



## Banburismus and the Story So Far

- WWII Codebreaking
- Double Deltas
- Machines
- Banburismus
- Tree Sorting
- Course Roadmap

Pick Up Graded Problem Sets  
Or Possibly Lose Points!



## Outline

- WWII Codebreaking
- Double Deltas
- Machines
- Banburismus
- Tree Sorting
- Course Roadmap

Pick Up Graded Problem Sets  
Or Possibly Lose Points!



## Intercepting Traffic

- Set up listening post to intercept traffic from 12 Lorenz (Fish) links
  - Different links between conquered capitals
  - Slightly different coding procedures, and different configurations
- 600 people worked on intercepting traffic



#5

## One-Slide Summary

- British codebreakers used **cribs** (guesses), brute force, and **analysis** to break the Lorenz cipher. Guessed wheel settings were likely to be correct if they resulted in a message with the right linguistic properties for German.
- If you've guessed the right wheel settings, two adjacent letters are more likely to be the same than they are to be different letters: **Double Deltas**.
- We can tell if two messages were encrypted using the same wheel settings (= same key) because the output letters will match when the input letters match. So we can try to "line them up" using **Banburismus** to look for matches.
- Tree sorting is only efficient if the trees are **balanced**. If not, it's  $\Theta(n^2)$ . The best possible sorting is  $\Theta(n \log n)$ .

#2

## Reverse Engineering Lorenz

- From the 2 intercepted messages, Col. John Tiltman worked on guessing cribs to find M1 and M2: 4000 letter messages, found 4000 letter key K1
- Bill Tutte (recent Chemistry graduate) given task of determining machine structure
  - Already knew it was 2 sets of 5 wheels and 2 wheels of unknown function
  - Six months later new machine structure likely to generate K1

#4

## Breaking WWII Traffic

- Knew machine structure, but a different initial configuration was used for each message
- Need to determine wheel setting:
  - Initial position of each of the 12 wheels
  - 1271 possible starting positions
  - Needed to try them fast enough to decrypt message while it was still strategically valuable

This is what you did for PS4 (except with fewer wheels)

#6

## Recognizing a Good Guess

- Intercepted Message (divided into 5 channels for each Baudot code bit)

$$Z_c = z_0 z_1 z_2 z_3 z_4 z_5 z_6 z_7 \dots$$

$$Z_{c,i} = M_{c,i} \oplus X_{c,i} \oplus S_{c,i}$$

Message Key (parts from S-wheels and rest)

- Look for statistical properties

- How many of the  $Z_{c,i}$ 's are 0?  $\frac{1}{2}$  (not useful)
- How many of  $(Z_{c,i+1} \oplus Z_{c,i})$  are 0?  $\frac{1}{2}$

#7

## Actual Advantage

- Probability of repeating letters

$$\text{Prob}[\Delta M_{1,i} \oplus \Delta M_{2,i} = 0] \sim 0.614$$

3.3% of German digraphs are repeating

- Probability of repeating S-keys

$$\text{Prob}[\Delta S_{1,i} \oplus \Delta S_{2,i} = 0] \sim 0.73$$

$$\text{Prob}[\Delta Z_{1,i} \oplus \Delta Z_{2,i} \oplus \Delta X_{1,i} \oplus \Delta X_{2,i} = 0]$$

$$= 0.614 * 0.73 + (1-0.614) * (1-0.73)$$

$\Delta M$  and  $S$  are 0  $\Delta M$  and  $S$  are 1

= **0.55** if the wheel settings guess is correct (0.5 otherwise)

#9

## Heath Robinson Machine

- Dec 1942: Decide to build a machine to do these  $\oplus$  quickly, due June 1943

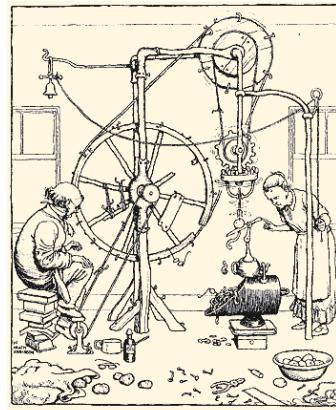
- Apr 1943: first **“Heath Robinson” machine** is delivered!

