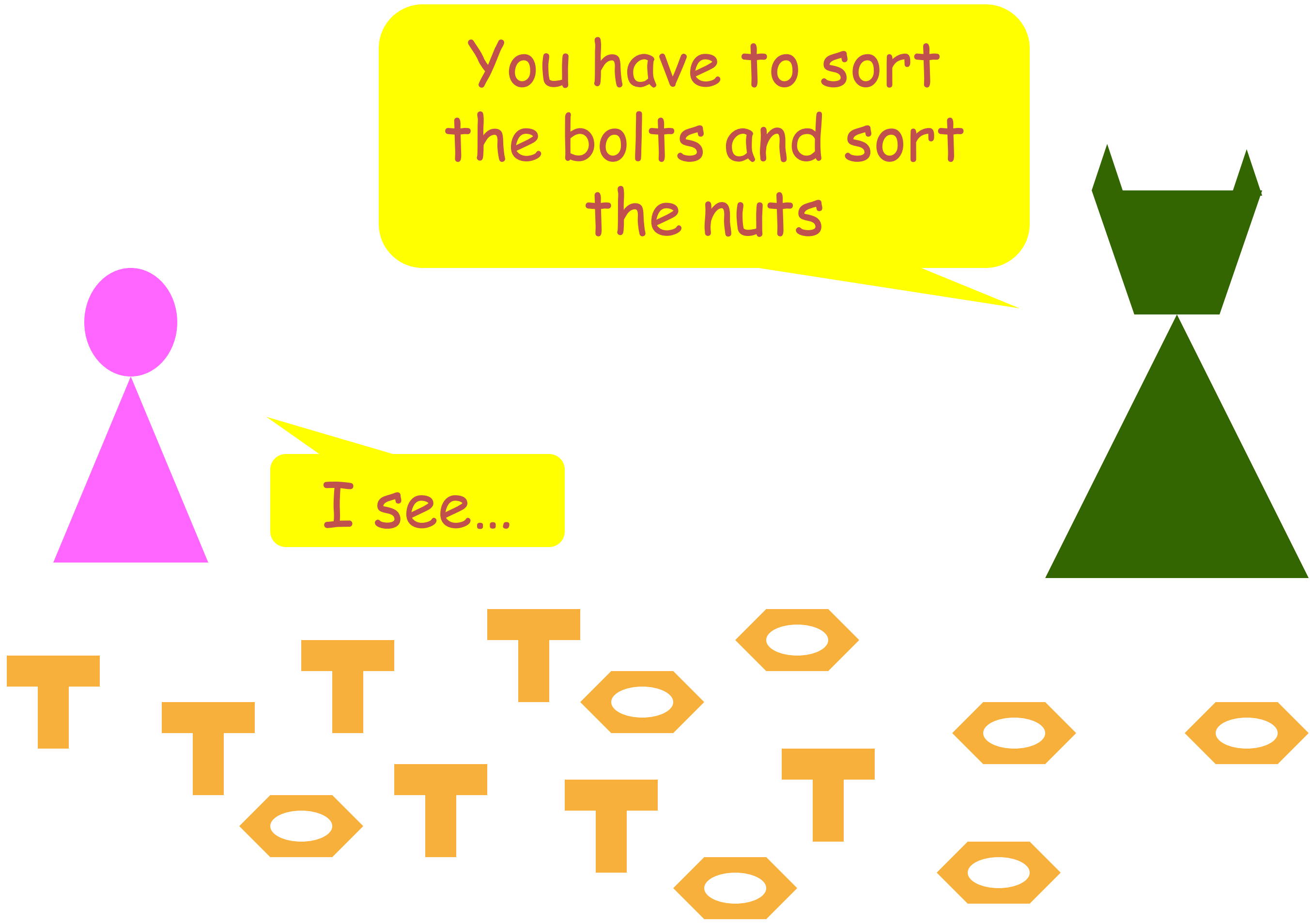
Chapter 7 Quicksort

About this lecture

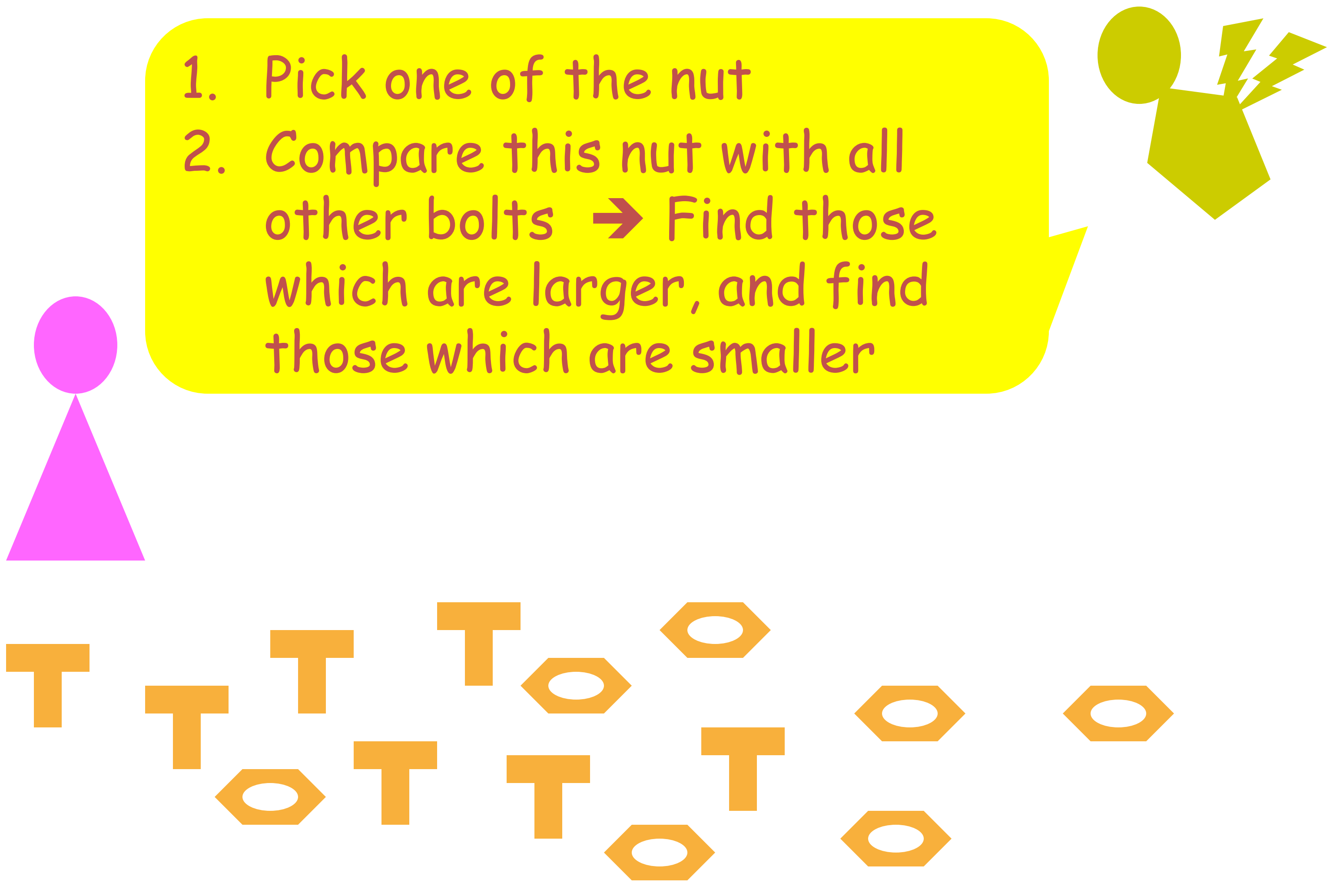
- Introduce Quicksort

- Running time of Quicksort
  - Worst-Case
  - Average-Case

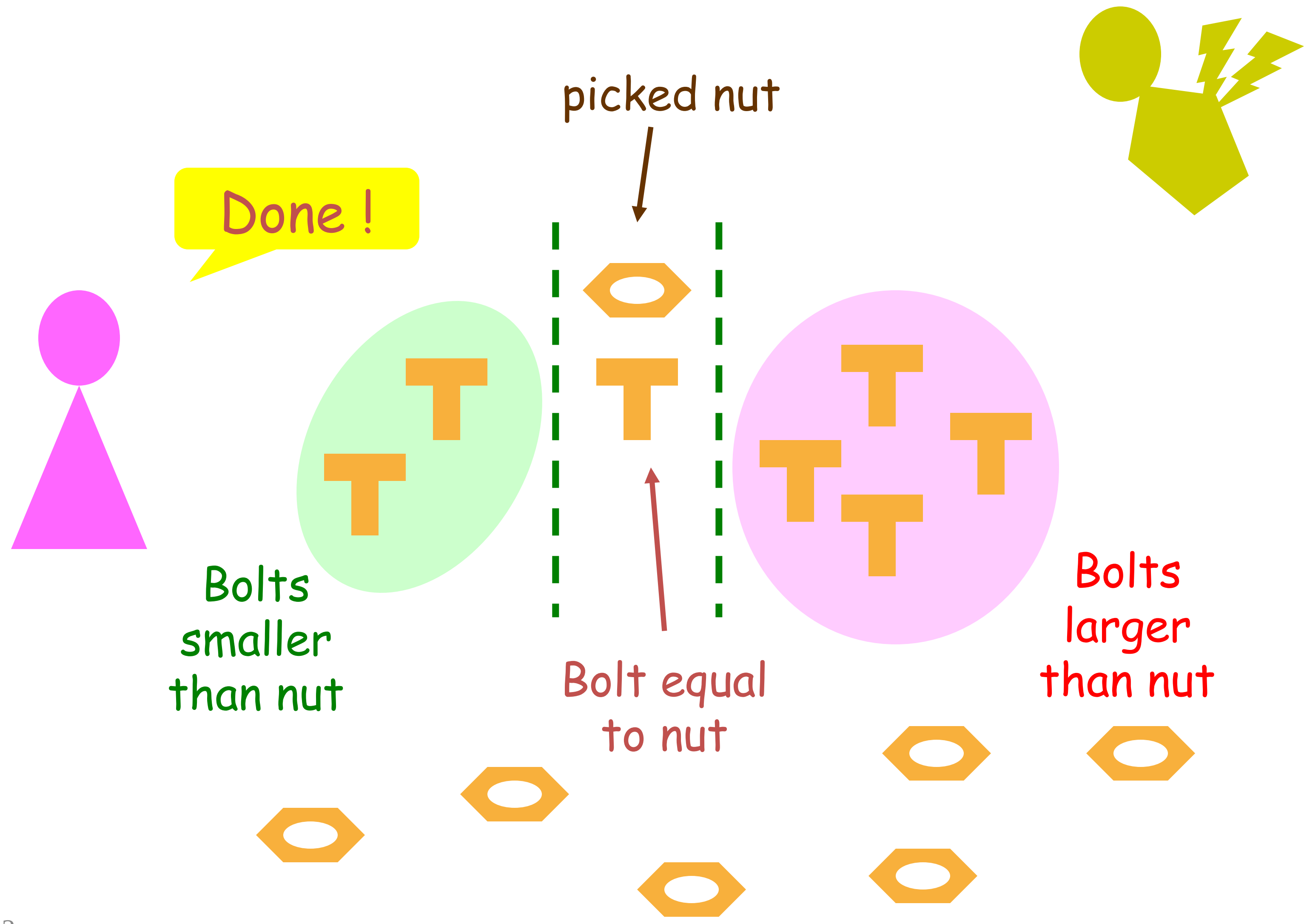
Cinderella’s New Problem



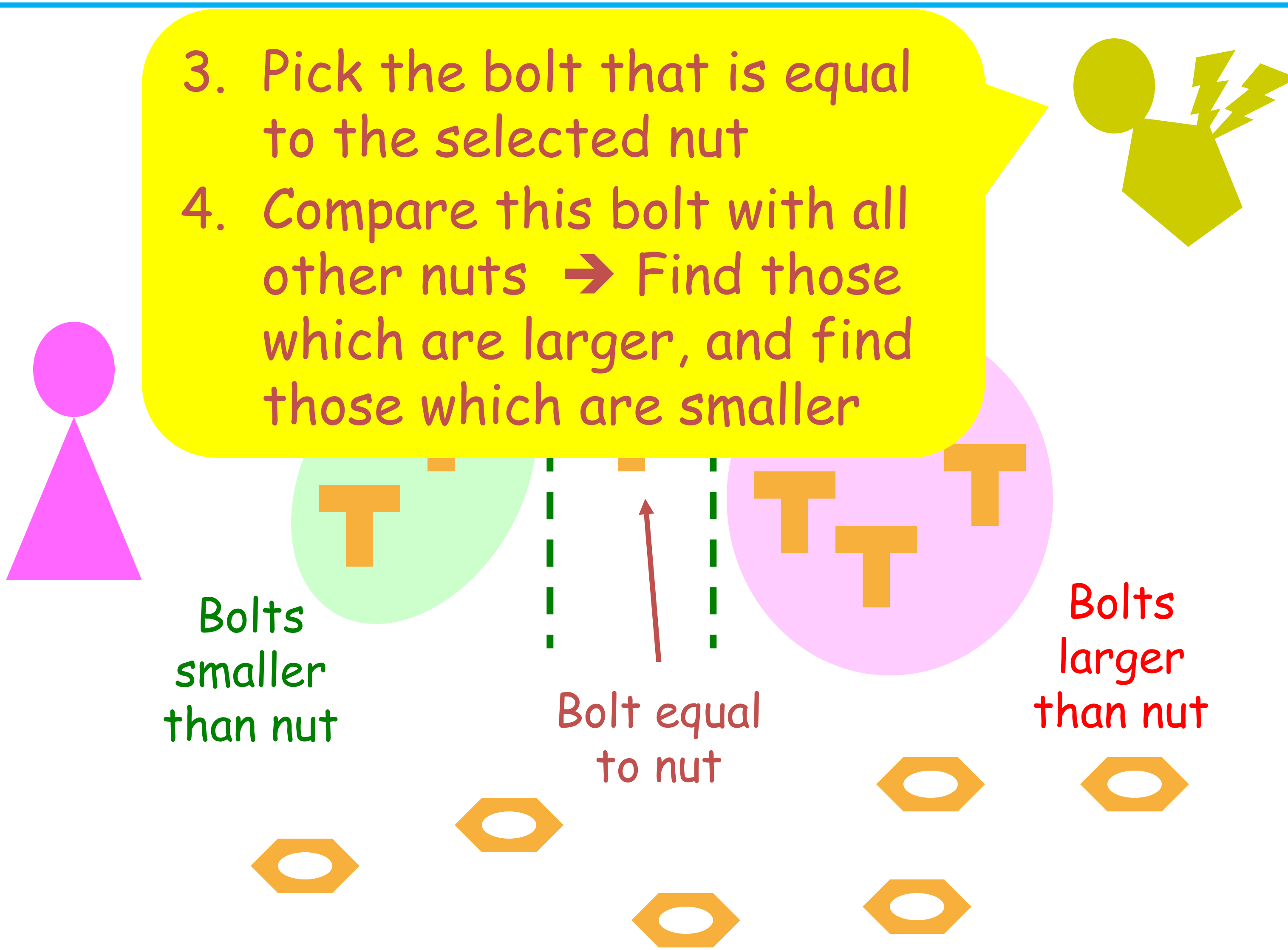
