YSC2229: Introductory Data Structures and Algorithms

Ilya Sergey

ilya.sergey@yale-nus.edu.sg

Some Terminology

- *Data* represents *information*
- *Computations* represent *data processing*
- An *algorithm* is a sequence of computational steps that transform the *input* data (given) into the *output* data (wanted).
- A *data structure* is a *representation* of *data* that makes it suitable for algorithmic treatment.

What this course is about?

Algorithms in a Nutshell

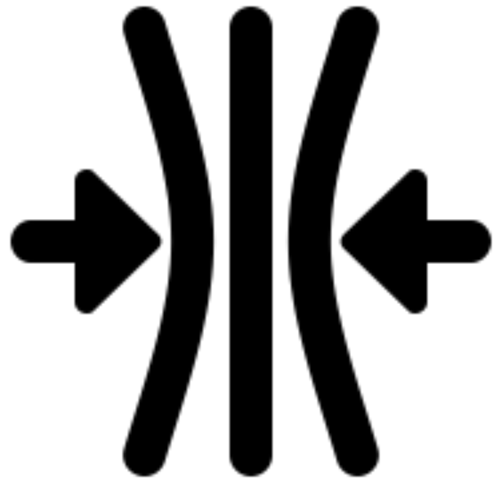
Desired Guarantee:
for every input, the algorithm must provide the *correct* output *quickly*.

Solving computational problems

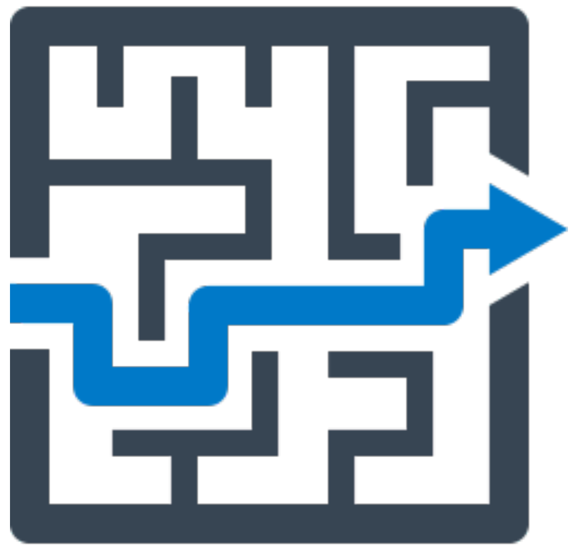
- **Searching:**
  finding a word in a text or an article to buy on Amazon

- **Storing and retrieving data:**
  representing files in you computer

- **Data compression/decompression:**
  transferring files on the internet

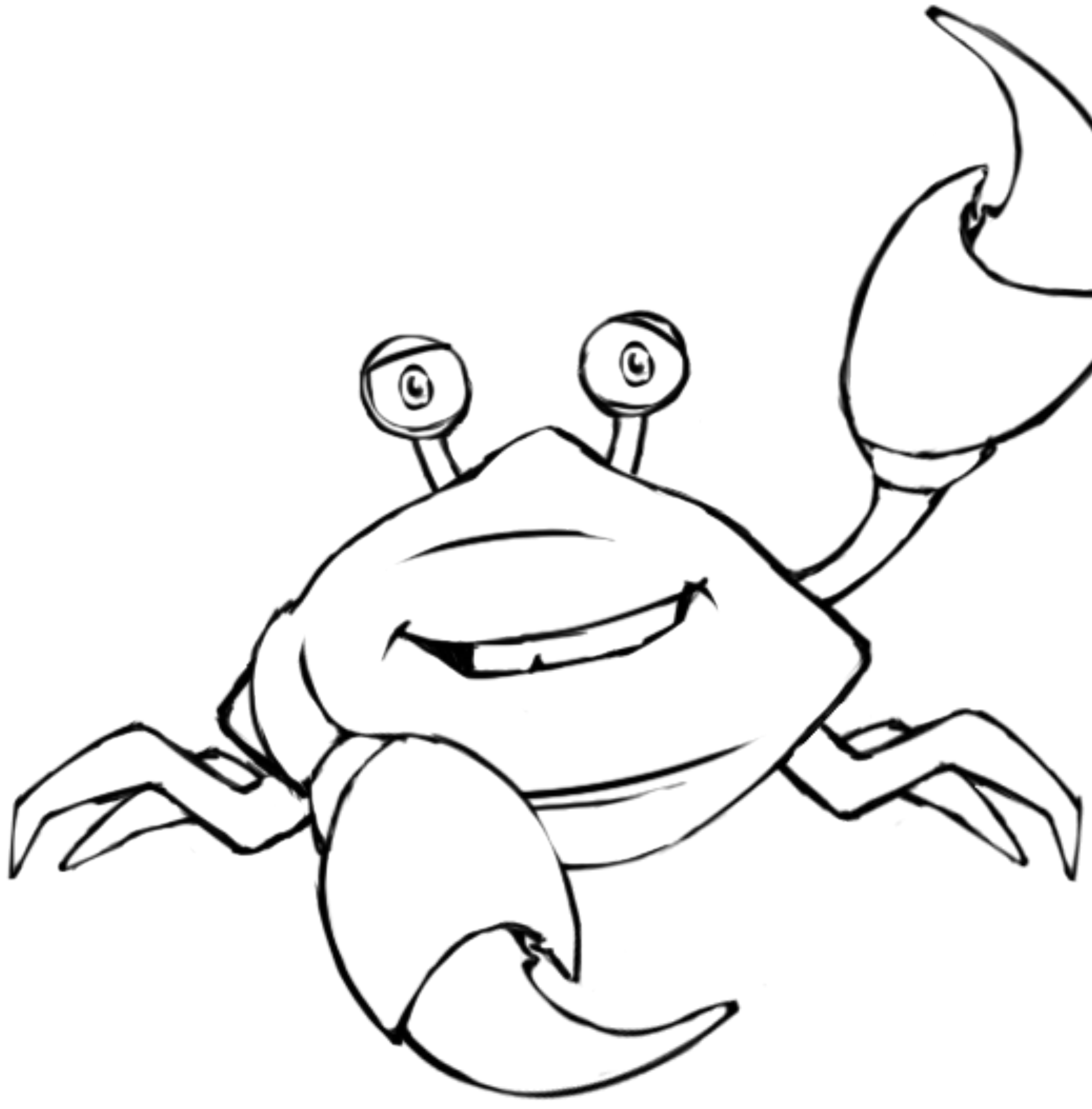


- **Path finding:**
  getting from a point A to point B in the most efficient way



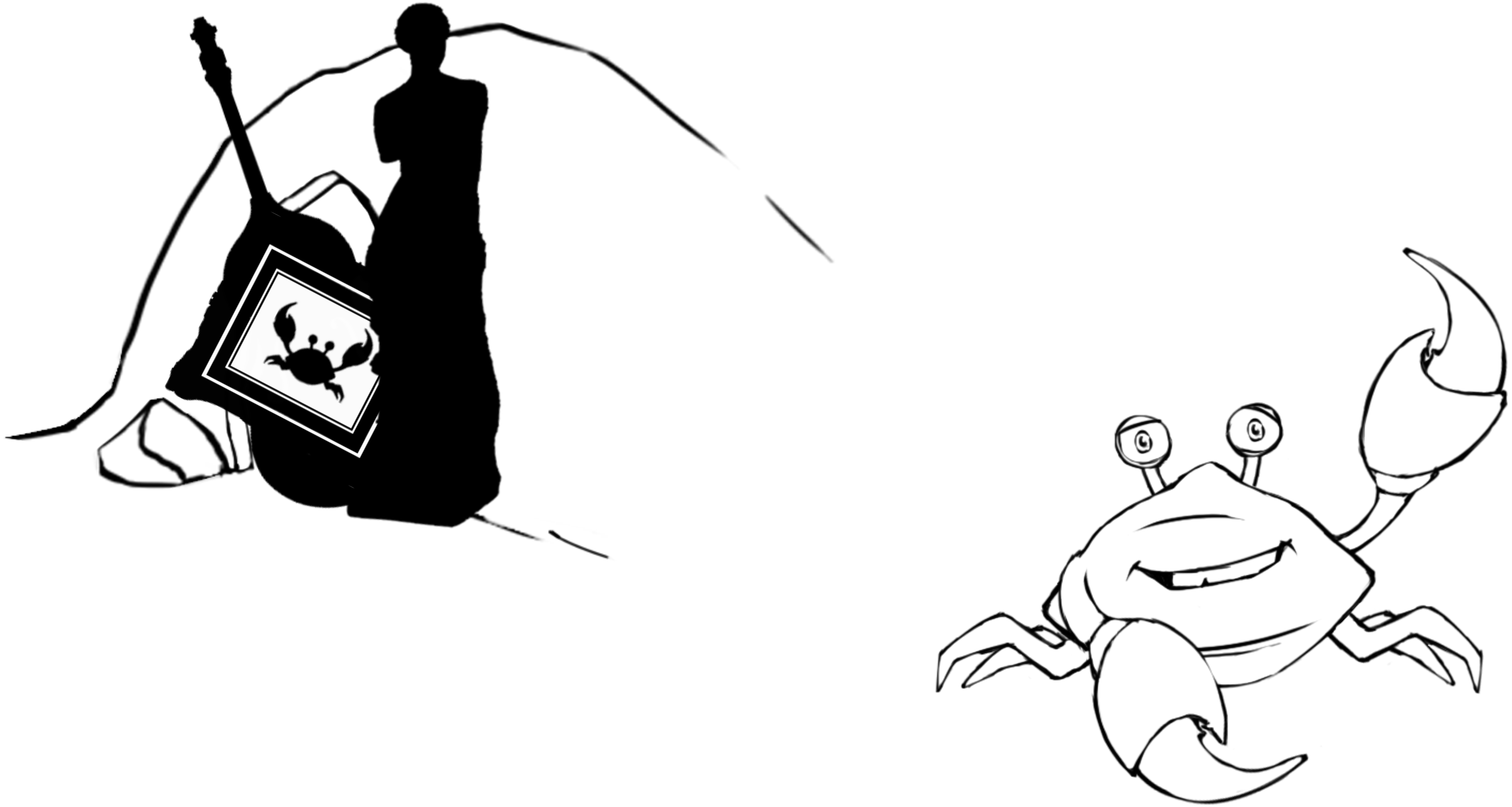
- **Geometric problems**:
  finding the closes fuel station, shape intersection

