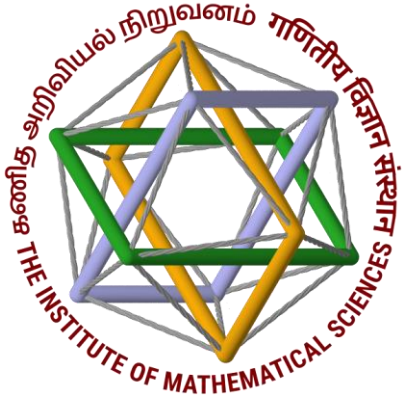
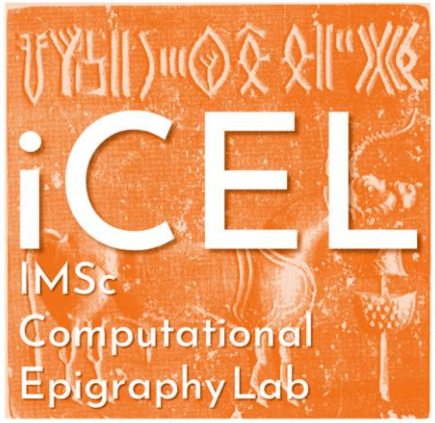



Bits & Scripts




Tutorial:

# Cipher & Decipher

Sitabhra Sinha

IMSc, Chennai

A cipher is an algorithm – a series of well-defined steps that can be followed as a procedure – for performing encryption or decryption *Kryptós,* Gr. "hidden, secret"

Image: mathdiscoveries.wordpress.com

Ché Guevara one-time pad cipher

# Codes vs. Ciphers

- Codes generally substitute different length strings of characters in the output.
  A code maps Imageone meaning with another.
  Words and phrases can be coded as letters or numbers.
  operates by substituting using a codebook which links random strings of characters or numbers to a word or phrase.
  For example, "QQQQ" → "Get out of the auditorium fast"

- Ciphers generally substitute the same number of characters as are input.
  The given input must follow the cipher's algorithmic process to be solved.
  The ciphertext message contains all the information of the plaintext message, but is not in a format readable by a human or computer without the proper mechanism to decrypt it.

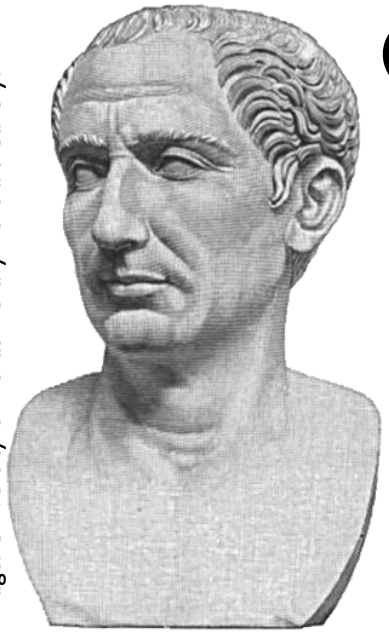


Image: University of Texas Library Portrait Gallery.

# Caesar cipher

A substitution cipher in which each letter in the plaintext is replaced by a letter some fixed number of positions down the alphabet.

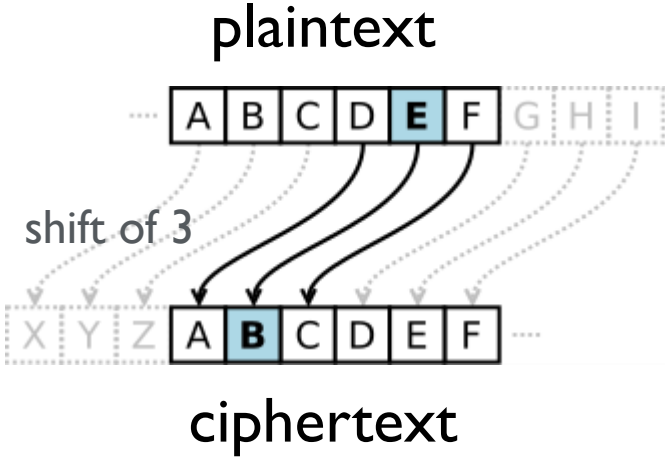
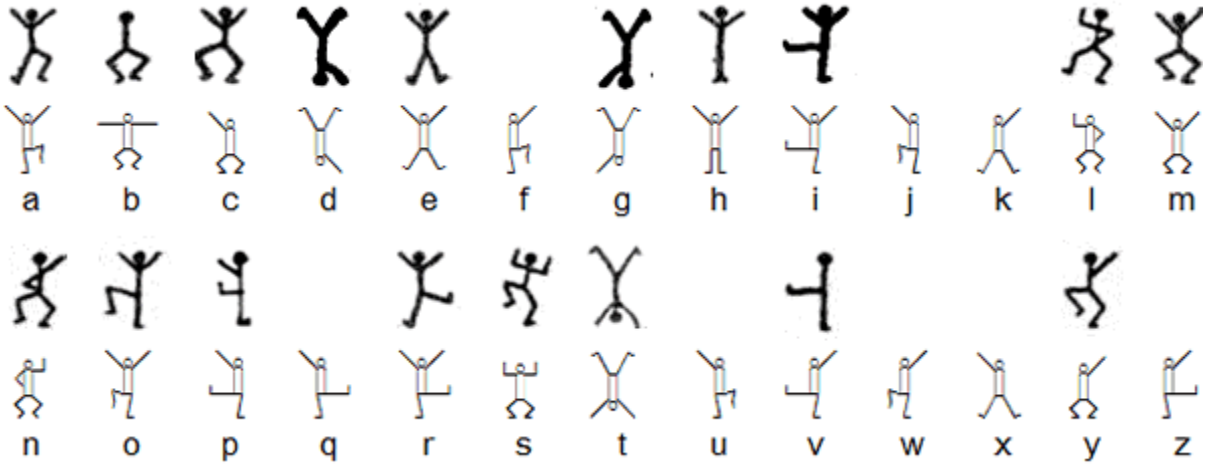



