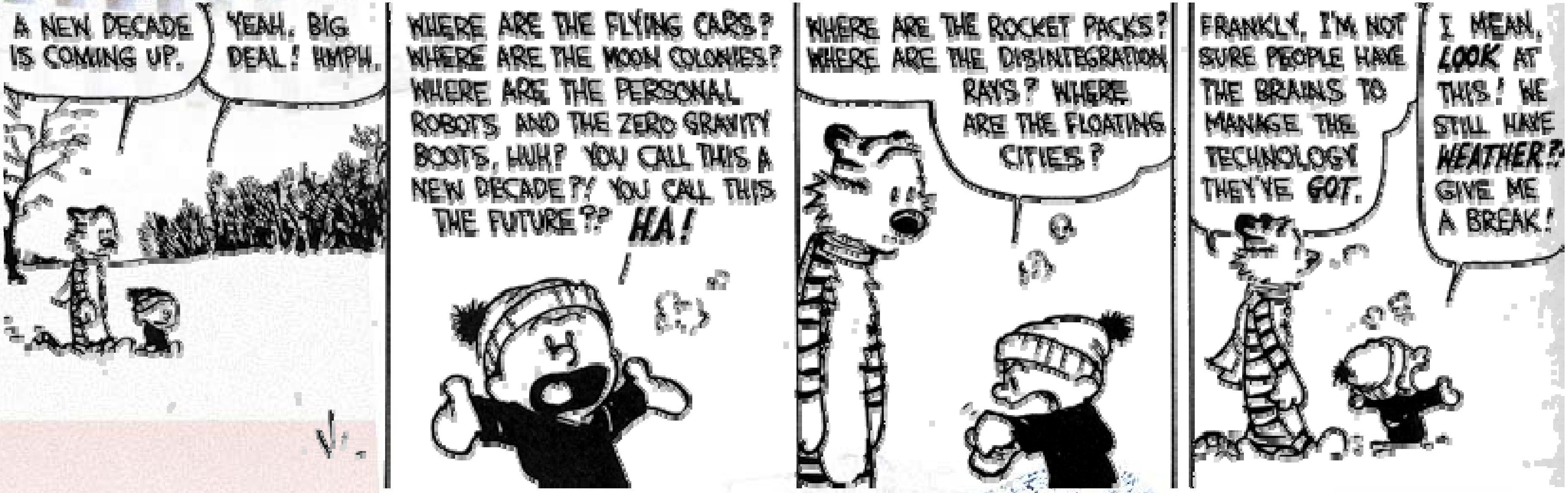
Introduction to Informatics

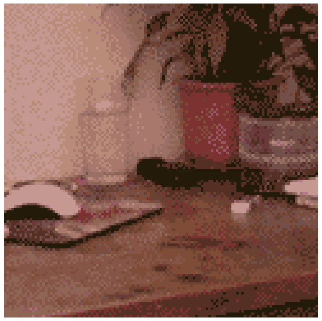
Lecture 28:

Review


A NEW DECADE IS COMING UP.
YEAH. BIG DEAL! HMPH.
WHERE ARE THE FLYING CARS? WHERE ARE THE MOON COLONIES? WHERE ARE THE PERSONAL ROBOTS AND THE ZERO GRAVITY BOOTS, HUH? YOU CALL THIS A NEW DECADE?! YOU CALL THIS THE FUTURE?? HA!
WHERE ARE THE ROCKET PACKS? WHERE ARE THE DISINTEGRATION RAYS? WHERE ARE THE FLOATING CITIES?
FRANKLY, I'M NOT SURE PEOPLE HAVE THE BRAINS TO MANAGE THE TECHNOLOGY THEY'VE GOT.
I MEAN, LOOK AT THIS! WE STILL HAVE WEATHER?! GIVE ME A BREAK!

Luis M.Rocha and Santiago Schnell

Readings until now



- Lecture notes
  - Posted online
    - http://informatics.indiana.edu/rocha/i101
      - *The Nature of Information*
      - *Technology*
      - *Modeling the World*
  - *@ infoport*
    - *http://infoport.blogspot.com*
  - From course package
    - Von Baeyer, H.C. [2004]. *Information: The New Language of Science*. Harvard University Press.
      - Chapters 1, 4 (pages 1-12)
      - Chapter 10 (pages 13-17)
    - From Andy Clark's book "*Natural-Born Cyborgs*"
      - Chapters 2 and 6 (pages 19 - 67)
    - From Irv Englander's book "*The Architecture of Computer Hardware and Systems Software*"
      - Chapter 3: Data Formats (pp. 70-86)
    - Klir, J.G., U. St. Clair, and B.Yuan [1997]. Fuzzy Set Theory: foundations and Applications. Prentice Hall
      - Chapter 2: Classical Logic (pp. 87-97)
      - Chapter 3: Classical Set Theory (pp. 98-103)
    - Norman, G.R. and D.L. Streinrt [2000]. *Biostatistics: The Bare Essentials*.
      - Chapters 1-3 (pages 105-129)
      - OPTIONAL: Chapter 4 (pages 131-136)
      - Chapter 13 (pages 147-155)
      - Chapter 5 (pages 141-144)
    - Igor Aleksander, "Understanding Information Bit by Bit"
      - Pages 157-166
    - Ellen Ullman, "Dining with Robots"
      - Pages 167-172

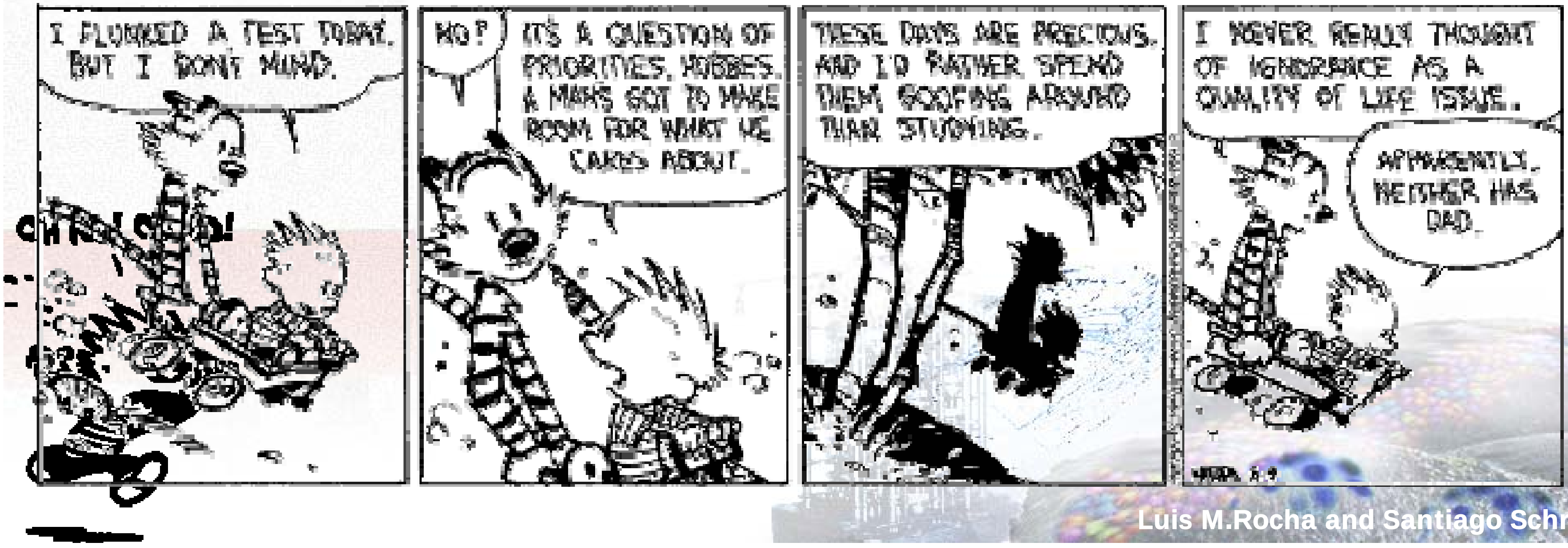
Assignment Situation

- Labs
  - Past
    - Lab 1: Blogs
      - Closed (Friday, January 19): Grades Posted
    - Lab 2: Basic HTML
      - Closed (Wednesday, January 31): Grades Posted
    - Lab 3: Advanced HTML: Cascading Style Sheets
      - Closed (Friday, February 2): Grades Posted
    - Lab 4: More HTML and CSS
      - Closed (Friday, February 9): Grades Posted
    - Lab 5: Introduction to Operating Systems: Unix
      - Closed (Friday, February 16): Grades Posted
    - Lab 6: More Unix and FTP
      - Closed (Friday, February 23): Grades Posted
    - Lab 7: Logic Gates
      - Closed (Friday, March 9): Grades Posted
    - Lab 8: Intro to Statistical Analysis using Excel
      - Closed (Friday, March 30): Grades Posted
    - Lab 9: Data analysis with Excel (linear regression)
      - Closed (Friday, April 6): Grades Posted
    - Lab 10: Simple programming in Excel and Measuring Uncertainty
      - April 12 and 13, Due April 20

- Assignments
  - Individual
    - First installment
      - Closed: February 9: Grades Posted
    - Second Installment
      - Past: March 2: Grades Posted
    - Third installment
      - Past: Grades Posted
    - Fourth Installment
      - Presented April 10th, Due April 20th
  - Group
    - First Installment
      - Past: March 9th, graded
    - Second Installment
      - Past: April 6th Graded
    - Third Installment
      - Presented Thursday, April 12; Due Friday, April 27

Exam Schedule

- 11595
  - Midterm
    - March 1st (Thursday)
      - Regular Class time
  - Final Exam
    - May 3rd (Thursday)
      - 7:15-9:15 p.m.


I FLUNKED A TEST TODAY, BUT I DON'T MIND.
HO?
IT'S A QUESTION OF PRIORITIES, HOBBES. A MAN'S GOT TO MAKE ROOM FOR WHAT HE CARES ABOUT.
THESE DAYS ARE PRECIOUS, AND I'D RATHER SPEND THEM GOOFING AROUND THAN STUDYING.
I NEVER REALLY THOUGHT OF IGNORANCE AS A QUALITY OF LIFE ISSUE.
APPARENTLY, NEITHER HAS DAD.