Thinking Algorithmically is Fun

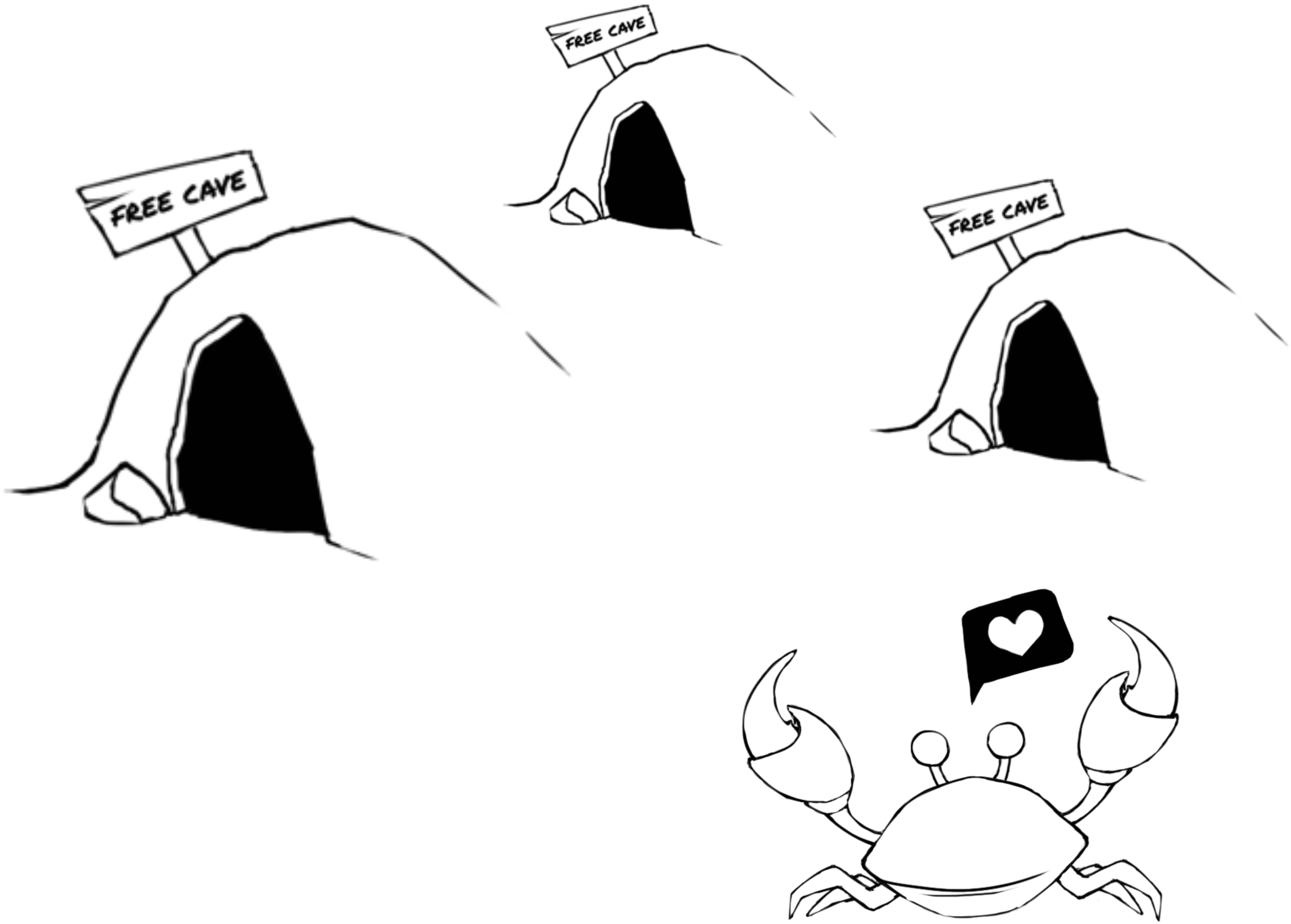


Torpe

(the prosperous crab)