Image: Matt_Crypto, Cepheus

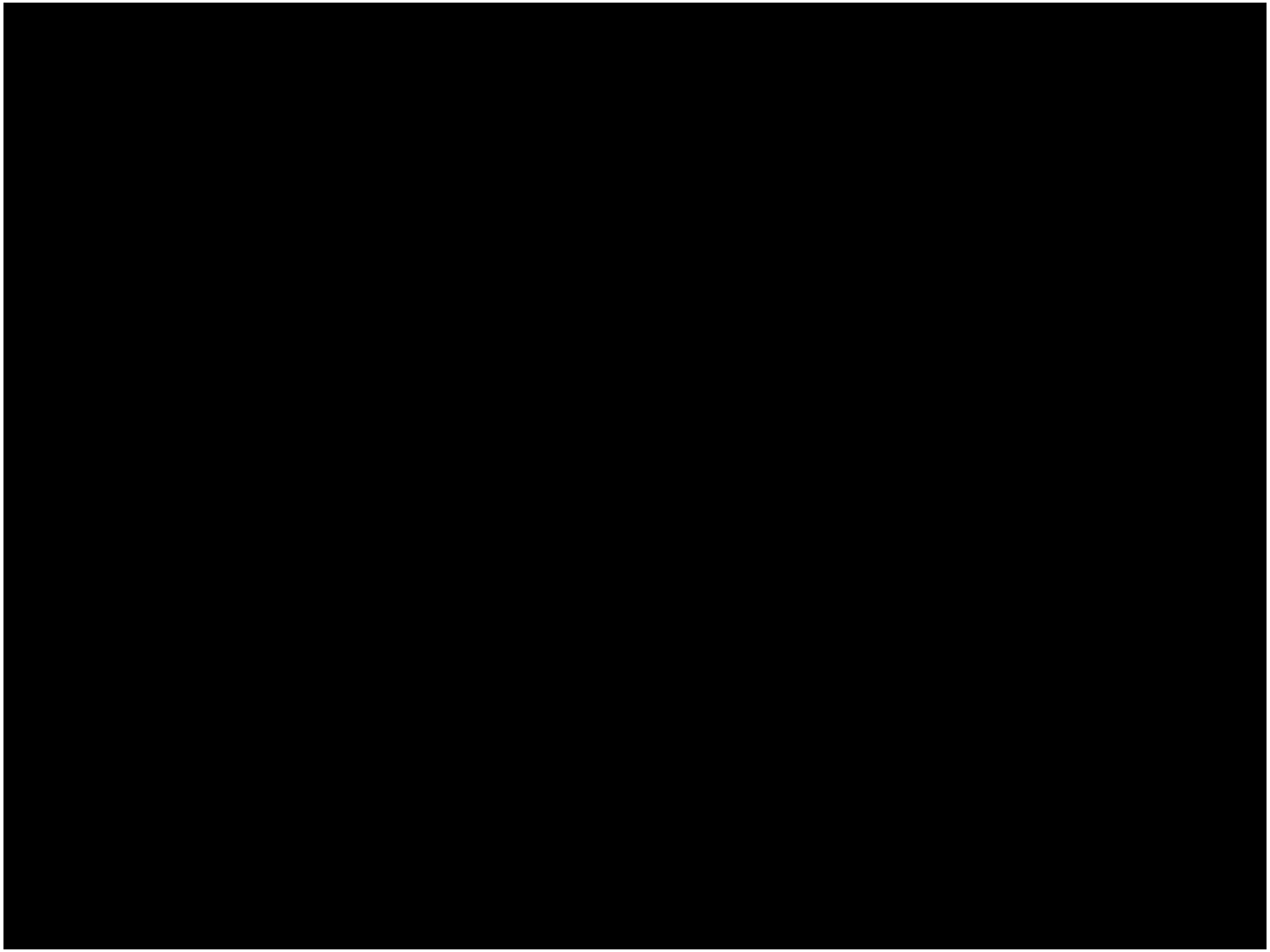
Plaintext: THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG
Ciphertext: QEB NRFZH YOLTK CLU GRJMP LSBO QEB IXWV ALD

One can also use another set of symbols to replace each letter of the alphabet




https://kid-celephais.fandom.com/wiki/Dancing_Men_Code

Arthur Conan Doyle,
*The Adventure of the Dancing Men*, in Return of SherlockHolmes (1903)



https://www.dailymotion.com/video/x6cnebx

Granada Television 1984

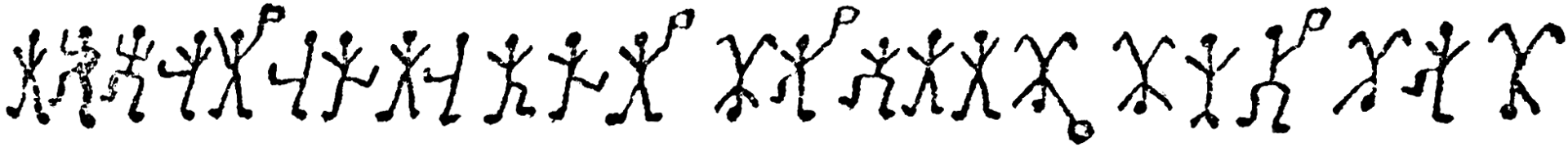
Criminal's (Abe Slaney staying at Elrige's Farm) first message

Criminal's second message

Criminal's third message

Elsie's response

Criminal's final message

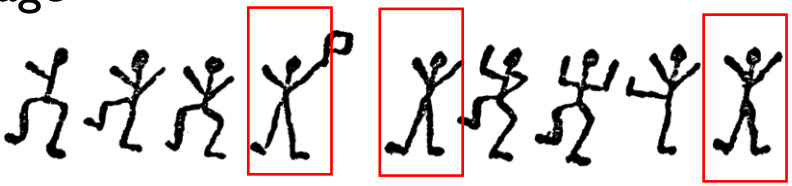


Sherlock Holmes' message to the criminal

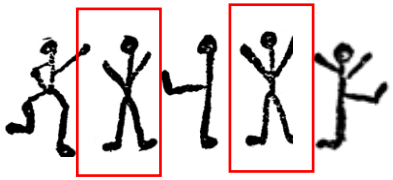
Criminal's (Abe Slaney staying at Elrige's Farm) first message