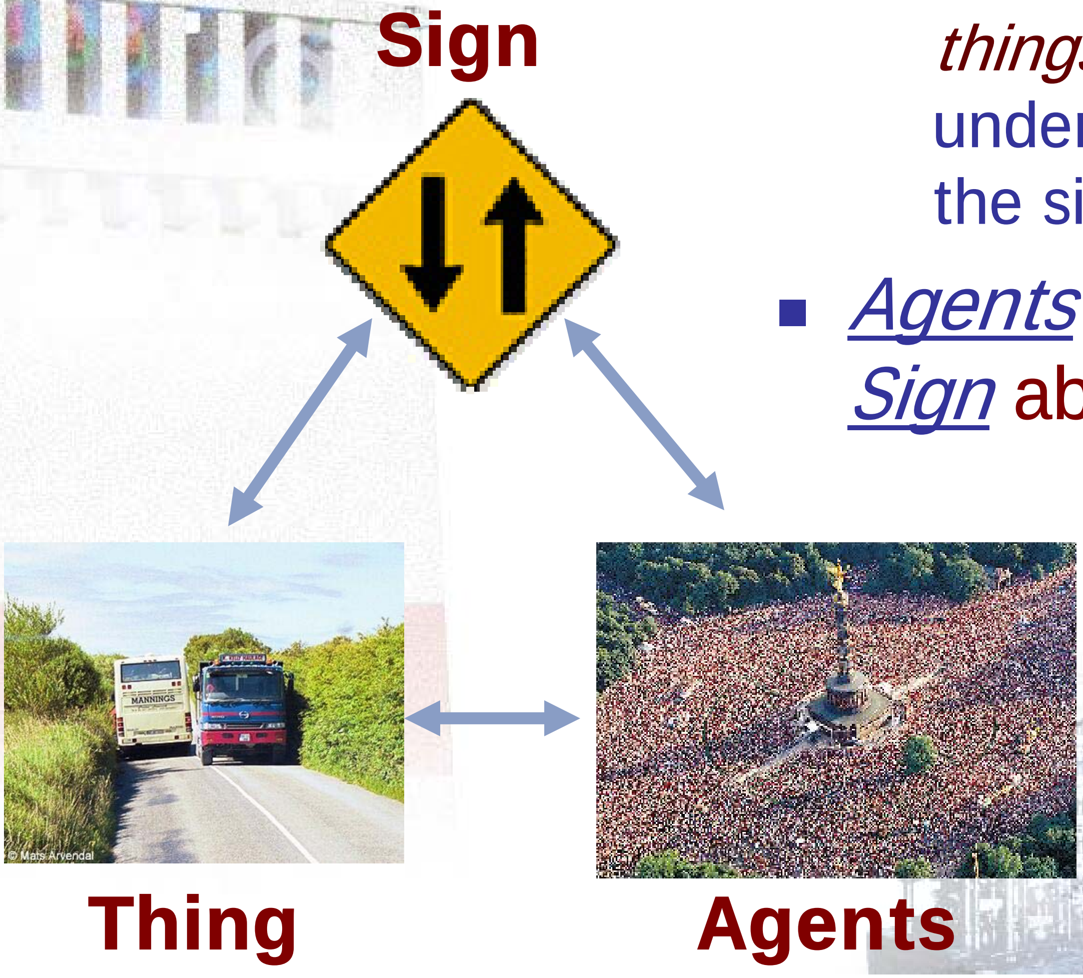
The Nature of Information

Symbols and Information Quantity

Luis M.Rocha and Santiago Schnell

Information is a Relation!

- The central structure of information is a *relation*
  - among *signs*, *objects or things*, and *agents* capable of understanding (or decoding) the signs.
- *Agents* are informed by a *Sign* about some *Thing*.


Sign

Thing

Agents

Semiotics and Informatics

Information

- Semantics
  - the content or **meaning** of the *Sign* of a *Thing* for an *Agent*
    - Relations between signs and objects for an agent
    - the study of meaning.
- Syntax ⟸ **Information Technology**
  - the characteristics of *signs* and symbols devoid of meaning
    - Relations among signs such as their rules of operation, production, storage, and manipulation.
- Pragmatics
  - the context of signs and repercussions of sign-systems in an environment
    - it studies how context influences the interpretation of signs and how well a signs-system represents some aspect of the environment

Informatics

(Peirce’s) Typology of Signs

- ***Icons*** are direct representations of objects.
  - Similar to the thing they represent.
  - Pictorial road signs, scale models, computer icons.
    - A footprint on the sand is an icon of a foot.
- ***Indices*** are indirect representations of objects, but necessarily related.
  - Smoke is an index of fire, the bell is an index of the tolling stroke
    - a footprint is an index of a person.
- ***Symbols*** are arbitrary representations of objects
  - Require exclusively a social convention to be understood
    - Convention establishes a code, agreed by a group of agents, for understanding (decoding) the information contained in symbols.
    - Smoke is an index of fire, but if we agree on an appropriate code (e.g. Morse code) we can use smoke signals to communicate symbolically.

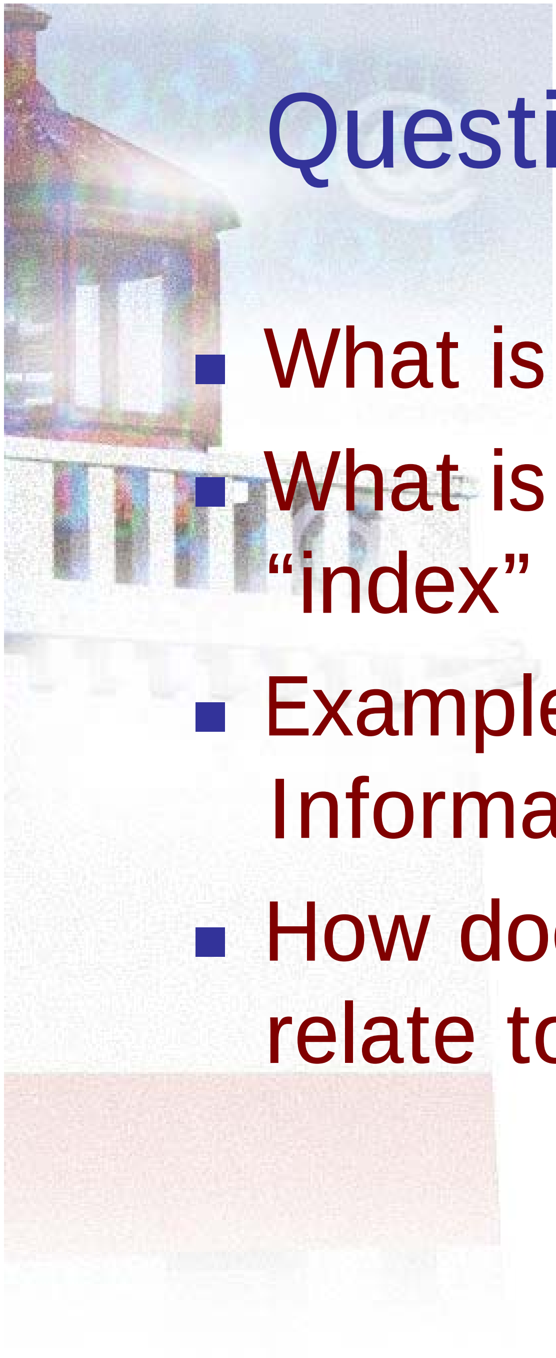
The Bit

- Shannon used the binary system because it is the most economical
  - Uses less memory
  - *Information quantity* depends on the number of alternative choices of messages encoded in the binary system
- Bit (short for *binary digit*) is the most elementary choice one can make
  - Between two items: “0’ and “1”, “heads” or “tails”, “true” or “false”, etc.
  - Bit is equivalent to the choice between two equally likely choices
    - Example, if we know that a coin is to be tossed, but are unable to see it as it falls, a message telling whether the coin came up heads or tails gives us one bit of information

Digital versus Analog

- *Digital* is used to convey the notion of discrete objects/values
  - Things we can count
  - The word digit comes from the Latin word for finger (*digitus*)
  - Digital information is equivalent to symbolic information
    - Any symbol system requires a set of discrete symbols for setting up an arbitrary semantic relation

- Analog (or Analogue)
  - Information transmission via electrical, mechanical, hydraulic, and sound signals
    - Continuously varying signals which are not countable
    - What was used up until Shannon
  - Instead of messages being arbitrarily encoded, analog signals rely on some physical property of the medium
  - It implies an analogy between cause and effect, input and output
    - Voltage as an “analogy” to sound in analog synthesizer
    - But it cannot encode any sound whatsoever!
    - Sounds depend on the physical properties of electricity, the transducer and equipment used

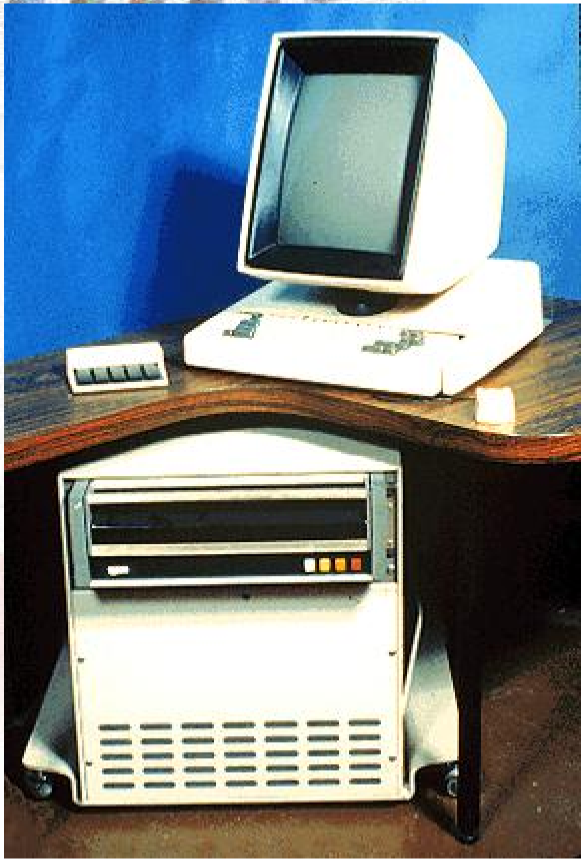


Questions

- What is informatics?
- What is the difference between an “index” and an “symbol”?
- Examples of Analogue vs. Digital Information?
- How does Information Technology relate to semiotics?

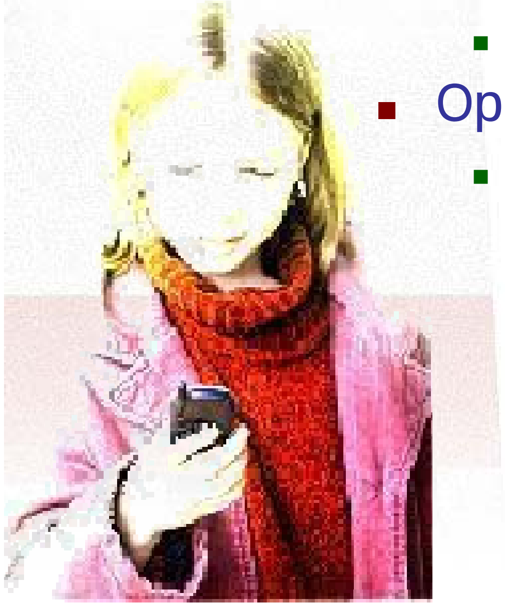
Technology

Tools, Cyborgs and History of IT



Transparent Technology

- So well fitted to, and integrated with, our own lives, biological capacities, and projects as to become almost invisible in use (Andy Clark)
  - Glasses, wrist-watches, driving cars, mobile phones, pens, sports and musical equipment: human-centered
    - Not the same as easy to understand
- Opaque Technology
  - Highly visible in use: technology-centered
    - Computers, industrial machines
  - Opaque technology can become transparent with practice
    - But it works better when biologically suited
      - Natural fit, ergonomics



http://www.baddesigns.com/examples.html

http://www.jnd.org/

(Donald Norman)

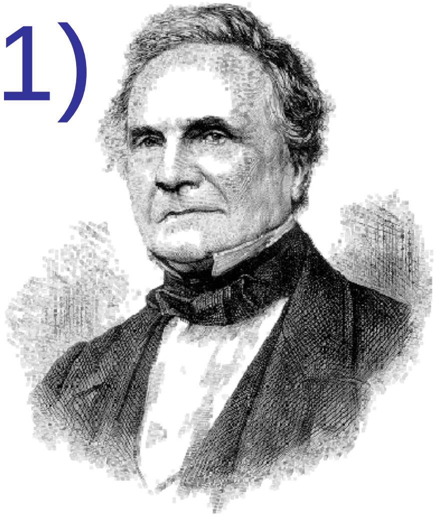
Natural-born Cyborgs?

