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 Element 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 $a \notin \{\{a\}\}$.

Solutions

Solution to Exercise 1.

- $\{\text{Bruce, Marlene, Bela}\}$
- $\{\text{red, orange, yellow, green, blue, indigo, violet}\}$
- $\{2, 4, 6, 8, 10\}$

Explanation.

1. Suppose that your favorite actors are Bruce Campbell (*Evil Dead*), Marlene Dietrich (*The Blue Angel*, *Witness for the Prosecution*), and Bela Lugosi (*Dracula*). Their first names are *Bruce*, *Marlene*, and *Bela*, so the set that contains their first names, and nothing else, is $\{\text{Bruce, Marlene, Bela}\}$.

2. If you're like me, you probably had to look this one up, but rainbows have six colors, which are red, orange, yellow, green, blue, indigo, and violet. So the set we're looking for contains these six colors. We can specify that set using the names of these colors, but it would also be okay to use colored squares instead of names, colored lines, anything that conveys clearly that this is the set that contains these six colors.

3. A number is even if and only if it is a multiple of 2. Between 1 and 11, that's 2, 4, 6, 8, and 10.

Solution to Exercise 2.

1. $5 \in \{1, 2, 4, 5, 8\}$
2. $6 \notin \{1, 2, 4, 5, 8\}$
3. $\{5\} \notin \{1, 2, 4, 5, 8\}$
4. $5 \in \{1, 2, 4, 5, 8\} \notin 6$ or $5 \in \{1, 2, 4, 5, 8\} \not\ni 6$.

Explanation.

1. The set $\{1, 2, 4, 5, 8\}$ has 5 as one of its members, which we write as $5 \in \{1, 2, 4, 5, 8\}$.
2. On the other hand, 6 is nowhere to be found in $\{1, 2, 4, 5, 8\}$, its only elements are 1, 2, 4, 5, and 8. Hence $6 \notin \{1, 2, 4, 5, 8\}$.
3. This one is a bit tricky. We already know that $\{1, 2, 4, 5, 8\}$ contains 5, but that's not the same thing as containing $\{5\}$, i.e. the set containing 5. None of the elements of $\{1, 2, 4, 5, 8\}$ is $\{5\}$. By the way, this is also why $5 \notin \{1, 2, 4, \{5\}, 8\}$. Never confuse an object with a set containing that object, the two are very different things.
4. We already know that $\{1, 2, 4, 5, 8\}$ contains 5, so the first gap must be $\in$. For the second gap, there are two options. If we fill the gap with $\notin$, we are saying that $\{1, 2, 4, 5, 8\}$ is not a member of 6, which is true, but it's an odd thing to say as 6 isn't a set to begin with. Instead, it makes more sense to fill

the gap with $\not\ni$, which states that 6 is not a member of $\{1,2,4,5,8\}$.

Solution to Exercise 3.

1. $5 \in \{1, \{2,4\}, 5, 8\}$
2. $6 \notin \{1, \{2,4,5,8\}\}$
3. $\{5\} \in \{1, \{2,4\}, \{5\}, 8\}$
4. $5 \notin \{\{1,2,4,5,8\}\} \not\ni 6$

Explanation.
This isn't all too different from the previous exercise, we just have to be careful not to get confused by all the set brackets.

1. We now have the $\{1, \{2,4\}, 5, 8\}$ instead of $\{1,2,4,5,8\}$ from the previous exercise, but this difference doesn't matter for the set membership of 5.
2. This is another straightforward one where the change between the two exercises doesn't make a difference.
3. Here the change is important. Before, we saw that $\{5\} \notin \{1,2,4,5,8\}$ because the latter contains only 5, which is distinct from $\{5\}$. But now we do have a set that actually contains the set containing 5.
4. Here it is really important to read carefully. We are not dealing with $\{1,2,4,5,8\}$, which is the set containing 1, 2, 4, 5, 8 and nothing else. Instead, we have $\{\{1,2,4,5,8\}\}$. This set has only one element, which is $\{1,2,4,5,8\}$.

Solution to Exercise 4.

1. $\{a,b\} = \{a,b\}$
2. $\{b,a\} = \{a,b\}$
3. $\{b,a,c,d\} \neq \{e,a,b,d\}$ (assuming that c and e are distinct objects)

Explanation.