Criminal's second message

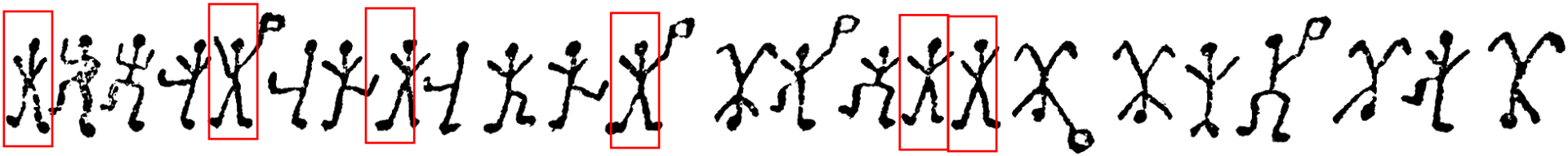
Criminal's third message



Elsie's response

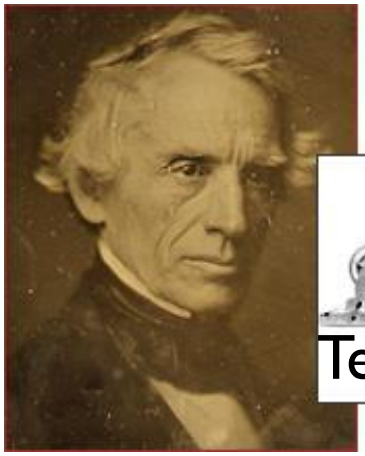


Criminal's final message

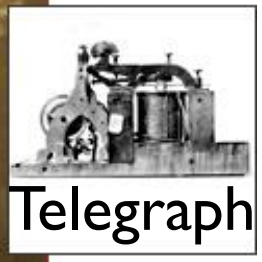


Sherlock Holmes' message to the criminal

# Frequency of English Characters



Samuel Morse
1791-1872



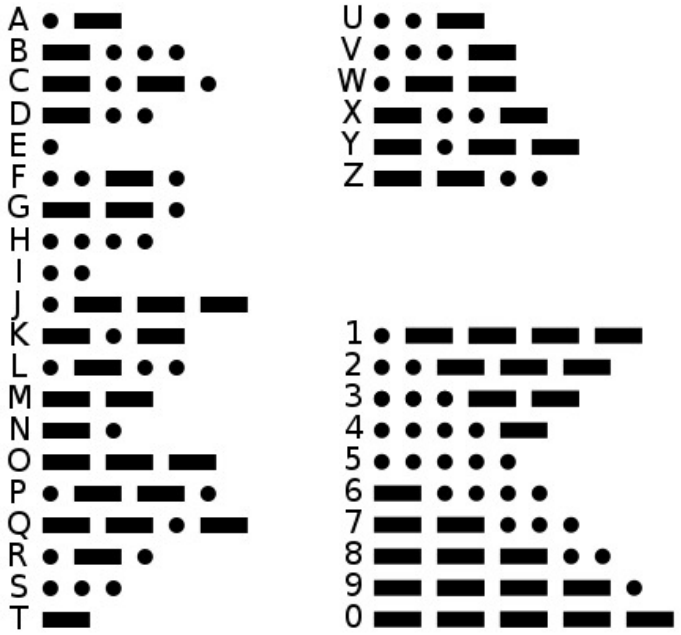
Telegraph

Highest frequency characters **ETAOIN SHRDLU**

To increase the efficiency of encoding, Morse code was originally designed so that the length of each symbol is approximately inverse to the frequency of occurrence of the character that it represents in text of the English language.

But Morse's character frequency counts are based on pieces of text considers the commonest words multiple times

## Morse code




If one considers all distinct words so that word frequency does not produce artefacts, e.g., by considering words from a dictionary, the highest frequency characters are

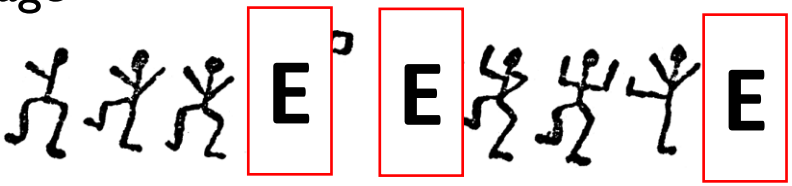
**EARIOT NSLCUD**

In either case, E is the commonest letter

Criminal's (Abe Slaney staying at Elrige's Farm) first message

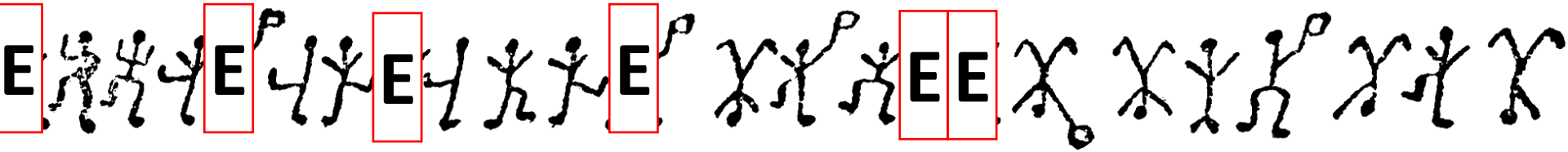

Criminal's second message


Criminal's third message




Elsie's response


Criminal's final message

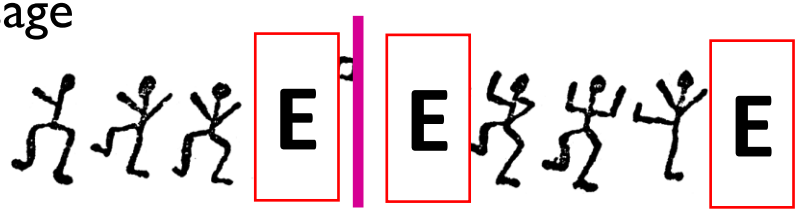



Sherlock Holmes' message to the criminal

Criminal's (Abe Slaney staying at Elrige's Farm) first message


Criminal's second message


Criminal's third message




Elsie's response


There are no word boundaries – could the flags indicate word endings ?

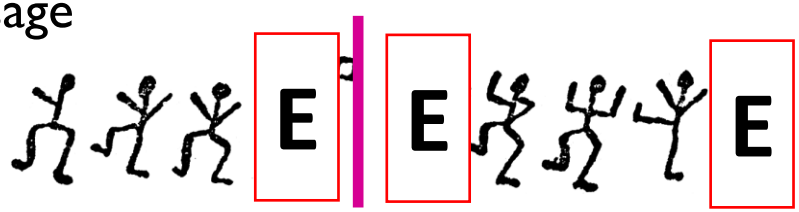
Criminal's final message


Sherlock Holmes' message to the criminal

Criminal's (Abe Slaney staying at Elrige's Farm) first message


Criminal's second message


Criminal's third message




Elsie's response


No flags – single word!

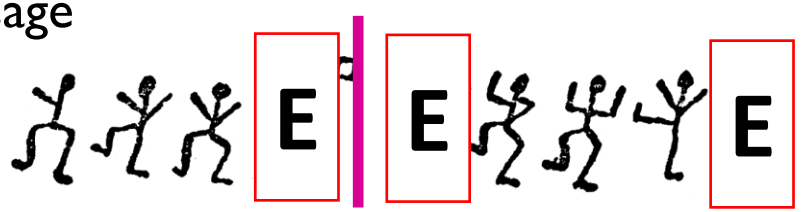
Criminal's final message


Sherlock Holmes' message to the criminal

Criminal's (Abe Slaney staying at Elrige's Farm) first message


Criminal's second message


Criminal's third message




Elsie's response


Possible English words of the form * E * E * are

SEVER
LEVER
NEVER

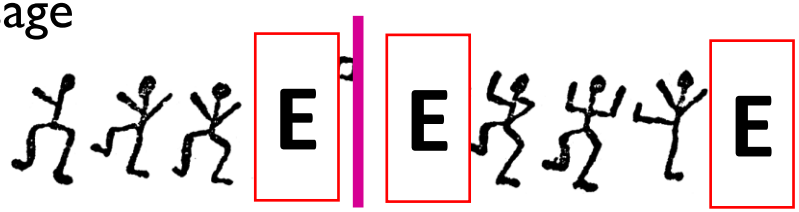
Criminal's final message


Sherlock Holmes' message to the criminal

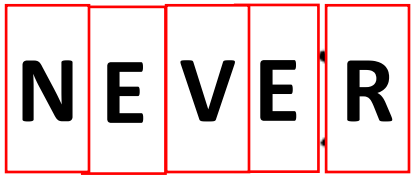
Criminal's (Abe Slaney staying at Elrige's Farm) first message


