\}=\frac{17}{120}+\frac{17}{120}+\frac{1}{10} .$$

Entropy*

For an experiment of sample space $\{s_1,...,s_n\}$, a distribution of probability

$$P\{s_1\},...,P\{s_n\}$$

is a set of numbers. This set of numbers allows us to calculate other numbers. Here we define a number called entropy.

Entropy of a distribution of probability. Let the sample space of an experiment be $\{s_1,...,s_n\}$, and a distribution of probability be $P(s_1),...,P(s_n)$. Define the *entropy* S of the distribution of probability by

$$S = -\sum P(s_i)\log P(s_i).$$

The sum is over all samples. As a convention we will use the logarithm of the natural base. Because $x\log x \to 0$ as $x \to 0$, we set $0\log 0 = 0$.

Properties of entropy. The history that led to the definition of entropy took more than hundred years, and was interesting and confusing. But once entropy is defined, we can try to deduce its properties. Properties of entropy make it suitable as a single-number measure of the randomness of an experiment. Here we deduce a few elementary properties.

Entropy is dimensionless and nonnegative number. Because the distribution of probability consists of a list of dimensionless numbers, $P\{s_1\},...,P\{s_n\}$, the entropy is also dimensionless.

Because $0 \le P(s_i) \le 1$, the number $-\log P(s_i)$ is nonnegative. Thus, entropy is always nonnegative.

An experiment of equally probable outcomes. Consider an experiment of n outcomes, $\{s_1,...,s_n\}$. Every outcome is equally probable. Thus, the probability of each outcome is

$$P(s_i) = 1/n.$$

The entropy of the distribution of probability is

$$S = \log n.$$

For the experiment "tossing a fair coin", the sample space has two equally probable samples, and the entropy of this experiment is $\log 2$.

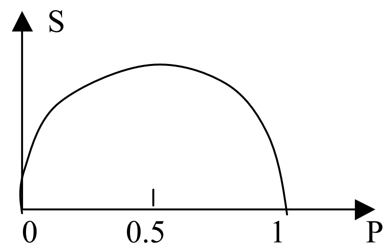
For the experiment "rolling a fair die", the sample space has six equally probable samples, and the entropy of this experiment is $\log 6$.

The experiment "rolling a fair die" is more random than the experiment "tossing a fair coin". Among experiments having equally probable outcomes, the more outcomes an experiment has, the more random the experiment is.

An experiment of two outcomes. An experiment has two outcomes, one with probability P, and other with probability $1-P$. The entropy of the experiment is

$$S = -P\log P - (1-P)\log(1-P).$$

The entropy depends on the distribution of probability, or the number P. When $P = 0$, the outcome is certain, the experiment has no



randomness, and the entropy vanishes. The same is true when $P = 1$. Of all values of *P*, the value $1/2$ maximizes entropy.

More generally, of all experiments of a fixed number of outcomes, the experiment of equally probable outcomes maximizes the entropy.

Entropy in the theory of probability. To calculate the entropy, all we need is a distribution of probability. No more, no less. The physical nature of the experiment is irrelevant, be it tossing a coin, rolling a die, or enjoying a glass of wine.

We do not introduce entropy by following the accidents in the history of thermodynamics. Entropy is a mathematical concept independent of the concept of energy and temperature. The entropy of rolling a die is just as valid a concept as the entropy of a glass of wine. Entropy is a number calculated from a distribution of probability.

The situation is like integrating a function. An integral is a number calculated from a function. You may integrate a function in engineering, or a function in economics. There is no difference between an engineer's integration and an economist's integration. The difference is in the applications, not in the integration.

Random Variable*

Random variable. A random variable *f* is a map from the sample space *S* to some other set *X*. When a trial of the experiment produces a sample *s* in *S*, the random variable *f* sends *s* to an element *x* in *X*. We write the map in several usual ways:

$$f: S \to X,$$

$$x = f(s),$$

$$s \mapsto f(s),$$

$$S \xrightarrow{f} X.$$

The domain of the map *f*, the sample space *S*, is an unordered set. The codomain *X* can be either an unordered set (e.g., a collection of things), or an ordered set (e.g., a collection of numbers).

The map is deterministic, but the outcome of a trial of the experiment is random. Consequently, each trial produces a random sample *s* in *S* and a random element *x* in *X*. The map *f* sends the sample *s* to its image *x*: $x = f(s)$.

Example. Consider a die with faces labeled as $\{a,b,c,d,e,f\}$. Before rolling, we agree on a rule of winning: \$200 for face *a*, \$600 for face *b*, \$100 for face *c*, \$400 for face *d*, \$700 for face *e*, \$0 for face *f*. This rule is a map that sends each face of the die to an amount of winning. The domain of the map is the sample space—the set of the six faces, which is an unordered set. The codomain of the map is a set of numbers—the amounts of winning, which is an ordered set. Whereas the rule is deterministic, the face obtained from each roll of the die is random, and the amount of winning after each roll is random.