- Predecessor to Colossus

- Intercepted ciphertext on tape:

- 2000 characters per second (12 miles per hour)

- Needed to perform 7  $\oplus$  operations each  $\frac{1}{2}$  ms



Heath Robinson, British Cartoonist (1872-1944)

#11

## Double Delta

$$\Delta Z_{c,i} = Z_{c,i} \oplus Z_{c,i+1}$$

Combine two channels:

$$\Delta Z_{1,i} \oplus \Delta Z_{2,i} = \Delta M_{1,i} \oplus \Delta M_{2,i} > \frac{1}{2} \text{ Yippee!}$$

$$\oplus \Delta X_{1,i} \oplus \Delta X_{2,i} = \frac{1}{2} \text{ (key)}$$

$$\oplus \Delta S_{1,i} \oplus \Delta S_{2,i} > \frac{1}{2} \text{ Yippee!}$$

Why is  $\Delta M_{1,i} \oplus \Delta M_{2,i} > \frac{1}{2}$

Message is in German, more likely following letter is a repetition than random

Why is  $\Delta S_{1,i} \oplus \Delta S_{2,i} > \frac{1}{2}$

S-wheels only turn when M-wheel is 1

#8

## Using the Advantage

- If the guess of  $X$  is correct, should see higher than  $\frac{1}{2}$  of the double deltas are 0

- Try guessing different configurations to find highest number of 0 double deltas

### • Problem:

# of double delta operations to try one config

= length of  $Z$  \* length of  $X$

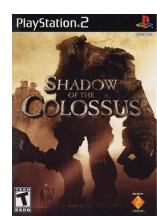
= for 10,000 letter message = 12 M for each setting  
\* 7  $\oplus$  per double delta

= 89 M  $\oplus$  operations  
(that's a lot!)



#10

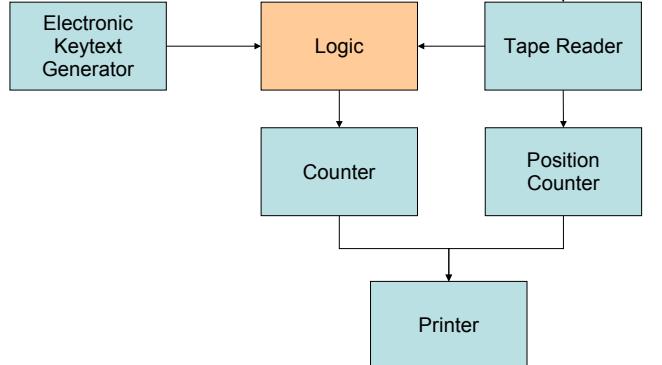
## Colossus



- Heath Robinson machines were too slow
- **Colossus** designed and first built in Jan 1944
- Replaced keytext tape loop with electronic keytext generator
- Speed up ciphertext tape:
  - 5,000 chars per second = 30 mph
  - Perform 5 double deltas simultaneously
  - Speedup = 2.5X for faster tape \* 5X for parallelism

#12

## Colossus Design



#13

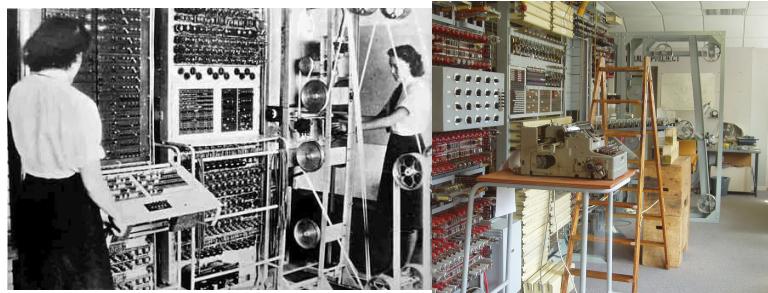
## Impact on WWII

- 10 Colossus machines operated at Bletchley park
  - Various improvements in speed
- Decoded 63 million letters in Nazi command messages
- Learned German troop locations to plan D-Day (knew the deception was working)

