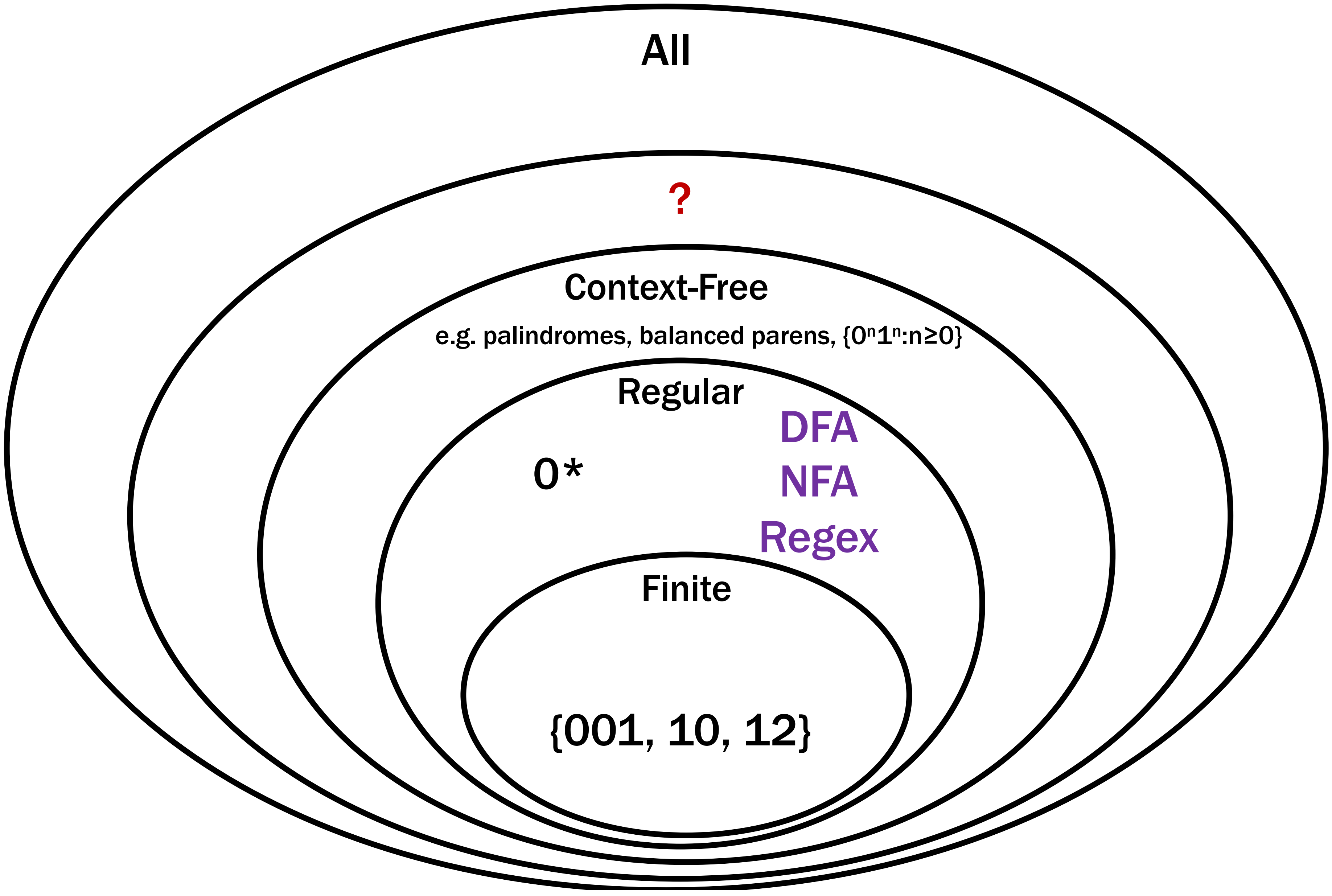
CSE 311: Foundations of Computing

Topic 11: Undecidability

Last time: Languages and Representations



Computers from Thought

Computers as we know them grew out of a desire to avoid bugs in mathematical reasoning.

Hilbert in a famous speech at the International Congress of Mathematicians in 1900 set out the goal to mechanize all of mathematics.

In the 1930s, work of Gödel and Turing showed that Hilbert's program is impossible.

Gödel's Incompleteness Theorem
Undecidability of the Halting Problem

Both of these employ an idea we will see called diagonalization.

The ideas are simple but so revolutionary that their inventor Georg Cantor was initially shunned by the mathematical leaders of the time:

Poincaré referred to them as a "grave disease infecting mathematics."

Kronecker fought to keep Cantor's papers out of his journals.

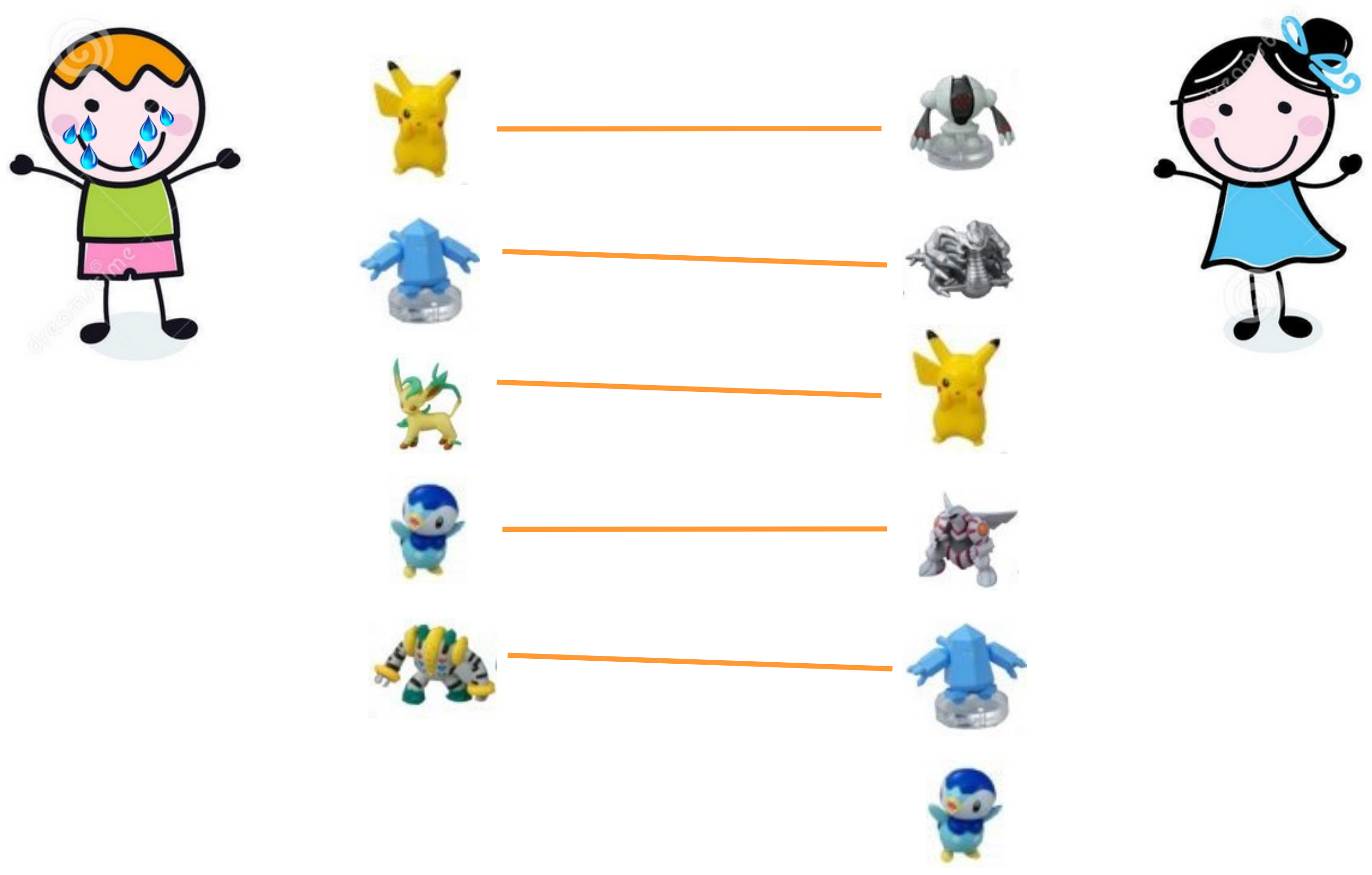
Full employment for mathematicians and computer scientists!!