Fairy Godmother’s Proposal



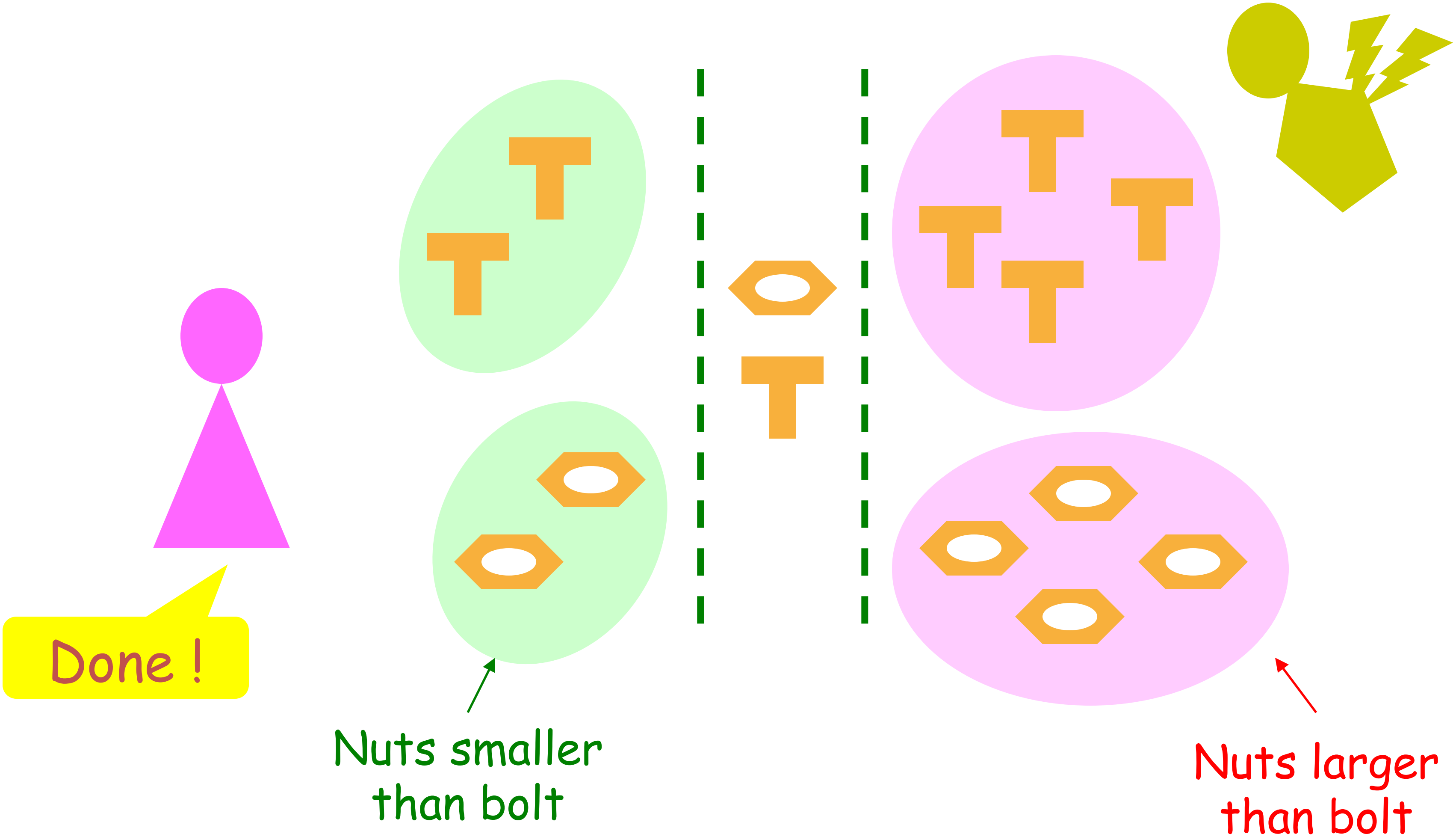
Fairy Godmother's Proposal



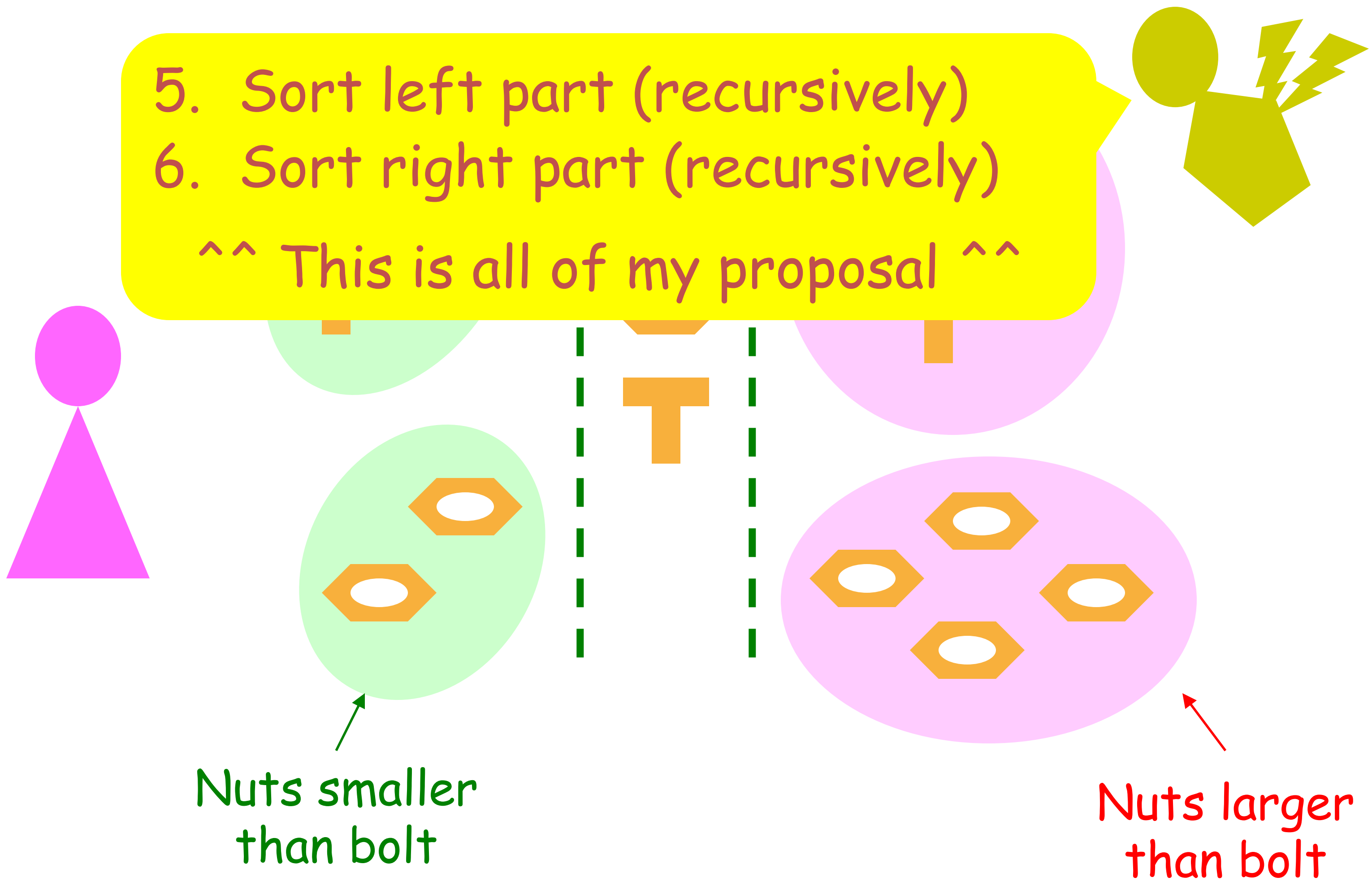
Fairy Godmother's Proposal



Fairy Godmother's Proposal



Fairy Godmother's Proposal



Fairy Godmother's Proposal

- Can you see why Fairy Godmother's proposal is a correct algorithm?
- What is the running time ?
  - Worst-case: $\Theta(n^2)$ comparisons
  - No better than the brute force approach !!
- Though worst-case runs badly, the average case is good: $\Theta(n \log n)$ comparisons

Quicksort uses Partition

- The previous algorithm is exactly Quicksort, which makes use of a Partition function:

```
Partition(A,p,r)     /* to partition array A[p..r] */
    1. Pick an element, say A[t]  (called pivot)
    2. Let q = #elements less than pivot
    3. Put elements less than pivot to A[p..p+q-1]
    4. Put pivot to A[p+q]
    5. Put remaining elements to A[p+q+1..r]
    6. Return q
```