Partition a sample space using a random variable. A common way to partition a sample space S is through the use of a random variable. Suppose a gambler decides that a win for her is an amount above $300. The random variable defined above partition the sample space into two events:

$$\text{Win} = \{b,d,e\},$$

$$\text{Lose} = \{a,c,f\}.$$

A random variable is typically a non-injective map—that is, the map often sends multiple samples in S to a single element in X.

Label the faces of a die by numbers. Usually we label the six faces of a die by numbers 1, 2, 3, 4, 5, and 6. We should regard this labeling as a random variable, a map from the sample space (the six faces) to the set of numbers $\{1,2,3,4,5,6\}$. We usually allow the numbers to obey the arithmetic rules. This practice allows textbooks on probability to pose many questions.

When you roll a die once, what is the event of getting an even number? Answer: $\{2,4,6\}$.

When you roll two dice simultaneously, what is the event of getting a sum 7? Answer: $\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$.

Map from sample space to a real variable. Let S be a sample space and R be the set of real numbers. Write a map f from S to R as

$$f : S \to R .$$

Such random variables benefit from properties of numbers. Even though we cannot add and multiply the augments of the map, we can add and multiple the values of the map.

Expectation, Variance, and Standard Deviation*

Let the sample space of an experiment be $\{s_1,...,s_n\}$, and a distribution of probability be $P(s_1),...,P(s_n)$. When the outcome of the experiment is s_i, a random variable takes value $f(s_i)$. Here we restrict the value of the function to be a real variable. The two sets of numbers, $P(s_i)$ and $f(s_i)$, let us calculate other numbers.

Expectation. Define the *mean* or the *expectation* of the random variable *f* by

$$E(f) = \sum P(s_i) f(s_i).$$

The sum is taken over all samples.

Variance. Define the *variance* of a random variable *f* by

$$Var(f) = \sum P(s_i)\left[f(s_i) - E(f)\right]^2.$$

The sum is taken over all samples.

The calculation of fluctuation is often aided by the identity

$$Var(f) = E(f^2) - 2E(f)E(f) + \left[E(f)\right]^2 = E(f^2) - \left[E(f)\right]^2.$$

Standard deviation. Define the standard deviation of the random variable by

$$\sigma(f) = \sqrt{Var(f)}.$$

The unit of the mean is the same as that of the random variable. The unit of the variance is the square of that the random variable. The fluctuation of the random variable *f* can be measured by a dimensionless ratio

$$\frac{\sqrt{Var(f)}}{E(f)}.$$

If we double the random variable, the mean will double and the variance will quadruple, but the fluctuation remains the same.

Example. A fair die has six faces labeled as $\{a,b,c,d,e,f\}$. A random variable maps the sample space to a set of six numbers as follows:

$$f(a) = 1^2, f(b) = 2^2, f(c) = 3^2, f(d) = 4^2, f(e) = 5^2, f(f) = 6^2,$$

Thus,

$$E(f) = \frac{1^2}{6} + \frac{2^2}{6} + \frac{3^2}{6} + \frac{4^2}{6} + \frac{5^2}{6} + \frac{6^2}{6} = \frac{91}{6}$$

$$E(f^2) = \frac{1^4}{6} + \frac{2^4}{6} + \frac{3^4}{6} + \frac{4^4}{6} + \frac{5^4}{6} + \frac{6^4}{6} = \frac{2275}{6},$$

$$Var(f) = E(f^2) - \left[E(f)\right]^2 = \frac{2275}{6} - \left(\frac{91}{6}\right)^2 = 149.08.$$

$$\frac{\sqrt{Var(f)}}{E(f)} = 0.81 .$$

More Examples of Maps

Composition. Consider two maps:

$$f : X \to Y ,$$
$$g : Y \to Z .$$

The two maps in succession send an element in the set X to an element in the set Y, and then to an element in the set Z.

For every element x in the set X, there exists a unique element y in the set Y such that $(x,y) \in f$, and we write

$$y = f(x) .$$

For every element y in the set Y, there exists a unique element z in the set Z such that $(y,z) \in g$, and we write

$$z = g(y) .$$

Consequent, for every element x in the set X, the two maps in succession define a unique element z in the set Z. Write

$$z = g(f(x)).$$

This procedure defines a map from the set X to the set Z. This map is called the *composition* of the maps f and g, and is sometimes written as $g \circ f$.

The composition is a special kind of ternary relation over the three sets X, Y, and Z. For every element x in X, there exist a unique element y in Y and a unique element z in Z such that the triple (x,y,z) belongs to the composition.