Cardinality

What does it mean that two sets have the same size?

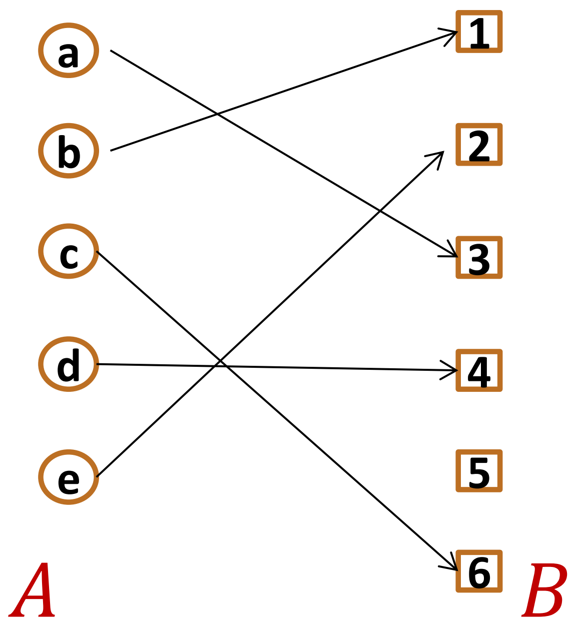
Cardinality

What does it mean that two sets have the same size?



1-to-1 Functions

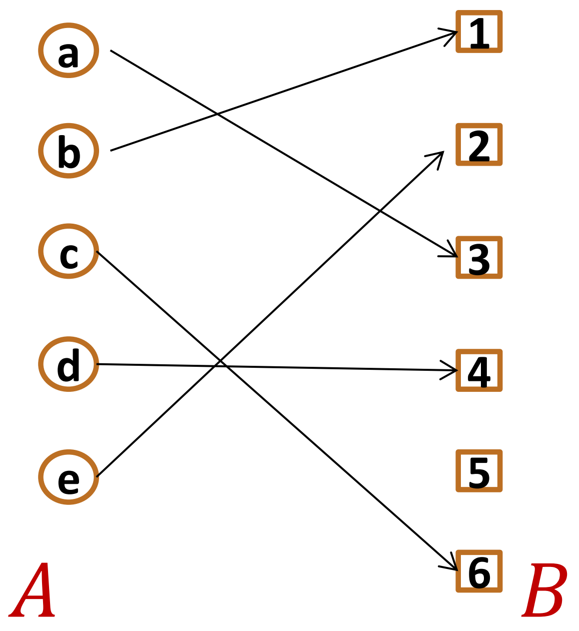
A **function** $f : A \rightarrow B$ is **one-to-one** (1-to-o1) if every output corresponds to at most one input;
i.e. $f(x) = f(x') \Rightarrow x = x'$ for all $x, x' \in A$.



1-1

Onto Functions

A function $f : A \to B$ is **onto** if every output gets hit;
i.e. for every $y \in B$, there exists $x \in A$ such that $f(x) = y$.



not onto

1-to-1 Functions

A function $f : A \to B$ is **one-to-one** (1-to-o1) if every output corresponds to at most one input;
i.e. $f(x) = f(x') \Rightarrow x = x'$ for all $x, x' \in A$.

Pigeonhole Principle: if A is larger than B (i.e., $|A| > |B|$),
then there is **no 1-to-1** function $A \to B$

e.g., the machine M takes two strings to the same state

Contrapositive: if there is a **1-to-1** function $A \to B$,
then A is not larger than B (i.e., $|A| \leq |B|$)

Onto Functions

A function $f : A \to B$ is onto if every output gets hit;
i.e. for every $y \in B$, there exists $x \in A$ such that $f(x) = y$.

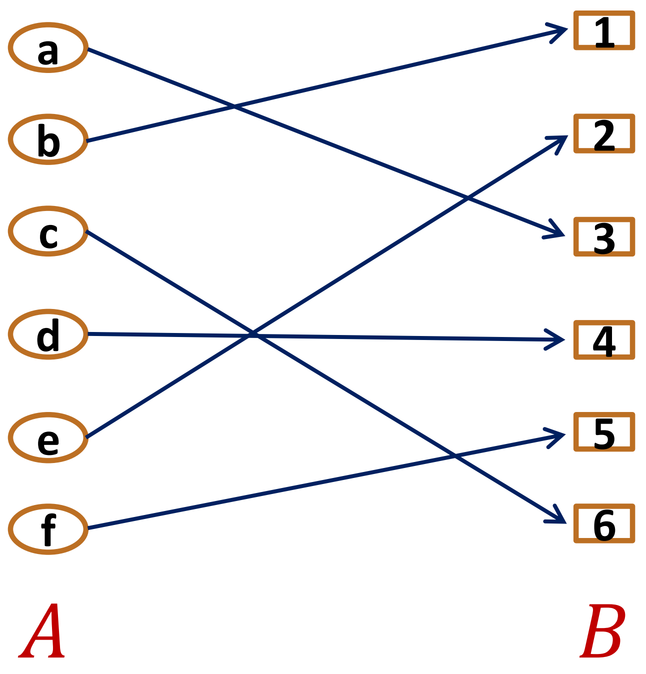
New Principle: if A is smaller than B (i.e., $|A| < |B|$),
then there is **no onto** function $A \to B$

there is some $y \in B$ such that no $x \in A$ is taken to y by f

Contrapositive: if there is an **onto** function $A \to B$,
then A is not smaller than B (i.e., $|A| \geq |B|$)

Cardinality

Definition: Two sets A and B have the same **cardinality** if there is a one-to-one correspondence between the elements of A and those of B. More precisely, if there is a **1-1 and onto** function $f : A \to B$.