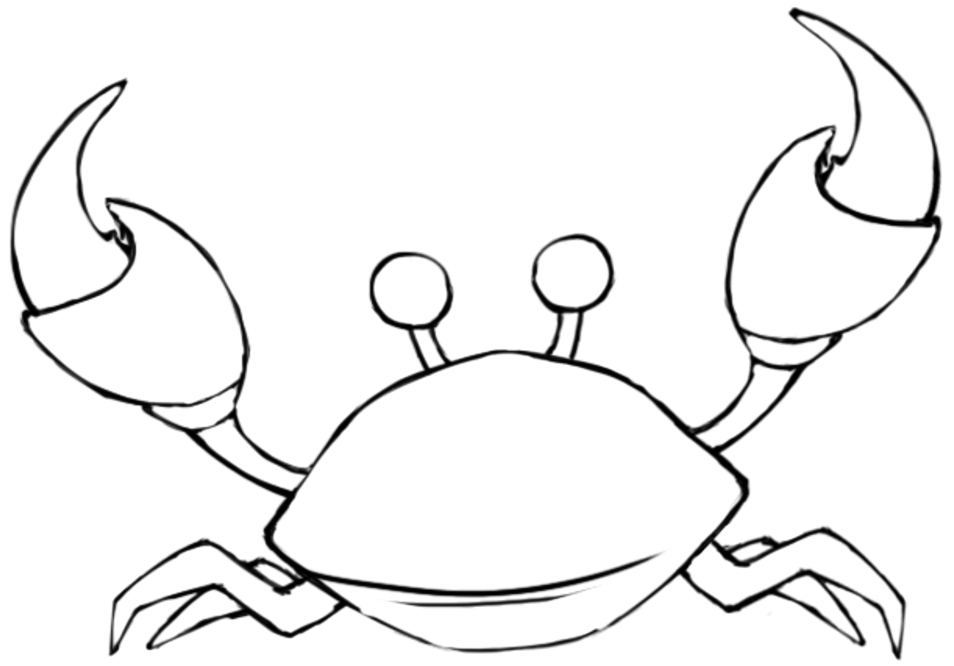

FREE CAVE

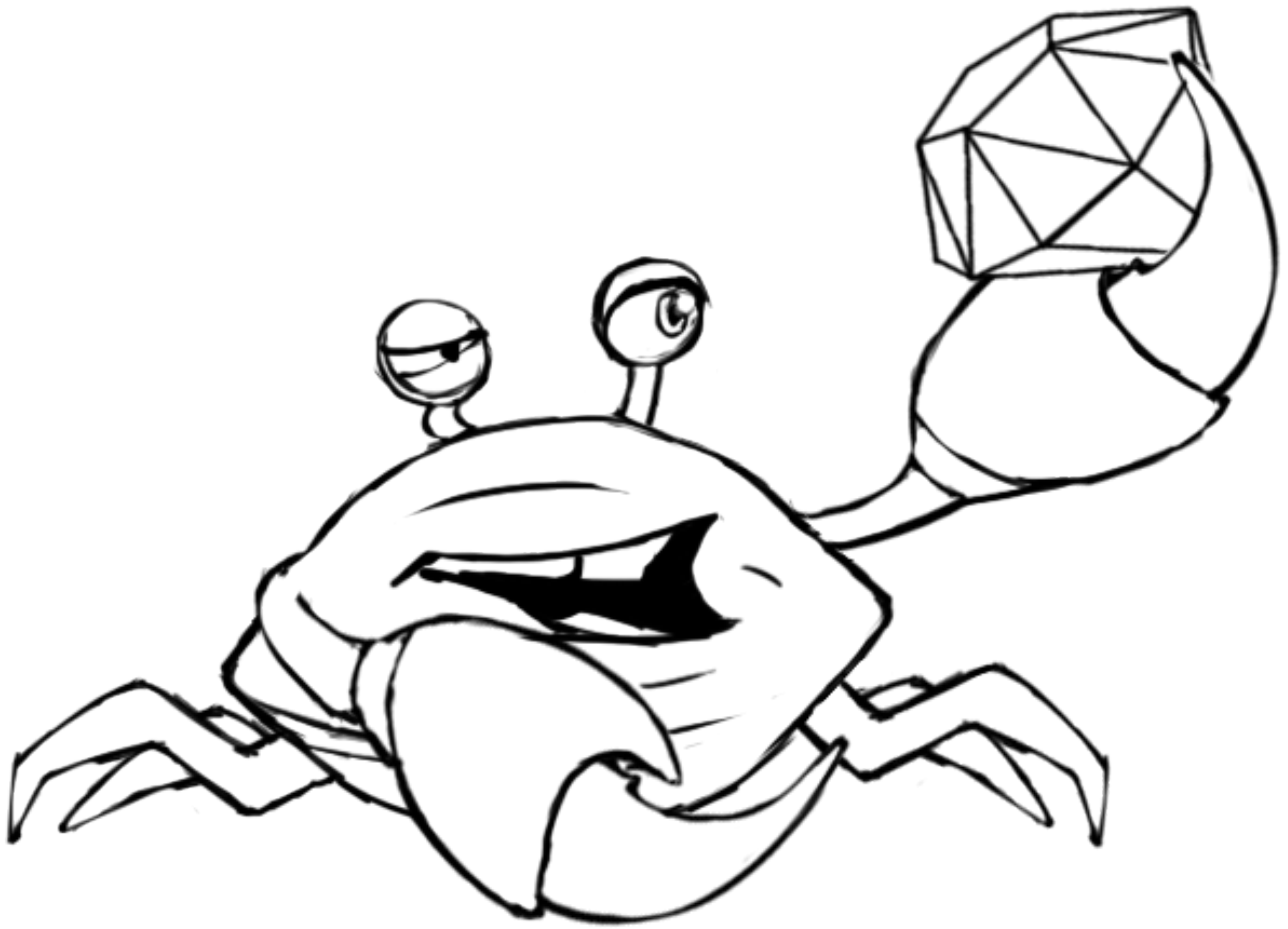

FREE CAVE
FREE CAVE
FREE CAVE


FREE CAVE

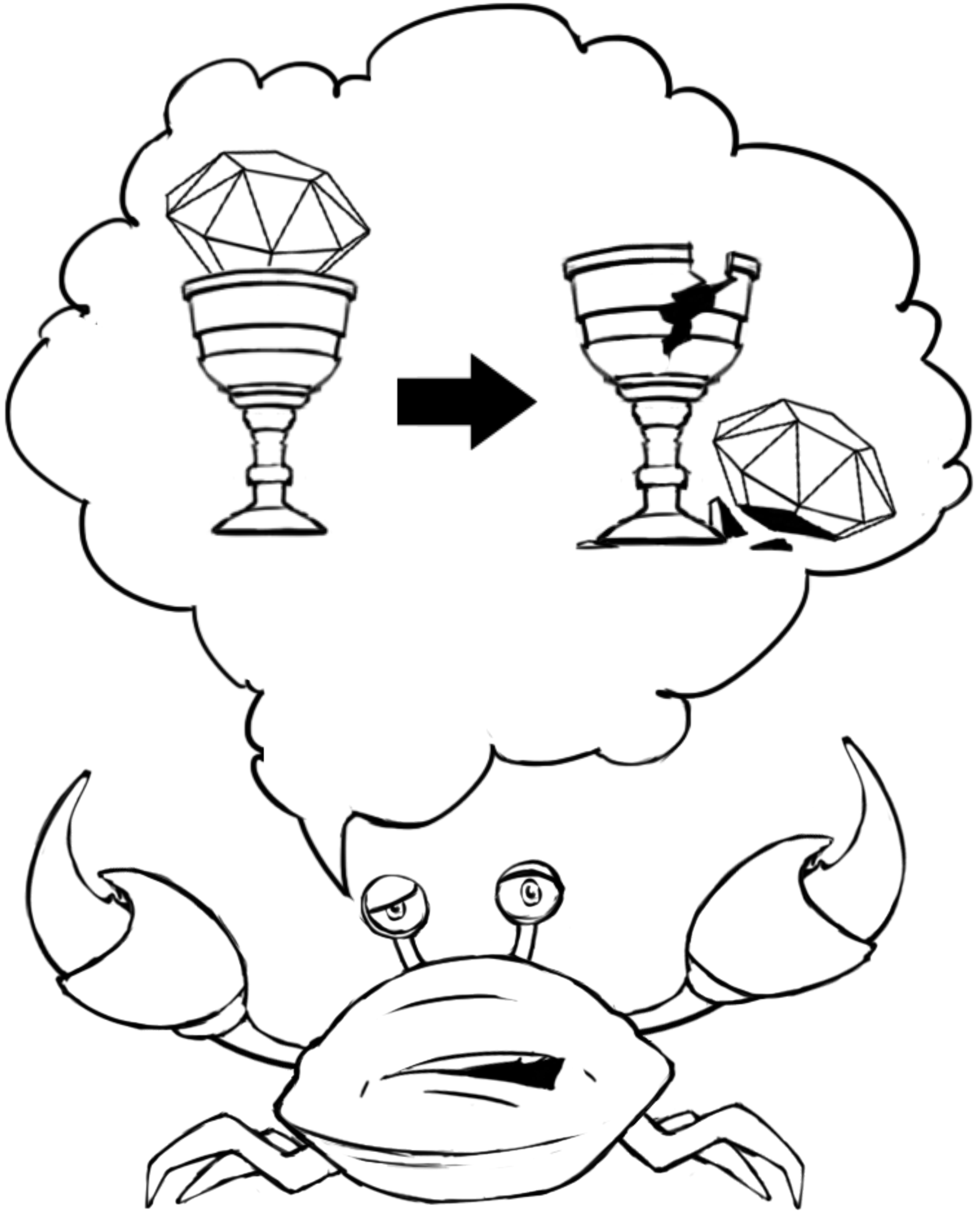

FREE CAVE

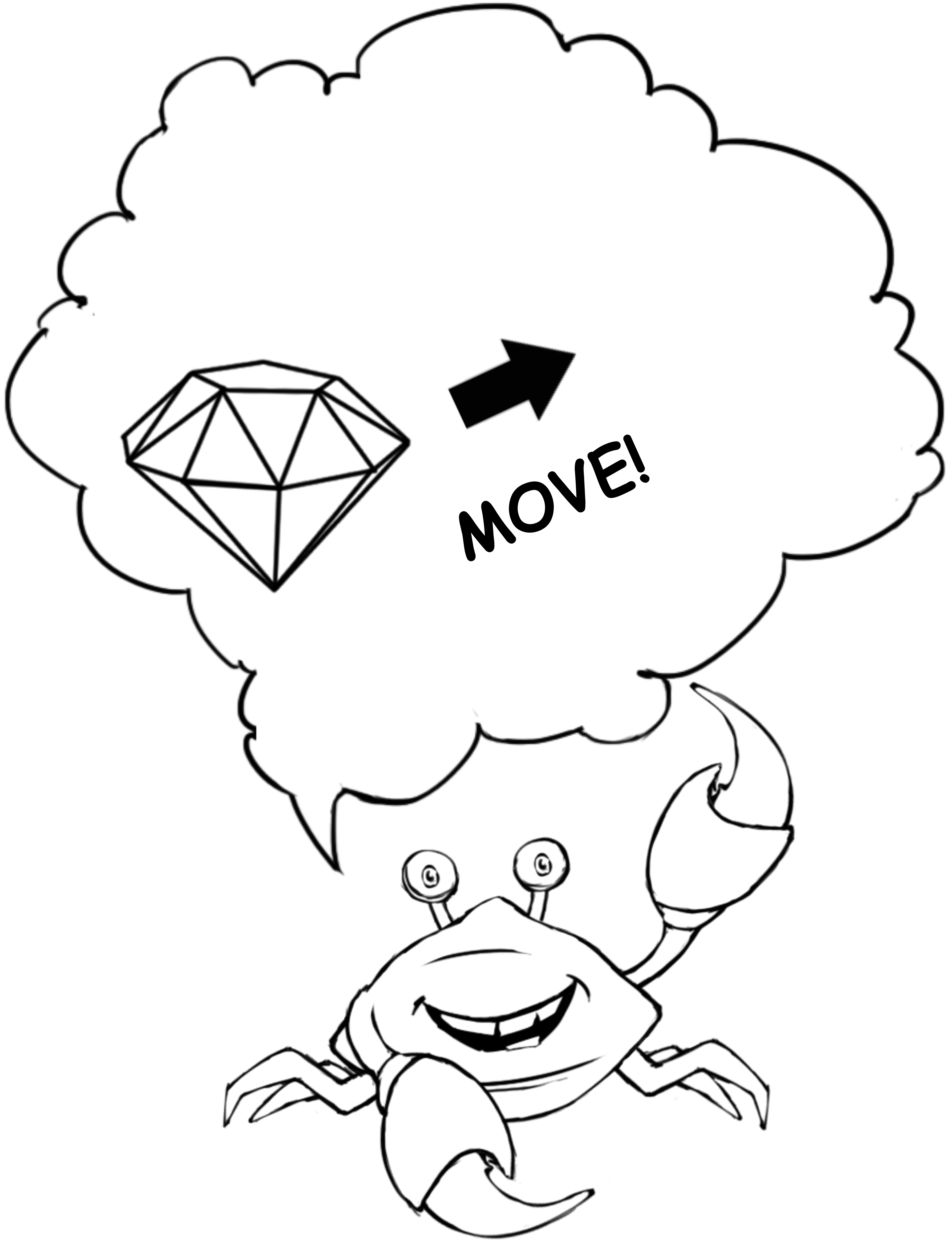

FREE CAVE

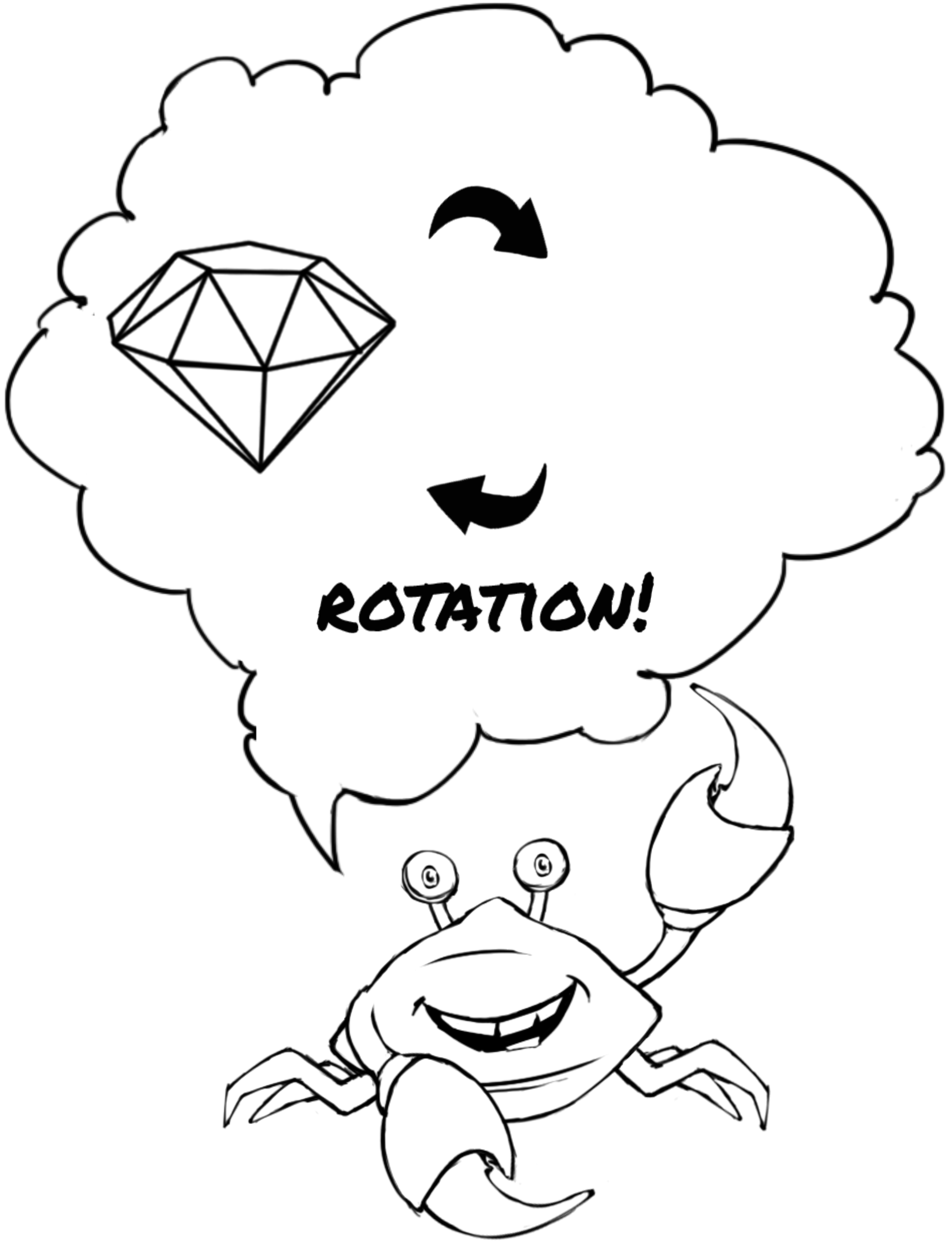


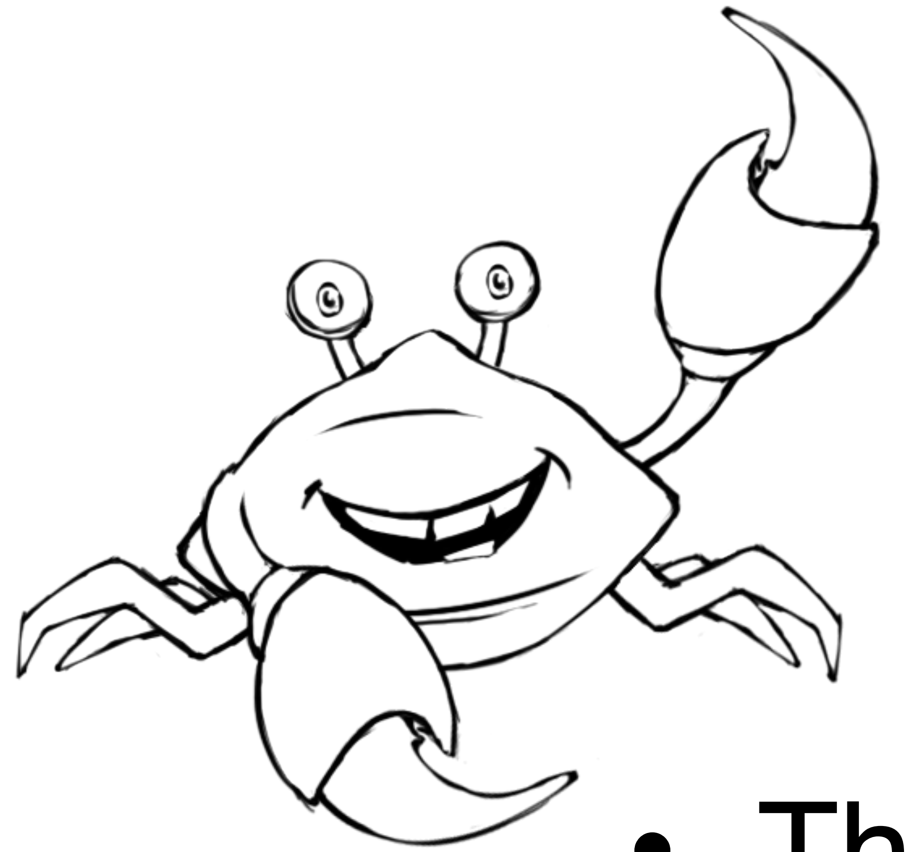



MAX $$$




MOVE!


