Equivalence Relations

Often we want to express the notion that two elements of a set S are "related" in some way. For example, if $S = \mathbf{Z}$ we might say that m is related to n if $m < n$; or we might want to say that any two even integers are related to each other and any two odd integers are related to each other. To define carefully just what we can mean by "related", without getting into semantic or philosophical difficulties, we look at the logical essence of what's involved in the intuitive notion of "relation": the ability to take an ordered pair of elements of a set, and determine whether a certain statement about that ordered pair is true. If the given statement is true about a pair (x, y), we say that x is related to y. We thereby obtain a set of pairs (x, y)—i.e. a subset of the Cartesian product $S \times S$—for which the given statement is true. This leads us to the following definition.

Definition. A *relation* R on a set S is a subset of $S \times S$. We sometimes use the notation "xRy" as short-hand for the statement "$(x, y) \in R$". We often read the notation xRy as "x is related to y".

Some people prefer the term "binary relation".

Note that any function is a relation, but not every relation is a function. For a relation, we do not require that for all $x \in S$, there exist a $y \in S$ for which x is related to y (in fact, we do not even require a relation to be nonempty), and we allow a single element x to be related to more than one element y.

Henceforth I'll use a squiggle ($\sim$) instead of an R, and write $x \sim y$ instead of xRy.

We define several nice properties a relation can have.

Definition. A relation $\sim$ on S is
(a) *reflexive* if $x \sim x$, $\forall x \in S$,
(b) *symmetric* if $x \sim y$ implies $y \sim x$,
(c) *transitive* if whenever $x, y, z \in S$ are distinct and $x \sim y$ and $y \sim z$, we have $x \sim z$.

You can find relations that have none of these properties, or that have any combination of these properties but not remaining properties.

Examples. The relation "$<$" on $\mathbf{N}$ is transitive, but not reflexive or symmetric. The relation "$\leq$" on the same set is transitive and reflexive, but not symmetric. The relation "is a first cousin of" on the set of people is symmetric but not transitive or reflexive.

The best relations are those that have all three nice properties.

Definition An *equivalence relation* on a set S is a relation that is reflexive, symmetric, and transitive.

Note that since an equivalence relation is reflexive, it is automatically nonempty, provided S is nonempty.

Examples of equivalence relations.

1. Let $S = \mathbf{Z}$. Declare $x \sim y$ iff $x - y$ is divisible by 2. This equivalence relation is called *congruence modulo* 2; in place of $x \sim y$ one usually writes $x \equiv y \mod 2$. Similarly we can define congruence modulo any positive integer.

2. Let S be the set of points on dry land on the earth's surface. Declare two points related if you can get from one to the other without crossing water. Obviously this is an equivalence relation.

3. Let $\mathcal{A}$ be the collection of all sets. (Note: I am using the word "collection" here instead of "set" because there are dangers and subtleties in talking about the "set of all sets", which I don't want to get into here. But for the purposes of this handout, "collection" means the same thing as "set.) Declare two sets X, Y related if there exists a bijection $f : X \to Y$. It is easy to check that this is an equivalence relation.

4. Let $\sim_1$ be *any* relation on a nonempty set S. We define relations $\sim_2, \sim_3$ on S as follows. We declare $x \sim_2 y$ if either $x = y$, $x \sim_1 y$, or $y \sim_1 x$. (Thus $\sim_2$ is reflexive and symmetric, but it may not transitive.) We declare $x \sim_3 y$ if there exists a finite set of elements $\{z_1, z_2, ..., z_n\}$ such that $x \sim_2 z_1$, $z_n \sim_2 y$, and $z_i \sim_2 z_{i+1}$ for $1 \leq i \leq n-1$ (i.e. if there is a "chain" beginning at x, ending at y, with each element related by $\sim_2$ to the next one in the chain).[1] Then $\sim_3$ is an equivalence relation, called the *equivalence relation generated by the relation* $\sim_1$. Some examples are:

   (a) Let S be the set of animals. Declare two animals related if they can breed to produce fertile offspring. The equivalence relation generated by this is (by definition) "being in the same species".

   (b) Let S be the set of human beings. Declare two people related if they have met each other. I'll call the equivalence relation generated by this the "networking" relation.

By now you will have noticed that an equivalence relation partitions a set into a bunch of disjoint subsets, each of which consists of mutually equivalent elements. These subsets are called *equivalence classes*. The precise theorem is

[1] The relation $\sim_3$ can be defined more quickly, if somewhat less intuitively, without introducing $\sim_2$ at all: declare $x \sim_3 y$ if and only if either $x = y$ or there exists a finite set of elements $\{z_1, z_2, ..., z_n\}$ in S, with $x = z_1$ and $y = z_n$, such that $z_i \sim_1 z_{i+1}$ or $z_{i+1} \sim_1 z_i$ for $1 \leq i \leq n-1$.

Theorem. Let $\sim$ be an equivalence relation on a nonempty set S. Then there exists a unique collection C of nonempty subsets of S with the following properties.

(i) If $V_1, V_2 \in C$ and $V_1 \neq V_2$, then $V_1 \bigcap V_2 = \emptyset$ (i.e. distinct equivalence classes are disjoint).

(ii) $\bigcup_{V \in C} V = S$ (i.e. every element of S lies in some equivalence class).

(iii) For all $x, y \in S$, there exists $V \in C$ containing both x and y iff $x \sim y$.

I leave the proof to you as an exercise.

Let's re-examine the examples of equivalence relations above. In example 1, there are exactly two equivalence classes, one consisting of the even integers, and the other the odd integers. In example 2, the equivalence classes are the land masses such as continents and islands. In example 3 the equivalence classes are *cardinalities*, extending the notion of "number of elements in set" from finite sets to sets that may be finite or infinite. In example 4a, of course, the equivalence classes are the species. In example 4b (variants of which are relevant to the spread of infectious diseases), I don't know what to call the equivalence classes, but it's amusing to ponder how many there have been as a function of time. Clearly there were many equivalence classes thousands of years ago. Today there is probably only one class, unless there are still some isolated groups somewhere in the world as yet undiscovered by the rest of us.