#14

## Colossus History

Kept secret after the war, all machines destroyed



During WWII

Rebuild, Bletchley Park,  
Summer 2004

#15

How could the folks at Bletchley Park solve a problem ~ 1 quintillion times harder than ps4?

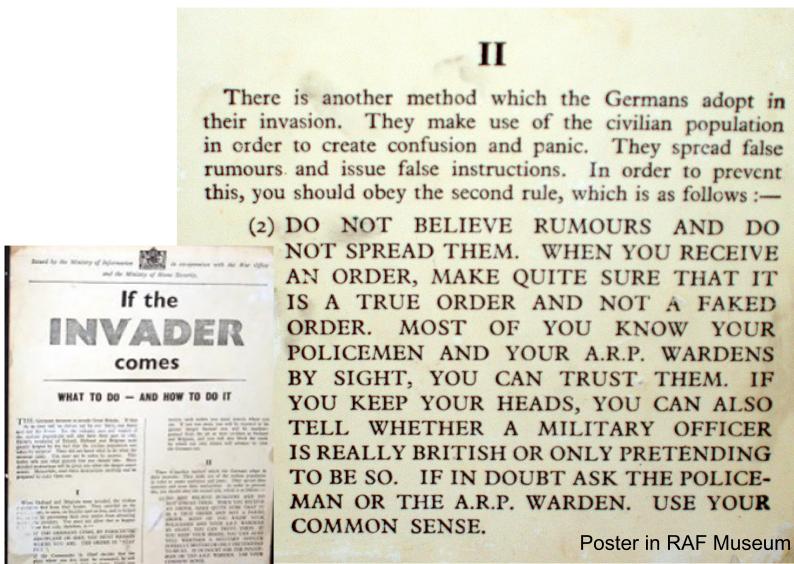
#17

## Gender and WWII

- “By the end of the war, the Testery had grown to 9 cryptographers, 24 ATS girls [sic] with a total staff of 118, organised in 3 shifts working round the clock.”
  - ATS = “Auxiliary Territorial Service”, the women’s branch of the British Army during WWII
- Example: Helen Currie joined the Auxiliary Territorial Service (ATS) in 1938, training as an intercept operator in 1942. This led to her transfer to Bletchley Park, where she joined the Tunny-breaking section called the “Testery”. Initially her job was to translate punched paper tape into German by sight; later she operated a British copy of the German Tunny machine which produced the German plaintext automatically. After the war she entered local government and served as a magistrate.



#16



Poster in RAF Museum

#18

## Motivation Helps...

Confronted with the prospect of defeat, the Allied cryptanalysts had worked night and day to penetrate German ciphers. It would appear that **fear was the main driving force**, and that adversity is one of the foundations of successful codebreaking.

Simon Singh, *The Code Book*

## Liberal Arts Trivia: Maritime Law

- A letter of marque is an official government document authorizing an agent to search, seize, or destroy specified assets or personnel belonging to a foreign party beyond the borders of the nation ("marque" or frontier). They are usually used to authorize private parties to raid and capture merchant shipping of an enemy nation. In the past, a ship operating under a letter of marque and reprisal was privately owned and was called a "private man-of-war" or ... what?

#19

#20

## Liberal Arts Trivia: Geography

- This capital city of Uttar Pradesh, the most populous state of India, is popularly known as the The City of Nawabs. It is also known as the Golden City of the East, Shiraz-i-Hind and The Constantinople of India. It is a center of Hindi and Urdu literature, and the birthplace of Kathak, a classic Indian dance form. The city was besieged during the Indian Rebellion of 1857.