- Humans more than using, incorporate technology
  - We know we “know” the time, simply because we are equipped with a watch
  - As more portable computing devices become available, will we incorporate easily accessible collective knowledge as our own?
    - Transparent ***knowledge*** technology
      - Example: Google SMS
      - Adaptive Knowledge Technology (Clark, Chapter 6)

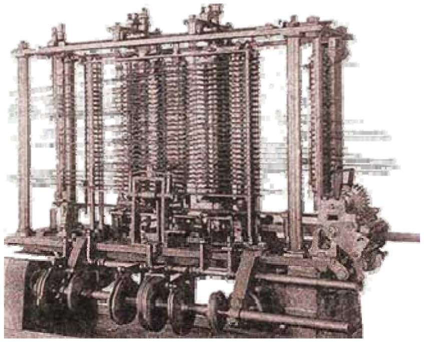
(1of2)
Google Local:
Oliveri Jos Pizza
253 Canal St
New York,
NY 10013
(212) 226-9817

http://www.google.com/sms/howtouse.html#top

Charles Babbage (1791 – 1871)



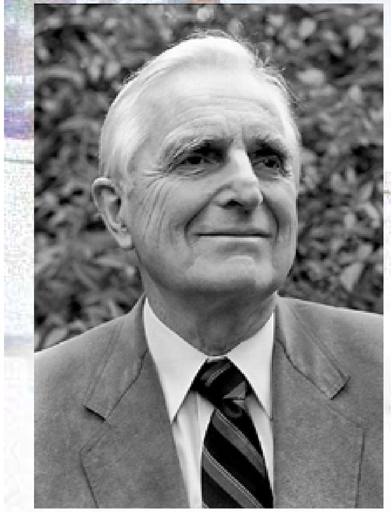
- Analytical Engine
  - Working with Ada Lovelace (daughter of Lord Byron) designed what was to have been a general-purpose mechanical digital computer.
    - With a memory store and a central processing unit (or ‘mill’) and would have been able to select from among alternative actions consequent upon the outcome of its previous actions
      - Conditional branching: Choice, information
    - Programmed with instructions contained on punched cards



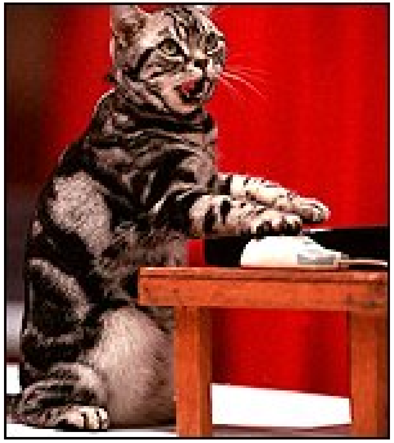
The First Personal Computer

- In 1971, Intel released the first microprocessor.
  - Able to process four bits of data at a time!
- The Altair 8800 (1975)
  - by a company called *Micro Instrumentation and Telementry Systems* (MITS) sold for $397
  - Came as a kit for assembly who had to to write software for the machine
    - in machine code!
  - 256 byte memory --about the size of a paragraph
- Microsoft
  - Was born to create a BASIC compiler for the Altair
    - Beginners All-purpose Symbolic Instruction Code

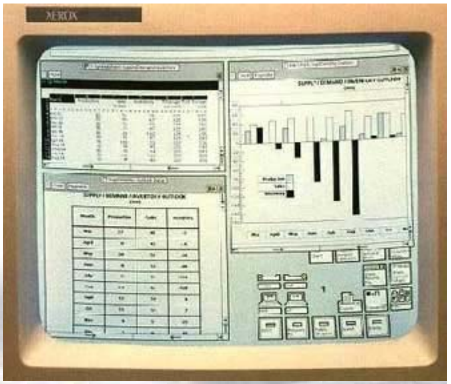
graphical user interface (GUI)



- On-Line System (NLS) (1960s)
  - Doug Engelbart's *Augmentation of Human Intellect* project @ Stanford Research Institute
    - pioneer of human-computer interaction
    - also developed *hypertext*
  - Incorporated a mouse-driven cursor and multiple windows.
    - WIMP (windows, icons, menus and pointers)
    - See his demo
      - http://sloan.stanford.edu/MouseSite/1968Demo.html


A P

- XEROX PARC
  - Xerox Alto (1973)
    - first computer to use the *desktop metaphor* and GUI

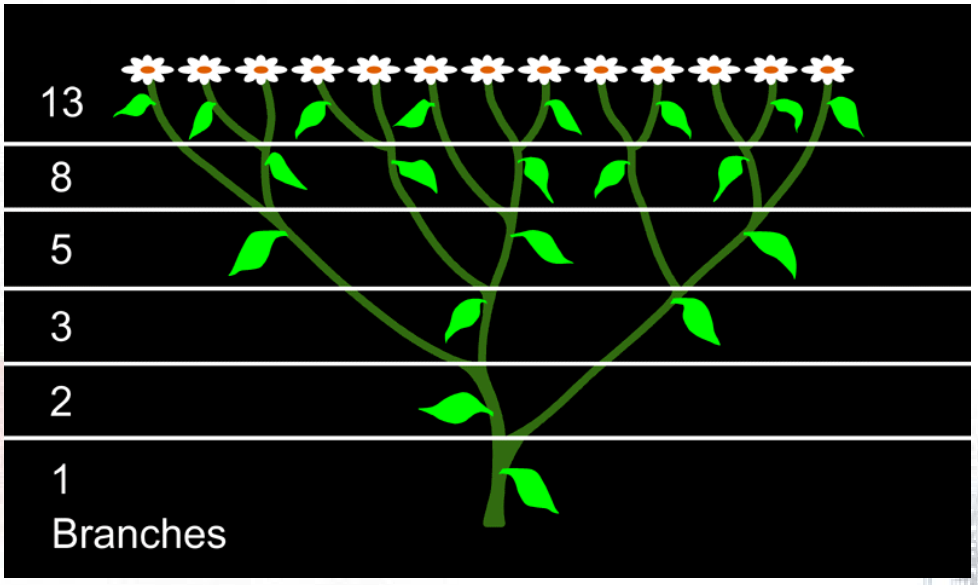


Questions

- Transparent vs Opaque Technology?
- Describe two computing devices used before the XX century.
- Which communications protocol marks the transition from the ARPANET to the Internet?

Modeling

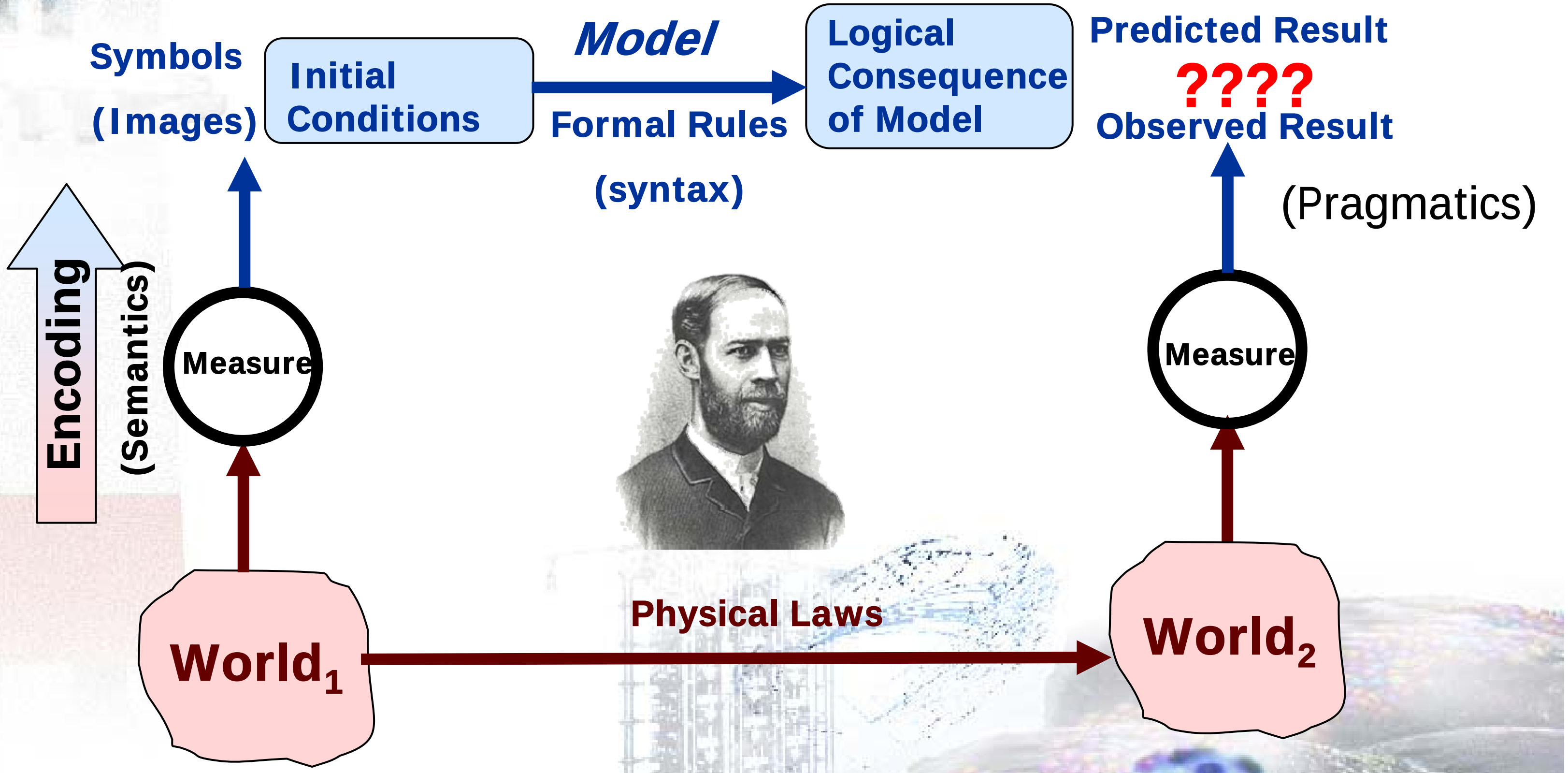
Describing and Understanding the World



Luis M.Rocha and Santiago Schnell

The Hertzian Modeling Paradigm

“The most direct and in a sense the most important problem which our conscious knowledge of nature should enable us to solve is the ***anticipation of future events***, so that we may arrange our present affairs in accordance with such anticipation”. (Hertz, 1894)



Symbols

(Images)

Initial Conditions

Model

Formal Rules

(syntax)

Logical Consequence of Model

Predicted Result

????