ROTATION!

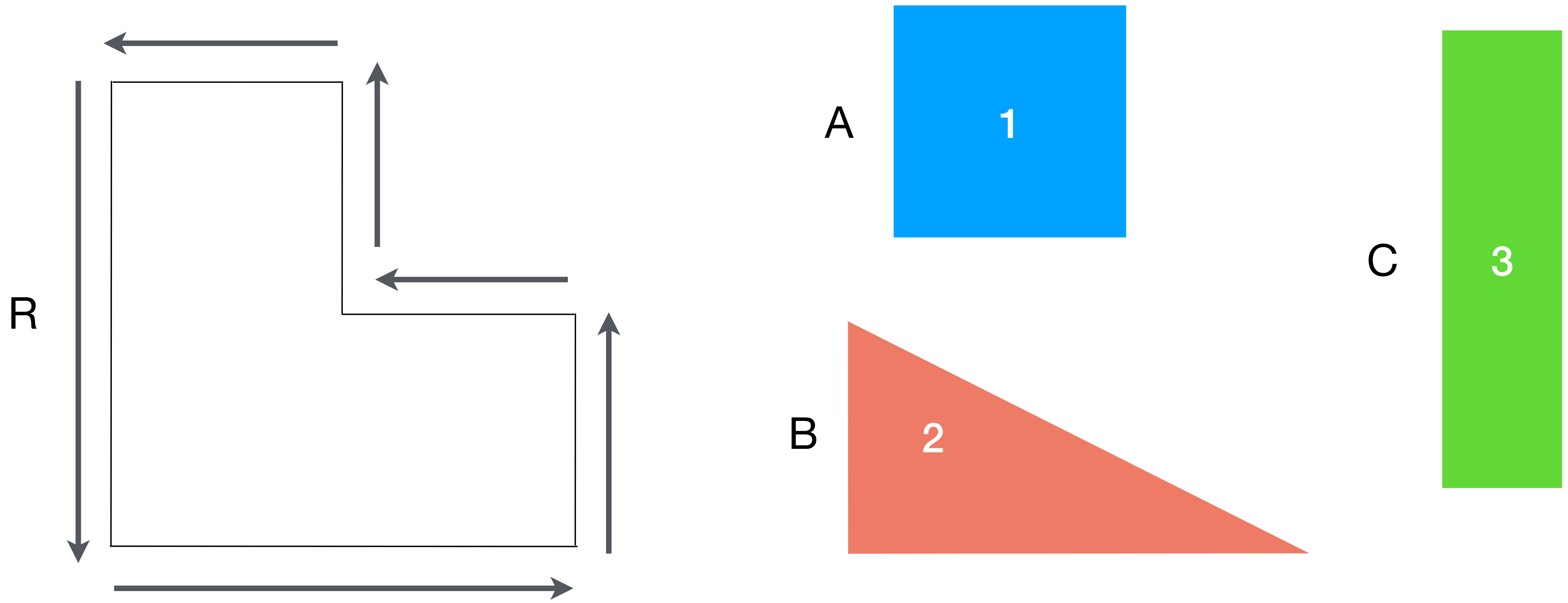


Rules of the Game

- **The problem**: pack Torpe's belongings into a cave (2D)
- **Requirements**:
  - No overlapping, all within the room, at least 30% covered
  - Try to find the best (maximal cost)
- **Available actions**:
  - Moving the furniture
  - Rotating the furniture

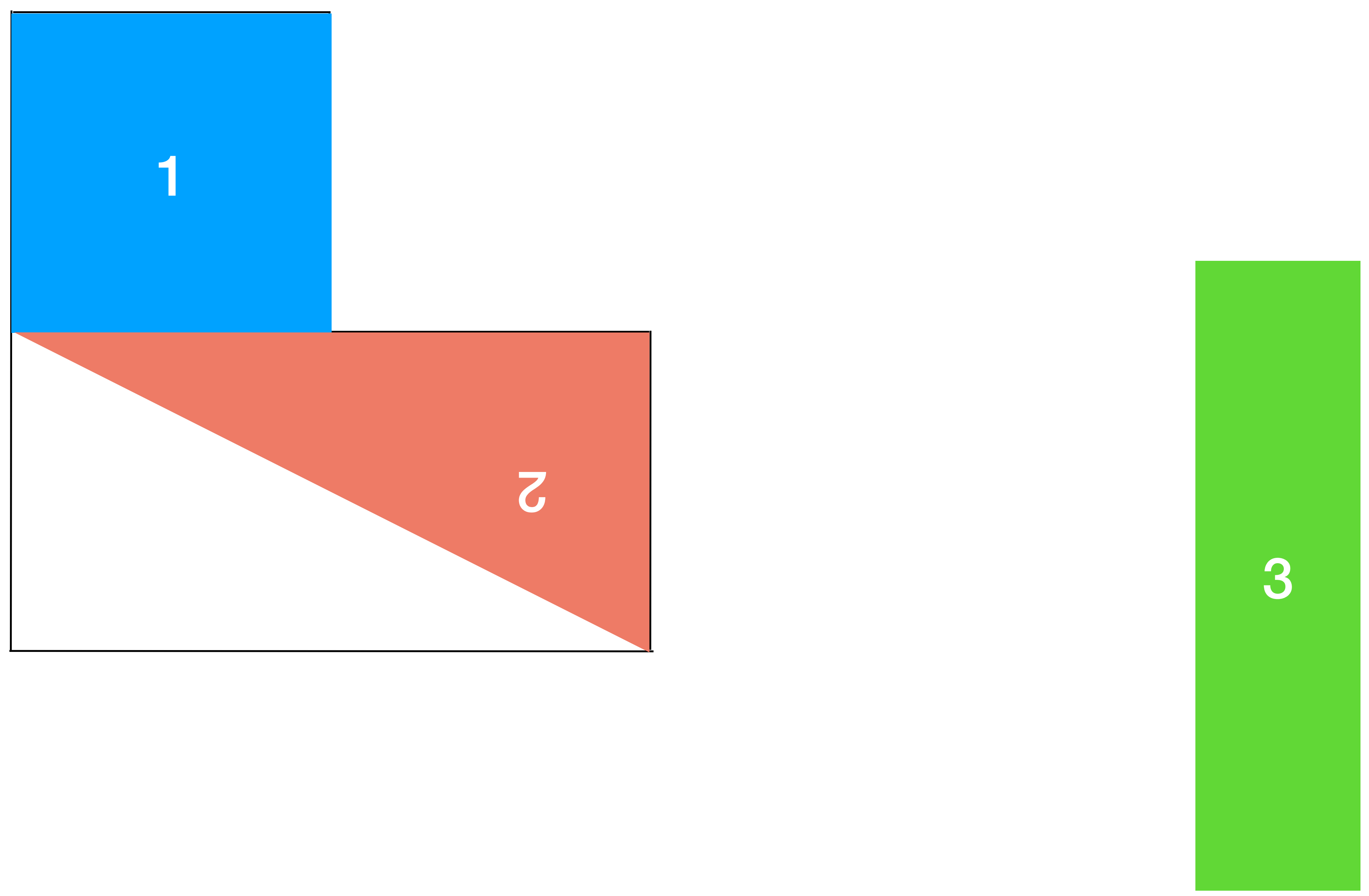
Q1: What Data Structures should We Use?

- Representing the cave
- Representing the furniture items
- Encoding the item costs
- Encoding the solutions



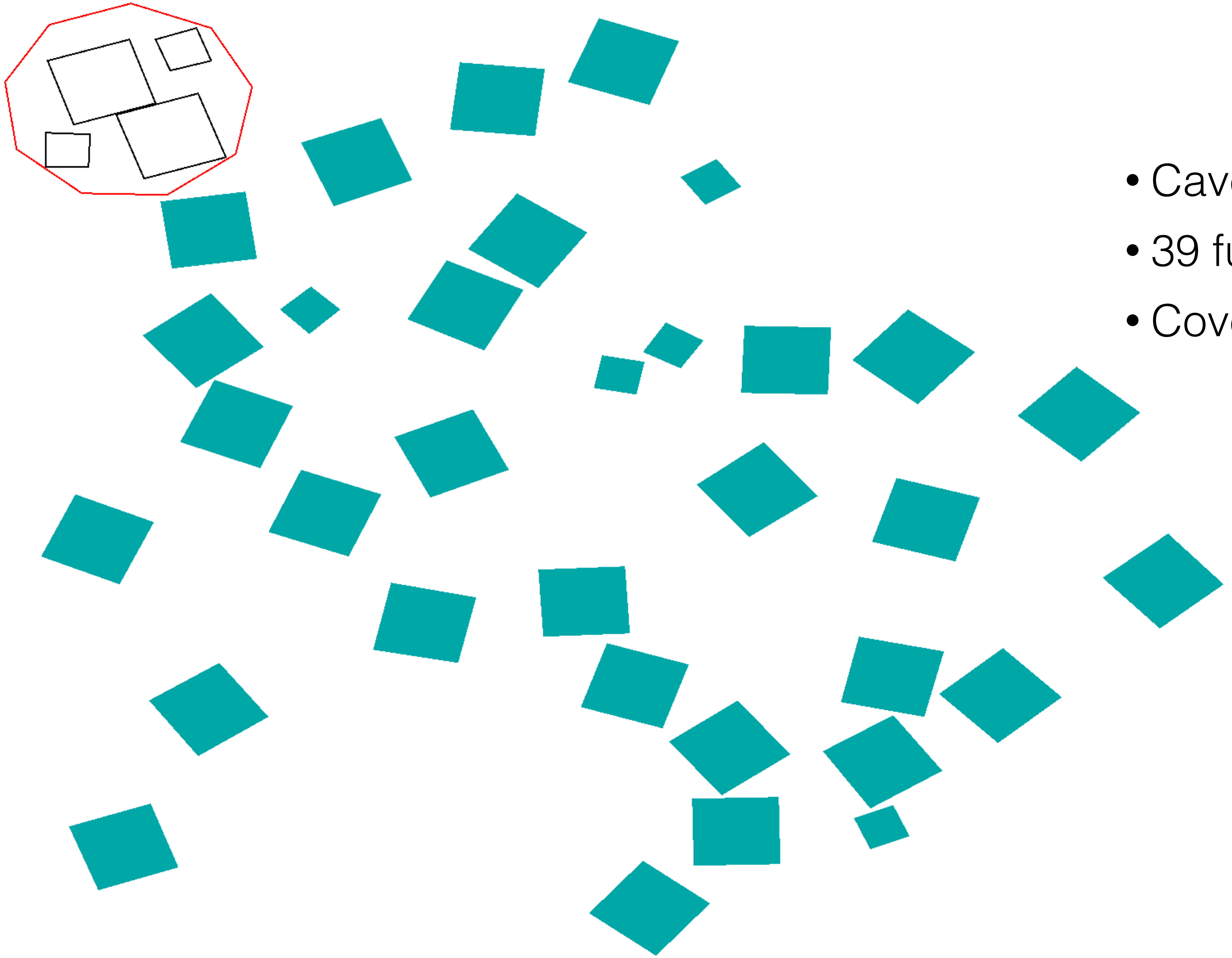

```
1: (0,0), (2,0), (2,1), (1,1), (1,2), (0,2) # 1:(0,0), (1,0), (1,1), (0,1); 2:(0,0), (2,0), (0,1); 3:(0,0), (0.5,0), (0.5,2), (0,2)
```