More on Partition

- After Partition(A,p,r), we obtain the value q, and know that
  - Pivot was A[p+q]
  - Before A[p+q] : smaller than pivot
  - After A[p+q] :   larger than pivot

- There are many ways to perform Partition. One way is shown in the next slides
  - It will be an in-place algorithm (using O(1) extra space in addition to the input array)

Ideas for In-Place Partition

- Step 1: Use A[r] (the last element) as pivot
- Step 2: Process A[p..r] from left to right
  - Use two counters:
    - ✓One for the length of the prefix
    - ✓One for the element we are looking
  - The prefix of A stores all elements less than pivot seen so far

In-Place Partition in Action

before running

Length of prefix = 0

1	3	7	8	2	6	4	5

next element

pivot

Because next element is less than pivot,
we shall extend the prefix by 1

In-Place Partition in Action

after 1 step

Length of prefix = 1

1	3	7	8	2	6	4	5

next element

pivot

Because next element is smaller than pivot, and is adjacent to the prefix, we extend the prefix

In-Place Partition in Action

after 2 steps

Length of prefix = 2

1	3	7	8	2	6	4	5

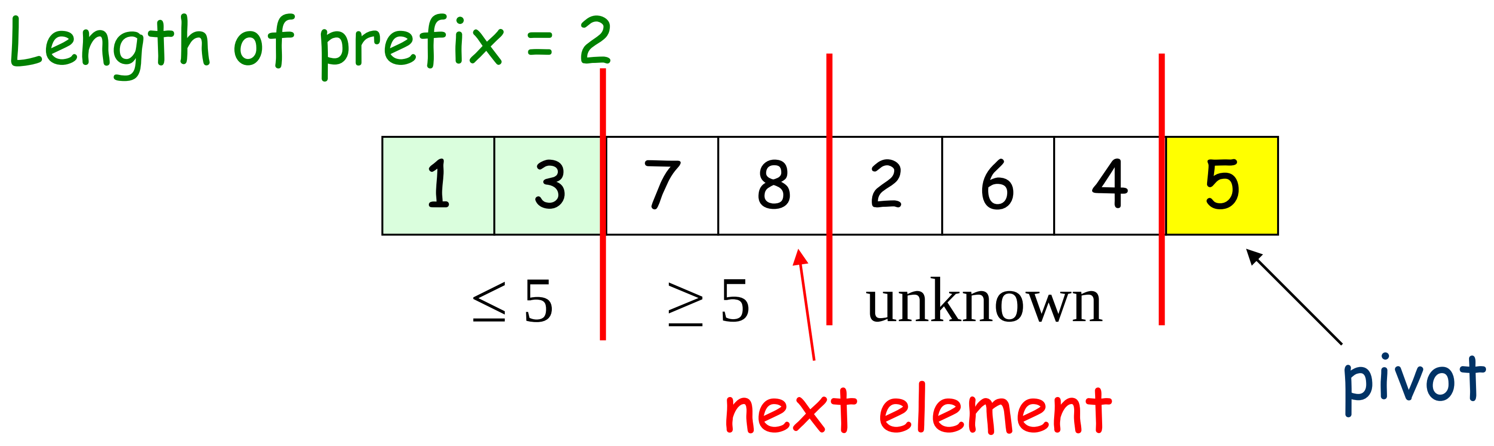
next element

pivot

Because next element is larger than pivot,
no change to prefix

In-Place Partition in Action

after 3 steps

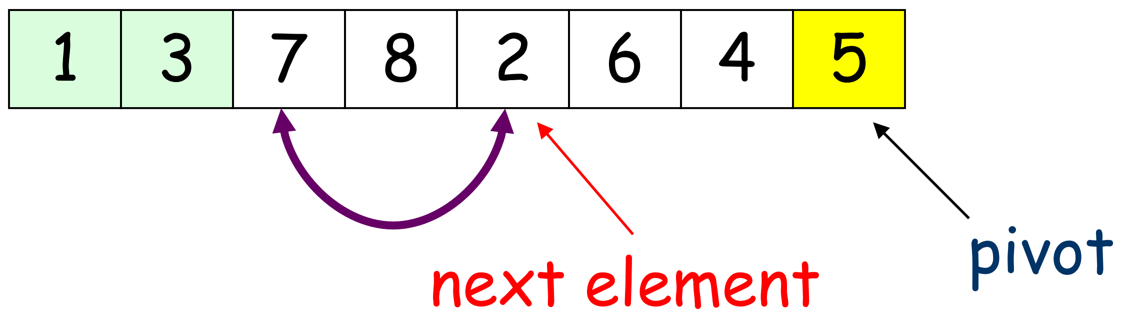


Again, next element is larger than pivot, no change to prefix

In-Place Partition in Action

after 4 steps

Length of prefix = 2



Because next element is less than pivot,
we shall extend the prefix by swapping

In-Place Partition in Action

after 5 steps

Length of prefix = 3

1	3	2	8	7	6	4	5

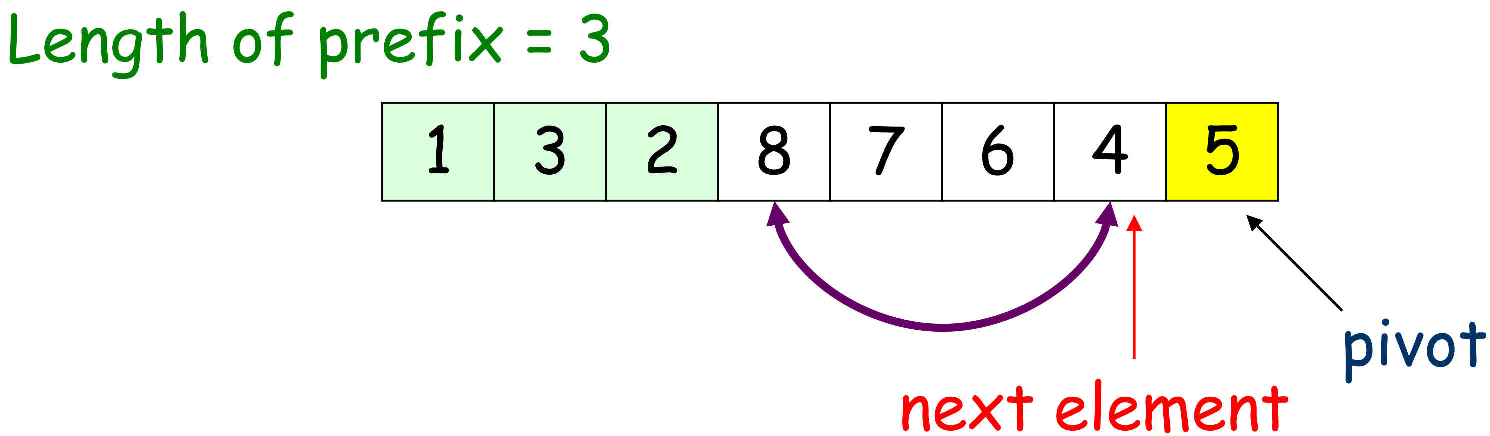
next element

pivot

Because next element is larger than pivot,
no change to prefix

In-Place Partition in Action

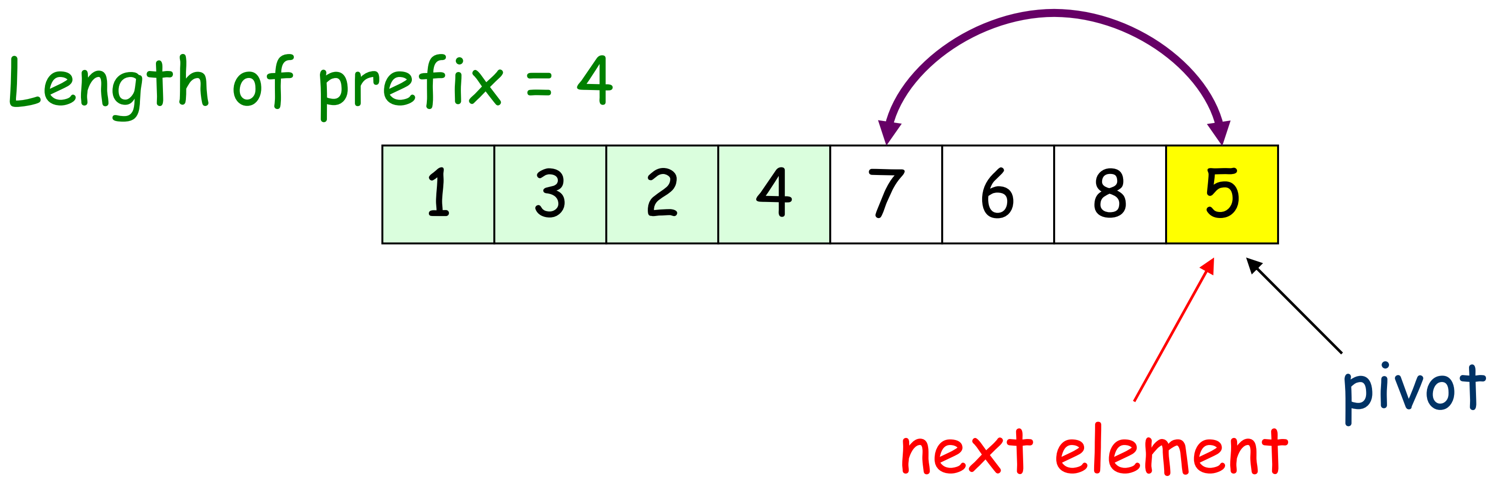
after 6 steps



Because next element is less than pivot,