Cartesian product as the domain or codomain of a map. The Cartesian product of two or more sets is also a set. Thus, a Cartesian product can serve as the domain or the codomain of a map.

When a bird flies in space, the bird defines a map from the set of moments in time to the set of places in space. The latter, the codomain of the map, is the Cartesian product of three one-dimensional lines.

When we identify parents of every individual person, we construct a map from the set all people S to the Cartesian product $S \times S$. This map is neither injective nor surjective.

A coin has two sides, and a die has six faces. A toss of the coin and die simultaneously results in one of twelve outcomes. These outcomes constitute the Cartesian product of two sets: the set of the two sides of the coin, and the set of the six faces of the die. For each outcome we may assign an amount of winning.

This assignment is a map from the Cartesian product to a set of twelve amounts of winning.

We call a map from the Cartesian product of two sets $X \times Y$ to another set Z a binary map. This map, of course, is also a special type of ternary relation on $X \times Y \times Z$.

A relation is a special kind of map. So far our narrative proceeds along the following line. A relation is a set of ordered pairs. A map is a set of ordered pair with a restriction: the first item of every ordered pair is unique. Thus, a map is a special kind of map.

Our obsession for uniqueness knows no bound. We will reword everything to make it unique. We can even make a relation a special kind of map. Here is yet another equivalent, if twisted, definition of relation. A relation *R* over sets *X* and *Y* is a map from $X \times Y$ to a set of two elements $\{\text{yes, no}\}$:

$$R: X \times Y \to \{\text{yes, no}\}.$$

The map *R* sends every ordered pair (x,y) to either "yes" or "no". When the map *R* sends an ordered pair (x,y) to "yes", this pair belongs to the relation *R*. When the map *R* sends an ordered pair (x,y) to "no", this pair does not belong to the relation *R*. We can picture this definition on the plane with *X* and *Y* as two coordinates, and mark every ordered pair that belongs to the relation *R*.

Operation. An *operation* on a set *S* is a map from the Cartesian product $S \times S$ to the set *S*. When an operation is defined on a set, we say that the set is *closed under the operation*.

Thus, to specify an operation, we need to specify a set *S*, as well as a map. The domain of the map is the Cartesian product $S \times S$, and the codomain of the map is *S*. Thus, we write

$$S \times S \xrightarrow{\text{operation}} S.$$

Operation is usually a surjection, but not an injection.

Let *a* and *b* be two elements in *S*. An operation maps the ordered pair (a,b) to an element *c* in *S*. The operation is a special map, and we use a special notation. Instead of writing $c = f(a,b)$, we write

$$c = a * b.$$

This notation better reflects the nature of an operation: it comes between two elements, rather than before them.

Relation as a domain or codomain of a map. A relation is a set, which can of course serve as a domain or codomain of a map. Consider an example.

The dates of weddings define a map *f*, whose domain is the relation *R* of all married couples, and whose codomain is the set of dates *D*. We write the map as

$$f : R \to D .$$

We can also represent the same data as a ternary relation on the Cartesian products of the three sets: the set of all people *S*, the set of all people *S* again, and the set of dates *D*. But this ternary relation does not take advantage a simple fact: every wedding has a unique date.

Furthermore, the set of married couples—the relation *R*—is a very small subset of $S \times S$. Most of people are not married to one another. The Cartesian product $S \times S \times D$ is such a vast space that most triples have nothing to do with the dates of weddings.

Also, we cannot use the Cartesian product $S \times S$ as a domain of the map *f*, because not every pair of people are married.

Map and Equivalence

An equivalence relation is a relation on one set, and a map is a relation over two sets. The two types of relations interplay in many ways.

Many examples above define an equivalence relation on a set *S* by using a map from *S* to other set *X*.

Bijection and equivalence. Given a family of sets, the bijection between sets defines an equivalence relation on the family. For example, consider the family of number systems,

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

"Being bijective" defines an equivalent relation, which partitions the family into two subfamilies:

$$\left\{ N, Z, Q \right\}, \quad \left\{ R, C \right\}.$$

We can find a bijection between any pair of sets in each subfamily, but cannot find a bijection between two sets from different subfamilies. Each subfamily is an equivalent class.

Map and Order

An order is a relation on one set, and a map is a relation over two sets. The two types of relations interplay in many ways.

Map from unordered set to unordered set. On identifying the countries of birth of a set of people, we map an unordered set (people) to another unordered set (countries).