1-1 proves ≤
onto proves ≥

The definition also makes sense for infinite sets!

Cardinality

Do the natural numbers and the even natural numbers have the same cardinality?

Yes!

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	…
0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	…

What's the map $\boldsymbol{f} : \mathbb{N} \to \mathbf{2}\mathbb{N}$? $\boldsymbol{f(n) = 2n}$

Countable sets

Definition: A set is **countable** iff it has the same cardinality as some subset of $\mathbb{N}$.

Equivalent: A set $\boldsymbol{S}$ is countable iff there is an *onto* function $\boldsymbol{g} : \mathbb{N} \to \boldsymbol{S}$

Equivalent: A set $\boldsymbol{S}$ is countable iff we can order the elements $\boldsymbol{S} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \dots\}$

The set $\mathbb{Z}$ of all integers

The set $\mathbb{Z}$ of all integers

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	...
0	1	-1	2	-2	3	-3	4	-4	5	-5	6	-6	7	-7	...

The set $\mathbb{Q}$ of rational numbers

We can’t do the same thing we did for the integers.

Between any two rational numbers there are an infinite number of others.

The set of positive rational numbers

1/1	1/2	1/3	1/4	1/5	1/6	1/7	1/8	…
2/1	2/2	2/3	2/4	2/5	2/6	2/7	2/8	…
3/1	3/2	3/3	3/4	3/5	3/6	3/7	3/8	…
4/1	4/2	4/3	4/4	4/5	4/6	4/7	4/8	…
5/1	5/2	5/3	5/4	5/5	5/6	5/7	…	
6/1	6/2	6/3	6/4	6/5	6/6	…		
7/1	7/2	7/3	7/4	7/5	….			
…	…	…	…	…				

The set of positive rational numbers

The set of all positive rational numbers **is countable.**

$$\mathbb{Q}^{+} = \{\mathbf{1/1, 2/1, 1/2, 3/1, 2/2, 1/3, 4/1, 2/3, 3/2, 1/4, 5/1, 4/2, 3/3, 2/4, 1/5,}\ \ldots\}$$

List elements in order of numerator+denominator, breaking ties according to denominator.

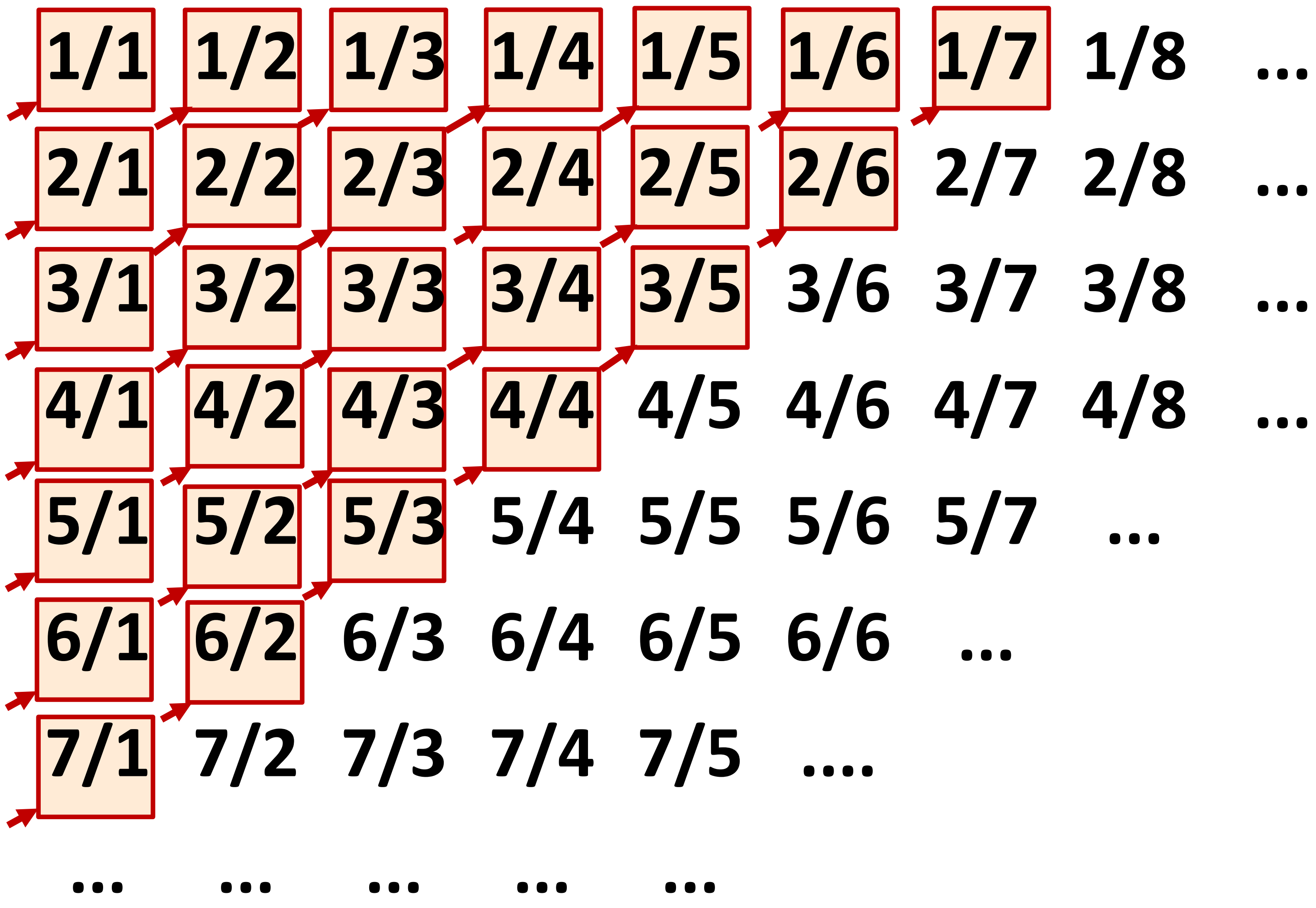
Only $\boldsymbol{k}$ numbers have total of sum of $\boldsymbol{k+1}$, so every positive rational number comes up some point.

The technique is called "**dovetailing**."

More generally:

- Put all elements into *finite* groups
- Order the groups
- List elements in order by group (arbitrary order within each group)

The set of positive rational numbers



Claim: Σ^* is countable for every finite Σ

Dictionary/Alphabetical/Lexicographical order is bad

- Never get past the A's
- A, AA, AAA, AAAA, AAAAA, AAAAAA, ….