#21

## Banburismus

Given two Enigma-encrypted messages, how can we determine if they were encrypted starting with the same wheel settings?



Enigma in Use,  
10 December 1943

#22

## Enigma



- Invented commercially, 1923
- German Navy, Army, Air Force
- About 50,000 in use (many were captured by Allies)
- Modified throughout WWII, Germans believed perfectly secure
- Kahn's *Codebreakers* (1967) didn't know it was broken
- Turing's 1940 Treatise on Enigma declassified in 1996

Enigma machine at Bletchley Park



## Reverse Engineering Enigma

"This fictional movie about a fictional U.S. submarine mission is followed by a mention in the end credits of those actual British missions. Oh, the British deciphered the Enigma code, too. Come to think of it, they pretty much did everything in real life that the Americans do in this movie."

Roger Ebert's review of *U-571*  
(2000 Academy Award Winner)

#23

#24

## Simple Substitution Ciphers

ABCDEFGHIJKLMNOPQRSTUVWXYZ



JIDKQACRSHLGNFEXUZVTPMYOB

HELLO  $\rightarrow$  RQGGF

#25

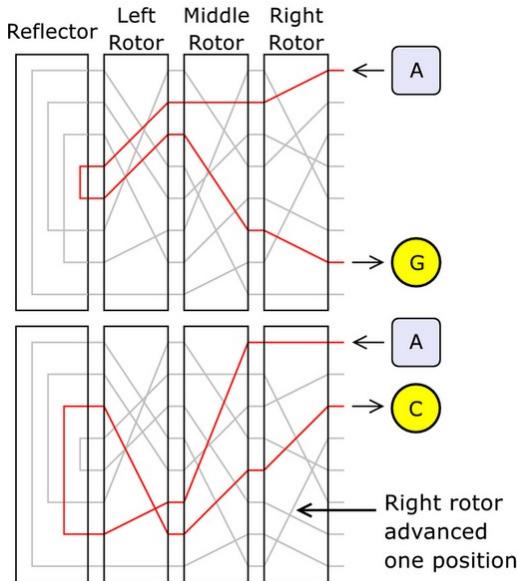
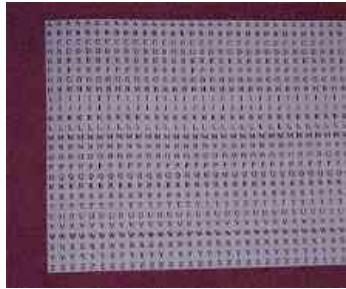