As another example, the six faces of a die is an unordered set. Before we mark the faces in any way, the faces look similar, and do not have any "natural" order. It is common to mark the faces by numbers 1, 2, 3, 4, 5, 6. If we regard

$\{1,2,3,4,5,6\}$ as an ordered set, the act of "marking" creates a map from an unordered set (the set of faces) to an ordered set (the set of numbers). If we do not wish to force an order on the set of faces, but still wish to tell them apart readily, we can mark the faces with a set of symbols, such as

$$\{\text{diamond, club, heart, spade, circle, triangle}\}.$$

This marking maps one unordered set to another unordered set.

Orders are seductive. We must guard ourselves against the impulse to impose an order on a set that does not need one.

Map from unordered set to ordered set. We often map people to numbers. Examples include social security numbers, passport numbers, and cell phone numbers. However, rarely these numbers are used to create an ordered list of people. One exception is the identification numbers for students. Such "IDs" are used to list students at universities in China.

On occasions, we need to disregard any existing order, such as the alphabetical order of names or the numerical order of IDs, and create a new order. For example, when we need to create an order in which a set of candidates appear on stage, we write numbers on pieces of paper, fold them, and ask each candidate to choose one piece. This process defines an order on the set of candidates. Alternatively, we say that the process defines a map from the set of candidates to the set of numbers.

Map from ordered set to unordered set. We list a set of words,

$$\{\text{crawl, fly, jump, run, swim, walk}\},$$

in the alphabetical order. A dictionary maps these words into their meanings—a set of five types of movements. We regard the five types of movements an unordered set. Thus, the dictionary is a map from an ordered set (words) to an unordered set (movements).

A computer prints out two columns, one consisting of a set of student IDs, and the other consisting of the Chinese names of the corresponding students. The printout defines a map from an ordered set (numbers) to an unordered set (people).

Map from ordered set to ordered set. Ordered sets thoroughly permeate "quantitative disciplines". In science and engineering, we deal with distance, duration, energy, mass, temperature, etc. In economics, we deal with supply, demand, commodities, labor, etc.

When both the domain and the codomain of a map are ordered sets, the graph of the map no long suffers the multiplicity due to the order in which we list the elements of the two sets. We can place the elements in the domain as an ordered list, and the elements in the codomain as another ordered list. We place the two ordered lists horizontally and vertically on a piece of paper, and represent the Cartesian product of the two sets by the points in the plane. Because both sets are ordered, the relative positions of points are fixed. We can represent the map by marking the ordered pairs.

Let X and Y be two ordered sets. A map $f: X \to Y$ is said to be a *monotonically increasing map* if, for every a and b in X,

$$a < b \Rightarrow f(a) < f(b).$$

That is, a monotonically increasing map preserves the order.

A map $f: X \to Y$ is said to be a *monotonically decreasing map* if, for every a and b in X,

$$a < b \Rightarrow f(a) > f(b).$$

That is, a monotonically decreasing map reverse the order.

For example, the map $x \mapsto x^3$ sends one real number to another real number. This map is a monotonically increasing map.

As another example, a university assigns identification numbers to its students. This process defines a map from an ordered set (the names of students in the alphabetical order) to another ordered set (the identification numbers). Usually this map is neither monotonically increasing, nor monotonically decreasing.

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 is (almost) a map. If we ignore a small number of exceptions, for every Chinese character, there is a unique spelling in the Latin alphabet. Thus the Pinyin system is a map that sends every Chinese character (ignoring the exceptions) to a unique spelling in the Latin alphabet.

Pinyin is a non-injective map. Because many different Chinese characters sound identical, these different Chinese characters have the same representation in Pinyin. Consequently, Pinyin is a non-injective map from Chinese Characters to spellings in the Latin alphabet.

This fact enables a popular genre of Chinese humor, which is impossible to translate to other languages. The large number of words with the same sound also makes Pinyin a difficult language for reading. It is impossible to determine the Chinese characters by looking at the Pinyin of a Chinese name of a person.

Composition of two maps. Parents in China name their children using Chinese characters. This naming process defines a map from a set of people to a set of Chinese names. The Pinyin system then maps the set of Chinese names to a set of names using the Latin alphabet. The composition of the two maps defines a map from the set of people to the set of names using the Latin alphabet. For example, a boy born in 1893 was named "毛泽东". Years later, his name was spelled in Pinyin as "Mao Zedong".

Pinyin maps an unordered set to an ordered set. 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 maps 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 defines a map from 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.

Model of Ranking*

Human beings rank everything instinctively, obsessively, and sometimes destructively. This instinct must be deeply rooted in our evolutionary past, in the struggle for survival, in the world of scarcity. Here we formulate the mathematical model of ranking.

Use and abuse of rank. Given a set of n beads, we can put them on a string in a total of $n!$ distinct sequences. Each sequence represents an order on the set of beads. For a set of many elements, the total number of possible orders on the set is vast. This mathematical fact, as well as our obsession for order, has spawned an equally vast business to select particular orders—that is, to put the beads on a string in some particular sequences.