R A B C

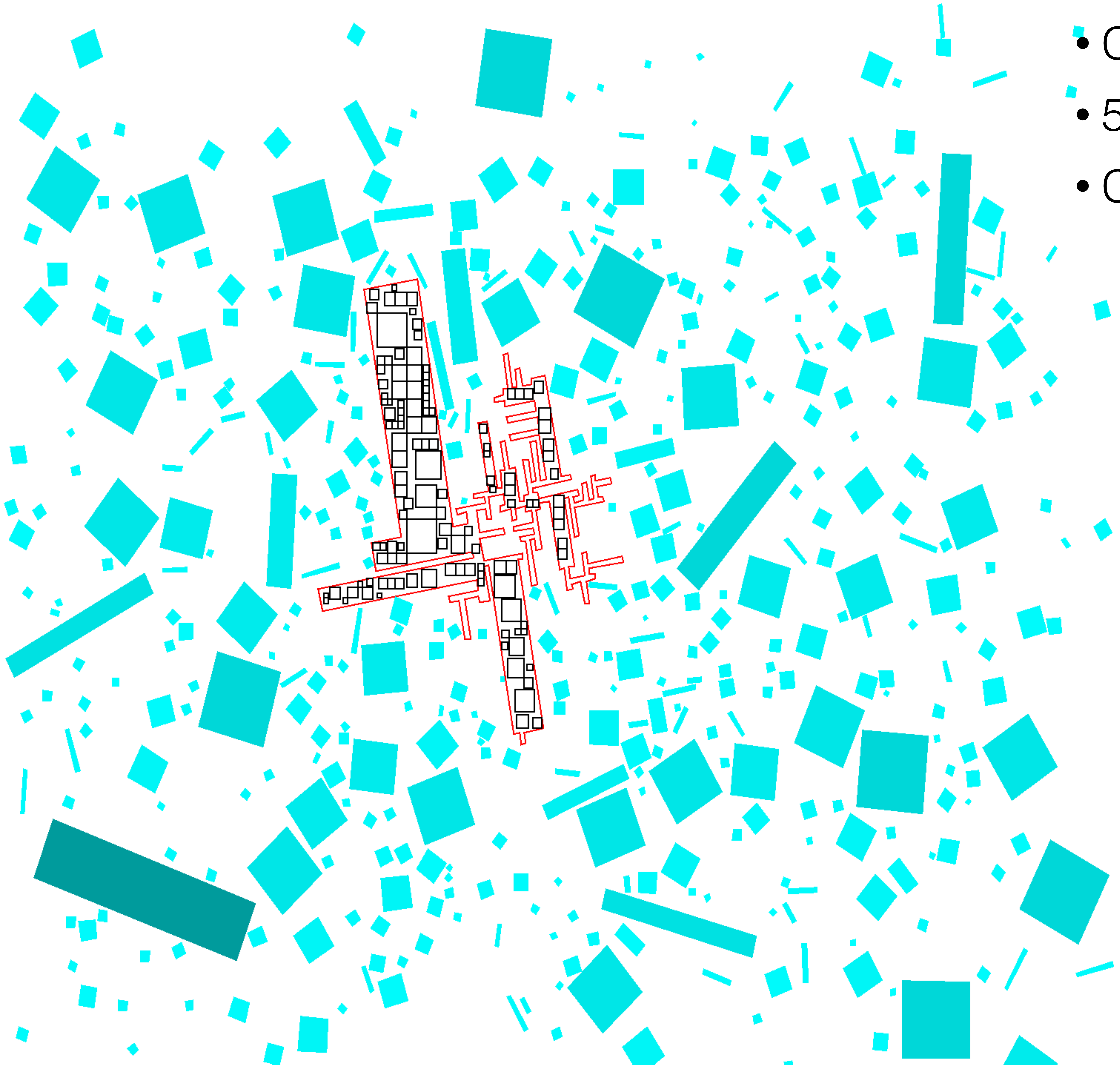

1
2
3

Q2: Algorithm for Checking Solutions

- What is an acceptable solution?
- How to check it using the data types we already have?