Image from <http://en.wikipedia.org/wik/Image:Enigma-action.png>

#27

## Alan Turing's Solution



M1: GXCYBGDSLWBDJLKWIPEHVYQZWDTHRQXIKEESQS

M2: YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCQAXVXDVU

#29

## Rotor Wheels



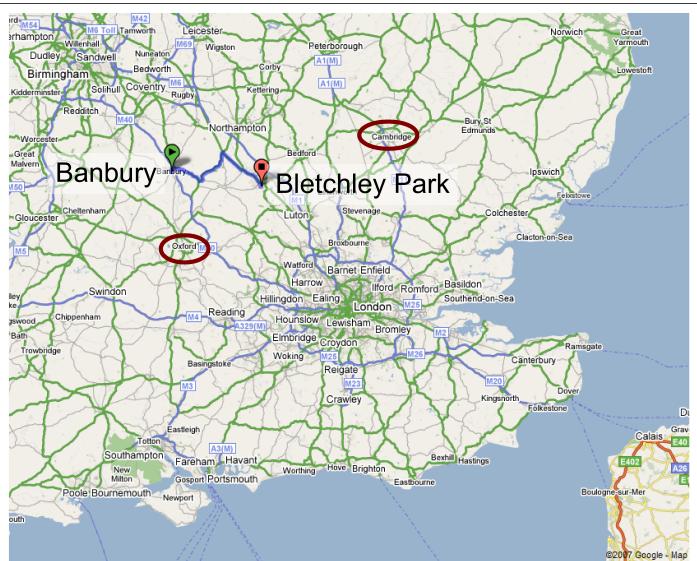
Simple substitution

Latch turns next rotor once per rotation

#26

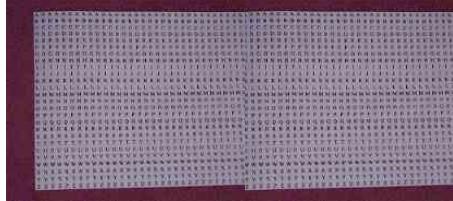
## Language is Non-Random

- Random strings: the probability of two letters in the two messages matching is 1/26 (number of letters in alphabet)
- Same-encrypted strings: the output letters will match when the input letters match
  - This happens much more frequently because some letters (e.g., "e" is ~13% of all English letters) are more common

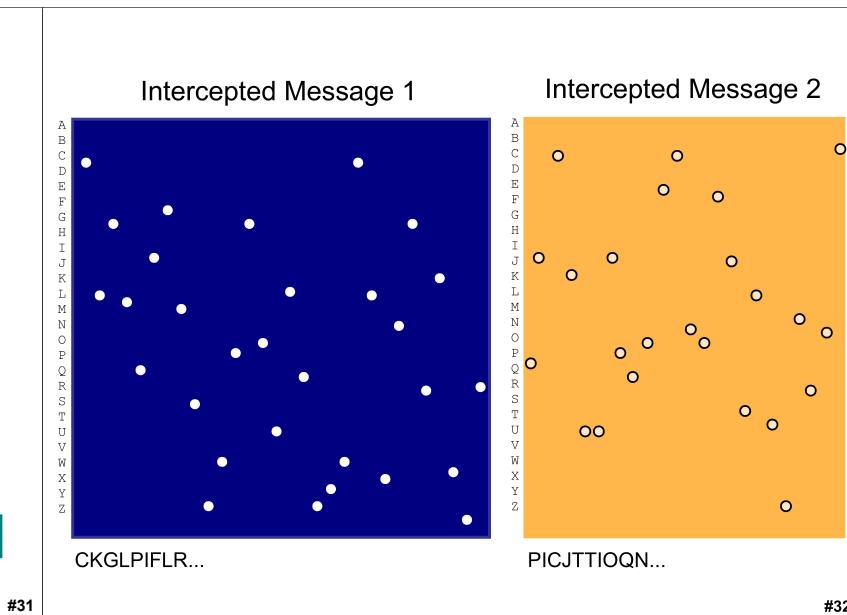


#30

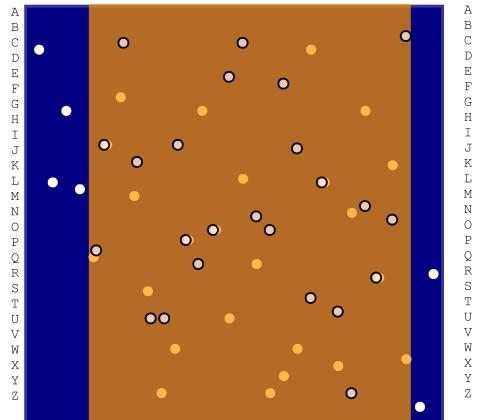
## Banburismus



M1: **GXCYBGDSLWBDJLKWIPEHVGQZWDTHRQXIKEESQS**  
M2: **YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU**



### Intercepted Message 1



### Trying Possible Alignments

GXCYBGDSLWBDJLKWIPEHVGQZWDTHRQXIKEESQS  
YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU  
YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU  
YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU  
...  
YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU

#33

#34

### Trying Possible Alignments

GXCYBGDSLWBDJLKWIPEHVGQZWDTHRQXIKEESQS  
YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU  
YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU  
YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU  
...  
YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU

### Trying Possible Alignments

GXCYBGDSLWBDJLKWIPEHVGQZWDTHRQXIKEESQS  
YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU  
YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU  
YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU  
...  
YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU

#35

#36

## Trying Possible Alignments

```
GXCYBGDSLWBDJLKWIPEHVYGOZWDTROXIKEESQS
YNSCFCCPVIPEMSGIZWFIHESCIYSPVRXMCFQAXVXDVU
YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU
YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU
...
YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU..
```

#37



Turing's Hut 8 at Bletchley Park

Don't complain about your working space.  
You can do good computer science anywhere.  
But find a quiet, undisturbed place to work on the exam.

#38

## Liberal Arts Trivia: Geology

- A stratovolcano or composite volcano is a tall, conical volcano made of many layers of lava, tephra and volcanic ash: they are characterized by steep sides and periodic eruptions. They are common in subduction zones where the ocean crust is drawn under the continental crust. Mount St. Helens and Mount Fuji are both stratovolcanos: name the country containing each one.

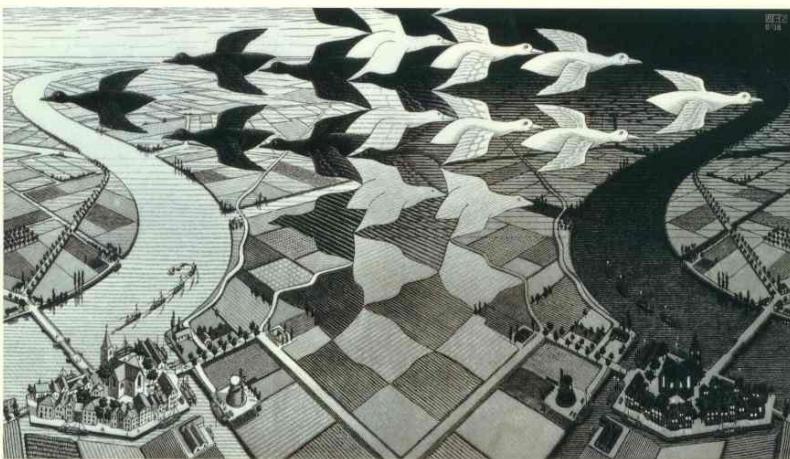
#39

## Liberal Arts Trivia: Mythology

- In Egyptian mythology, this falcon-headed son of Isis and Osiris fought with Seth for the throne of Egypt. In the battle his eye was wounded and later healed by Isis; this became an important symbol for renewal. He united Egypt and bestowed divinity on the pharaohs (who were viewed as his living incarnations). Name this sun, sky and war god, shown here in hieroglyphs:



#41



## The Story So Far

#40

## insert-one-tree

```
def insert_one_tree(cf, new_elt, tree):
    if not tree:
        return make_tree([], new_elt, [])
    element_here = get_element(tree)
    if cf(new_elt, element_here):
        return make_tree(
            insert_one_tree(cf, new_elt, get_left(tree)),
            element_here, get_right(tree))
    else:
        return make_tree(
            get_left(tree), element_here,
            insert_one_tree(cf, new_elt, get_right(tree)))
```

Each time we call insert-one-tree, the size of the tree approximately **halves** (if it is well balanced).

Each application is constant time.

The running time of insert-one-tree is in  $\Theta(\log n)$  where  $n$  is the number of elements in the input tree, which must be well-balanced.

#43

## quicker-insert-one

```
def quicker_insert_one(cf, lst):
    if not lst: return []
    return insert_one_tree(cf,
        lst[0],
        quicker_insert_one(cf, lst[1:]))
```

No change (other than using insert\_one\_tree)...but evaluates to a tree not a list!