Observed Result

(Pragmatics)

Encoding

(Semantics)

Measure

Measure

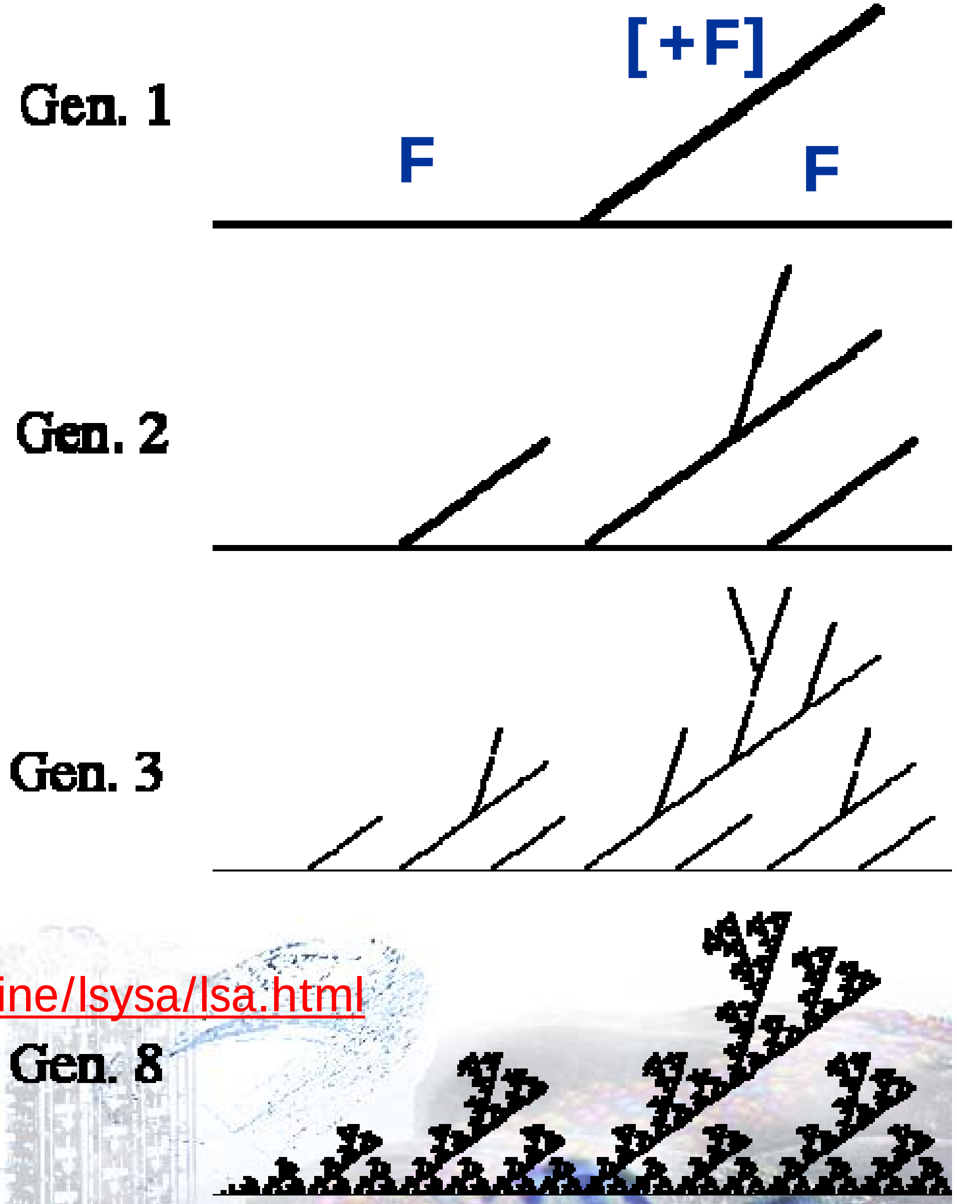
World$_1$

Physical Laws

World$_2$

Branching L-Systems

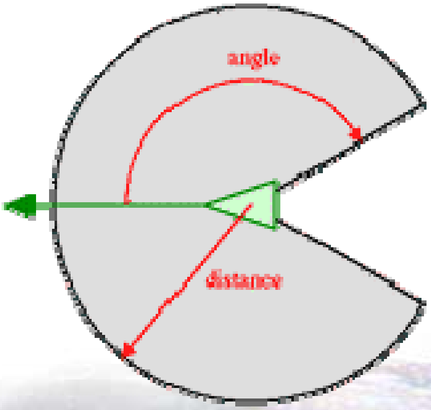
- Add branching symbols [ ]
  - simple example
    - Main trunk shoots off one side branch
  - Angle 10
  - Axiom F
  - F=F[+F]F

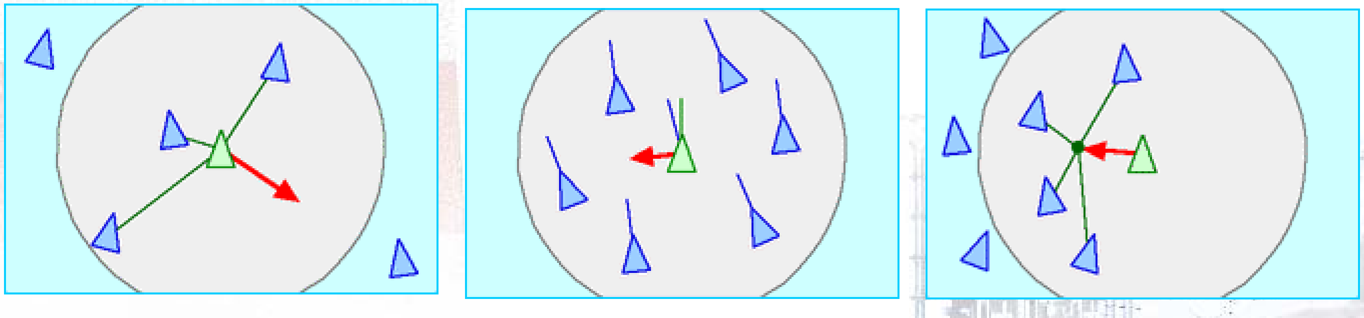


http://www.info.univ-angers.fr/pub/gh/vitrine/lsysa/lsa.html

Flocking Behavior

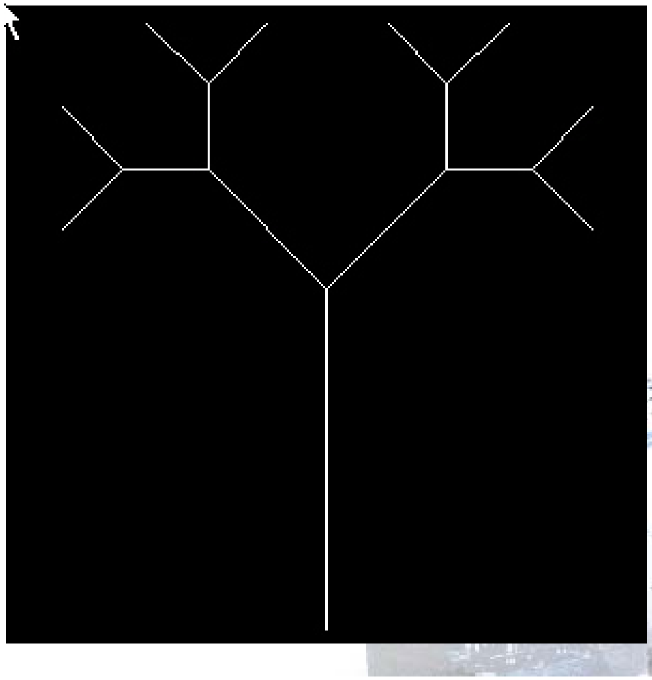
- Boids by Craig Reynolds (1986)
  - 3 Steering behaviors
    - **Separation**: steer to avoid crowding local flockmates
      - Maintain minimum distance to others
    - **Alignment**: steer towards the average heading of local flockmates
      - Adjust speed according to others in vicinity
    - **Cohesion**: steer to move toward the average position of local flockmates
  - Each boid sees only flockmates within a certain small neighborhood around itself.
  - http://www.red3d.com/cwr/boids/



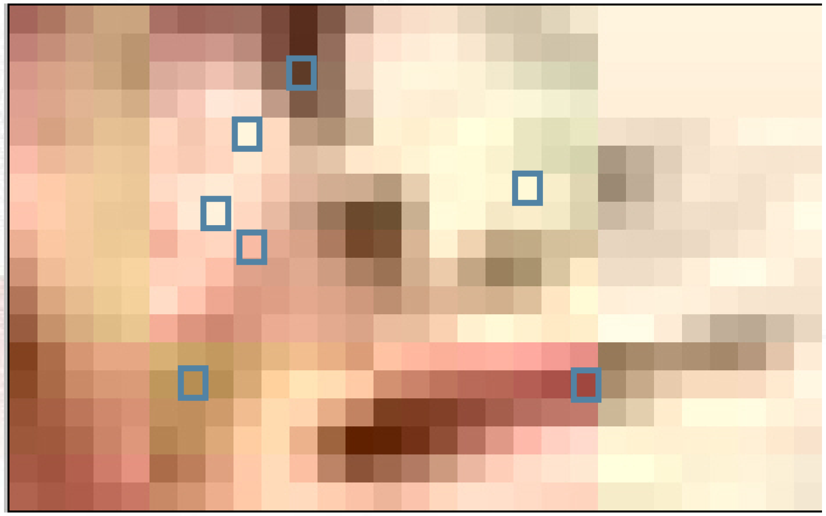


Questions

1. Describe the Hertz Modelling Process
2. What are Boids and how do they work?
3. Propose a L-System Rule to draw the following artificial plant