Claim: Σ^* is countable for every finite Σ

Dictionary/Alphabetical/Lexicographical order is bad

- Never get past the A's
- A, AA, AAA, AAAA, AAAAA, AAAAAA, ….

Instead, use same "dovetailing" idea, except that we group based on length: only $|\Sigma|^k$ strings of length k.

e.g. {0,1}* is countable:

{ε, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, 101, 110, 111, … }

The set of all Java programs is countable

Java programs are just strings in Σ^* where Σ is the alphabet of ASCII characters.

Since Σ^* is countable, so is the set of all Java programs.

More generally, any subset of a countable set is countable: it has same cardinality as an (even smaller) subset of $\mathbb{N}$

OK OK... Is Everything Countable ?!!

Are the real numbers countable?

Theorem [Cantor]:
The set of real numbers between 0 and 1 is **not** countable.

Proof will be by contradiction.
Uses a new method called diagonalization.

Real numbers between 0 and 1: $[0,1)$

Every number between 0 and 1 has an infinite decimal expansion:

1/2	=	0.50000000000000000000000…
1/3	=	0.33333333333333333333333…
1/7	=	0.14285714285714285714285…
π-3	=	0.14159265358979323846264…
1/5	=	0.19999999999999999999999…
	=	0.20000000000000000000000…

Representation is unique except for the cases that the decimal expansion ends in all 0's or all 9's.
We will never use the all 9's representation.

Proof that $[0,1)$ is not countable

Suppose, for a contradiction, that there is a list of them:

r_1	**0.50000000…**
r_2	**0.33333333…**
r_3	**0.14285714…**
r_4	**0.14159265…**
r_5	**0.12122122…**
r_6	**0.25000000…**
r_7	**0.71828182…**
r_8	**0.61803394…**
…	**…**

Proof that $[0,1)$ is not countable

Suppose, for a contradiction, that there is a list of them:

		1	2	3	4	5	6	7	8	9	...
r_1	0.	5	0	0	0	0	0	0	0	...	...
r_2	0.	3	3	3	3	3	3	3	3	...	...
r_3	0.	1	4	2	8	5	7	1	4	...	...
r_4	0.	1	4	1	5	9	2	6	5	...	...
r_5	0.	1	2	1	2	2	1	2	2	...	...
r_6	0.	2	5	0	0	0	0	0	0	...	...
r_7	0.	7	1	8	2	8	1	8	2	...	...
r_8	0.	6	1	8	0	3	3	9	4	...	...
...	....	...	....	....	...	...	...	...	...	...	

Proof that $[0,1)$ is not countable

Suppose, for a contradiction, that there is a list of them:

		1	2	3	4	5	6	7	8	9	...
r_1	0.	5	0	0	0	0	0	0	0	...	...
r_2	0.	3	3	3	3	3	3	3	3	...	...
r_3	0.	1	4	2	8	5	7	1	4	...	...
r_4	0.	1	4	1	5	9	2	6	5	...	...
r_5	0.	1	2	1	2	2	1	2	2	...	...
r_6	0.	2	5	0	0	0	0	0	0	...	...
r_7	0.	7	1	8	2	8	1	8	2	...	...
r_8	0.	6	1	8	0	3	3	9	4	...	...
...	....	...	....	....	...	...	...	...	...	...	

Proof that $[0,1)$ is not countable

Suppose, for a contradiction, that there is a list of them:

		1	2	3	4						
r_1	0.	5	0	0	0						
r_2	0.	3	3	3	3	3	3	3	3	...	...
r_3	0.	1	4	2	8	5	7	1	4	...	...
r_4	0.	1	4	1	5	9	2	6	5	...	...
r_5	0.	1	2	1	2	2	1	2	2	...	...
r_6	0.	2	5	0	0	0	0	0	0	...	...
r_7	0.	7	1	8	2	8	1	8	2	...	...
r_8	0.	6	1	8	0	3	3	9	4	...	...
...	....	...	....	....	...	...	...	...	...	...	

Flipping rule:

Only if the other driver deserves it.

Proof that $[0,1)$ is not countable

Suppose, for a contradiction, that there is a list of them:

		1	2	3	4						
r_1	0.	5^{1}	0	0	0						
r_2	0.	3	3^{5}	3	3						
r_3	0.	1	4	2^{5}	8	5	7	1	4	...	...
r_4	0.	1	4	1	5^{1}	9	2	6	5	...	...
r_5	0.	1	2	1	2	2^{5}	1	2	2	...	...
r_6	0.	2	5	0	0	0	0^{5}	0	0	...	...
r_7	0.	7	1	8	2	8	1	8^{5}	2	...	...
r_8	0.	6	1	8	0	3	3	9	4^{5}	...	...
...	....	...	....	....	...	...	...	...	...	...	

Flipping rule:

If digit is 5, make it 1.

If digit is not 5, make it 5.

Proof that $[0,1)$ is not countable

Suppose, for a contradiction, that there is a list of them:

		1	2	3	4						
r_1	0.	5^{1}	0	0	0						
r_2	0.	3	3^{5}	3	3						
r_3	0.	1	4	2^{5}	8	5	7	1	4	…	…
r_4	0.	1	4	1	5^{1}	9	2	6	5	…	…
r_5	0.	1	2	1	2	2^{5}	1	2	2	…	…
r_6	0.	2	5	0	0	0	0^{5}	0	0	…	…
r_7	0.	7	1	8	2	8	1	8^{5}	2	…	…

Flipping rule:

If digit is 5, make it 1.

If digit is not 5, make it 5.

If diagonal element is $\mathbf{0}.x_{11}x_{22}x_{33}x_{44}x_{55}\cdots$ then let's call the flipped number $\mathbf{0}.\hat{x}_{11}\hat{x}_{22}\hat{x}_{33}\hat{x}_{44}\hat{x}_{55}\cdots$

It cannot appear anywhere on the list!

Proof that $[0,1)$ is not countable

Suppose, for a contradiction, that there is a list of them:

Flipping rule:

If digit is 5, make it 1.

If digit is not 5, make it 5.

$$
\begin{array}{cccccccccccc}
 & & 1 & 2 & 3 & 4 & & & & & & \\
r_1 & 0. & 5^{1} & 0 & 0 & 0 & & & & & & \\
r_2 & 0. & 3 & 3^{5} & 3 & 3 & & & & & & \\
r_3 & 0. & 1 & 4 & 2^{5} & 8 & 5 & 7 & 1 & 4 & \ldots & \ldots \\
r_4 & 0. & 1 & 4 & 1 & 5^{1} & 9 & 2 & 6 & 5 & \ldots & \ldots \\
 & & & & & & 2^{5} & 1 & 2 & 2 & \ldots & \ldots \\
 & & & & & & 0 & 0^{5} & 0 & 0 & \ldots & \ldots \\
 & & & & & & 8 & 1 & 8^{5} & 2 & \ldots & \ldots
\end{array}
$$

For every $n \geq 1$:
$r_n \neq 0.\hat{x}_{11}\hat{x}_{22}\hat{x}_{33}\hat{x}_{44}\hat{x}_{55}\cdots$
because the numbers differ on the n-th digit!