1. Two sets are identical if they contain exactly the same elements. This is clearly the case here, both sets contain a, both of them also contain b, and they contain nothing else.
2. It may seem like $\{b,a\}$ is different from $\{a,b\}$, but remember that sets are unordered. It does not matter in what order we write down the elements, all that matters is whether the sets have the same members. And this is still the case here.
3. Each one of the sets contains four elements, three of which are a, b, and d. Only the first set contains c, and only the second set contains e. This suggests that the sets are not equivalent. However, we have to be careful here as c and e may just be different ways of referring to the same object. As a concrete example, suppose that we are talking about sets of actors, and we are looking at the sets {Arnold} and {Schwarzenegger}. These are not

actually distinct sets, they are just different ways of writing down the set that contains only the actor Arnold Schwarzenegger. So remember: what matters isn't how we choose to write down a set, what matters is what it actually contains. If both c and e are just ways of referring to, say, the number 5, then the two sets above are equivalent.

Solution to Exercise 5.

1. $\{a, b\} = \{a, a, b, b\}$
2. $\{b, a\} = \{a, b, a\}$
3. $\{c, b, a, a, d, c\} = \{a, a, b, d, c, c, c\}$
4. $\{a\} \neq \{a, a, a, a, a, a, c, a, a, a, a, a, a\}$

Explanation.

1. Remember, one and the same element cannot be contained in a set multiple times. Either it is a member of the set, or it is not a member, it cannot be a member two times, or three times, or anything like that. Hence $\{a, a, b, b\} = \{a, b, b\} = \{a, b\}$.
2. The same logic applies in this case: $\{a, b, a\} = \{a, b\}$, which is the same as $\{b, a\}$ because sets are not ordered.
3. Again we have two identical sets. The first set contains a, b, c, d, and nothing else, and the same is true for the second set.
4. Finally we have two sets that are distinct (if one assumes that $a \neq c$). The second set could be written more compactly as $\{a, c\}$, and that is not the same set as $\{a\}$ because only the former contains c.

Solution to Exercise 6.

$\{\text{if}, \text{police[N]}, \text{police[V]}, \text{then}\}$.

Explanation.

We can build this set incrementally by moving from left to right through the string *If police[N] police[V] police[N], then police[N] police[V] police[N]* (ignoring capitalization). The string starts with *if*, so at this point our set is $\{\text{if}\}$. Next we see *police[N]*, and since this is not in our set yet, we add it and obtain $\{\text{if}, \text{police[N]}\}$. After that we encounter *police[V]*, which is not an element of our set yet — the set only contains the noun *police*, not the verb *police*. Hence we add *police[V]*, yielding $\{\text{if}, \text{police[N]}, \text{police[V]}\}$. Now the next word is *police[N]*, but since that's already in our set we do not need to add it again (there's no point in writing down the same element multiple times). After that we add *then*, and at this point we are done. Our set is $\{\text{if}, \text{police[N]}, \text{police[V]}, \text{then}\}$, and our example sentence contains no words after *then* that aren't already members of this set.

Solution to Exercise 7.

1. $\{a, b\} \neq \{a, a, b, \{b\}\}$
2. $\{\{b\}, \{a\}\} = \{\{a\}, \{b\}, \{a\}\}$
3. $\{c, b, \{a, a\}, d, c\} = \{\{a\}, b, d, c, c, c\}$
4. $\{a\} \neq \{a, a, a, a, a, a, \{a\}, a, a, a, a, a, a\}$

Explanation.

1. Since b is not the same as $\{b\}$, $\{a, a, b, \{b\}\}$ contains an element that is not a member of $\{a, b\}$, which means that the two are distinct sets.
2. As sets do not contain duplicates, it holds that $\{\{a\}, \{b\}, \{a\}\} = \{\{a\}, \{b\}\}$. But sets are also unordered, so that $\{\{a\}, \{b\}\} = \{\{b\}, \{a\}\}$.
3. Again we have to keep in mind that sets do not contain duplicates, and as a result $\{a, a\} = \{a\}$. We thus can simplify the two sets as follows: $\{c, b, \{a, a\}, d, c\} = \{c, b, \{a\}, d, c\} = \{c, b, \{a\}, d\} = \{\{a\}, b, c, d\}$ and $\{\{a\}, b, d, c, c, c\} = \{\{a\}, b, d, c\} = \{\{a\}, b, c, d\}$.
4. We can reduce $\{a, a, a, a, a, a, \{a\}, a, a, a, a, a, a\}$ to $\{a, \{a\}\}$, but since $a \neq \{a\}$, we can see that $\{a, \{a\}\}$ is not the same set as $\{a\}$.