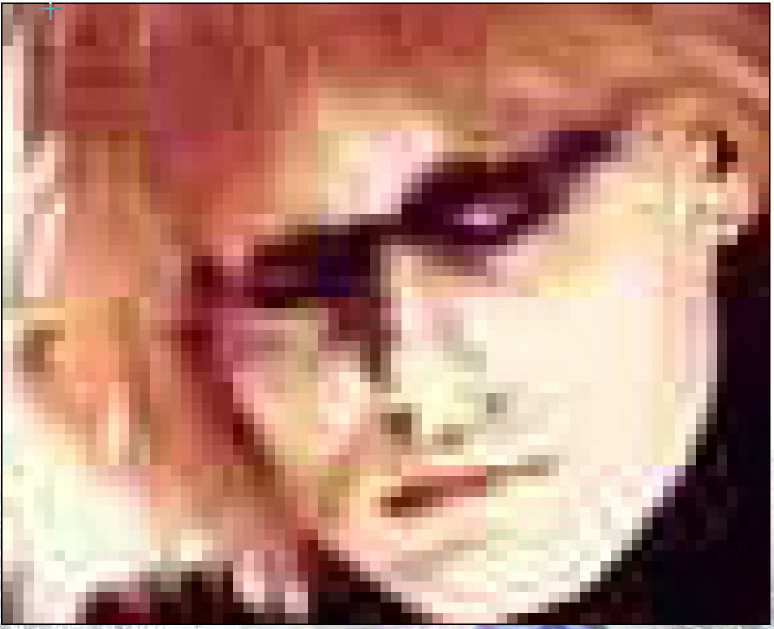


Data Representation

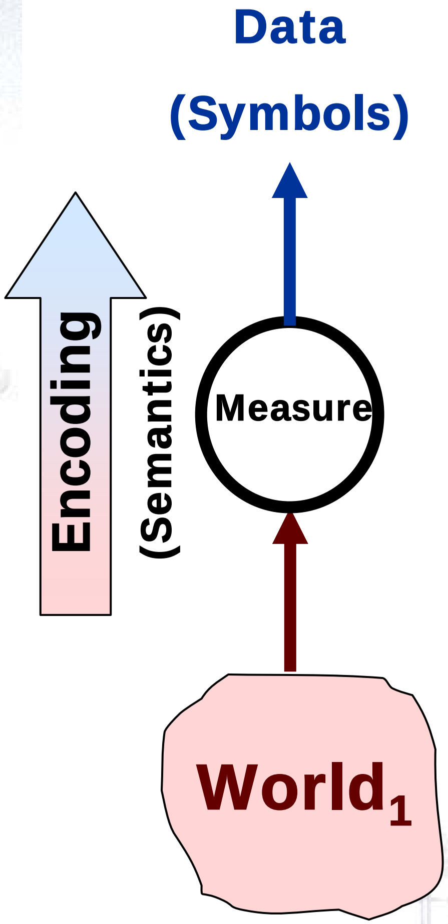


Pixels: picture elements

Encoding the World

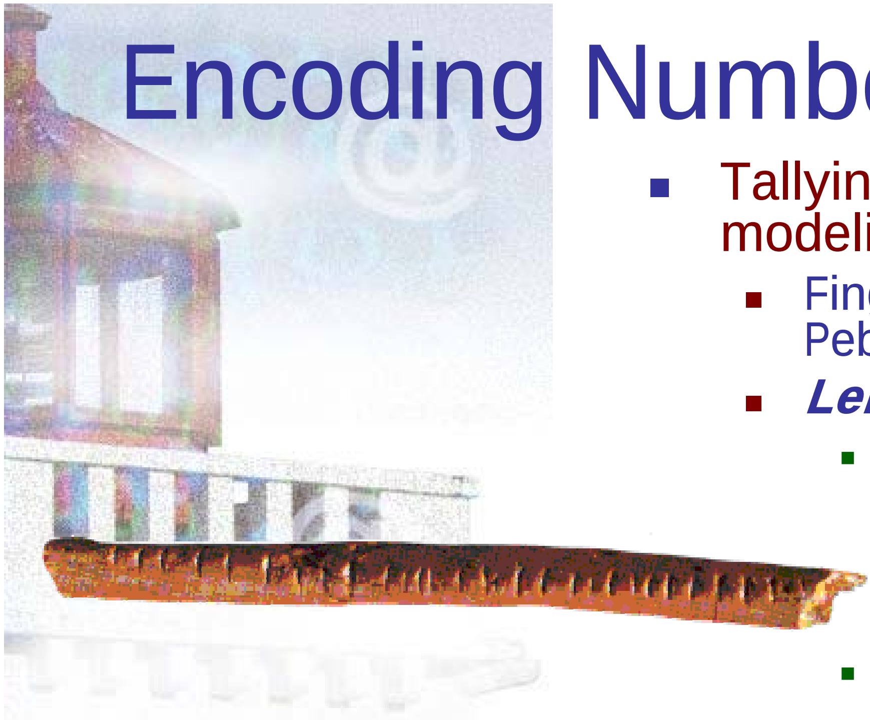


Encoding in the Modeling Relation

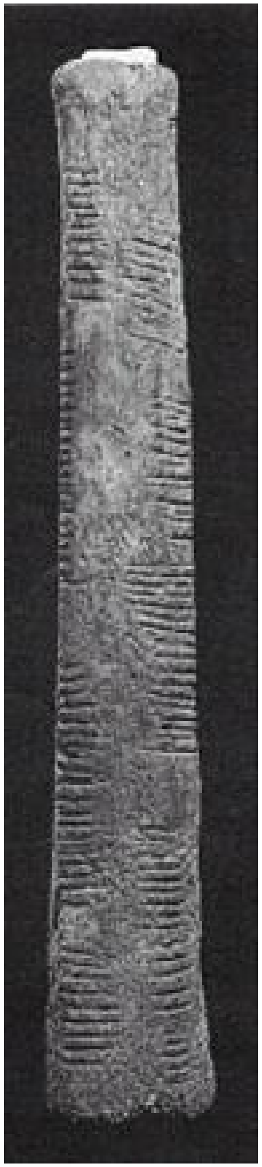


- How to encode data?
  - What is data?
    - Information without context and knowledge
    - Part of Syntax
  - Keeping Numbers
    - The most fundamental need for modeling and information

Encoding Numbers:Counting

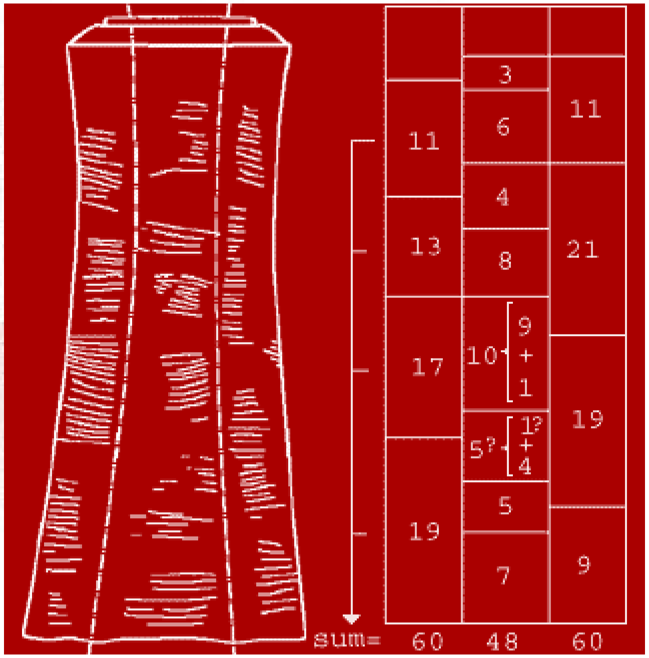


- Tallying is the earliest form of modeling
  - Fingers (digits), stones (Lt “calculus”= Pebble), bones
  - ***Lebombo bone***
    - Oldest counting tool is a piece of baboon fibula with 29 notches from 35,000 BC, discovered in the mountains between South Africa and Swaziland
    - Probably representing the number of days in a Moon Cycle (A Model!)
  - Czechoslovakian wolf's bone
    - with 55 notches in groups of 5, from 30,000 BC.



The ***Ishango Bone***

- Oldest Mathematical Artefact?
  - 10,000 BC, border of Zaire and Uganda
  - Used as a counting tool?
    - 9,11,13,17,19, 21: odd numbers
    - 11, 13, 17, 19: prime numbers
    - 60 and 48 are multiples of 12



http://www.simonsingh.com/The_Ishango_Bone.html

Converting Binary to Decimal

- $2^8 = 256$
- $2^7 = 128$
- $2^6 = 64$
- $2^5 = 32$
- $2^4 = 16$
- $2^3 = 8$
- $2^2 = 4$
- $2^1 = 2$
- $2^0 = 1$

2^8	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
0	1	1	0	0	1	0	0	1
	128	+ 64			+ 8			+ 1

201

$$\ldots d_4 d_3 d_2 d_1 d_0 =$$
$$\ldots + d_4\times 2^4 + d_3\times 2^3 + d_2\times 2^2 + d_1\times 2^1 + d_0\times 2^0$$

Base Conversion

- Decimal to Binary
  - Repeated Division by 2
    - Divide the decimal number by 2
    - If the remainder is 0, on the side write down a 0
    - If the remainder is 1, write down a 1
    - Continue until the quotient is 0
    - Remainders are written beginning at the least significant digit (right) and each new digit is written to more significant digit (the left) of the previous digit.

decimal	quotient	Remain.	binary
58	29	0	0
29	14	1	10
14	7	0	010
7	3	1	1010
3	1	1	11010
1	0	1	**111010**

Dealing with rational numbers

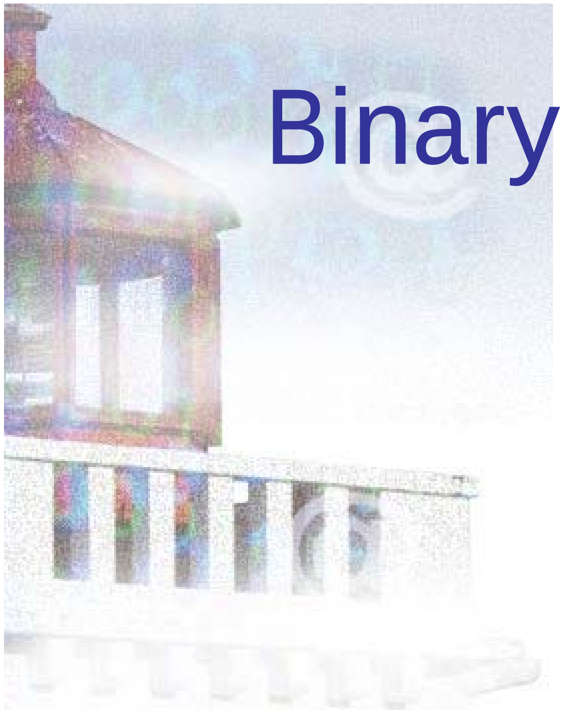
- $2^4 = 16$
- $2^3 = 8$
- $2^2 = 4$
- $2^1 = 2$
- $2^0 = 1$
- $2^{-1} = 0.5$
- $2^{-2} = 0.25$
- $2^{-3} = 0.125$

2^4	2^3	2^2	2^1	2^0	2^{-1}	2^{-2}	2^{-3}
1	1	0	0	1	0	0	1
16	+ 8			+ 1			+ .125

25.125

$$\ldots d_2 d_1 d_0 . d_{-1} d_{-2} \ldots =$$
$$\ldots + d_2 \times 2^2 + d_1 \times 2^1 + d_0 \times 2^0 + d_{-1} \times 2^{-1} + d_{-2} \times 2^{-2} + \ldots$$

Binary Arithmetic



- Addition Rules
  - 0+0 = 0, with no carry,
  - 1+0 = 1, with no carry,
  - 0+1 = 1, with no carry,
  - 1+1 = 0, and you carry a 1

```
     1____        111__
      1010         1010
      1100         1110
   -------      -------
     10110        11000
(10+12=22)   (10+14=24)
```

Binary Multiplication

```
    1010
    1110
 -------
111_____
    0000
   1010_
  1010__
 1010___
-----------
10001100
```

(10+14=140)

```
     1010
     1100
  -------
     0000
    0000_
   1010__
  1010___
------------
  1111000
```

(10×12=120)