Criminal's second message


Criminal's third message




Elsie's response

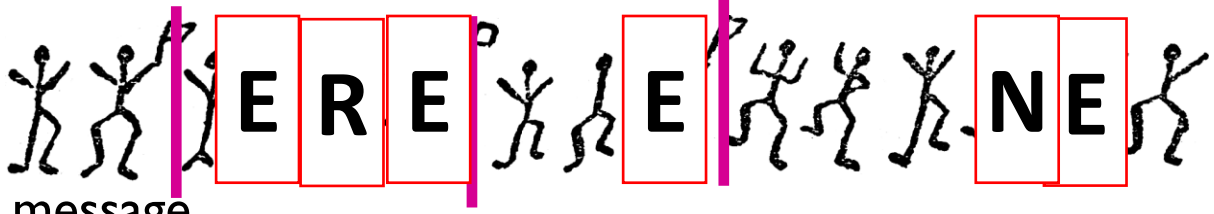



Assuming that this is a conversation between two people, very likely that it is NEVER

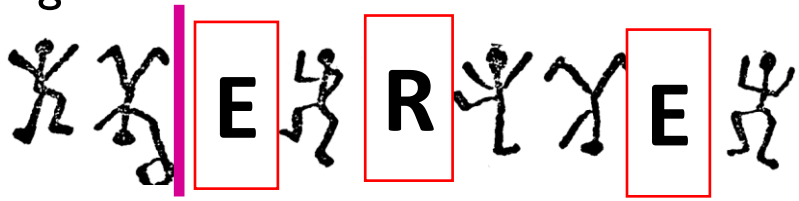
Criminal's final message


Sherlock Holmes' message to the criminal

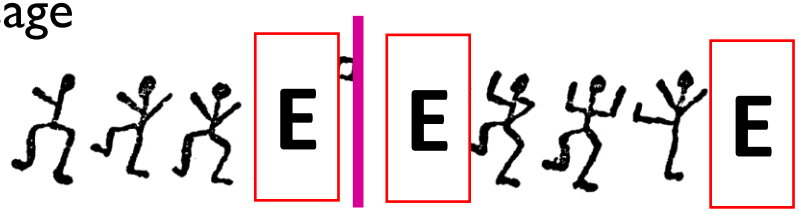
Criminal's (Abe Slaney staying at Elrige's Farm) first message



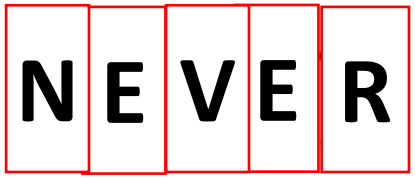

Criminal's second message




Criminal's third message




Elsie's response




Gives symbols for N, V, R

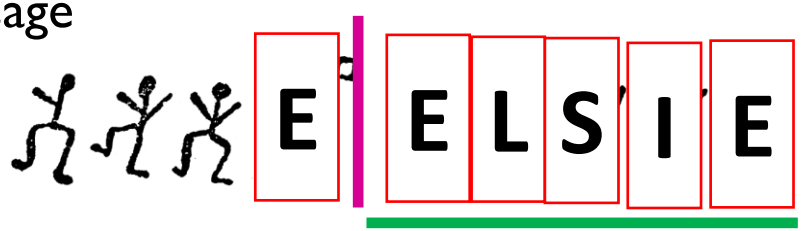
Criminal's final message


Sherlock Holmes' message to the criminal

Criminal's (Abe Slaney staying at Elrige's Farm) first message


Criminal's second message


Criminal's third message




E * * * E word occurs twice – is this a name ? ELSIE?

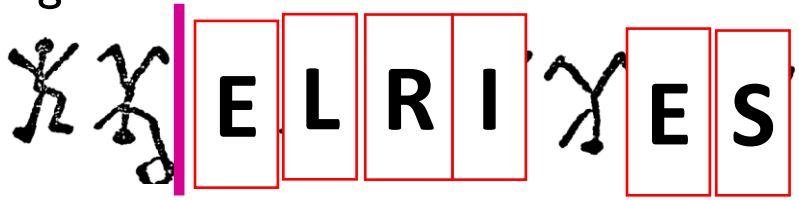
Elsie's response


Criminal's final message

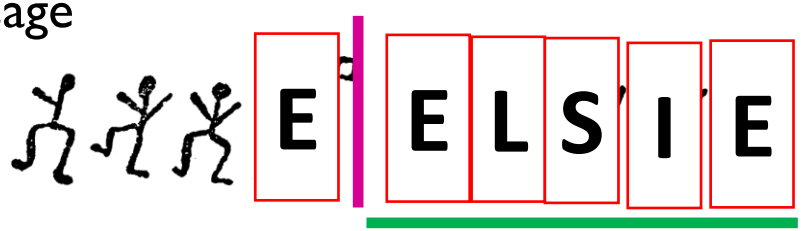

Sherlock Holmes' message to the criminal

Criminal's (Abe Slaney staying at Elrige's Farm) first message


Criminal's second message




Criminal's third message



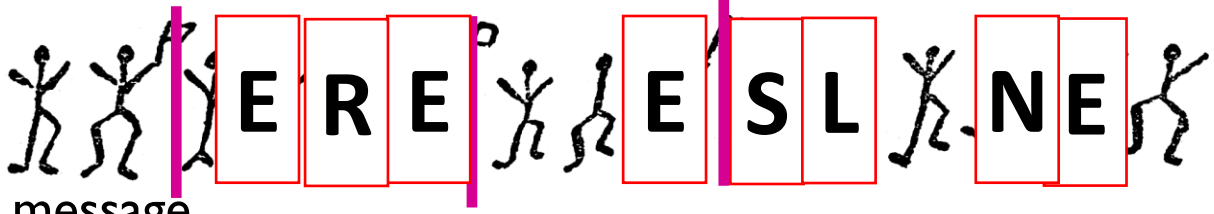

Elsie's response


Gives symbols for L, S, I

Criminal's final message

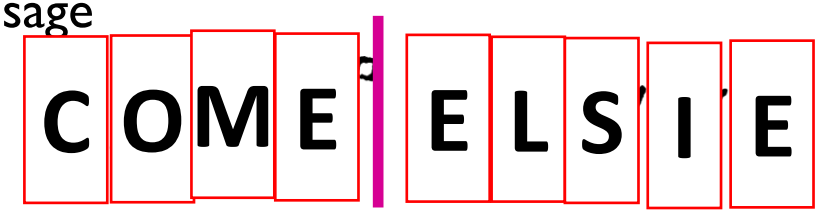

Sherlock Holmes' message to the criminal

Criminal's (Abe Slaney staying at Elrige's Farm) first message



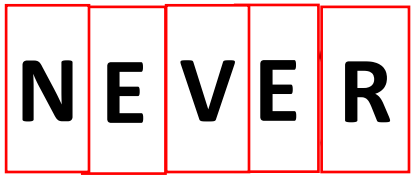

Criminal's second message


Criminal's third message




* * * E may be an instruction.
COME ?