If diagonal element is $0.x_{11}x_{22}x_{33}x_{44}x_{55}\cdots$ then let's call the flipped number $0.\hat{x}_{11}\hat{x}_{22}\hat{x}_{33}\hat{x}_{44}\hat{x}_{55}\cdots$

It cannot appear anywhere on the list!

Proof that $[0,1)$ is not countable

Suppose, for a contradiction, that there is a list of them:

$$\begin{array}{cccccccccccc}
 & & 1 & 2 & 3 & 4 & & & & & & \\
r_1 & 0. & 5^{1} & 0 & 0 & 0 & & & & & & \\
r_2 & 0. & 3 & 3^{5} & 3 & 3 & & & & & & \\
r_3 & 0. & 1 & 4 & 2^{5} & 8 & 5 & 7 & 1 & 4 & \ldots & \ldots \\
r_4 & 0. & 1 & 4 & 1 & 5^{1} & 9 & 2 & 6 & 5 & \ldots & \ldots \\
 & & & & & & 2^{5} & 1 & 2 & 2 & \ldots & \ldots \\
 & & & & & & 0 & 0^{5} & 0 & 0 & \ldots & \ldots \\
 & & & & & & 8 & 1 & 8^{5} & 2 & \ldots & \ldots
\end{array}$$

Flipping rule:

If digit is 5, make it 1.

If digit is not 5, make it 5.

For every $n \geq 1$:

$r_n \neq 0.\hat{x}_{11}\hat{x}_{22}\hat{x}_{33}\hat{x}_{44}\hat{x}_{55}\cdots$

because the numbers differ on the n-th digit!

So the list is incomplete, which is a contradiction.

Thus the real numbers between 0 and 1 are **not countable**: “uncountable”

Last time: Countable sets

Definition: A set is **countable** iff it has the same cardinality as some subset of $\mathbb{N}$.

Equivalent: A set $\boldsymbol{S}$ is countable iff there is an *onto* function $\boldsymbol{g} : \mathbb{N} \rightarrow \boldsymbol{S}$

Equivalent: A set $\boldsymbol{S}$ is countable iff we can order the elements $\boldsymbol{S} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \dots\}$

Last time: Countable sets

A set $\boldsymbol{S}$ is **countable** iff we can order the elements of $\boldsymbol{S}$ as

$$\boldsymbol{S} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \dots\}$$

Countable sets:

$\mathbb{N}$ - the natural numbers

$\mathbb{Z}$ - the integers

$\mathbb{Q}$ - the rationals

Σ^* - the strings over any finite Σ

The set of all Java programs

$\left.\right\}$ **Shown by "dovetailing"**

Last time: Not every set is countable

Theorem [Cantor]:
The set of real numbers between 0 and 1 is **not** countable.

Proof using “diagonalization”.

A note on this proof

- The set of rational numbers in [0,1) also have decimal representations like this
  - The only difference is that rational numbers always have repeating decimals in their expansions 0.33333… or .25000000…
- So why wouldn't the same proof show that this set of rational numbers is uncountable?
  - Given any listing we could create the flipped diagonal number d as before
  - However, d would not have a repeating decimal expansion and so wouldn't be a rational #

    It would not be a "missing" number, so no contradiction.

The set of all functions $f : \mathbb{N} \to \{0, \ldots, 9\}$ is uncountable

The set of all functions $f : \mathbb{N} \to \{0, \ldots, 9\}$ is uncountable

Supposed listing of all the functions:

	1	2	3	4	5	6	7	8	9	…
f_1	5	0	0	0	0	0	0	0	…	…
f_2	3	3	3	3	3	3	3	3	…	…
f_3	1	4	2	8	5	7	1	4	…	…
f_4	1	4	1	5	9	2	6	5	…	…
f_5	1	2	1	2	2	1	2	2	…	…
f_6	2	5	0	0	0	0	0	0	…	…
f_7	7	1	8	2	8	1	8	2	…	…
f_8	6	1	8	0	3	3	9	4	…	…
…	…	….	….	…	…	…	…	…	…	

The set of all functions $f : \mathbb{N} \to \{0, \ldots, 9\}$ is uncountable

Supposed listing of all the functions:

Flipping rule:

If $f_n(n) = 5$, set $D(n) = 1$

If $f_n(n) \neq 5$, set $D(n) = 5$

	1	2	3	4						
f_1	5^{1}	0	0	0						
f_2	3	3^{5}	3	3						
f_3	1	4	2^{5}	8	5	7	1	4	...	...
f_4	1	4	1	5^{1}	9	2	6	5	...	...
f_5	1	2	1	2	2^{5}	1	2	2	...	...
f_6	2	5	0	0	0	0^{5}	0	0	...	...
f_7	7	1	8	2	8	1	8^{5}	2	...	...
f_8	6	1	8	0	3	3	9	4^{5}	...	...
...	...	....	....	...	...	...	...	...	...	

The set of all functions $f: \mathbb{N} \to \{0, \dots, 9\}$ is uncountable

Supposed listing of all the functions:

	1	2	3	4						
f_1	5^{1}	0	0	0						
f_2	3	3^{5}	3	3						
f_3	1	4	2^{5}	8	5	7	1	4	...	...
f_4	1	4	1	5^{1}	9	2	6	5	...	...
f_5	1	2	1	2	2^{5}	1	2	2	...	...
f_6	2	5	0	0	0	0^{5}	0	0	...	...
f_7	7	1	8	2	8	1	8^{5}	2	...	...

Flipping rule:

If $f_n(n) = 5$, set $D(n) = 1$

If $f_n(n) \neq 5$, set $D(n) = 5$

For all n, we have $D(n) \neq f_n(n)$. Therefore $D \neq f_n$ for any n and the list is incomplete! $\Rightarrow$ $\{f \mid f: \mathbb{N} \to \{0,1, \dots, 9\}\}$ is **not** countable

