Definition: Unique decodability

In many situations, we may want to transmit or store more than a single symbol from the source. Instead, we want to transmit or store a list of symbols, or even a potentially infinite stream of symbols.

Definition: Extended symbol code

Let $C : \mathcal{X} \to \mathcal{A}^*$ be a symbol code. The extended code $C^* : \mathcal{X}^* \to \mathcal{A}^*$ is defined by concatenation:

$$C^*(x_1, \ldots, x_n) := C(x_1)| \cdots |C(x_n).$$

The injectivity of C ensures that we can always uniquely decode $C(x)$. However, if one transmits a sequence $x_1, \ldots, x_m \in \mathcal{X}$ (or stores them "sequentially") by sending the concatenation $C(x_1, \ldots, x_n)$, ambiguities may arise, namely in cases where it is possible to parse this long string in two consistent but different ways. Indeed, injectivity of the encoding function per se does not rule out that there exists a positive integer m' and elements $x_1', \ldots, x_{m'}' \in \mathcal{X}$ such that $C(x_1)| \cdots |C(x_m) = C(x_1')| \cdots |C(x_{m'}')$. Of course, this problem can be circumvented by introducing a special separation symbol. However, such a symbol might not be available, and maybe even more importantly, even if an additional symbol *is* available, then one can often create a better code by using it as an ordinary alphabet symbol (in addition to 0 and 1) rather than as a special separation symbol. This is why it is interesting to study the following class of symbol codes:

Definition: Uniquely decodable code

A binary symbol code $C : \mathcal{X} \to \{0, 1\}^*$ is uniquely decodable if C^* is injective as well.

In the quiz, you will see an example of a code that is not uniquely decodable.