Hexadecimal

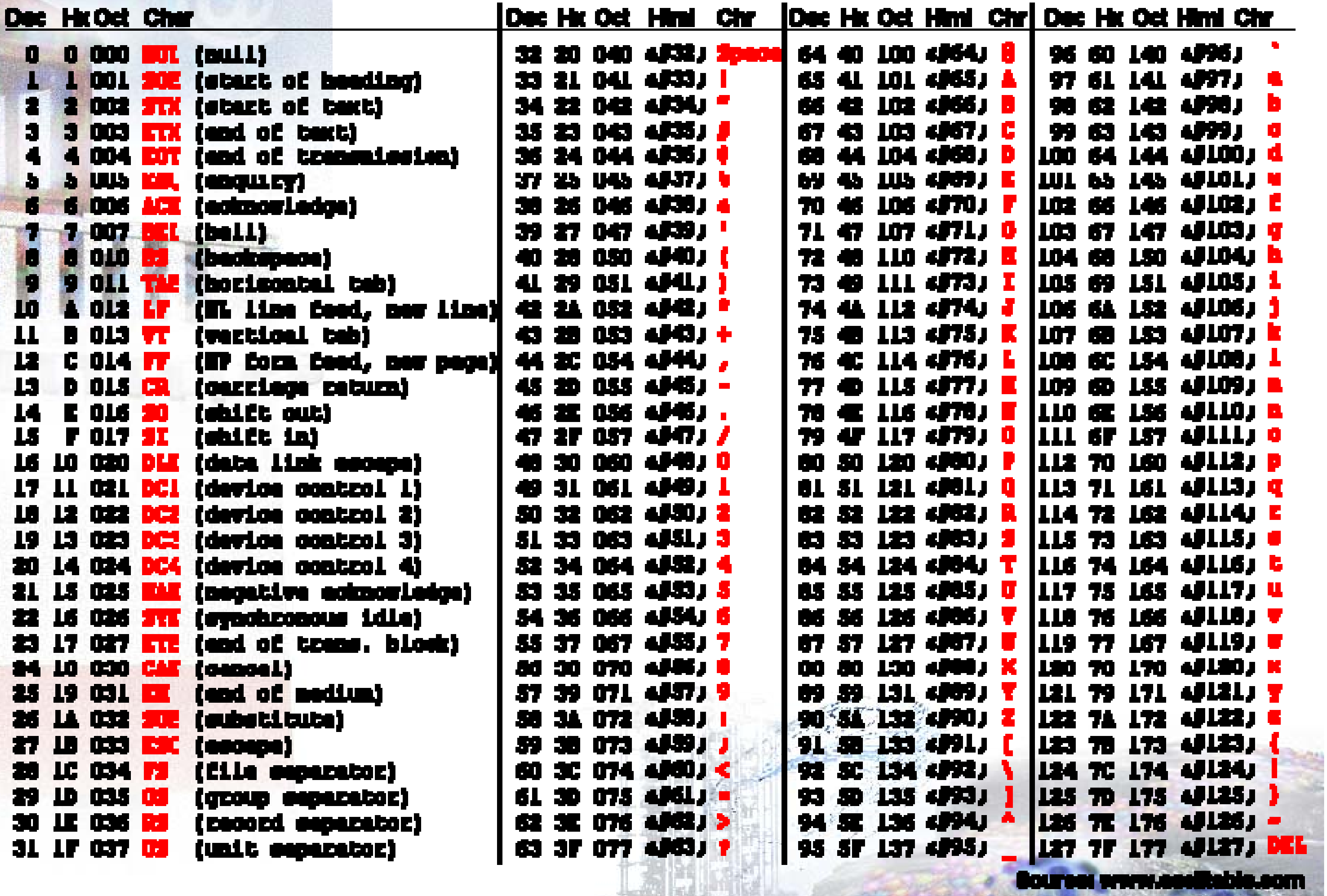
- Base 16
  - 16 symbols: 0,1,2,3,4,5,6,7,8,9, A, B, C, D, E, F
  - Easy to convert to and from Binary
    - 16 is a power of 2: 16 = 2^4
    - It takes 4 binary digits for every hexadecimal one
    - Good to represent binary in compressed form!

Hex	Bin	Hex	Bin	Hex	Bin	Hex	Bin
0	0000	4	0100	8	1000	C	1100
1	0001	5	0101	9	1001	D	1101
2	0010	6	0110	A	1010	E	1110
3	0011	7	0111	B	1011	F	1111

Encoding Text

- ASCII
  - American Standard Code for Information Interchange
    - between binary numbers and computer and roman symbols
    - Standard to allow computers to communicate textual data
  - Uses 7 bits to encode 128 symbols or characters
    - 2^7 = 128. It fills a byte, but the 8th bit is used to encode additional symbols for other languages and graphics
    - Usually described in hexadecimal
  - 4 groups of 32 characters
    - 00 to 1F: **control characters**
      - Mostly printer/display operations: *carriage return* (0Dh), *line feed* (0Ah), *back space* (08h), etc.
    - 20 to 3F: punctuation, numeric, and special characters
      - Space (20h), digits 0-9 (30h-39h)
        - Arranged so that by subtracting 30h from the ASCII code for any digit, we obtain the numeric equivalent of the digit
    - 40 to 5F: uppercase letters, plus some special characters
    - 60 to 7F: lowercase letters, plus some special characters and a control character (DEL)

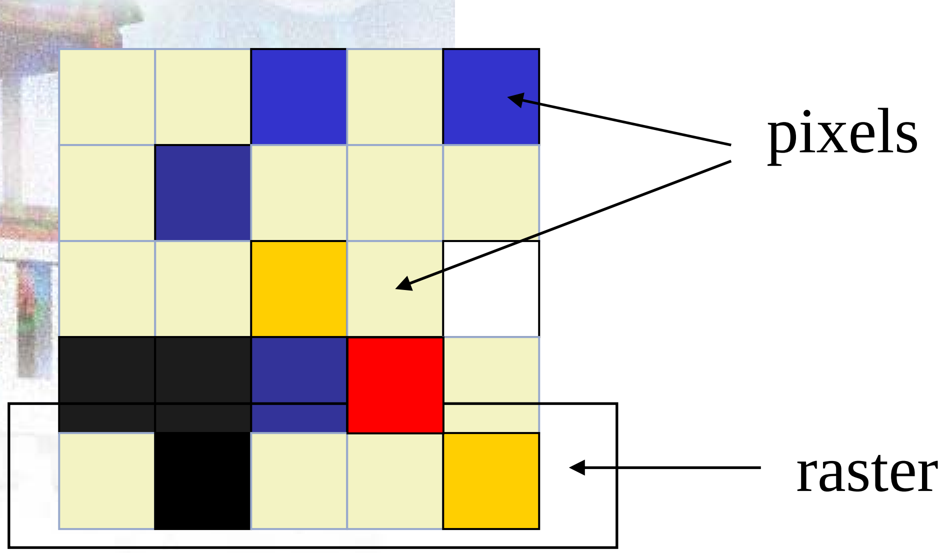
ASCII Table



Dec	Hx	Oct	Char	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr
0	0	000	NUL (null)	32	20	040	&#32;	Space	64	40	100	&#64;	@	96	60	140	&#96;	`
1	1	001	SOH (start of heading)	33	21	041	&#33;	!	65	41	101	&#65;	A	97	61	141	&#97;	a
2	2	002	STX (start of text)	34	22	042	&#34;	"	66	42	102	&#66;	B	98	62	142	&#98;	b
3	3	003	ETX (end of text)	35	23	043	&#35;	#	67	43	103	&#67;	C	99	63	143	&#99;	c
4	4	004	EOT (end of transmission)	36	24	044	&#36;	$	68	44	104	&#68;	D	100	64	144	&#100;	d
5	5	005	ENQ (enquiry)	37	25	045	&#37;	%	69	45	105	&#69;	E	101	65	145	&#101;	e
6	6	006	ACK (acknowledge)	38	26	046	&#38;	&	70	46	106	&#70;	F	102	66	146	&#102;	f
7	7	007	BEL (bell)	39	27	047	&#39;	'	71	47	107	&#71;	G	103	67	147	&#103;	g
8	8	010	BS (backspace)	40	28	050	&#40;	(	72	48	110	&#72;	H	104	68	150	&#104;	h
9	9	011	TAB (horizontal tab)	41	29	051	&#41;	)	73	49	111	&#73;	I	105	69	151	&#105;	i
10	A	012	LF (NL line feed, new line)	42	2A	052	&#42;	*	74	4A	112	&#74;	J	106	6A	152	&#106;	j
11	B	013	VT (vertical tab)	43	2B	053	&#43;	+	75	4B	113	&#75;	K	107	6B	153	&#107;	k
12	C	014	FF (NP form feed, new page)	44	2C	054	&#44;	,	76	4C	114	&#76;	L	108	6C	154	&#108;	l
13	D	015	CR (carriage return)	45	2D	055	&#45;	-	77	4D	115	&#77;	M	109	6D	155	&#109;	m
14	E	016	SO (shift out)	46	2E	056	&#46;	.	78	4E	116	&#78;	N	110	6E	156	&#110;	n
15	F	017	SI (shift in)	47	2F	057	&#47;	/	79	4F	117	&#79;	O	111	6F	157	&#111;	o
16	10	020	DLE (data link escape)	48	30	060	&#48;	0	80	50	120	&#80;	P	112	70	160	&#112;	p
17	11	021	DC1 (device control 1)	49	31	061	&#49;	1	81	51	121	&#81;	Q	113	71	161	&#113;	q
18	12	022	DC2 (device control 2)	50	32	062	&#50;	2	82	52	122	&#82;	R	114	72	162	&#114;	r
19	13	023	DC3 (device control 3)	51	33	063	&#51;	3	83	53	123	&#83;	S	115	73	163	&#115;	s
20	14	024	DC4 (device control 4)	52	34	064	&#52;	4	84	54	124	&#84;	T	116	74	164	&#116;	t
21	15	025	NAK (negative acknowledge)	53	35	065	&#53;	5	85	55	125	&#85;	U	117	75	165	&#117;	u
22	16	026	SYN (synchronous idle)	54	36	066	&#54;	6	86	56	126	&#86;	V	118	76	166	&#118;	v
23	17	027	ETB (end of trans. block)	55	37	067	&#55;	7	87	57	127	&#87;	W	119	77	167	&#119;	w
24	18	030	CAN (cancel)	56	38	070	&#56;	8	88	58	130	&#88;	X	120	78	170	&#120;	x
25	19	031	EM (end of medium)	57	39	071	&#57;	9	89	59	131	&#89;	Y	121	79	171	&#121;	y
26	1A	032	SUB (substitute)	58	3A	072	&#58;	:	90	5A	132	&#90;	Z	122	7A	172	&#122;	z
27	1B	033	ESC (escape)	59	3B	073	&#59;	;	91	5B	133	&#91;	[	123	7B	173	&#123;	{
28	1C	034	FS (file separator)	60	3C	074	&#60;	<	92	5C	134	&#92;	\	124	7C	174	&#124;	|
29	1D	035	GS (group separator)	61	3D	075	&#61;	=	93	5D	135	&#93;	]	125	7D	175	&#125;	}
30	1E	036	RS (record separator)	62	3E	076	&#62;	>	94	5E	136	&#94;	^	126	7E	176	&#126;	~
31	1F	037	US (unit separator)	63	3F	077	&#63;	?	95	5F	137	&#95;	_	127	7F	177	&#127;	DEL

Source: www.asciitable.com

Bitmap Image



- Representation of a two-dimensional image as a finite set of digital values
- Picture elements or ***pixels***
  - **Resolution:** number of pixels in an image
    - 1024 x 768
  - Each defined by one or more numbers
    - Color, intensity

2^{24} (about 16 million)