we shall extend the prefix by swapping

In-Place Partition in Action

after 7 steps



When next element is the pivot, we put it after the end of the prefix by swapping

In-Place Partition in Action

after 8 steps

Length of prefix = 4

1	3	2	4	5	6	8	7

pivot

Partition is done, and return length of prefix

Partitioning the array

```
PARTITION (A, p, r)
  x = A[r]
  i = p-1
  for j = p to r-1
     if A[j] <= x
          i = i+1
          exchange A[i] with A[j]
   exchange A[i+1] with A[r]
   return i + 1
```

Quicksort

- The Quicksort algorithm works as follows:

```
Quicksort(A,p,r)     /* to sort array A[p..r] */
   1.  if ( p < r )   return;
   2.  q = Partition(A, p, r);
   3.  Quicksort(A, p, p+q-1);
   4.  Quicksort(A, p+q+1, r);
```

To sort A[1..n], we just call Quicksort(A,1,n)

Randomized Versions of Partition

Randomized-Partition(A, p, r)

i = Random(p, r)

Exchange A[r] with A[i]

Return Partition(A, p, r)

Randomized-Quicksort(A, p, r)

If p < r

q = Randomized-Partition(A, p, r)

Randomized-Quicksort(A, p, q-1)

Randomized-Quicksort(A, q+1, r)

Worst-Case Running Time

- The worst-case running time of Quicksort can be expressed by:

$$T(n) = \max_{q=0 \text{ to } n-1} \big(T(q) + T(n-q-1)\big) + \Theta(n)$$

We prove $T(n) = O(n^2)$ by substitution method:

1. Guess $T(n) \le cn^2$ for some constant c
2. Next, verify our guess by induction

Basis: $n = 1$ hold, Assume $n \le k$ hold

Worst-Case Running Time

Inductive Case ($n = k + 1$):

$$T(n) = \max_{q=0 \text{ to } n-1} \big(T(q) + T(n-q-1)\big) + \Theta(n)$$

$$\leq \max_{q=0 \text{ to } n-1} \big(cq^2 + c(n-q-1)^2\big) + \Theta(n)$$

$$\leq c(n-1)^2 + \Theta(n)$$

$$= cn^2 - 2cn + c + \Theta(n)$$

Maximized when $q = 0$ or when $q = n-1$

$$\leq cn^2$$ when c is large enough

$q^2 + (n\text{-}q\text{-}1)^2 = (n\text{-}1)^2 + 2q(q\text{-}n\text{+}1)$ (1)