We rank people in many ways. That is, we define many ranks on the same set of people. We can readily rank people by age, height, or weight. Can we rank people by beauty or intelligence? The answer is no because we do not agree on what beauty or intelligence is. Can we rank people by their scholarly achievement? We are not sure, but do it anyway.

Who is the better physicist, Newton, Maxwell, or Einstein? Questions like this create controversy, and never reach a definitive answer. We can certainly rank them according to the number of citations, the number of published words, etc. But none of these ranks will tell us who is the better physicist.

We also rank universities. We rank universities by the number of students, the student-to-faculty ratio, the rate of admission, or the salary of the president. We also rank universities by the opinions of deans, or by some secret formula. What does it mean to say, "Harvard is better than Stanford?"

We rank nations. We rank them by the number of people, the area, the fraction of population graduated from college, or the gross domestic product. We even rank them by some "index of happiness".

We rank earthquakes and winds according to their strength. Someone (or maybe some computer) in the government seems to know how to rank the level of terrorist threat.

Rank on a set. A *rank* on a set S is a partition whose parts form an ordered set. Each part in the partition is called an equivalent class in the rank.

The set S is in general an unordered set. Recall that a partition P of a set S is a family of disjoint nonempty subsets of S whose union is S. The partition P is also a set, each element of which is a subset of S and is called a part of P. We can certainly define an order on the partition.

Rank students by grades. To see how we produce a rank, let us look at an example: a professor ranks the students in her class by grades. This sentence mentions two sets. Let S be the set of students:

$$S = \left\{ \text{Amy,..., Zak} \right\}.$$

We regard this set as an unordered set. The alphabetical order of the names of student has nothing to do with the intended rank. Let G be the set of grades:

$$G = \left\{ \text{A, B, C, D, F} \right\}.$$

This set is an ordered set. These two sets are given to the professor.

We now come to the time-consuming task: the verb of the sentence. How does the professor *rank* the students? She does grading, a messy process that takes much effort from her. The object of grading is simple: grading gives *every* student a *unique* grade. Thus, grading is a map from S to G. She records this map by a table, with the names of students in one column, usually in alphabetical order, and their grades in another column.

The grades of the students partition the set of students S into five subsets (equivalent classes): every student in the first subset receives A, …, and every student in the fifth subset receives F. The grades defines a rank on S.

The mathematical model of ranking. The above example is familiar to everyone, but contains many non-mathematical aspects. How we obtain the three ingredients, the unordered set, the ordered set, and the map? These non-mathematical aspects are distracting. We next formalize the procedure to produce a rank on a set.

Let S be an unordered set of n elements:

$$S = \left\{ s_1, \ldots, s_n \right\}.$$

Let G be an ordered set of k elements:

$$G = \left\{ g_1, \ldots, g_k \right\}.$$

The map f sends every element in S to a unique element in G,

$$f : S \to G.$$

Often f is neither an injective map nor a surjective map. That is, f may send multiple elements in S to the same element in G, and may send no elements in S to certain elements in G. The preimage of every element in G under the map f is a subset of S (an equivalent class). Write the list of preimages of all individual elements in G under f as

$$f^{-1}\left(g_1\right), \ldots, f^{-1}\left(g_k\right).$$

Thus, the map f defines a rank on S using the order on G. If f is a non-surjective map, some of the preimages under f are the empty set. We remove the empty set from the list.

Many ranks on one unordered set. Given a set of people, we can create many ranks on the set.

For example, we can measure the weights of people. This sentence identifies the three ingredients: an unordered set S (a set of people), and ordered set W (a set of weights), and a map f (the measurement). The measurement defines a map from the set of people to the set of weights, $f : S \to W$. We rank the set of people according to their weights.

As another example, the relation "people having the same birthday" also defines a map from a set of people S to an ordered set (days in a year). The latter is ordered. This relation thus generates a rank on S. This rank, of course, differs from the rank by weight.

Given a three-element set $\left\{ x, y, z \right\}$, we can create a total of thirteen ranks on the set:

$$
\begin{aligned}
&(\{x\},\{y\},\{z\}),\ (\{x\},\{z\},\{y\}),\ (\{y\},\{x\},\{z\}),\\
&(\{y\},\{z\},\{x\}),\ (\{z\},\{x\},\{y\}),\ (\{z\},\{y\},\{x\}),\\
&(\{x\},\{y,z\}),\ (\{y,z\},\{x\}),\\
&(\{y\},\{z,x\}),\ (\{z,x\},\{y\}),\\
&(\{z\},\{x,y\}),\ (\{x,y\},\{z\}),\\
&\{x,y,z\}.
\end{aligned}
$$

Order-preserving coarsening. 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 F. The five subintervals form a partition of the interval from 0 to 100.