- Cave size: 9
- 39 furniture pieces
- Coverage: 40%

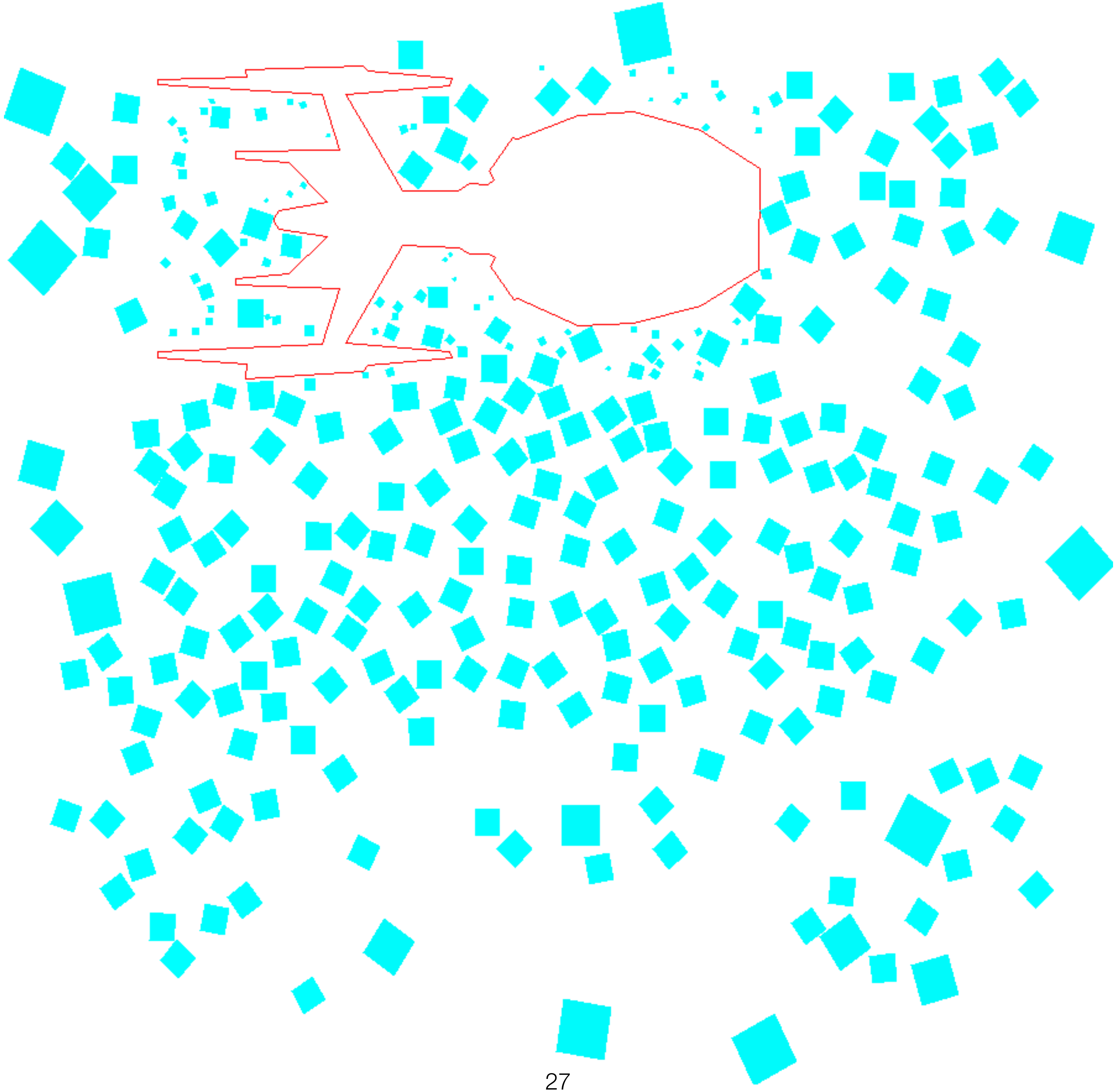


- Cave size: 180
- 500 furniture pieces
- Coverage: 46%

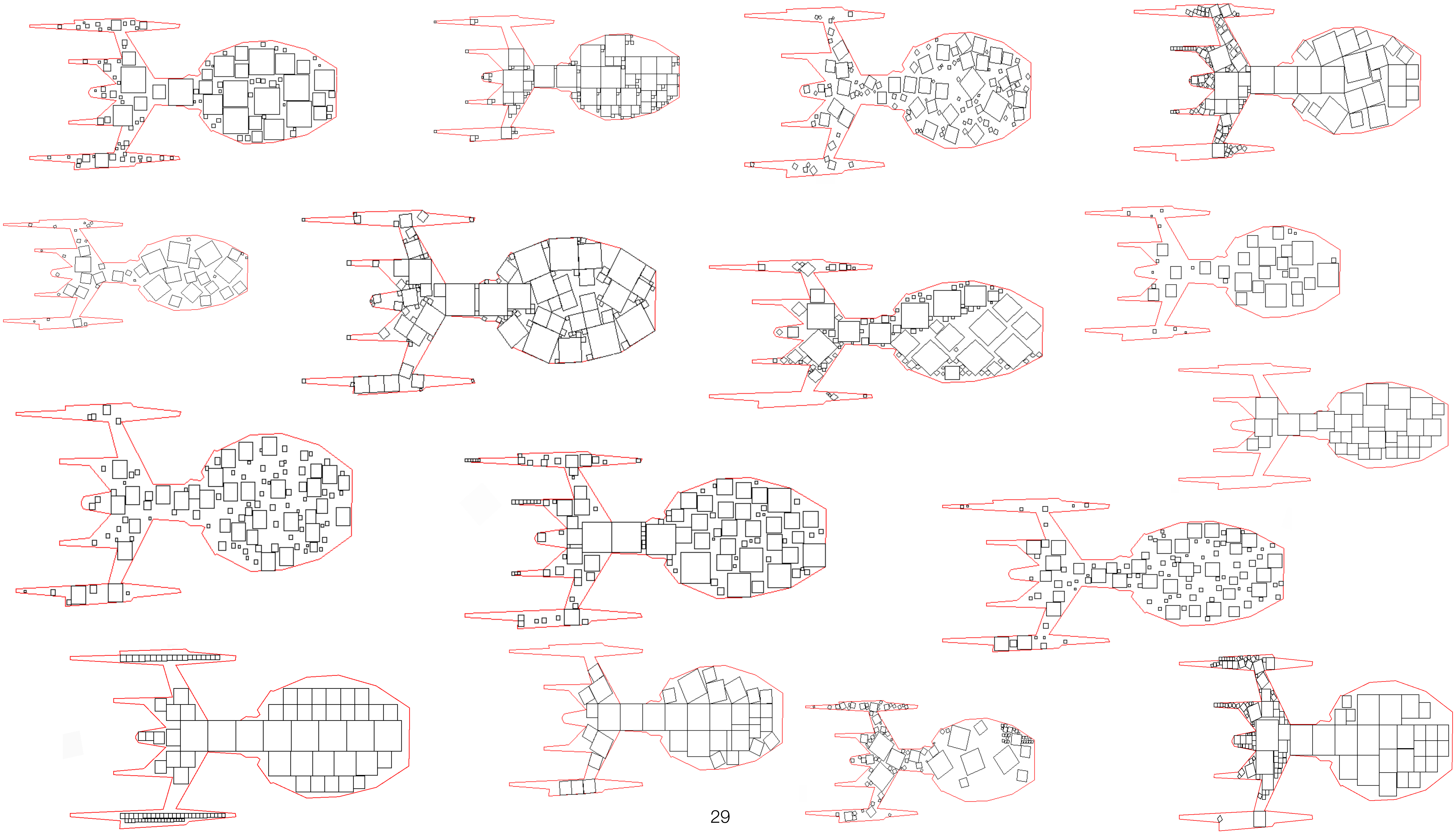
Q3: Algorithm for Solving the Problem

- What are the main steps?
- How to produce an acceptable solution?
- When should we stop?

Some solutions




USS ENTERPRISE
NCC-1701



Why take this class?

- You will learn:
  - To *understand* and *evaluate* some classic algorithms
  - How to design algorithms that are *fast*
  - How to choose the right data structures for your problems
  - How to exhaustively *test* your code
  - A little bit about *compilers* and *memory management*
  - More functional and imperative programming in OCaml
  - How to be a better programmer (not just in OCaml, but any language)

- Expect this to be a *very challenging*, implementation-oriented course (*duh*!)
  - Programming assignments might take up *tens of hours* per week…

Workload in 2020

(10 respondents)

B2: Please select the exact number of hours you spent on this course in a typical week, not including scheduled seminar or lecture time.

Name	1hr	2hr	3hr	4hr	5hr	6hr	7hr	8hr	9hr	10hr
YSC2229: Introductory Data Structures and Algorithms	0	0	0	0	1	0	0	0	0	3

Name	11hr	12hr	13hr	14hr	15hr	16hr	17hr	18hr	19hr	20hr	Mean
YSC2229: Introductory Data Structures and Algorithms	1	3	1	0	0	1	0	0	0	0	11.10

What else this course is about

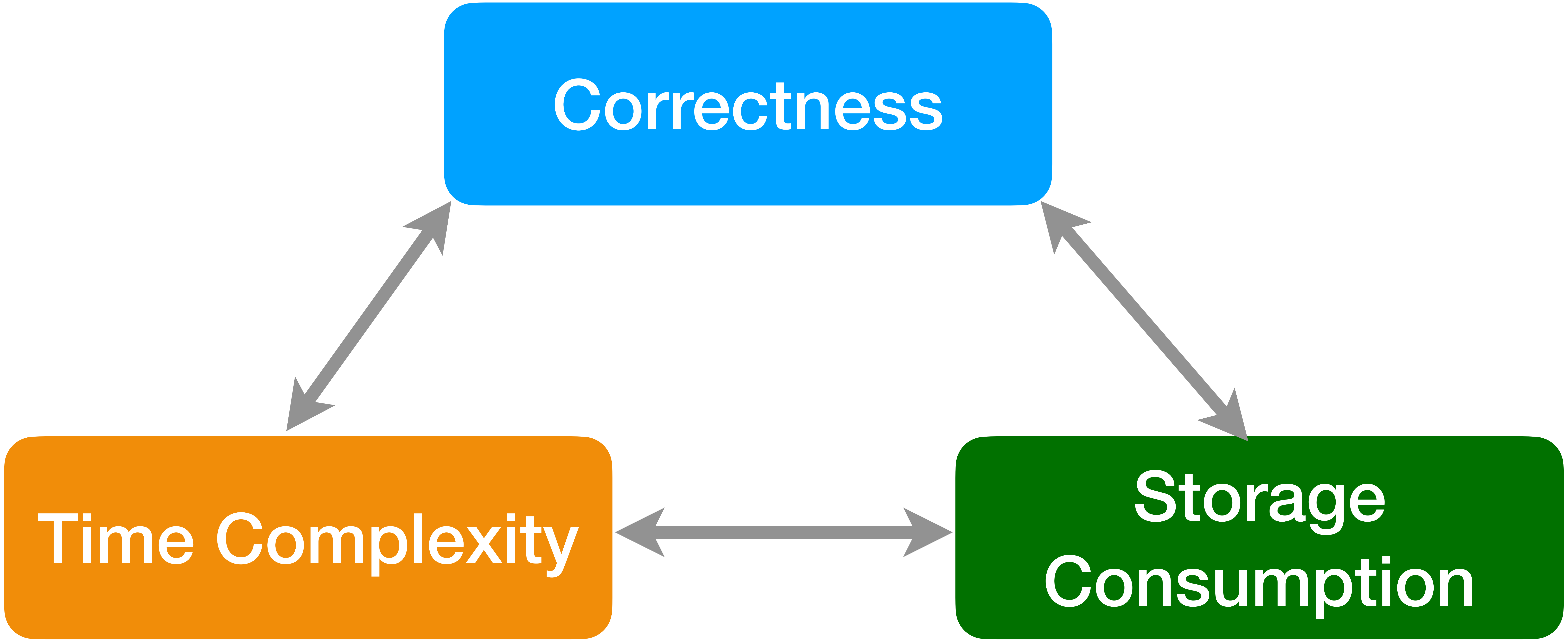
Analysis of Algorithms

Aspects that we will study

- Algorithm Correctness
- Algorithm Termination
- Time complexity
  - Worst case
  - Average case
  - Best Worst case

Aspects that we will study

- Algorithm Correctness — *Does my algorithm really do what it's supposed to do?*
- Algorithm Termination — *Does my algorithm always complete its work?*
- Time complexity — *How slow is my algorithm…*
  - Worst case — *… in the worst possible case?*
  - Average case — *… in an average case?*
  - Best Worst case — *… if I do my best to optimise it?*


Correctness
Time Complexity
Storage
Consumption

Algorithmic problems and Time Complexity

- **tractable problems** — admit solutions that run in “reasonable” time (e.g., sorting, searching, compression/decompression)
- **possibly intractable** — probably don’t have reasonable-time algorithmic solutions (e.g., SAT, graph isomorphism)
- **practically intractable** — definitely don’t have such solutions (e.g. the Towers of Hanoi)
- **non-computable** — can’t be solved algorithmically at all (e.g., the halting problem)

Why do we care about Time Complexity?

Example: Determinant of a matrix

Laplace expansion: $$|M| = \sum_{i=1}^{n} (-1)^{i-1} \overbrace{M_{1,i}}^{\text{M's element at row 1, column i}} \overbrace{|M^{1,i}|}^{\text{(1,i)-minor of M}}$$

For a 3x3 matrix: $$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} =$$

$$a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} =$$

$$a_{11}(a_{22} \cdot a_{33} - a_{23} \cdot a_{32}) - a_{12}(a_{21} \cdot a_{33} - a_{23} \cdot a_{31}) + a_{13}(a_{21} \cdot a_{32} - a_{22} \cdot a_{31})$$

Example: Determinant of a matrix

Laplace expansion: $$|M| = \sum_{i=1}^{n} (-1)^{i-1} M_{1,i} |M^{1,i}|$$

(in Haskell)

```
detLaplace :: Num a => Matrix a -> a

detLaplace m
    | size m == 1 = m ! (1,1)
    | otherwise   =
       sum [ (-1)^(i-1) * m ! (1,i) * detLaplace (minorMatrix 1 i m) |
             i <- [1 .. ncols m] ]
```

(demo)

Triangular matrices

$$L = \begin{pmatrix} a_{1,1} & 0 & \cdots & 0 \\ a_{2,1} & a_{2,2} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix} \qquad U = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ 0 & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{n,n} \end{pmatrix}$$

Determinant of a triangular matrix is a *product* of its diagonal elements.

For a 3x3 matrix: $\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{vmatrix} =$

$$a_{11}(a_{22} \cdot a_{33} - \cancel{a_{23} \cdot 0}) - a_{12}(\cancel{0 \cdot a_{33}} - \cancel{a_{23} \cdot 0}) + a_{13}(\cancel{0 \cdot a_{32}} - \cancel{a_{22} \cdot 0}) = a_{11} \cdot a_{22} \cdot a_{31}$$