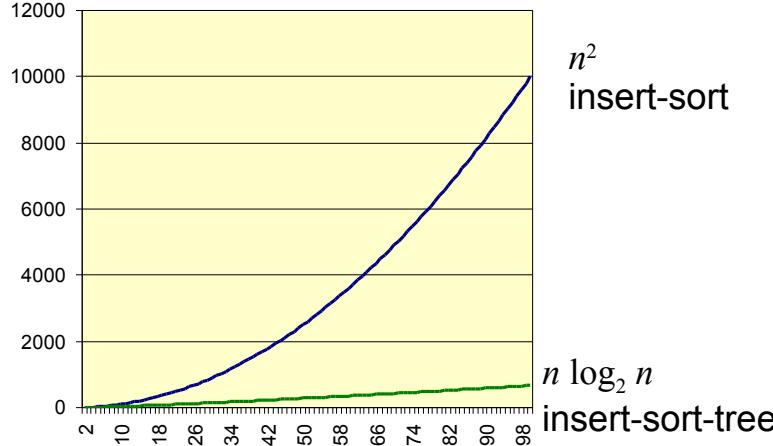
#44

## extract\_elements

We need to make a list of all the tree elements, from left to right.

```
def extract_elements(tree):
    if tree == []: return []
    return extract_elements(get_left(tree)) +
        [get_element(tree)] +
        extract_elements(get_right(tree))
```

Growth of time to sort random list



#45

## Running time of insert-sort-tree

```
def insert_one_tree(cf, new_elt, tree):
    if not tree:
        return make_tree([], new_elt, [])
    element_here = get_element(tree)
    if cf(new_elt, element_here):
        return make_tree(
            insert_one_tree(cf, new_elt, get_left(tree)),
            element_here, get_right(tree))
    else:
        return make_tree(
            get_left(tree), element_here,
            insert_one_tree(cf, new_elt, get_right(tree)))
```

$n$  = number of elements in tree

$\Theta(\log n)$

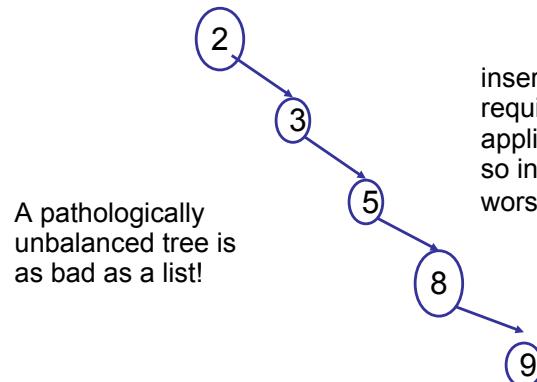
```
def insert_sort_tree(cf, lst):
    def quicker_insert_one(cf, lst):
        if not lst: return []
        return insert_one_tree(cf,
            lst[0],
            quicker_insert_one(cf, lst[1:]))
    return extract_elements(quicker_insert_one(cf, lst))
```

$n$  = number of elements in lst

$\Theta(n \log n)$

#46

## What if tree is not well-balanced?



insert-one worst case requires  $n$  recursive applications, so insert-sort-tree worst case is in  $\Theta(n^2)$