Because $(n\text{-}1)^2 > 0$ and $q < n$, maximum (1) is equal to minimize $q(q\text{-}n\text{+}1)$ Therefore, $q = 0$ or $q = n\text{-}1$

Worst-Case Running Time

Conclusion:

1. $T(n) = O(n^2)$
2. However, we can also show

$$T(n) = \Omega(n^2)$$

by finding a worst-case input

Let $T(n) = T(n-1) + \Theta(n)$

$\rightarrow$ $T(n) = \Theta(n^2)$

Balanced partitioning

- Quicksort's average running time is much closer to the best case than to the worst case
- Imagine that PARTITION always produces a 9-to-1 split. Get the recurrence
  - $T(n) = T(9n/10) + T(n/10) + c\,n$
- Any split of constant proportionality will yield a recursion tree of depth $\Theta(\lg n)$
- It's like the one for $T(n) = T(n/3) + T(2n/3) + O(n)$ in Section 4.4.

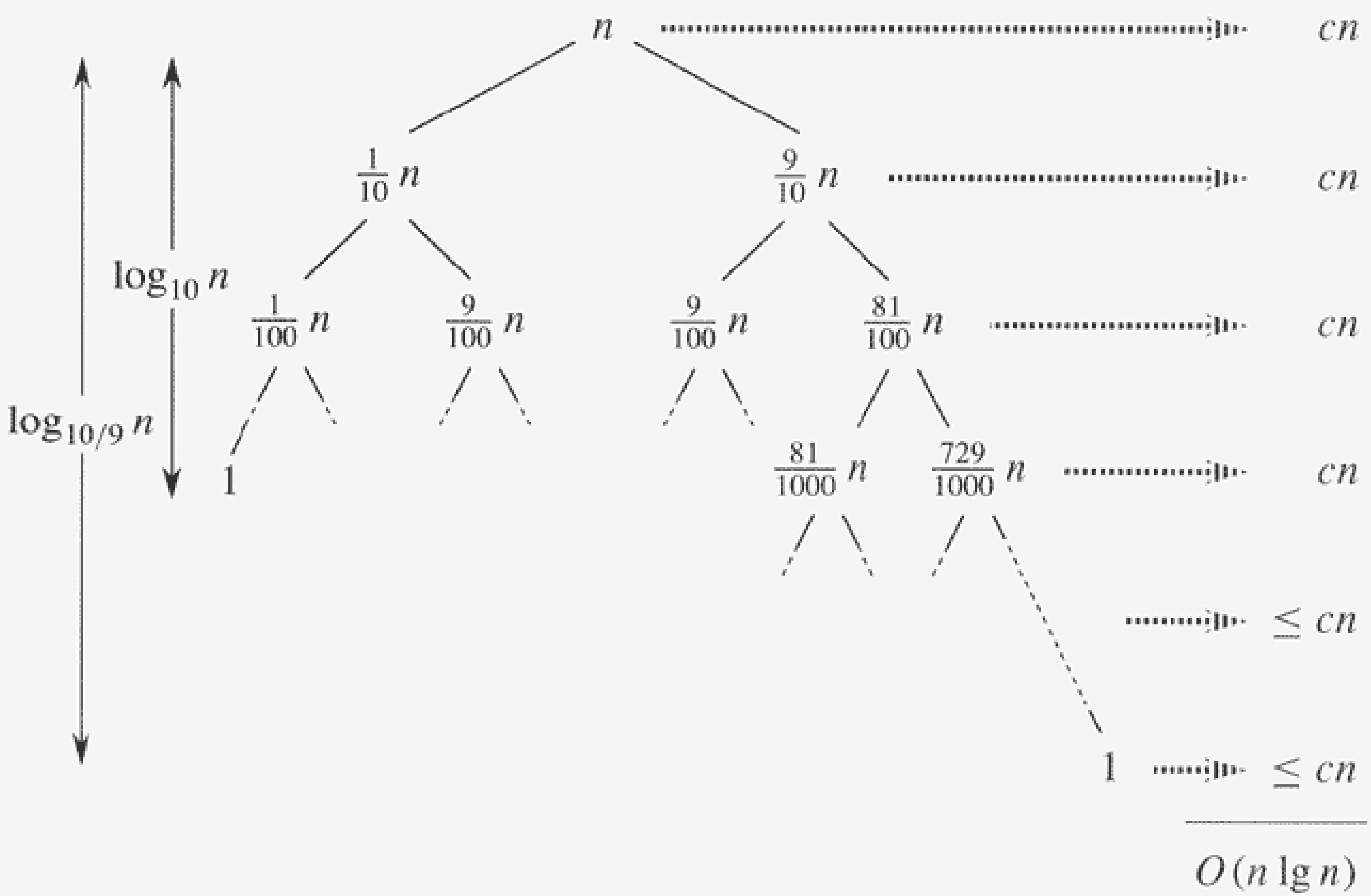


Figure 7.4 A recursion tree for QUICKSORT in which PARTITION always produces a 9-to-1 split, yielding a running time of $O(n \lg n)$. Nodes show subproblem sizes, with per-level costs on the right. The per-level costs include the constant c implicit in the $\Theta(n)$ term.

Average-Case Running Time

So, Quicksort runs badly for some input…

But suppose that when we store a set of n numbers into the input array, each of the $n!$ permutations are equally likely

➔ Running time varies on input

What will be the "average" running time ?