Determinants via LU-decomposition

LU-decomposition: any square matrix M, such that its top-left element is non-zero can be represented in a form

$$M = LU$$

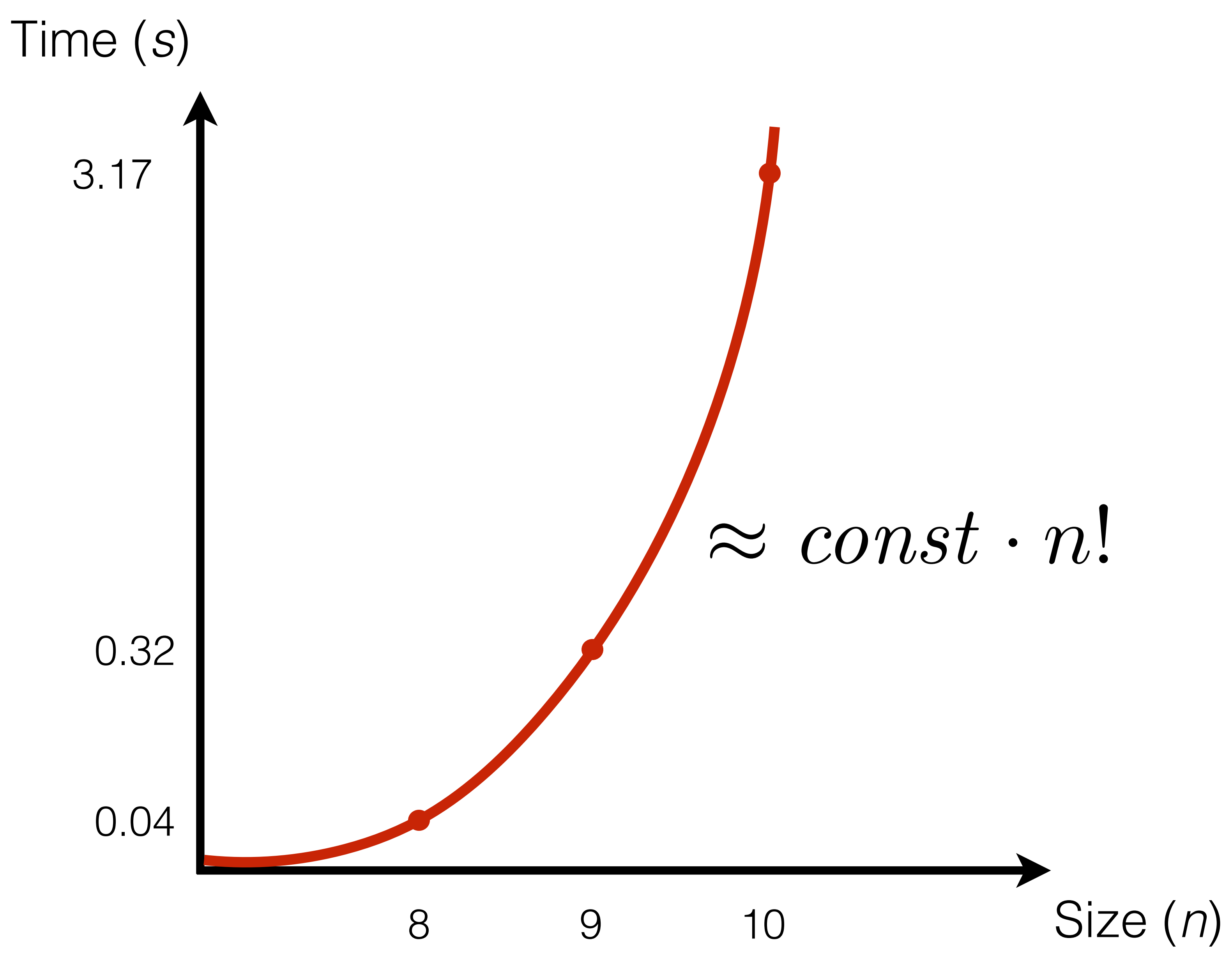
where L and U are lower- and upper-triangular matrices.

Therefore, $|M| = |L| \cdot |U|$

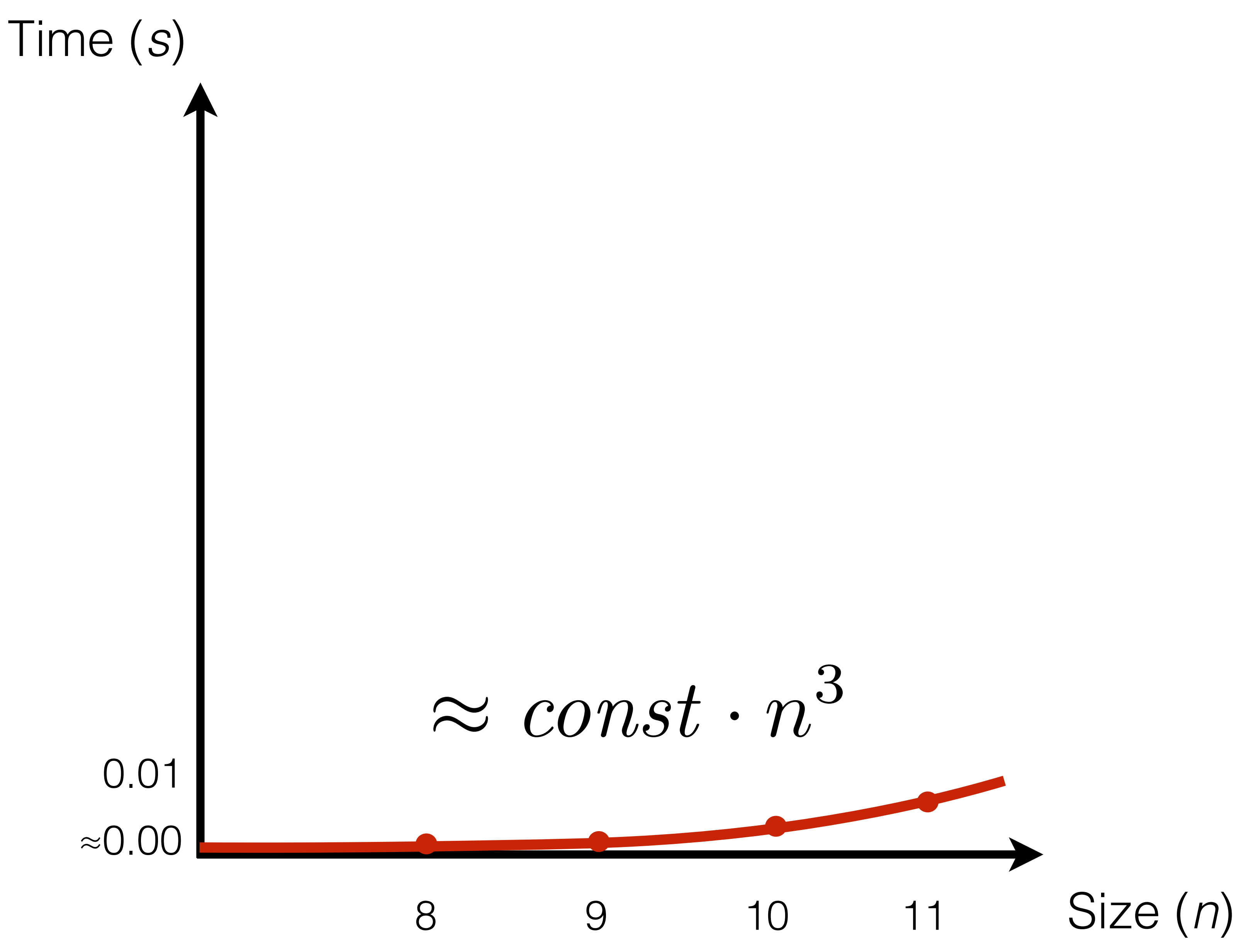
```
detLU :: Num a => Matrix a -> a
detLU m = case luDecomp m of
            (l, u) -> diagProd l * diagProd u
```

(demo)

Running time as a function of size



Determinant via
Laplace expansion



Determinant via
LU-decomposition

Time demand depends on problem size

	Problem size			
Function	10	10^2	10^3	10^4
$\log_2 n$	3.3	6.6	10	13.3
n	10	100	1000	10^4
$n \log_2 n$	33	700	10^4	1.3×10^5
n^2	100	10^4	10^6	10^8
n^3	1000	10^6	10^9	10^{12}
2^n	1024	1.3×10^{30}	$> 10^{100}$	$> 10^{100}$
$n!$	3×10^6	$> 10^{100}$	$> 10^{100}$	$> 10^{100}$

http://en.wikipedia.org/wiki/Googol

“Sizes” of different problems

Problem	**Input size, *n***
sorting	number of items to be sorted
searching	size of the set to query
determinant calculation	number of rows and columns in the matrix
finding a shortest path	number of “checkpoints” to choose from

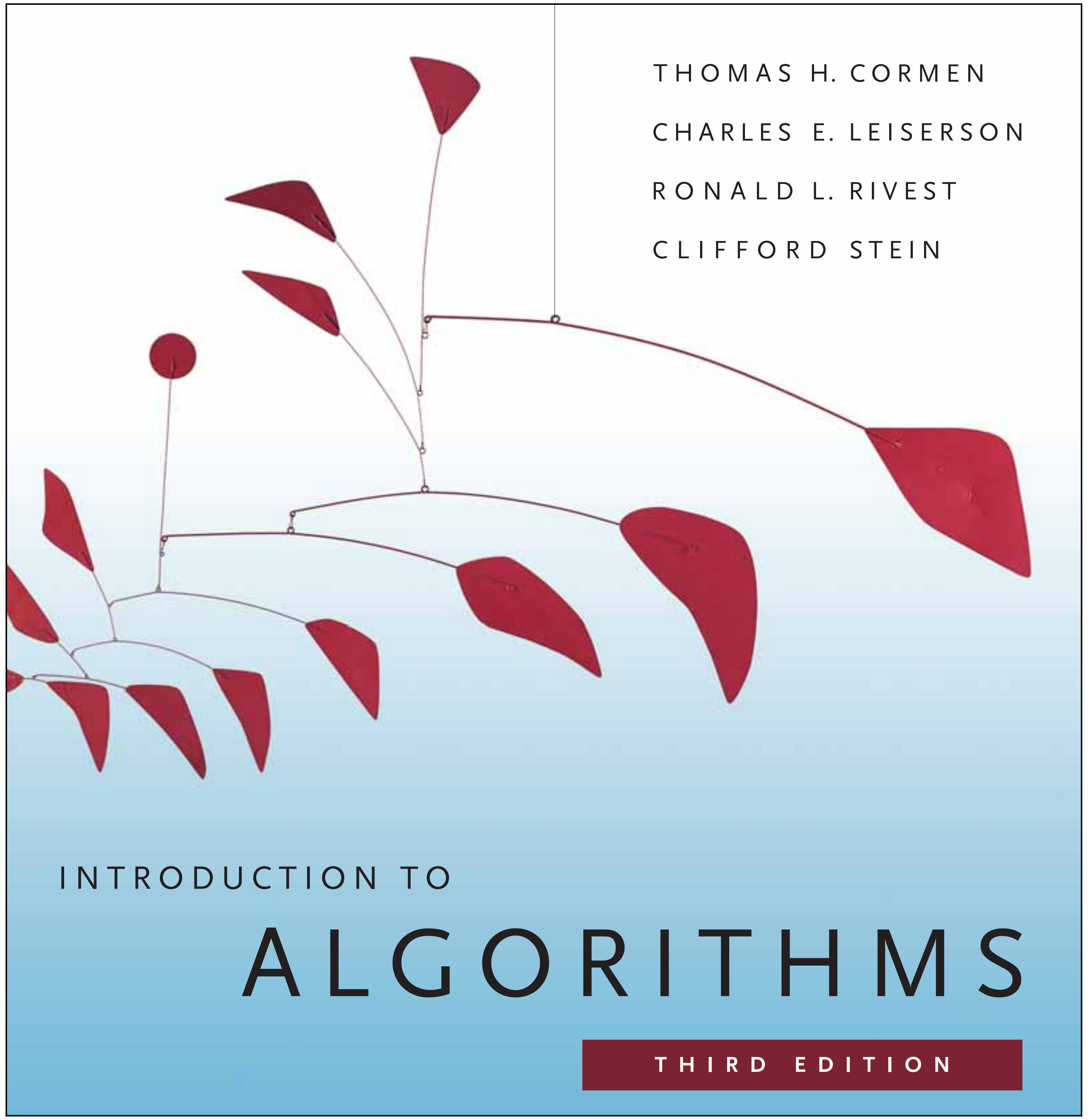
Two ways to analyse algorithms

- **Empirical** — repeatedly run algorithm with different inputs to get some idea of behaviour on different inputs
  - was our selection of inputs representative?
  - this consumes the very resource (time) we are trying to conserve!
- **Theoretical** — analysis of a “paper” version of the algorithm
  - can deal with all cases (even impractically large input instances);
  - machine-independent.