2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
1	1	0	0	1	0	0	1

2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
1	1	0	0	1	0	0	1

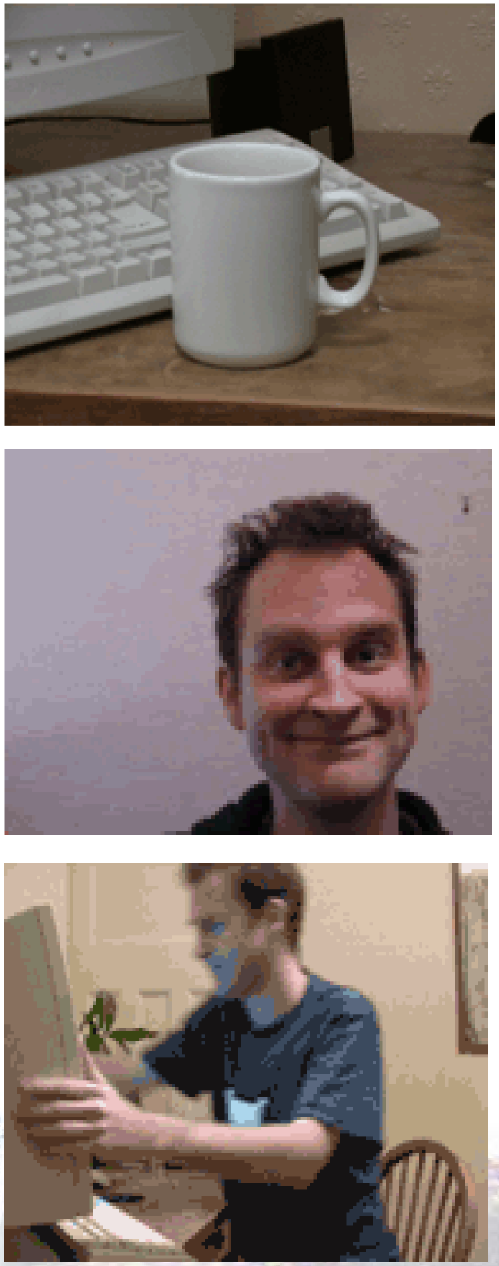
2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
1	1	0	0	1	0	0	1

RGB: 3 Bytes, one for each primary colors: Red, Green, and Blue: *16,777,216* colors total

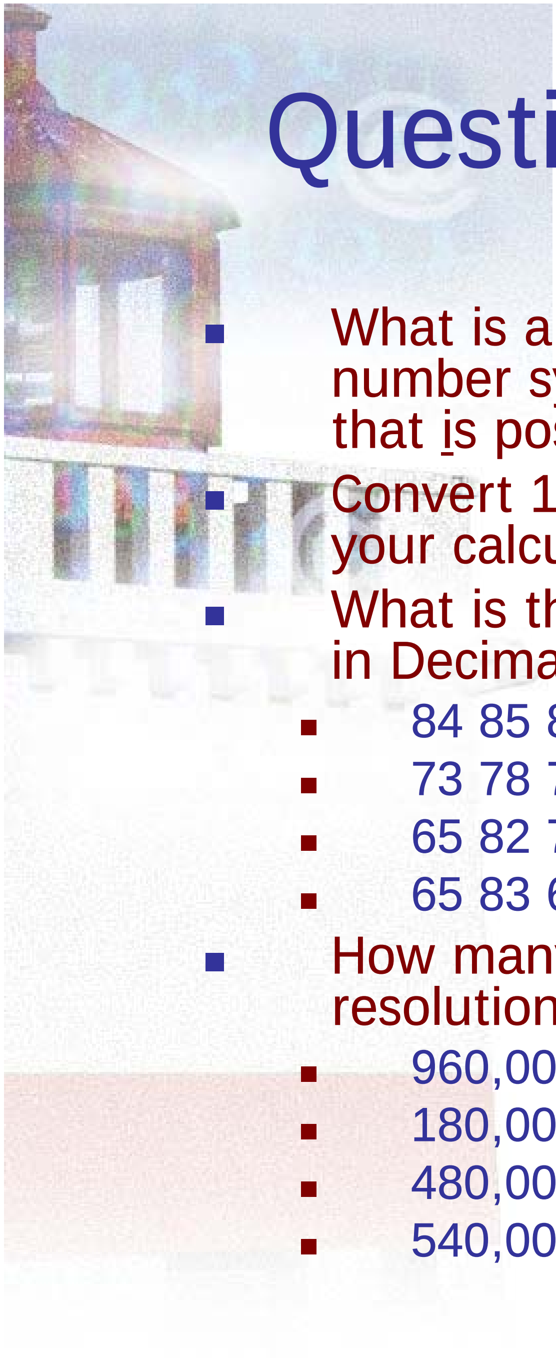
Adapted from Cathy Wyss (I308)

Graphics Interchange Format

- Developed by CompuServe in 1987 (GIF87a)
  - Developed to facilitate exchange across computing platforms
  - Allows transparency
  - GIF89a
    - 1989: allows animated GIF images
- Uses LZW data compression
  - More efficient than plain bitmaps
  - Large images downloaded quicker
  - Patent owned by Unisys until 2003
    - Compuserve did not know that LZW was covered by a patent.
    - Before 1994, Unisys was not aware that GIF used LZW.
  - Builds a dictionary of previously seen strings in the information being compressed.
    - The dictionary does not have to be transmitted



http://sheepfilms.co.uk



Questions

- What is a positional number system? Give an example of a number system that is not positional, and an example of one that is positional.
- Convert 1001001101.01 from binary to decimal. Please show your calculations.
- What is the ASCII encoding of the word TURING (Uppercase) in Decimal?
  - 84 85 82 73 78 71
  - 73 78 70 79 82 77 65 84 73 67 83
  - 65 82 73 83 84 79 84 76 69
  - 65 83 67 73 73 50
- How many bytes do you need to encode a bitmap figure with resolution 300 x 600 using the RGB format?
  - 960,000
  - 180,000
  - 480,000
  - 540,000

Deductive Modeling

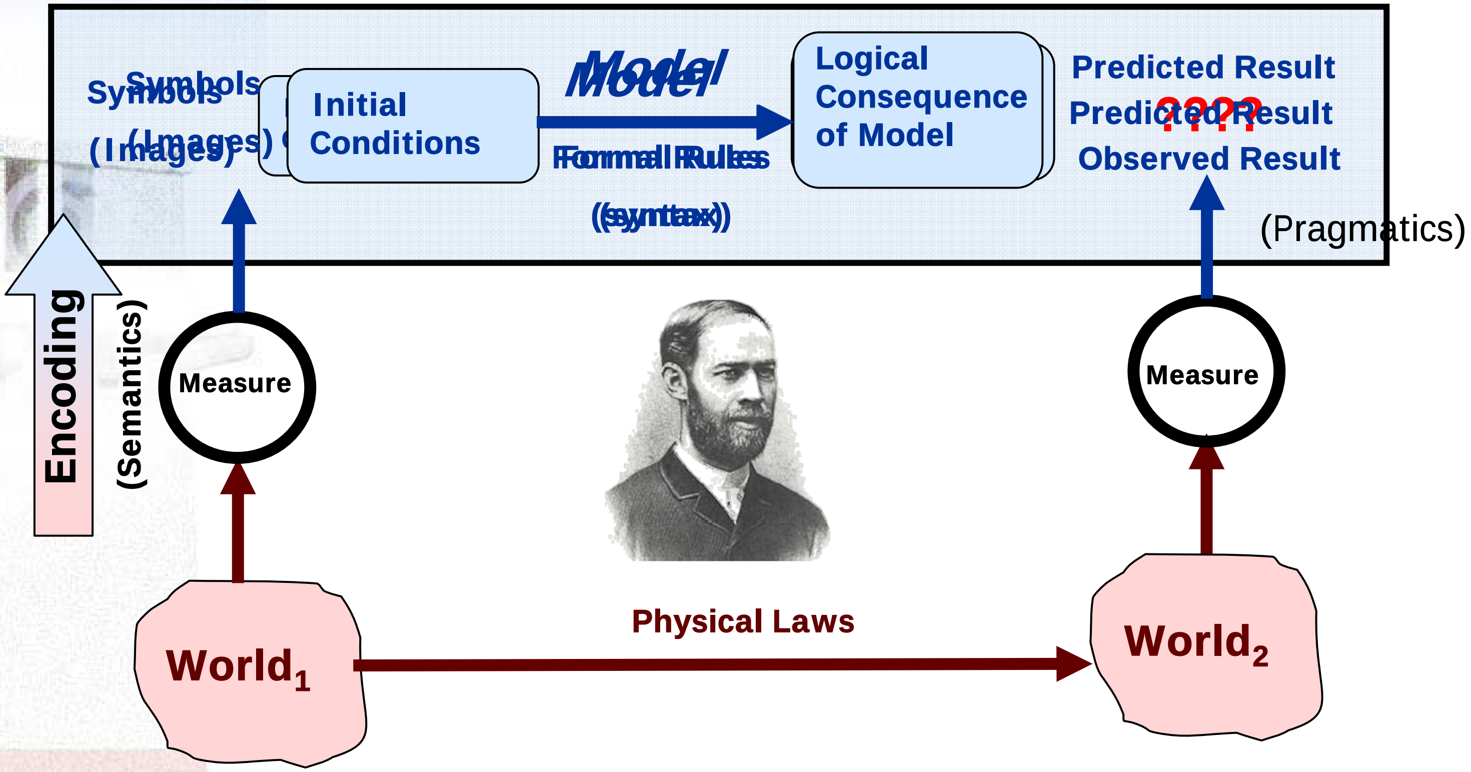
Logic, Sets, Deduction

PENGUINS ARE BLACK AND WHITE.
SOME OLD TV SHOWS ARE BLACK AND WHITE.
THEREFORE, SOME PENGUINS ARE OLD TV SHOWS.
GLASBERGEN

Logic: another thing that penguins aren't very good at.

The Modeling Relation

Hertz' Modeling Paradigm



Symbols (Images)

Initial Conditions

Model

Formal Rules (syntax)

Logical Consequence of Model

Predicted Result

????