Average Running Time

- Let X = # comparisons in all Partition
- Later, we will show that

Running time = $O(n + X)$ varies on input

Finding average of X (i.e. #comparisons) gives average running time

Our first target: Compute average of X

Average # of Comparisons

- We define some notation to help the analysis:
- Let $a_1, a_2, \ldots, a_n$ denote the set of n numbers initially placed in the array
- Further, we assume $a_1 < a_2 < \ldots < a_n$ (So, a_1 may not be the element in $A[1]$ originally)
- Let X_{ij} = # comparisons between a_i and a_j in all Partition calls

Average # of Comparisons

- Then, X = # comparisons in all Partition calls

$$= X_{12} + X_{13} + \ldots + X_{n-1,n}$$

➔ Average # comparisons

$$= E[X]$$

$$= E[X_{12} + X_{13} + \ldots + X_{n-1,n}]$$

$$= E[X_{12}] + E[X_{13}] + \ldots + E[X_{n-1,n}]$$

Average # of Comparisons

The next slides will prove: $E[X_{ij}] = 2/(j-i+1)$

Using this result,

$$E[X] = \sum_{i=1 \text{ to } n-1} \sum_{j=i+1 \text{ to } n} 2/(j-i+1)$$

$$= \sum_{i=1 \text{ to } n-1} \sum_{k=1 \text{ to } n-i} 2/(k+1) \quad \text{(let } k = j-i)$$

$$< \sum_{i=1 \text{ to } n-1} \sum_{k=1 \text{ to } n} 2/k$$

$$= \sum_{i=1 \text{ to } n-1} O(\log n) = O(n \log n)$$

Comparison between a_i and a_j

Question: # times a_i be compared with a_j ?

Answer: At most once, which happens only if a_i or a_j are chosen as pivot before they are partitioned to either side

1	3	2	4	5	6	8	7

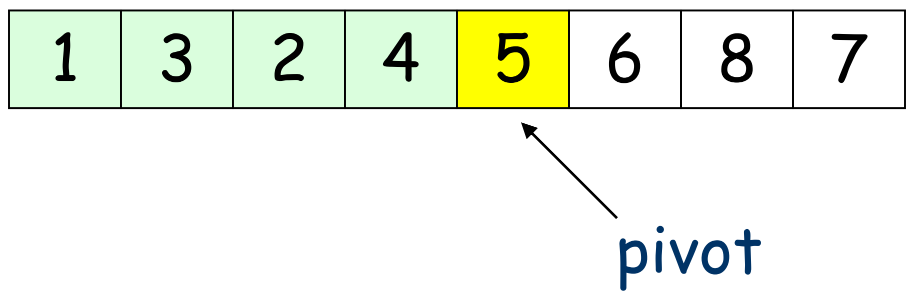
pivot

After that, the pivot is fixed and is never compared with the others

Comparison between a_i and a_j

Question: Will a_i always be compared with a_j ?

Answer: No.



we will separately Quicksort the first 4 elements, and then the last 3 elements

➔ 3 is never compared with 8

Comparison between a_i and a_j

Observation:

Consider the elements a_i, a_{i+1}, …, a_{j-1}, a_j

(i) If a_i or a_j is first chosen as a pivot, then a_i is compared with a_j

(ii) Else, if any element of a_{i+1}, …, a_{j-1} is first chosen as a pivot, then a_i is never compared with a_j

Comparison between a_i and a_j

- When the $n!$ permutations are equally likely to be the input,

$Pr(a_i \text{ compared with } a_j \text{ once}) = 2/(j-i+1)$
$Pr(a_i \text{ not compared with } a_j) = (j-i-1)/(j-i+1)$

$$\Rightarrow E[X_{ij}] = 1 * 2/(j-i+1) + 0 * (j-i-1)/(j-i+1)$$
$$= 2/(j-i+1)$$

Proof: Running time = $O(n+X)$

- Observe that in the Quicksort algorithm:

  ✓Each Partition fixes the position of pivot

  ➔ at most n Partition calls

- After each Partition, we have 2 Quicksort
- Also, all Quicksort (except 1st one: Quicksort(A,1,n)) are invoked after a Partition

  ➔ total $\Theta(n)$ Quicksort calls

Proof: Running time = $O(n+X)$

- So, if we ignore the comparison time in all Partition calls, the time used = $O(n)$

- Thus, we include back the comparison time in all Partition calls,

$$\text{Running time} = O(\ n + X\ )$$

Homework

- Exercises: 7.1-2, 7.2-3, 7.3-1, 7.4-5
- Problem 7-2

Homework

2. (Challenging) You have just finished sorting an array $A[1..n]$ of n distinct numbers into increasing order. When you go out to have a break, your mischievous friend, John, has divided your array into two parts $A_{\text{left}} = A[1..i]$ and $A_{\text{right}} = A[i + 1..n]$, and re-arrange the array so that he puts A_{right} in front of A_{left}; precisely, the array now becomes $A[i+1..n]A[1..i]$. See Figure 1 for an example.

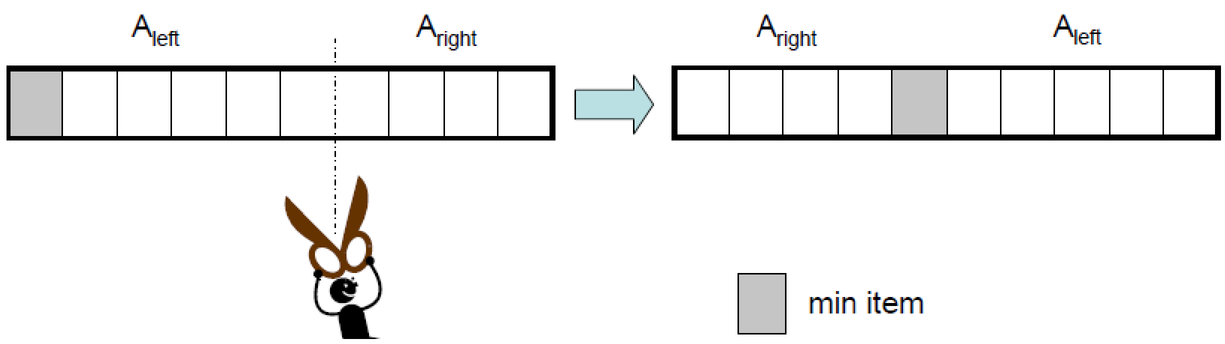


Figure 1: John's modification to the array.

After you come back, John tells you about what he has done, but without telling you the value of i. To reverse the change, you want to locate the entry with the minimum item, as this will be the boundary between A_{right} and A_{left}.

Design an $O(\log n)$-time algorithm to find the position of the minimum item. Show that your algorithm is correct.