Order vs. rank. In producing a rank on a set *S*, the set *S* itself is usually an unordered set. As noted above, even though we often list students in the alphabetical order of their names, this order on the set of students has nothing to do with the rank on the set—the grades of the students. We should simply regard the set of students as an unordered set.

Thus, to produce a rank on an unordered set *S* is to define a map *f* from *S* to an ordered set *G*.

The relation "people born in the same country" defines a map from a set of people *S* to a set of countries. We have not defined an order on the set of countries. Consequently, this map does not generate a rank on *S*.

How to debunk an abuse of rank? An abuse of rank becomes evident when we answer the question, "How is the rank produced?" That is, in addition to the unordered set *S*, we should identify the ordered set *G* and the map *f*.

For example, we have ranked a set of people by their weights. This rank of people, of course, does not tell us how they will rank according to height, beauty, or intelligence.

Similarly, the rate of admission maps the set of universities to a set of numbers. The latter is an ordered set. We can certainly rank universities according to the rate of admission, but this rank does not tell us how the universities will rank according to the rate of graduation, the average salary of the graduates, etc. In particular, this rank does not tell us which university is best for our children.

The number of citations maps a set of scholars to a set of numbers. The latter is an ordered set. We can certainly rank the set of scholars according to the numbers of citations, and this rank does provide data. But we cannot use this rank to judge the accomplishments of the scholars. To do so will require us to define a map and an ordered set associated with the nebulous notion of "accomplishment".

A type of ranking emits an air of authority when we give an incomprehensible answer to the question, "How is the rank produced?" We may use a cumbersome formula to map a set of universities to a set of numbers. Often, the cumbersome formula draws on questionable inputs, such as opinions of deans and of the general public.

Temperature*

In search of an order. Just like any language, mathematics has been used to describe all things. As we will see, making up another rank of universities and discovering temperature follow the same mathematical model: a map from an unordered set to an ordered set.

It is perhaps frivolous to generate yet another rank of universities or of scholars. By contrast, it is often a long, painstaking effort to discover in nature a map from an unordered set to an ordered set.

What is temperature? How do we know that all values of temperature form an ordered set? What is the corresponding unordered set? How do we map the unordered set to temperature, an ordered set? We outline the empirical observations that lead to the discovery of temperature.

Feeling hot. Temperature is synonymous to *hotness*. A value of temperature means a *place of hotness*. But what is hotness?

Our feeling of hotness comes from everyday experience. We use the adjectives hot, warm, cool, and cold to indicate places of hotness. But the four words are insufficient to indicate all places of hotness. Everyday experience tells us that many places of hotness exist, and that all places of hotness can be represented by a real variable.

Why is hotness so different from happiness? Most our feelings however, cannot be represented by real numbers. Think of happiness, love, and anxiety. It is remarkable that this particular type of feeling—hotness—can be represented by real numbers. What makes hotness so different from happiness? This question is hard to answer because we do not know happiness in the same way as we know hotness. Instead, we ask a simpler question, What do we know about hotness?

We now form the concept of hotness from empirical observations of thermal contact. These observations are milestones in a long march toward a profound discovery of humankind: we can name all places of hotness by a real variable.

Translation. We organize our narrative of this long march using the mathematical model of ranking. Here is a dictionary that translates the words in the mathematical model to the words specific to the science of temperature:

- Unordered set: the set of states of equilibrium
- Ordered set: the set of all places of hotness
- A map from the unordered set to the ordered set: thermometry, the temperature of a state of equilibrium.

We next describe the three ingredients—the unordered set, the ordered set, and the map—in turn.

System. We can regard any part of the world as a *system*. For example, a proton and an electron constitute a system, which we call a hydrogen atom. A glass of wine is also a system. This system is composed of water, ethanol, and other molecules. Do we include the glass as a part of the system? Maybe, if we decide to study the reaction between water and glass. The decision is ours. We can regard any part of the world as a system. Even the empty space can be a system; the vacuum hosts electromagnetic field.

System interacts with the rest of the world. The hydrogen atom interacts with the rest of the world, for example, by absorbing or emitting photons. When the hydrogen atom absorbs a photon, the electron cloud changes its shape.

A system like the glass of wine may interact with the rest of the world in many ways. We see the wine because it reflects light. We smell the wine because molecules jump off the wine and enter our noses. We hear the wine when we move a wet finger around the edge of the glass. We warm up the wine when the vibration of the molecules in our hands transfers energy to the vibration of the molecules in the wine. We taste the wine when we drink it.