Elsie's response



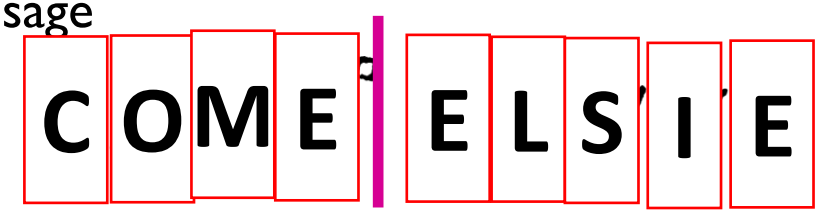

Criminal's final message


Sherlock Holmes' message to the criminal

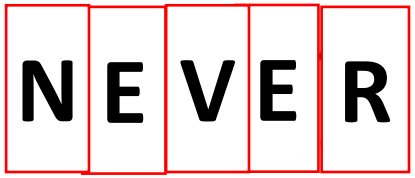
Criminal's (Abe Slaney staying at Elrige's Farm) first message


Criminal's second message


Criminal's third message




Elsie's response




Gives symbols for C, O, M

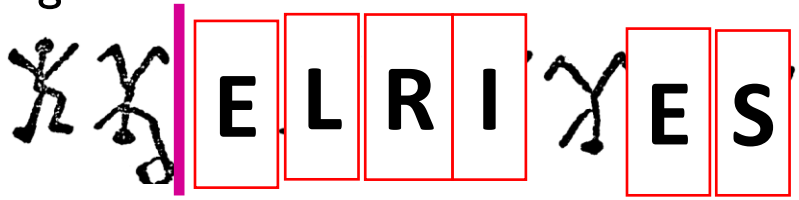
Criminal's final message


Sherlock Holmes' message to the criminal

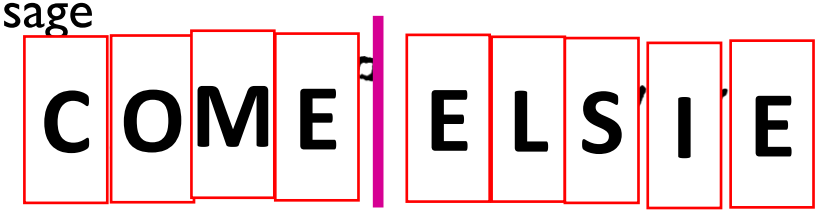
Criminal's (Abe Slaney staying at Elrige's Farm) first message

"AM HERE" ?

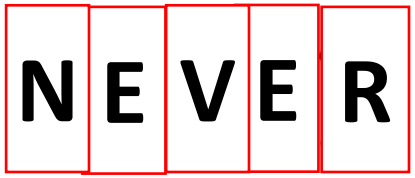

Criminal's second message




Criminal's third message




Elsie's response



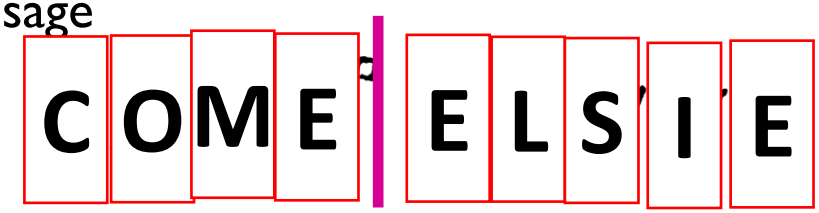

Criminal's final message


Sherlock Holmes' message to the criminal

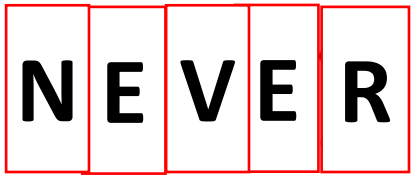
Criminal's (Abe Slaney staying at Elrige's Farm) first message


Criminal's second message


Criminal's third message




Elsie's response




Gives symbols for A, H

Criminal's final message

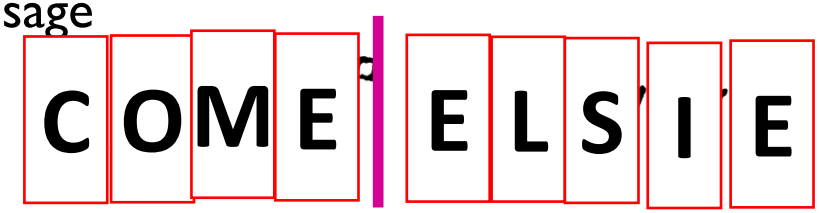

Sherlock Holmes' message to the criminal

Criminal's (Abe Slaney staying at Elrige's Farm) first message


Criminal's second message

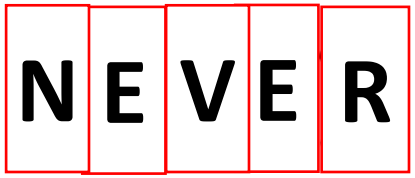

A* ELRI*ES
A* = AT ?

Criminal's third message

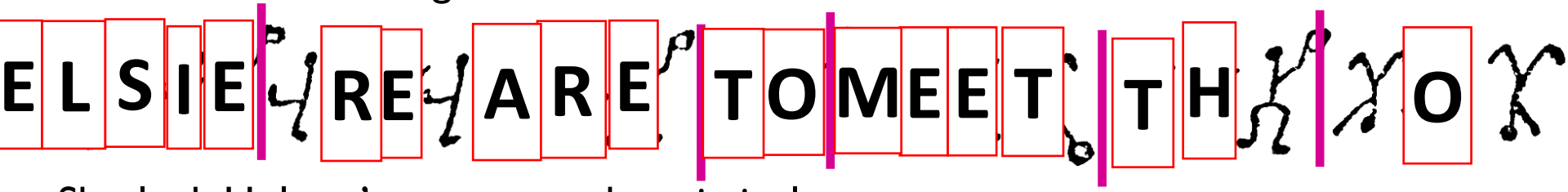



Gives T

Elsie's response




Criminal's final message

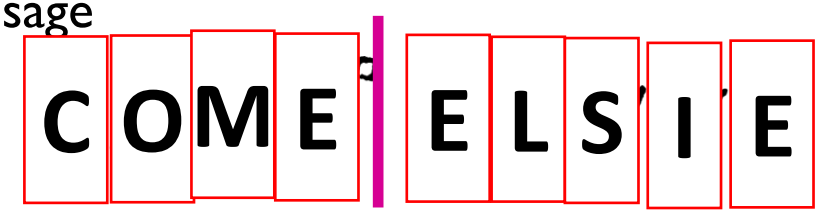



Sherlock Holmes' message to the criminal

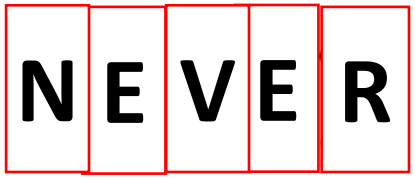
Criminal's (Abe Slaney staying at Elrige's Farm) first message


