Sets: The basics

Sets are the fundamental building block of modern mathematics. Intuitively, a set is a collection of objects, but with two important twists:

1. Sets are unordered.
2. Sets contain no duplicates.

Example 1.
Suppose you want to keep a record of which words occur in a text. You aren't interested in how often a given word occurred, just whether it occurs at all. Nor do you care in which order the words occurred in the text. So you are actually interested in the *set* of words that occur in the text.

Each property is explained in detail below, but let's first put some helpful notation in place.

1 List notation

Sets are often written as lists with curly braces around them. So $\{a, b, c, d\}$ denotes the set containing a, b, c, d. Here a, b, c, d are some arbitrary objects. This is known as **list notation**. More complex sets are defined with **set-builder notation**, which will be covered in a later unit.

Example 2.
Consider the string *If John slept, then Mary left*. Its set of words (ignoring sentence-initial capitalization) is $\{\text{if}, \text{John}, \text{left}, \text{Mary}, \text{slept}, \text{then}\}$.

Exercise 1.
Write the following as a set:

- the first names of your three favorite actors/actresses,
- the colors of the rainbow,
- all even numbers between 1 and 11

2 Elements and set membership

The objects contained in a set are called its **elements** or **members**. One writes $e \in S$ to indicate that e is an element of S. The opposite is denoted $e \notin S$: e is not an element of S. The symbol $\in$ thus indicates **set membership**.

Example 3.
Let W be the set of words in the string *If John slept, then Mary left*. Then it holds that $left \in W$ and $right \notin W$. But it is not the case that $then \notin W$ or $awake \in W$.

You might be wondering why we use the symbol $\in$ for set membership. The following mnemonic, while historically inaccurate, may help you with remembering the notation: $\in$ looks like a stylized E, and $x \in S$ means that x is an *E*lement of S.

Sometimes $\ni$ is used as the mirror image of $\in$. For example, $a \in S$ could also be written as $S \ni a$.

Example 4.
Continuing the previous example, it is true that $left \in W \ni then$. That is to say, both $left \in W$ and $then \in W$ are true.

Exercise 2.
Put $\in$, $\ni$, $\notin$, $\not\ni$ in the gaps below as appropriate:

- $5\_\{1,2,4,5,8\}$
- $6\_\{1,2,4,5,8\}$
- $\{5\}\_\{1,2,4,5,8\}$
- $5\_\{1,2,4,5,8\}\_6$

Sets can contain arbitrary objects, including other sets. However, the members of a set are just the elements immediately contained by the set, not what might in turn be contained inside of those elements.

Example 5.
The set $\{a,\{b\}\}$ has two members: a and $\{b\}$. While b is an element of $\{b\}$, it is not an element of $\{a,\{b\}\}$.

Exercise 3.
Put $\in$, $\ni$, $\notin$, $\not\ni$ in the gaps below as appropriate:

- $5\_\{1,\{2,4\},5,8\}$
- $6\_\{1,\{2,4,5,8\}\}$
- $\{5\}\_\{1,\{2,4\},\{5\},8\}$
- $5\_\{\{1,2,4,5,8\}\}\_6$

3 Lack of order

Even though we may write sets in a linear fashion as lists, they have no internal order. The set $\{a, b\}$ could also be written as $\{b, a\}$. So we have $\{a, b\} = \{b, a\}$, and $\{a, b, c\} = \{a, c, b\} = \{b, a, c\} = \{b, c, a\} = \{c, a, b\} = \{c, b, a\}$.

Example 6.
Consider the strings *If John slept, then Mary left* and *If Mary left, then John slept*. While they are clearly distinct sentences, their sets of words are identical.

Exercise 4.
For each one of the following, fill the gap with $=$ or $\neq$ as appropriate:

- $\{a, b\}\_\{a, b\}$
- $\{b, a\}\_\{a, b\}$
- $\{b, a, c, d\}\_\{e, a, b, d\}$

4 Lack of duplicates/Idempotency

Sets are **idempotent**, which means that duplicates are ignored. So $\{a, b\} = \{a, a, b\} = \{a, b, b, a, b, a, b, a, a\}$. It also holds that $\{a\} = \{a, a\} = \{a, a, a\}$, and so on.

Example 7.
Linguists distinguish between **word types** and **word tokens**. The sentence *dogs love dogs* contain two tokens of the type *dogs*, and one token of the type *love*. The sentences *dogs love* and *dogs love dogs* are different with respect to word tokens, but identical with respect to word types. So if you care about word types rather than word tokens, you’re dealing with a set because the only thing that matters is which words the text contains, not how many tokens of each word.

Example 8.
Consider the sentence *If police police police, then police police police*. Its set of words (ignoring capitalization) is $\{\text{if}, \text{police}, \text{then}\}$.

Exercise 5.
For each one of the following, fill the gap with $=$ or $\neq$ as appropriate:

- $\{a, b\}\_\{a, a, b, b\}$
- $\{b, a\}\_\{a, b, a\}$
- $\{c, b, a, a, d, c\}\_\{a, a, b, d, c, c, c\}$
- $\{a\}\_\{a, a, a, a, a, a, c, a, a, a, a, a, a\}$

Exercise 6.
The sentence *If police police police, then police police police* actually uses two different word types. It just just so happens that both are pronounced and spelled *police*. But one is the **noun** *police*, the other one the **verb** *police*. We can use the **parts of speech** N and V, respectively, to distinguish between the noun *police* and the verb *police*. Let us add parts of speech to all the words in the string (we use C for the **complementizers** *if* and *then*): *If[C] police[N] police[V] police[N], then[C] police[N] police[V] police[N]*. What would be the corresponding set of words for this string (i.e. where we now distinguish between *police[N]* and *police[V]*)?

It is important to keep in mind that idempotency only holds with respect to the elements of a set, not what may be contained by those elements.

Example 9.
Even though $\{a\} = \{a, a\}$, it is not the case that $\{a\} = \{a, \{a\}\}$. The former is the set containing a, the other is the set containing a and the set of a.

Exercise 7.
For each one of the following, fill the gap with $=$ or $\neq$ as appropriate:

- $\{a, b\}\_\{a, a, b, \{b\}\}$
- $\{\{b\}, \{a\}\}\_\{\{a\}, \{b\}, \{a\}\}$
- $\{c, b, \{a, a\}, d, c\}\_\{\{a\}, b, d, c, c, c\}$
- $\{a\}\_\{a, a, a, a, a, a, \{a\}, a, a, a, a, a, a\}$

5 Recap

- Sets are collections of arbitrary objects.
- Sets are unordered and idempotent (= duplicates are ignored).
- Sets can be defined with list notation, e.g. $\{a, b\}$.
- The objects contained in a set are called its *elements* or *members*.
- The symbols $\in$ and $\notin$ are used to indicate membership and non-membership, respectively.
- Occasionally, $\ni$ is used as the mirror image of $\in$.
- While sets may contain objects that are themselves collections of other objects, these objects inside objects are not considered for set membership. Remember: $a \in \{a\}$ and $\{a\} \in \{\{a\}\}$, but $a \notin \{\{a\}\}$.