We can augment our five senses by using other things to interact with the wine. We subject the wine to a pressure to learn how tightly the molecules pack. We subject the wine to an electric field to find out how electric charges on the molecules move around. We pour fruit juice into the wine and watch they mix. We place the wine on a flame and smell the vapor.

Isolated system. An isolated system is a system that does not interact with the rest of the world.

To make a glass of wine an isolated system, we seal the glass to prevent molecules from escaping, we place the glass in a thermos to prevent energy from escaping by heat, and we make the seal unmovable so that the external pressure cannot do work to the wine. We are alert to any other modes of interaction between the wine and the rest of the world. Does the magnetic field of the earth affect the wine? If it does, we will find a way to shield the glass of wine from the magnetic field also.

Of course, nothing is perfectly isolated. Like any idealization, the isolated system is a useful approximation of the reality, so long as the interaction between the system and the rest of the world negligibly affects a phenomenon that we choose to study.

For example, it may be too much trouble for us to isolate the wine from gravity. Few have the luxury to study the wine under the zero-gravity condition. Gravity is important if we send the wine around, but is unimportant if we want to study the interactions between the molecules in the wine.

Observation 1: A system isolated for a long time reaches a state of equilibrium. When we fill half of a bottle with water, initially water is turbulent, and some water molecules leave the liquid and become vapor. The half bottle of water consists of liquid and vapor. We then make the half bottle of water into an isolated system, and wait for some time. The water becomes quiescent. Of course, molecules still move incessantly in the liquid and the vapor. Some water molecules jump off the liquid and join the vapor, and others do in the opposite directions. But the number of molecules in the vapor becomes independent of time. We say that the half bottle of water has reached a state of equilibrium.

With this observation, we will regard the two phrases "a system isolated for a long time" and "a state of equilibrium" as synonymous.

All states of equilibrium form a set S, which we regard as an unordered set.

Thermal contact. When two systems can exchange energy by heat, we say that the two systems are in *thermal contact*.

We will make the transfer of energy by heat the only mode of interaction between the two systems. We block all other modes of interaction between them. The two systems do not exchange molecules, and do not exchange energy by work. Furthermore, we put the two systems together as an isolated system.

Two systems can be in thermal contact without touching each other; for example, energy can transfer from one system to the other by electromagnetic radiation.

Thermal equilibrium. Observation 1 tells us that two systems in thermal contact for a long time will reach a state of equilibrium and stop transferring energy. The two systems are said to have reached *thermal equilibrium*.

For example, a glass of wine has been kept in a refrigerator for a long time and is then isolated, and a piece of cheese is kept in a room for a long time and is then isolated. When the glass of wine and the piece of cheese come into thermal contact, the vibration of the molecules in the wine will interact with the vibration of the molecules in the cheese, through the vibration of the molecules in the glass. After some time, energy re-allocates between the wine and the cheese, and stops transferring.

Observation 2: If two systems are separately in thermal equilibrium with a third system, the two systems are in thermal equilibrium with each other. This observation is known as *the zeroth law of thermodynamics*. This observation establishes that thermal equilibrium is an equivalence relation on the set of all states of equilibrium.

Places of hotness. Consider many isolated systems. We discover places of hotness by experiments. We bring two systems into thermal contact, and check if they exchange energy by heat. If the two systems in thermal contact do not exchange energy by heat, we say that they are at the same place of hotness. If the two systems in thermal contact exchange energy by heat, we say that they are at different places of hotness. All states of equilibrium at the same place of hotness form an equivalent class.

All places of hotness form a set, which we call the set of temperature, T.

Name places of hotness any way you like. Everyday experience indicates that many places of hotness exist. To talk about them individually, we need to give each place of hotness a name. Places of hotness are real: they exist in the experiment of thermal contact. How to name the places of hotness is our choice. The names exist between our lips, and in our ears and books. All naming schemes are arbitrary decisions of human beings (or committees), but some naming schemes are more convenient than others.

The situation is analogous to naming streets in a city. The streets are real: they exist regardless how we name them. We can name the streets by using names of presidents, or names of universities. We can use numbers. We can be inconsistent, naming streets in one part of the city by numbers, and those in another part of the city by names of presidents. We can even give the same street several names, using English, Chinese, and Spanish.

Thermometry. Thermometry is the art of measuring hotness. The art has become sophisticated, but its foundation remains simple: thermometry relies on thermal equilibrium. By "measuring the hotness of a state of equilibrium X" we mean matching X in thermal equilibrium with a state of equilibrium of a known place of hotness. Thermometry is a map that sends every state of equilibrium to a level of hotness. Thermometry enables us to partition the set of all states of equilibrium into a library of equivalent classes.