Criminal's second message


Criminal's third message

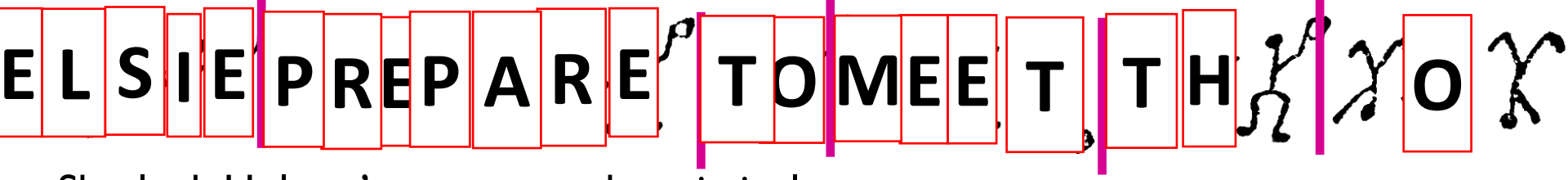



Elsie's response




* R E * A R E
PREPARE ?
Gives P

Criminal's final message



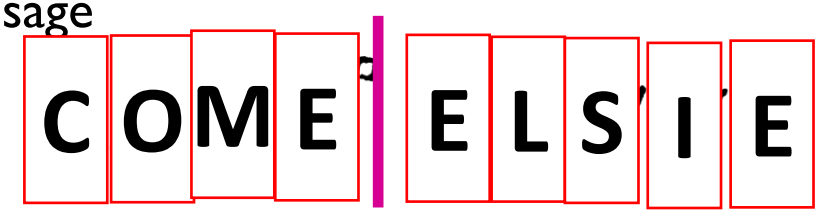

Sherlock Holmes' message to the criminal

Criminal's (Abe Slaney staying at Elrige's Farm) first message

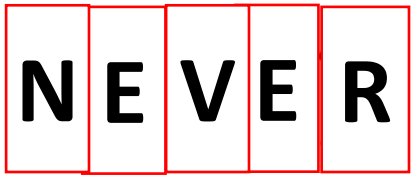

Criminal's second message


ELRI*ES
ELRIGES ?
Gives G

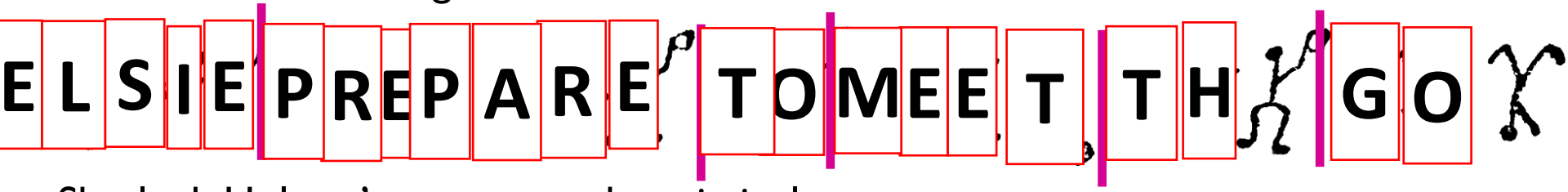
Criminal's third message




Elsie's response




Criminal's final message




Sherlock Holmes' message to the criminal

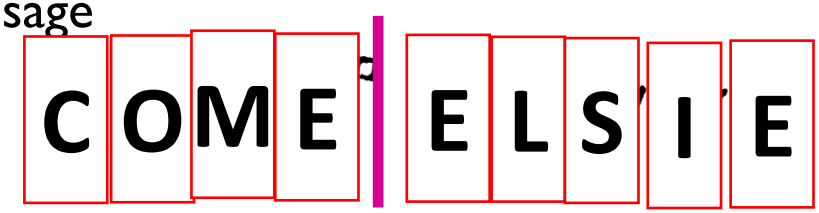
Criminal's (Abe Slaney staying at Elrige's Farm) first message


Criminal's second message

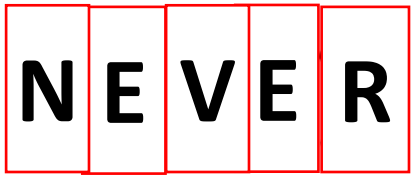

SLANE* & TH*

SLANEY & THY
Or
SLANEE & THE?

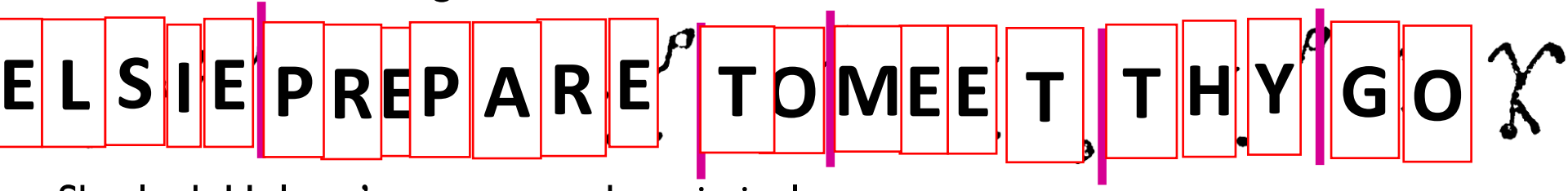
Criminal's third message




Elsie's response




Criminal's final message

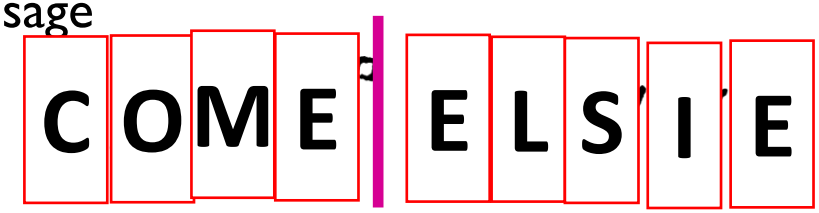



Sherlock Holmes' message to the criminal

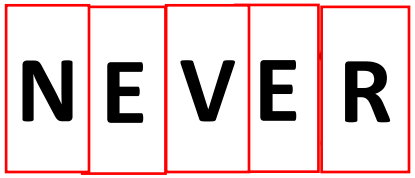
Criminal's (Abe Slaney staying at Elrige's Farm) first message