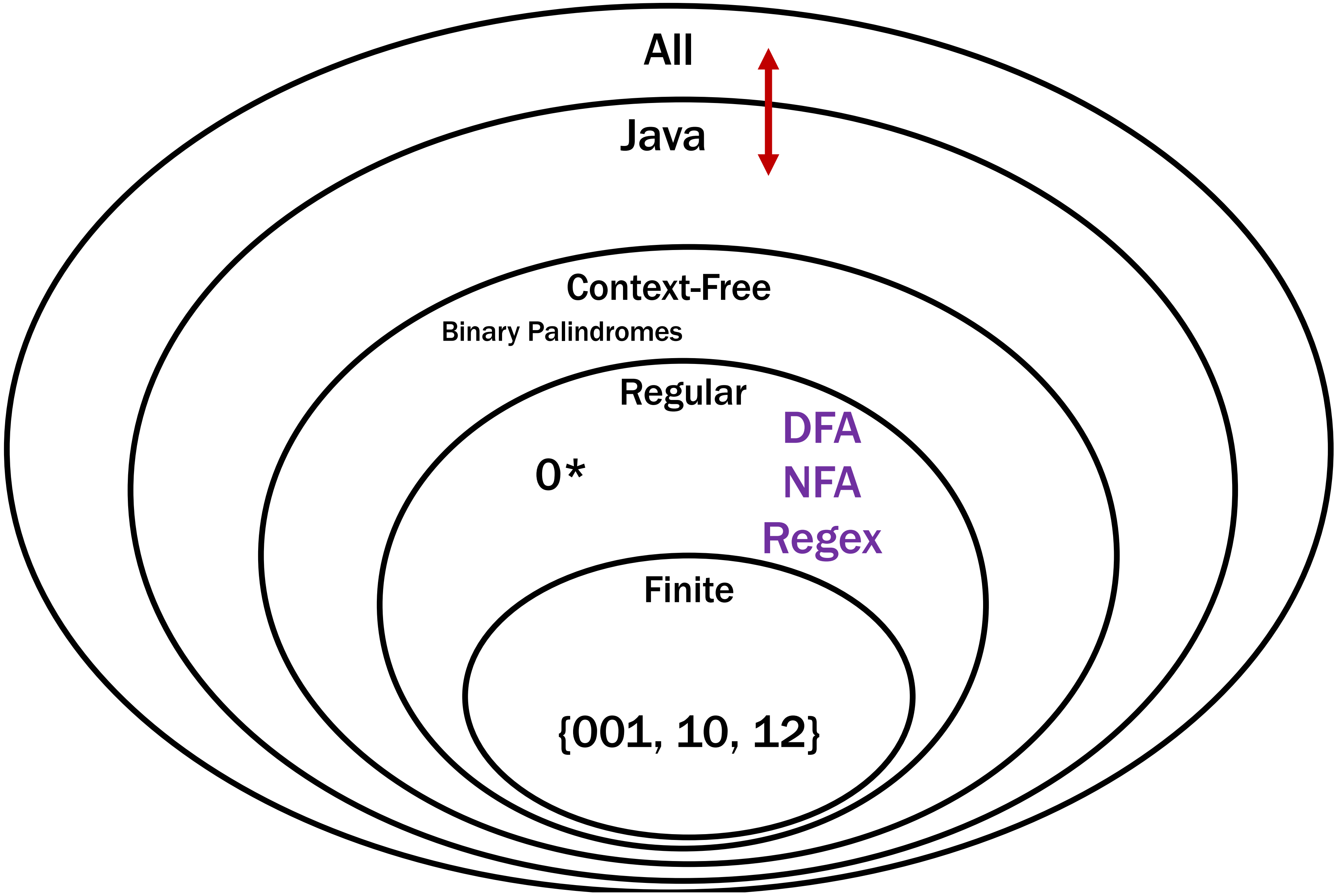
Uncomputable functions

We have seen that:

- The set of all (Java) programs is countable
- The set of all functions $f : \mathbb{N} \to \{0, \ldots, 9\}$ is not countable

So: There must be some function $f : \mathbb{N} \to \{0, \ldots, 9\}$ that is not computable by any program!

Recall our language picture



Uncomputable functions

Interesting… maybe.

Can we come up with an explicit function that is uncomputable?

A “Simple” Program

```
public static void collatz(n) {
    if (n == 1) {
        return 1;
    }
    if (n % 2 == 0) {
        return collatz(n/2)
    }
    else {
        return collatz(3*n + 1)
    }
}
```

What does this program do?

… on n=11?

… on n=10000000000000000001?

11
34
17
52
26
13
40
20
10
5
16
8
4
2
1

A “Simple” Program

```
public static void collatz(n) {
    if (n == 1) {
        return 1;
    }
    if (n % 2 == 0) {
        return collatz(n/2)
    }
    else {
        return collatz(3*n + 1)
    }
}
```

Nobody knows whether or not this program halts on all inputs!

What does this program do?

… `on n=11?`

… `on n=100000000000000000001?`

Some Notation

We're going to be talking about *Java code*.

CODE(**P**) will mean "the code of the program **P**"

So, consider the following function:

```
public String P(String x) {
    return new String(Arrays.sort(x.toCharArray());
}
```

What is **P**(CODE(**P**))?

"(((())))..;AACPSSaaabceeggghiiiilnnnnnooprrrrrrrrrrrsssttttttuuwxxyy{}"

The Halting Problem

`CODE(P)` means “the code of the program **P**”

The Halting Problem

Given: - CODE(**P**) for any program **P**
- input **x**

Output: **true** if **P** halts on input **x**
false if **P** does not halt on input **x**

Undecidability of the Halting Problem

CODE(**P**) means "the code of the program **P**"

The Halting Problem

Given: - CODE(**P**) for any program **P**
- input **x**

Output: **true** if **P** halts on input **x**
false if **P** does not halt on input **x**

Theorem [Turing]: **There is no program that solves the Halting Problem**

Proof by contradiction

Suppose that H is a Java program that solves the Halting problem.

Proof by contradiction

Suppose that H is a Java program that solves the Halting problem.

Then we can write this program:

```
public static void D(String s) {
    if (H(s,s)) {
        while (true);  // don’t halt
    } else {
        return;        // halt
    }
}

public static bool H(String s, String x) { ... }
```

Does D(CODE(D)) halt?

Does **D**(CODE(**D**)) halt?

```
public static void D(s) {
    if (H(s,s)) {
        while (true);   // don’t halt
    } else {
        return;         // halt
    }
}
```

Does **D**(CODE(**D**)) halt?

```
public static void D(s) {
    if (H(s,s)) {
        while (true);   // don't halt
    } else {
        return;         // halt
    }
}
```

H solves the halting problem implies that

H(CODE(**D**),s) is **true** iff **D**(s) halts, **H**(CODE(**D**),s) is **false** iff not

Does **D**(CODE(**D**)) halt?

```
public static void D(s) {
    if (H(s,s)) {
        while (true);   // don’t halt
    } else {
        return;         // halt
    }
}
```

H solves the halting problem implies that

H(CODE(**D**),s) is **true** iff **D**(s) halts, **H**(CODE(**D**),s) is **false** iff not

Suppose that **D**(CODE(**D**)) **halts**.

Then, by definition of **H** it must be that

H(CODE(**D**), CODE(**D**)) is **true**

Which by the definition of **D** means **D**(CODE(**D**)) **doesn’t halt**

Does **D**(CODE(**D**)) halt?

```
public static void D(s) {
    if (H(s,s)) {
        while (true);   // don’t halt
    } else {
        return;         // halt
    }
}
```

H solves the halting problem implies that

H(CODE(**D**),s) is **true** iff **D**(s) halts, **H**(CODE(**D**),s) is **false** iff not

Suppose that **D**(CODE(**D**)) **halts**.