What we will learn about

- Correctness and Invariants
- Time Complexity and Order Notation
- Reasoning about Recursive Algorithms
- Searching Algorithms
- InsertSort, MergeSort, QuickSort
- Sorting in Linear Time
- Binary Heaps and HeapSort
- Abstract Data Types: Stacks, Queues
- Hash-Tables
- Memory Allocation
- Randomised Structures and False Positives
- Substring Search Algorithms
- Constraint Solving and Backtracking
- Optimisation and Dynamic Programming
- Input/Output and Binary Encodings
- Data Compression and Huffman Encoding
- Union-Find
- Representing Sets, Binary Search Trees
- Representing Graphs
- Shortest Paths, Spanning Trees
- Basics of Computational Geometry
- Convex Hulls

The Textbook



Lecture Notes

ilyasergey.net/YSC2229

Code from Lectures

github.com/ysc2229/ysc2229-2021

every week is a new branch

Working Tools

- OCaml
- Emacs/Aquamacs
  - https://ilyasergey.net/YSC2229/prerequisites.html
- GitHub for homework assignments
  - Make sure to make yourself an account (it's free)
  - Also, ask for students benefits (also free)

Assessment

- 65% — homework exercises (10 assignments)
- 15% — mid-term project (12 code, 3 report)
- 15% — final project (12 code, 3 report)
- 5% class participation (attendance, questions)

Homework

- Two types: theoretical and programming assignments
- To be completed *individually*
- Deliverables:
  - a GitHub release with an OCaml project (programming)
  - a PDF with typeset answers (theory)
- Each assignment is graded out of 20 points
- Coding assignments that don’t compile will get 0 points

Collaboration

- Permitted:

  Talking about the homework problems with other students; using other textbooks; using the Internet to improve understanding of the problems.

- Not permitted:

  Obtaining the answer directly from anyone or anything else in any form.

Homework Policies

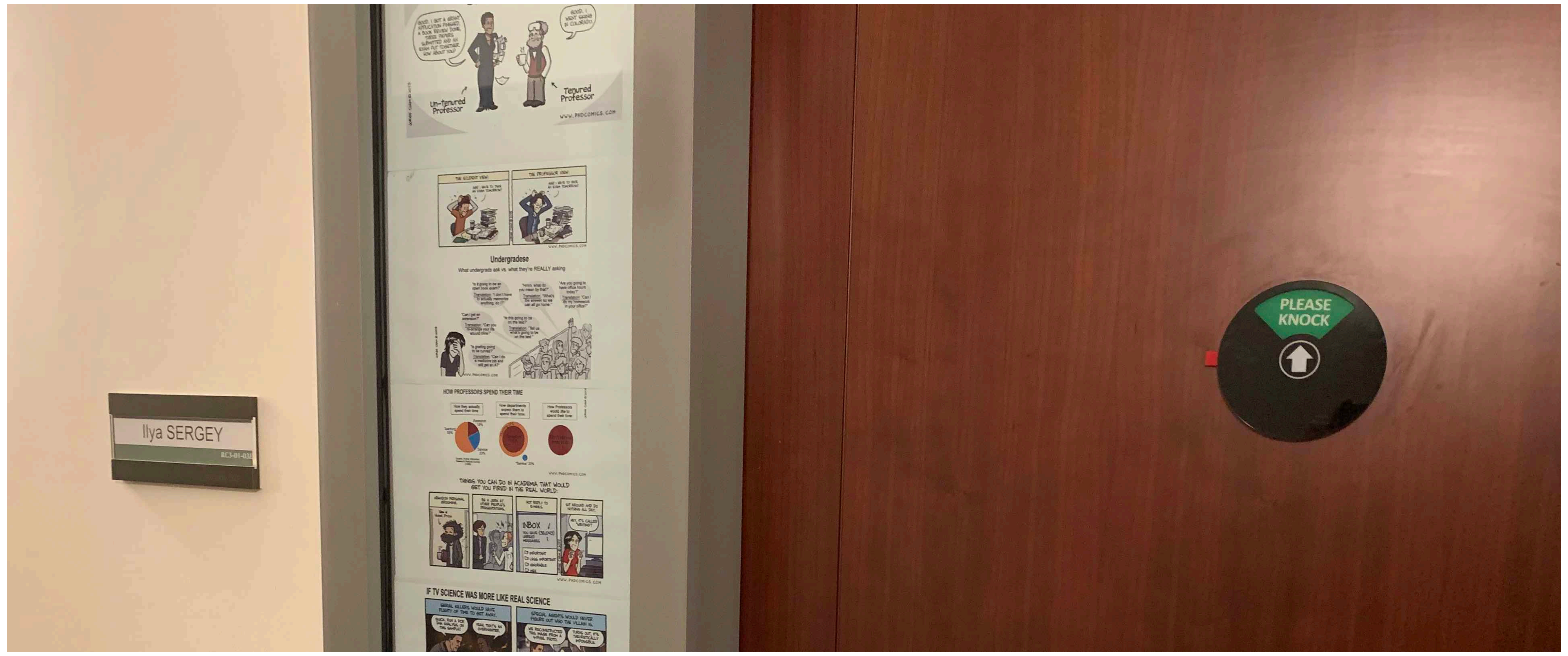
- Work submitted *before* the deadline and receiving less than 18 points can be resubmitted *within one week* after the grades are posted on Canvas.
- The amended grade will not be higher than 18
- Late submissions will be penalised by subtracting (full days after deadline + 2) points from the maximal score (20).
- Late submissions cannot be resubmitted.

Mid-term and Final Projects

- Done in **teams of two**
  (possibly one team of three, with some extra tasks)
- Graded out of 15 points (each counts towards 15% of final score)
- Deliverables:
  - GitHub release
  - PDF report, submitted *individually* by each member of the team.

Getting Help

- Office Hours (#RC3-01-03E, Cendana):
  Wednesdays 17:00-19:00
  Please, email me upfront!



- **E-mail policy:** questions about homework assignments sent less than 24 hours before submission deadline won't be answered.

- Exception: bug reports.

Peer Tutors

Tram Hoang

tram.hoang@u.yale-nus.edu.sg

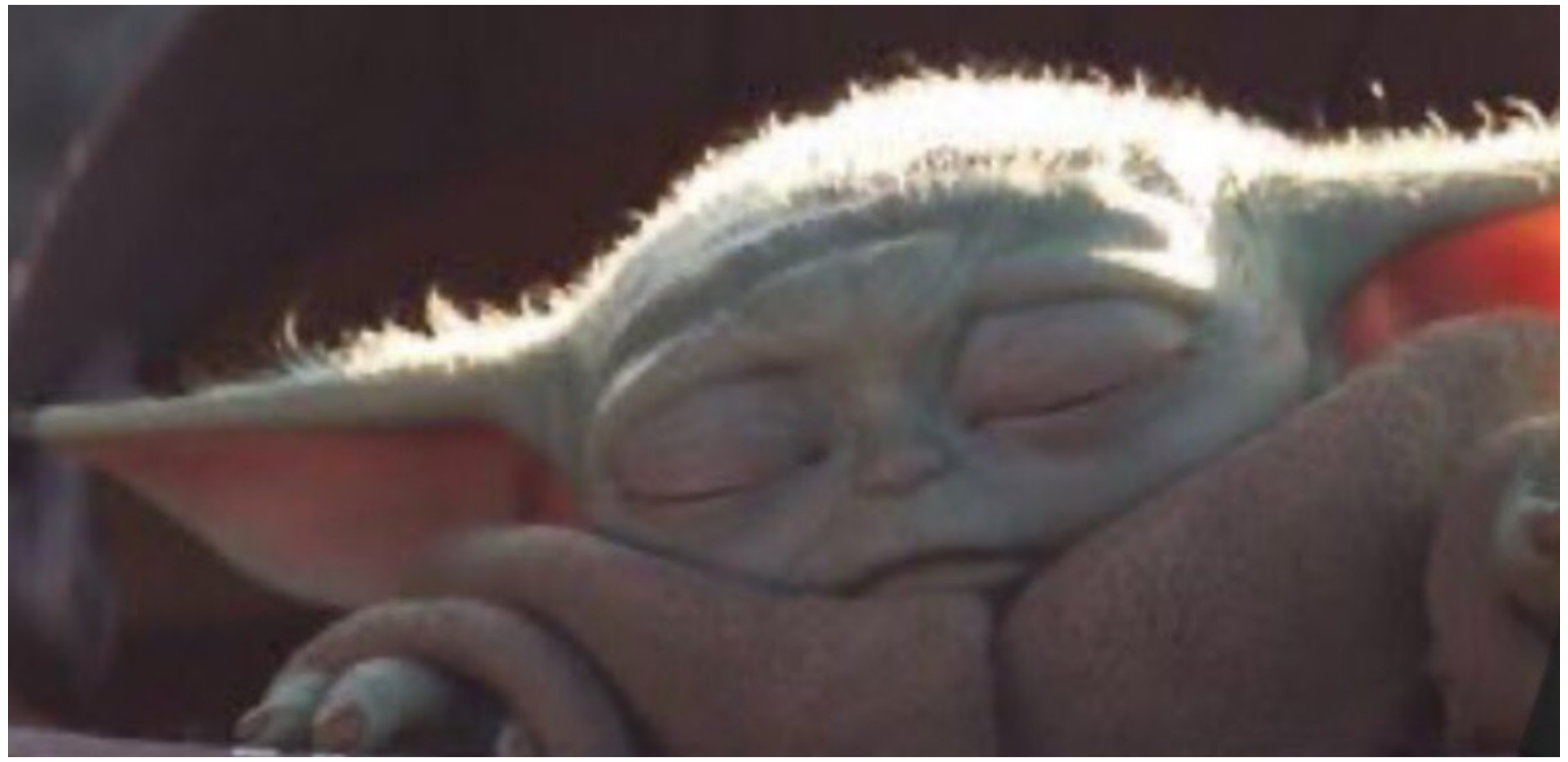
Wednesdays, 7pm-9pm, Location: CR20

Gabriel Petrov

gabrielphoenixpetrov@u.yale-nus.edu.sg

Thursdays, 6pm-8pm, Location: CR20

General Advice



- Friday afternoon class, many of you will be tired. Try to make class livelier by asking questions and participating in discussions.

- Lecture notes will contain exercises. Please, please try these! No better practice than actually solving problems.

Time for a short break