Criminal's second message


Criminal's third message



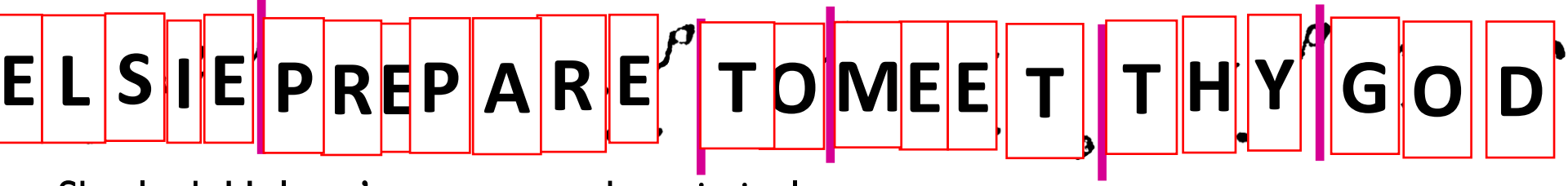

Elsie's response




GO*

GOD?

Criminal's final message




Sherlock Holmes' message to the criminal

Criminal's (Abe Slaney staying at Elrige's Farm) first message

AMHEREABESLANEY

Criminal's second message

ATELRIGES

A*E is a first name

ABE?

Criminal's third message

COMEELSIE

Elsie's response

NEVER

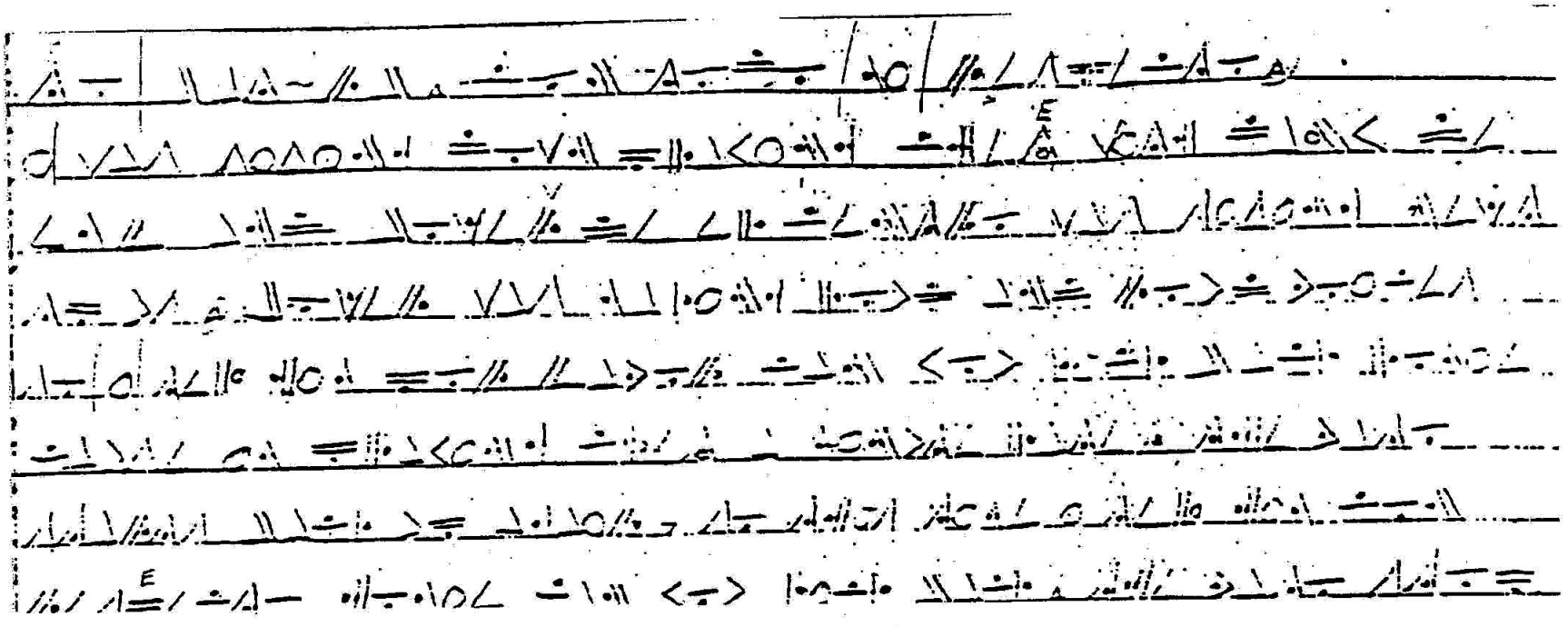
Criminal's final message

ELSIEPREPARETOMEETTHYGOD

Sherlock Holmes' message to the criminal

# Substitution ciphers in real life

A typical example of coded messages found in prisons in California sent to Stanford's Statistics Department for decryption



How to find the unknown function f such f : {code space} → {English alphabet}

# Markov chain Monte Carlo

A standard approach to deciphering is to use the statistics of written English to guess at probable choices for f, try these out, and see if the decrypted messages make sense.

The statistics is obtained from a text corpus. The proportion of consecutive text symbols from x to y gives a matrix M(x; y) of 1st order transitions from one symbol to another..
Then the plausibility of the mapping f

$$Pl(f) = \Pi_i M(f(s_i), f(s_{i+1}))$$

where $s_i$ runs over consecutive symbols in the coded message.

Mappings f which have high values of Pl(f) are good candidates for deciphering.
To maximize the plausibility we can run the following Markov Chain Monte Carlo algorithm

- Start with a preliminary guess, say $f$.
- Compute $\mathrm{Pl}(f)$.
- Change to $f_*$ by making a random transposition of the values $f$ assigns to two symbols.
- Compute $\mathrm{Pl}(f_*)$; if this is larger than $\mathrm{Pl}(f)$, accept $f_*$.
- If not, flip a $\mathrm{Pl}(f_*)/\mathrm{Pl}(f)$ coin; if it comes up heads, accept $f_*$.
- If the coin toss comes up tails, stay at $f$.