Then, by definition of **H** it must be that

H(CODE(**D**), CODE(**D**)) is **true**

Which by the definition of **D** means **D**(CODE(**D**)) **doesn’t halt**

Suppose that **D**(CODE(**D**)) **doesn’t halt**.

Then, by definition of **H** it must be that

H(CODE(**D**), CODE(**D**)) is **false**

Which by the definition of **D** means **D**(CODE(**D**)) **halts**

Does **D**(CODE(**D**)) halt?

```
public static void D(s) {
    if (H(s,s)) {
        while (true);   // don’t halt
    } else {
        return;         // halt
    }
}
```

H solves the halting problem implies that
    H(CODE(**D**),s) is **true** iff **D**(s) halts, **H**(CODE(**D**) ot

Suppose that **D**(CODE(**D**)) **halts**.
    Then, by definition of **H** it mus
        H(CODE(**D**), CO
    Which by the defi s **D**(CODE(**D**)) **doesn’t halt**

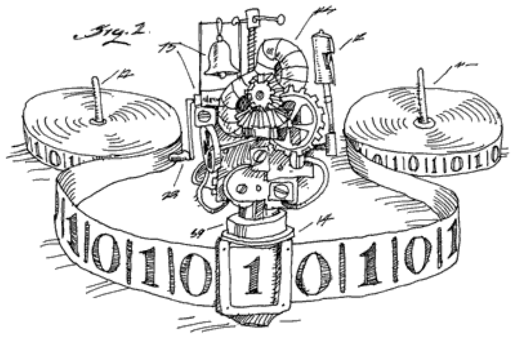
Suppose th **doesn’t halt**.
    Th of **H** it must be that
        ODE(**D**), CODE(**D**)) is **false**
    Whi by the definition of **D** means **D**(CODE(**D**)) **halts**

The ONLY assumption was that the program H exists so that assumption must have been false.

Contradiction!

Done

- We proved that there is no computer program that can solve the Halting Problem.
  - There was nothing special about Java*
    [Church-Turing thesis]



- This tells us that there is no compiler that can check our programs and guarantee to find any infinite loops they might have.

Terminology

- **With Java programs / general computation, we say that the computer "decides" the language L iff**
  - **it halts with output 1 on input $x \in \Sigma^*$ if $x \in L$**
  - **it halts with output 0 on input $x \in \Sigma^*$ if $x \notin L$**

    (difference is the possibility that machine doesn't halt)

- **If no machine decides L, then L is "undecidable"**

Where did the idea for creating D come from?

```
public static void D(s) {
    if (H(s,s) == true) {
        while (true);  // don’t halt
    } else {
        return;        // halt
    }
}
```

D halts on input code(P) iff H(code(P),code(P)) outputs false
iff P doesn’t halt on input code(P)

Connection to diagonalization

Write <**P**> **for** CODE(**P**)

Some possible inputs **x**

$$\begin{array}{c|cccccc} & \langle P_1\rangle & \langle P_2\rangle & \langle P_3\rangle & \langle P_4\rangle & \langle P_5\rangle & \langle P_6\rangle\ \ldots. \\ \hline P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \\ P_6 \\ P_7 \\ P_8 \\ P_9 \\ . \\ . \end{array}$$

All programs **P**

This listing of all programs really does exist since the set of all Java programs is countable

The goal of this “diagonal” argument is not to show that the listing is incomplete but rather to show that a “flipped” diagonal element is not in the listing

Connection to diagonalization

Write <P> for CODE(P)

Some possible inputs **x**

All programs **P**

	$<P_1>$	$<P_2>$	$<P_3>$	$<P_4>$	$<P_5>$	$<P_6>$	....					
P_1	0	1	1	0	1	1	1	0	0	0	1	...
P_2	1	1	0	1	0	1	1	0	1	1	1	...
P_3	1	0	1	0	0	0	0	0	0	0	1	...
P_4	0	1	1	0	1	0	1	1	0	1	0	...
P_5	0	1	1	1	1	1	1	0	0	0	1	...
P_6	1	1	0	0	0	1	1	0	1	1	1	...
P_7	1	0	1	1	0	0	0	0	0	0	1	...
P_8	0	1	1	1	1	0	1	1	0	1	0	...
P_9	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.												

(P,x) entry is **1** if program **P** halts on input **x**
and **0** if it runs forever

Connection to diagonalization

Write <P> for CODE(P)

Some possible inputs **x**

All programs **P**

	$\langle P_1\rangle$	$\langle P_2\rangle$	$\langle P_3\rangle$	$\langle P_4\rangle$	$\langle P_5\rangle$	$\langle P_6\rangle$	....					
P_1	0^{1}	1	1	0	1							
P_2	1	1^{0}	0	1	0							
P_3	1	0	1^{0}	0	0							
P_4	0	1	1	0^{1}	1	0	1	1	0	1	0	...
P_5	0	1	1	1	1^{0}	1	1	0	0	0	1	...
P_6	1	1	0	0	0	1^{0}	1	0	1	1	1	...
P_7	1	0	1	1	0	0	0^{1}	0	0	0	1	...
P_8	0	1	1	1	1	0	1	1^{0}	0	1	0	...
P_9	.	.	. . .	.	.	.	.	.	.	.	.	
.	.	.	. . .	.	.	.	.	.	.	.	.	
.												

Want behavior of program ***D*** to be like the flipped diagonal, so it can't be in the list of all programs.

(P,x) entry is **1** if program **P** halts on input **x** and **0** if it runs forever

Where did the idea for creating D come from?

```
public static void D(s) {
    if (H(s,s) == true) {
        while (true); /* don't halt */
    }
    else {
        return;       /*   halt   */
    }
}
```

D halts on input code(P) iff H(code(P),code(P)) outputs false
iff P doesn't halt on input code(P)

Therefore, for any program P, D differs from P on input code(P)

Recall: Undecidability of the Halting Problem

CODE(**P**) means “the code of the program **P**”

The Halting Problem

Given: - CODE(**P**) for any program **P**
- input **x**

Output: **true** if **P** halts on input **x**
false if **P** does not halt on input **x**

Theorem [Turing]: **There is no program that solves the Halting Problem**

Recall: Connection to diagonalization

Write <P> for CODE(P)

Some possible inputs **x**

All programs **P**

	$<P_1>$	$<P_2>$	$<P_3>$	$<P_4>$	$<P_5>$	$<P_6>$	....					
P_1	0^{1}	1	1	0	1							
P_2	1	1^{0}	0	1	0							
P_3	1	0	1^{0}	0	0							
P_4	0	1	1	0^{1}	1	0	1	1	0	1	0	...
P_5	0	1	1	1	1^{0}	1	1	0	0	0	1	...
P_6	1	1	0	0	0	1^{0}	1	0	1	1	1	...
P_7	1	0	1	1	0	0	0^{1}	0	0	0	1	...
P_8	0	1	1	1	1	0	1	1^{0}	0	1	0	...
P_9	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.												