#47

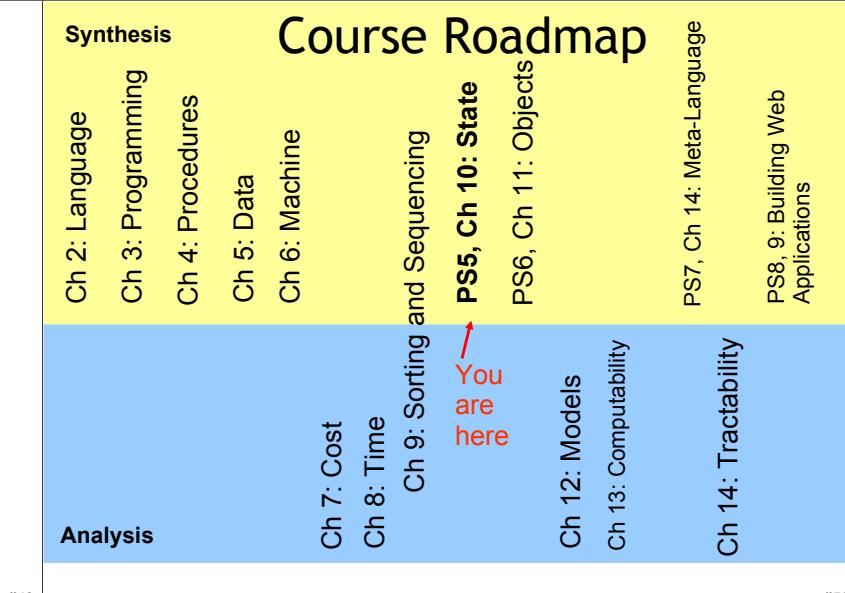
#48

## Can we do better?

- Making all those trees is a lot of work
- Can we divide the problem in two halves, without making trees?

This is the famous “**Quicksort**” algorithm invented by Sir Tony Hoare. See Course Book.

There are lots of ways to do a little bit better, but **no** way to do asymptotically better. **All** possible sort procedure have running times in  $\Omega(n \log n)$ . (We'll explain why later in the course...)



#49

#50

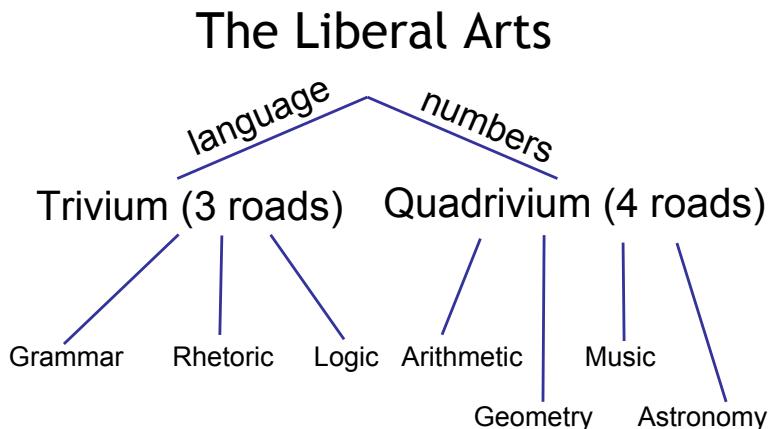
## Computer Science 1120 so far

- How to describe **information processes** by defining **procedures**
  - Programming with procedures, lists, **recursion**
  - Chapters 3, 4, 5
- How to **predict properties** about information processes
  - Predicting running time,  $\Theta$ ,  $O$ ,  $\Omega$
- How to elegantly and **efficiently implement** information processes
  - Chapter 3 (rules of evaluation)
  - Chapter 6 (machines)

#51

#52

From Lecture 1:



#53

#54

## CS1120 upcoming

- How to describe information processes by defining procedures
  - Programming with state, objects, networks
  - You will learn Java as well as Python
- How to predict properties about information processes
  - What is the fastest process that can solve a given problem?
  - Are there problems which can't be solved by algorithms?
- How to elegantly and efficiently implement information processes
  - How to implement a Python interpreter

#52

#53

## Liberal Arts Checkup

- Grammar: study of meaning in written expression
  - BNF replacement rules for describing languages, rules of evaluation for meaning
- Rhetoric: comprehension of verbal and written discourse
  - Not much yet... interfaces between components (PS6-9), program and user (PS8-9)
- Logic: argumentative discourse for discovering truth
  - Rules of evaluation, if, recursive definitions
- Arithmetic: understanding numbers
  - Not much yet... wait until last third
- Geometry: quantification of space
  - Curves as procedures, fractals (PS3)
- Music: number in time
  - Yes, listen to “Hey Jude!”
- Astronomy

Read Neil deGrasse Tyson's essay

# Homework

- **Exam 1 Out**

#55