# “Reps do it”

The space of possible f s is extremely large – can the Monte Carlo guided random walk achieve the “correct” f in reasonable time, or at all?

```
ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS
```

The text is scrambled at random and the Monte Carlo algorithm run

```
 100 ER ENOHDLAE OHDLO UOZEOUNORU O UOZEO HD OITO HEOQSET IUROFHE HENO ITORUZAEN
 200 ES ELOHRNDE OHRNO UOVEOULOSU O UOVEO HR OITO HEOQAET IUSOPHE HELO ITOSUVDEL
 300 ES ELOHANDE OHANO UOVEOULOSU O UOVEO HA OITO HEOQRET IUSOFHE HELO ITOSUVDEL
 400 ES ELOHINME OHINO UOVEOULOSU O UOVEO HI OATO HEOQRET AUSOWHE HELO ATOSUVMEL
 500 ES ELOHINME OHINO UODEOULOSU O UODEO HI OATO HEOQRET AUSOWHE HELO ATOSUDMEL
 600 ES ELOHINME OHINO UODEOULOSU O UODEO HI OATO HEOQRET AUSOWHE HELO ATOSUDMEL
 900 ES ELOHANME OHANO UODEOULOSU O UODEO HA OITO HEOQRET IUSOWHE HELO ITOSUDMEL
1000 IS ILOHANMI OHANO RODIORLOSR O RODIO HA OETO HIOQUIT ERSOWHI HILO ETOSRDMIL
1100 ISTILOHANMITOHANOT ODIO LOS TOT ODIOTHATOEROTHIOQUIRTE SOWHITHILOTEROS DMIL
1200 ISTILOHANMITOHANOT ODIO LOS TOT ODIOTHATOEROTHIOQUIRTE SOWHITHILOTEROS DMIL
1300 ISTILOHARMITOHAROT ODIO LOS TOT ODIOTHATOENOTHIOQUINTE SOWHITHILOTENOS DMIL
1400 ISTILOHAMRITOHAMOT OFIO LOS TOT OFIOTHATOENOTHIOQUINTE SOWHITHILOTENOS FRIL
1600 ESTEL HAMRET HAM TO CE OL SOT TO CE THAT IN THE QUENTIOS WHETHEL TIN SOCREL
1700 ESTEL HAMRET HAM TO BE OL SOT TO BE THAT IN THE QUENTIOS WHETHEL TIN SOBREL
1800 ESTER HAMLET HAM TO BE OR SOT TO BE THAT IN THE QUENTIOS WHETHER TIN SOBLER
1900 ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS NOBLER
2000 ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS NOBLER
```

The simple optimization process converges to original text in a few thousand steps

# Substitution ciphers in real life

Decipherment of the typical example of coded messages found in prisons in California sent to Stanford's Statistics Department for decryption

to bat-rb. con todo mi respeto. i was sitting down playing chess with danny de emf and boxer de el centro was sitting next to us. boxer was making loud and loud voices so i tell him por favor can you kick back homie cause im playing chess a minute later the vato starts back up again so this time i tell him con respecto homie can you kick back. the vato stop for a minute and he starts up again so i tell him check this out shut the f**k up cause im tired of your voice and if you got a problem with it we can go to celda and handle it. i really felt disrespected thats why i told him. anyways after i tell him that the next thing I know that vato slashes me and leaves. dy the time i figure im hit i try to get away but the c.o. is walking in my direction and he gets me right dy a celda. so i go to the hole. when im in the hole my home boys hit doxer so now "b" is also in the hole. while im in the hole im getting schoold wrong and

# The Voynich Manuscript (c.15th century ?)

an illustrated codex, hand-written in an unknown script – language unknown

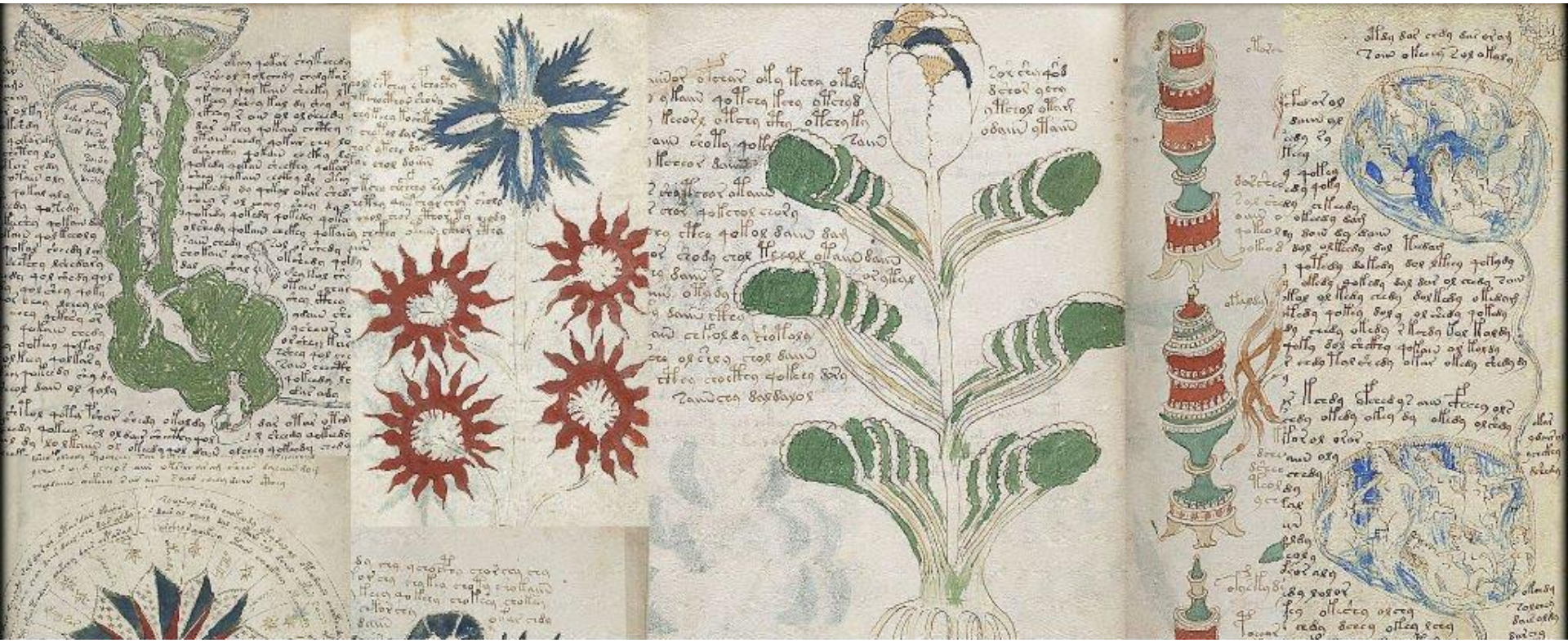


Image: Beinecke Rare Book & Manuscript Library

The vellum has been carbon-dated to 15th century

named after Wilfrid Voynich, a Polish book dealer who acuired it in 1912

Hypotheses range from a script for a natural language or constructed language, an unread code, cypher, or other form of cryptography, or perhaps a hoax!