Want behavior of program ***D*** to be like the flipped diagonal, so it can't be in the list of all programs.

(P,x) entry is **1** if program **P** halts on input **x** and **0** if it runs forever

Recall: Where did the idea for creating D come from?

```
public static void D(s) {
    if (H(s,s) == true) {
        while (true); /* don’t halt */
    }
    else {
        return;         /*    halt    */
    }
}
```

Shows that, if ***H*** is computable, then ***D*** is computable, so it **should be** in the list of programs!

D halts on input code(P) iff H(code(P),code(P)) outputs false
iff P doesn’t halt on input code(P)

Therefore, for any program P, D differs from P on input code(P)

More Undecidable Problems…

We know can answer almost any question about REs

- Do two RegExps recognize the same language?

But many problems about CFGs are undecidable!

- **Do two CFGs generate the same language?**
- **Is there any string that two CFGs both generate?**
  - **more general: "CFG intersection" problem**
- **Does a CFG generate *every* string?**

Proving Problems Undecidable

- Can use the fact that the Halting Problem is undecidable to show that other problems are undecidable

General method (a “reduction”):

Prove that, if there were a program deciding B, then there would be a program deciding the Halting Problem.

“B decidable $\rightarrow$ Halting Problem decidable”

Contrapositive:

“Halting Problem undecidable $\rightarrow$ B undecidable”

Therefore, B is undecidable

A CSE 142 assignment

Students should write a Java program that:

- Prints "Hello" to the console
- Eventually exits

GradeIt, PracticeIt, etc. need to grade these

How do we write that grading program?

WE CAN'T: THIS IS IMPOSSIBLE!

Another undecidable problem

- **CSE 142 Grading problem:**
  - **Input: CODE(Q)**
  - **Output:**

    True if **Q** outputs "HELLO" and exits

    False if **Q** does not do that

- **Theorem: The CSE 142 Grading is undecidable.**
- **Proof idea: Show that, if there is a program T to decide CSE 142 grading, then there is a program H to decide the Halting Problem for code(P) and input x.**

Another undecidable problem

Theorem: The CSE 142 Grading is undecidable.

Proof: Suppose there is a program T that decide CSE 142 grading problem. Then, there is a program H to decide the Halting Problem for code(P) and input x by

- transform P (with input x) into the following program Q…

Another undecidable problem

```
public class Q {
  private static String x = "...";

  public static void main(String[] args) {
    PrintStream out = System.out;
    System.setOut(new PrintStream(
        new WriterOutputStream(new StringWriter()));
    System.setIn(new ReaderInputStream(new StringReader(x)));

    P.main(args);

    out.println("HELLO");
  }
}

class P {
  public static void main(String[] args) { ... }
  ...
}
```

Another undecidable problem

Theorem: The CSE 142 Grading is undecidable.

Proof: Suppose there is a program T that decide CSE 142 grading problem. Then, there is a program H to decide the Halting Problem for code(P) and input x by

- transform P (with input x) into the following program Q
- run T on code(Q)
  - if it returns true, then P halted

    must halt in order to print "HELLO"
  - if it returns false, then P did not halt

    program Q can't output anything other than "HELLO"

More Reductions

- Can use undecidability of these problems to show that other problems are undecidable.

- For instance:

$\text{EQUIV}(P, Q)$:	**True**	if $P(x)$ and $Q(x)$ have the same behavior for every input x
	False	otherwise

Rice’s theorem

Not *every* problem on programs is undecidable!

Which of these is decidable?

- Input `CODE(P)` and x
  Output: true if P prints “ERROR” on input x after less than 100 steps
  false otherwise

- Input `CODE(P)` and x
  Output: true if P prints “ERROR” on input x after more than 100 steps
  false otherwise

Rice’s Theorem:

Any “non-trivial” property of the **input-output behavior** of Java programs is undecidable.

Rice’s theorem

Not *every* problem on programs is undecidable!

Which of these is decidable?

- Input `CODE(P)` and x
  Output: true if P prints “ERROR” on input x after less than 100 steps
  false otherwise

- Input `CODE(P)` and x
  Output: true if P prints “ERROR” on input x after more than 100 steps
  false otherwise

Rice’s Theorem (a.k.a. Compilers ARE DIFFICULT):
Any “non-trivial” property of the **input-output behavior** of Java programs is undecidable.

Proof of Rice's Theorem

Rice's Theorem:

Any "non-trivial" property of the **input-output behavior** of Java programs is undecidable.

- "Non-trivial" means there is at least one program with the property and at least one program without it.
  - if all programs have (or don't have) the property, it's easy
  - suppose that program A has the property and B does not

- Consider the infinite loop program L:

```
public static void main(String[] args) {
  while (true);
}
```

Proof of Rice's Theorem

- Suppose that program A has the property and B does not
- Suppose that L doesn't have the property
  - so A as the property and L doesn't
  - (other case is similar)
- Suppose, for a contradiction, that T checks for the property
- Given a program P, construct this program Q...

Proof of Rice's Theorem

```
public class Q {
  private static String x = “...”;

  public static void main(String[] args) {
    PrintStream out = System.out;
    System.setOut(new PrintStream(
        new WriterOutputStream(new StringWriter()));
    System.setIn(new ReaderInputStream(new StringReader(x)));

    P.main(args);  // If P doesn't halt, then it acts like L
    A.main(args);  // If P does halt, then it acts like A
  }
}

class P {
  public static void main(String[] args) { ... }
  ...
}
```

Proof of Rice's Theorem

- Suppose that program A has the property and B does not
- Suppose that L has the property (other case is similar).
  - so L has the property and B does not
- Suppose, for a contradiction, that T checks for the property
- Construct this program Q that
  - acts like A (has the property) if P halts on x
  - acts like L (doesn't have property) if P doesn't halt on x
- Ask T if Q has the property
  - if it does, then P halts on x
  - if it doesn't, then P doesn't halt on x

Recall: Computers from Thought

Computers as we know them grew out of a desire to avoid bugs in mathematical reasoning.

Hilbert in a famous speech at the International Congress of Mathematicians in 1900 set out the goal to mechanize all of mathematics.

In the 1930s, Gödel and Turing showed that Hilbert’s program is impossible!

Full employment for mathematicians and computer scientists!!

Takeaways From Undecidability

- Reasoning About / Understanding Math (311) cannot be fully mechanized
  - no program can reliably tell you if your theorem is true

- Reasoning About / Understanding Code (331) cannot be fully mechanized
  - no program can reliably tell you if your code is correct

- "Proving" and "Programming" are analogous
  - will require human brains forever
  - in fact, they are "the same" activity in a deep sense

    see HW1 & HW3 extra credit (Curry-Howard isomorphism)

Questions?

UW CSE's Steam-Powered Turing Machine

Original in Sieg Hall stairwell