Here is how we build a library of places of hotness. Name the places of the known hotness in the library as $\left\{ \mathrm{A,B,C,...} \right\}$. We bring state X in thermal contact with a state at hotness A, and observe if the two systems exchange energy by heat. If they do not exchange energy by heat, we have just measured the hotness of system X—it is at hotness A. If they do exchange energy by heat, we know system X is at a place of hotness different from A. We then bring system X in thermal contact with a system at hotness B. We repeat this procedure until we match system X in thermal equilibrium with a system of a place of hotness in the library.

What if we cannot match system X with any place of hotness in the library? We have just discovered a new place of hotness! We are the Columbus in the new world of hotness. We add this new place of hotness to the library, and name the place X.

Now we have discovered the set of places of hotness. Thermometry is a map that sends a state of equilibrium to a place of hotness:

$$\text{themometry} : S \rightarrow T \ .$$

The set of all states of equilibrium S is and unordered set. How do we discover that the set of all places of hotness T is an ordered set?

Observation 3: When a system of hotness A and a system of hotness B are brought into thermal contact, if energy goes from the system of hotness B to the system of hotness A, energy will not go in the opposite direction. This observation is a version of *the second law of thermodynamics*, known as the Clausius statement.

When the two systems of different places of hotness are in thermal contact, the flow of energy is *unidirectional*. When two systems are in thermal contact, the system gaining energy is said to be at a lower place of hotness than the system losing energy. By this definition, this observation says that heat goes from hot place to a cold place.

We use the word "hot" strictly within the context of thermal contact. In this usage, it makes no sense to say that one movie is hotter than the other, because the two movies cannot exchange energy.

Observation 4: If hotness A is lower than hotness B, and hotness B is lower than hotness C, then hotness A is lower than hotness C.

This observation generalizes the zeroth law of thermodynamics. By making thermal contact, we can order any library of places of hotness one after another—that is,

places of hotness form an ordered set T.

We call an ordered library of places of hotness a *scale of hotness*.

Observation 5: Between any two places of hotness there exists another place of hotness. The experimental demonstration goes like this. We have two systems at hotness A and B, respectively, where hotness A is lower than hotness B. We can always find another system, which loses energy when in thermal contact with A, but gains energy when in thermal contact with B. This observation indicates that places of hotness are dense.

Name all places of hotness by a real variable. All places of hotness are ordered, so that we can name them by using a set of numbers. Places of hotness are dense, so that we cannot name them by using a set of integers, but we can name them by using a real variable. A map from places of hotness to a real variable is called a *numerical scale of hotness*.

Around 1720, Fahrenheit assigned the number 32 to the melting point of water, and the number 212 to the boiling point of water. What would he do for other places of hotness? Mercury is a liquid within this range of hotness and beyond, sufficient for most purposes for everyday life. When energy is added to mercury by heat, mercury expands. The various volumes of a given quantity of mercury could be used to name the places of hotness.

What would Fahrenheit do for a high place of hotness when mercury is a vapor, or a low place of hotness when mercury is a solid? He could switch to materials other than mercury; for example, he could use a flask of gas. He could also use phenomena other than thermal expansion, such as a change in electrical resistance of a metal due to heat.

Long march toward mapping hotness to a real variable. By now we have completed this long march. In the beginning of this long march, we have invoked the analogy of naming streets in a city. Now note two differences between naming streets and naming places of hotness. First, we do not have a useful way to name all streets by an ordered array. In what sense one street is higher than the other? Second, we do not need a real variable to name all the streets. A city has a finite number of streets.

By contrast, observations 4 and 5 enable us to name all places of hotness by a ordered, continuous variable. Most textbooks state that the zeroth law (observation 2) establishes the concept of hotness. This statement is wrong. Zeroth law does not enable us to name all places of hotness by an ordered, continuous variable.

Map one numerical scale of hotness to another. Once a numerical scale of hotness is set up, any monotonically increasing function maps this scale to another scale of hotness. For example, in the Celsius scale, the freezing point of water is set to be 0C, and the boiling point of water 100C. We further require that the Celsius (C) scale be linear in the Fahrenheit (F) scale. These prescriptions give a map from one scale of hotness to the other:

$$C = \frac{5}{9}\left(F - 32\right).$$

In general, the map from one numerical scale of hotness to another need not be linear. Any increasing function will preserve the order of the places of hotness. Any smooth function will preserve the continuity of the places of hotness.

Numerical values of hotness do not obey arithmetic rules. Using numbers to name places of hotness does not authorize us to apply arithmetic rules: adding two places of hotness has no empirical significance. It is as meaningless as adding the addresses of two houses on a street. House number 2 and house number 7 do not add up to become house number 9. Also, raising the hotness from 0C to 50C is a different process from raising hotness from 50C to 100C. For instance, for a given amount of water, raising the hotness from 0C to 50C takes different amount of energy from raising hotness from 50C to 100C.