Predicted Result

Observed Result

(Pragmatics)

Encoding

(Semantics)

Measure

Measure

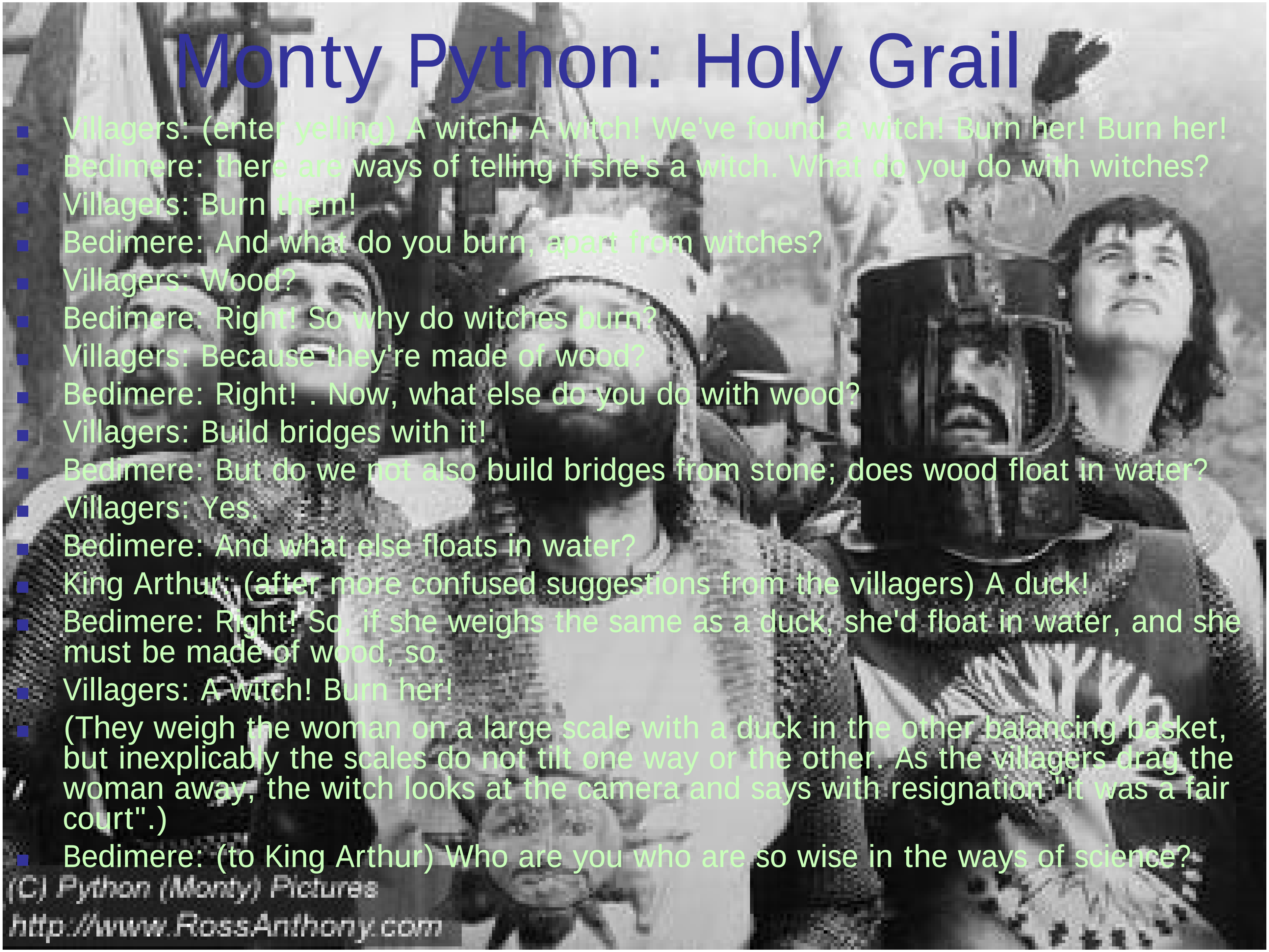
Physical Laws

World$_1$

World$_2$

- Formal Rules
  - From symbolic representations of observables
  - Produce Conclusions

Monty Python: Holy Grail



- Villagers: (enter yelling) A witch! A witch! We've found a witch! Burn her! Burn her!
- Bedimere: there are ways of telling if she's a witch. What do you do with witches?
- Villagers: Burn them!
- Bedimere: And what do you burn, apart from witches?
- Villagers: Wood?
- Bedimere: Right! So why do witches burn?
- Villagers: Because they're made of wood?
- Bedimere: Right! . Now, what else do you do with wood?
- Villagers: Build bridges with it!
- Bedimere: But do we not also build bridges from stone; does wood float in water?
- Villagers: Yes.
- Bedimere: And what else floats in water?
- King Arthur: (after more confused suggestions from the villagers) A duck!
- Bedimere: Right! So, if she weighs the same as a duck, she'd float in water, and she must be made of wood, so.
- Villagers: A witch! Burn her!
- (They weigh the woman on a large scale with a duck in the other balancing basket, but inexplicably the scales do not tilt one way or the other. As the villagers drag the woman away, the witch looks at the camera and says with resignation "it was a fair court".)
- Bedimere: (to King Arthur) Who are you who are so wise in the ways of science?

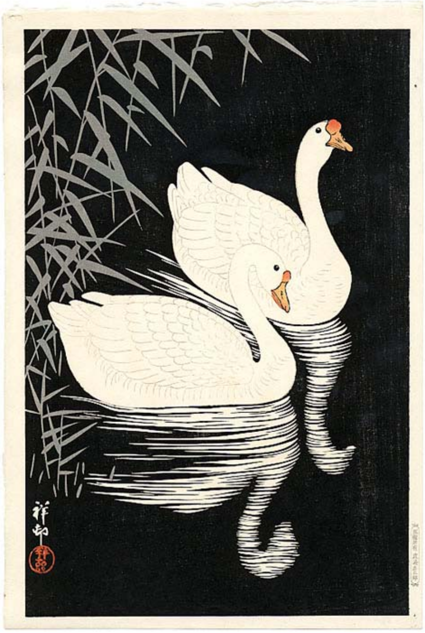
(C) Python (Monty) Pictures
http://www.RossAnthony.com

Deduction vs. Induction

- Propositional Logic is used to study ***inferences***
  - Lists of propositions
  - How conclusions can be reached from premises
- Deductive Inference ← **Logic**
  - If the premises are true, we have absolute ***certainty*** of the conclusion
    - February has 29 days only in leap years
    - Today is February 29th
    - This year is a leap year
- Inductive Inference ← **Uncertainty**
  - Conclusion supported by ***good evidence*** (significant number of examples/observations) but not full certainty -- ***likelihood***
    - Ran WhiteBox for 1000 cycles, “dead box” observed
    - Ran WhiteBox for 1000 cycles, “dead box” observed
    - Ran WhiteBox for 1000 cycles, “dead box” observed
    - ....
    - Ran WhiteBox for 1000 cycles, “dead box” observed
    - “Dead Box” always appears after 1000 cycles

Uncertainty in Induction

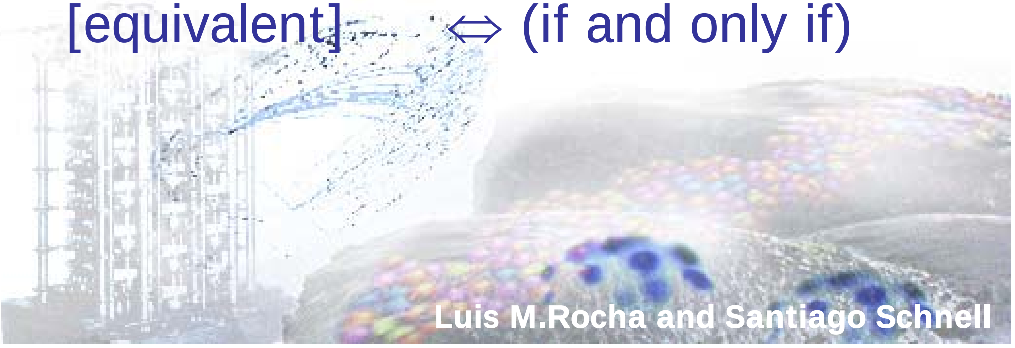
- Via Induction
  - Europeans could have thought that all Swans are White
    - by observing instance after instance
  - But black swans exist
    - From Australia

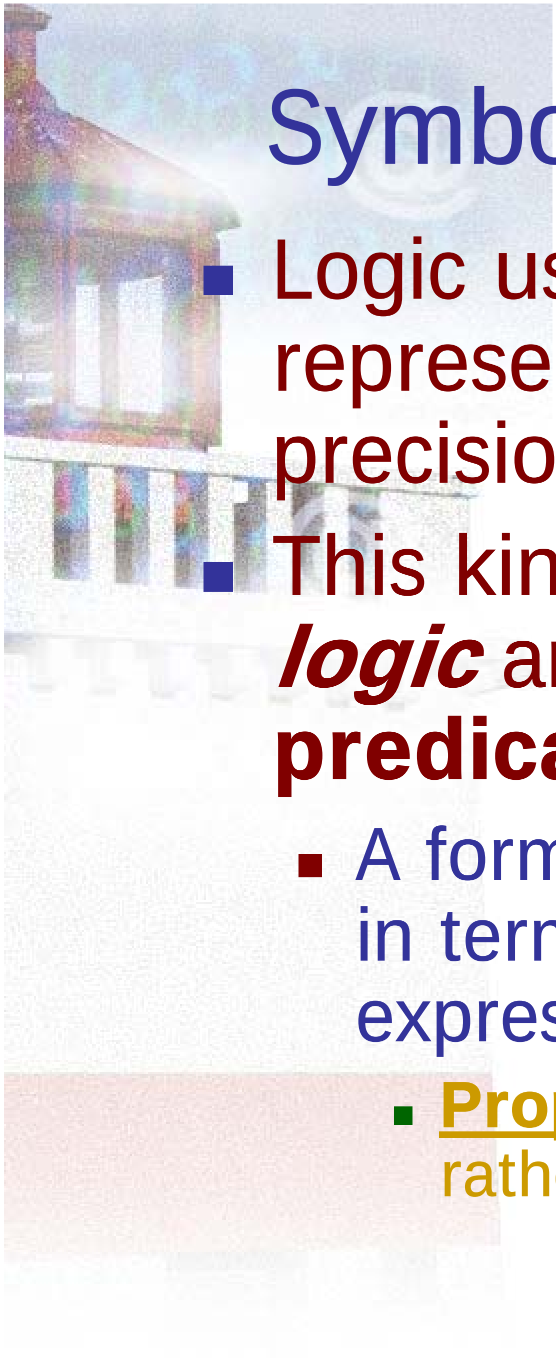


The structure of propositional logic

- *Simple propositions* are represented by single, lower case letters
  - Bloomington is a town – *p*
  - Indiana is a state - *q*
- *Complex propositions* are constructed by applying logical operations to simple propositions
  - Bloomington is a town and Indiana is a state – *p* and *q*
- *Logic Operations*

Conjunction	[and]	$\wedge$
Disjunction	[or]	$\vee$
Negation	[not]	$\neg$
Conditional	[implies]	$\Rightarrow$ (if, then)
Biconditional	[equivalent]	$\Leftrightarrow$ (if and only if)





Symbolic Logic

- Logic uses a set of symbols and rules to represent the structure of reasoning with precision.
- This kind of logic is known as ***symbolic logic*** and divides in **propositional** and **predicate** logic.
  - A formal system for representing knowledge in terms of ***declarative sentences*** that express ***propositions***
    - **Proposition** is the meaning of the sentence, rather than the sentence itself

Conjunction

- Conjunction: $p \wedge q$ corresponds to English "and."
- Proposition $p \wedge q$ is true when p and q are both true.

Example – Uma is blond and clever

Truth table for conjunction:

p	q	$p \wedge q$
0	0	0
0	1	0
1	0	0
1	1	1

Adapted from C. Heeren

Disjunction

- Disjunction: $p \vee q$ corresponds to English “or.”
- Proposition $p \vee q$ is true when p or q (or both) is true.

Example – Madonna is blond or clever

Truth table for disjunction:

p	q	$p \vee q$
0	0	0
0	1	1
1	0	1
1	1	1

Adapted from C. Heeren

Implication or Conditional

Implication: $p \Rightarrow q$ corresponds to English "if…then…"
or "*p* only if q"

Example: Brandon will pass I101, only if he is awake in classes

Truth table for implication:

Antecedent	Consequent	
p	q	$p \Rightarrow q$
0	0	1
0	1	1
1	0	0
1	1	1

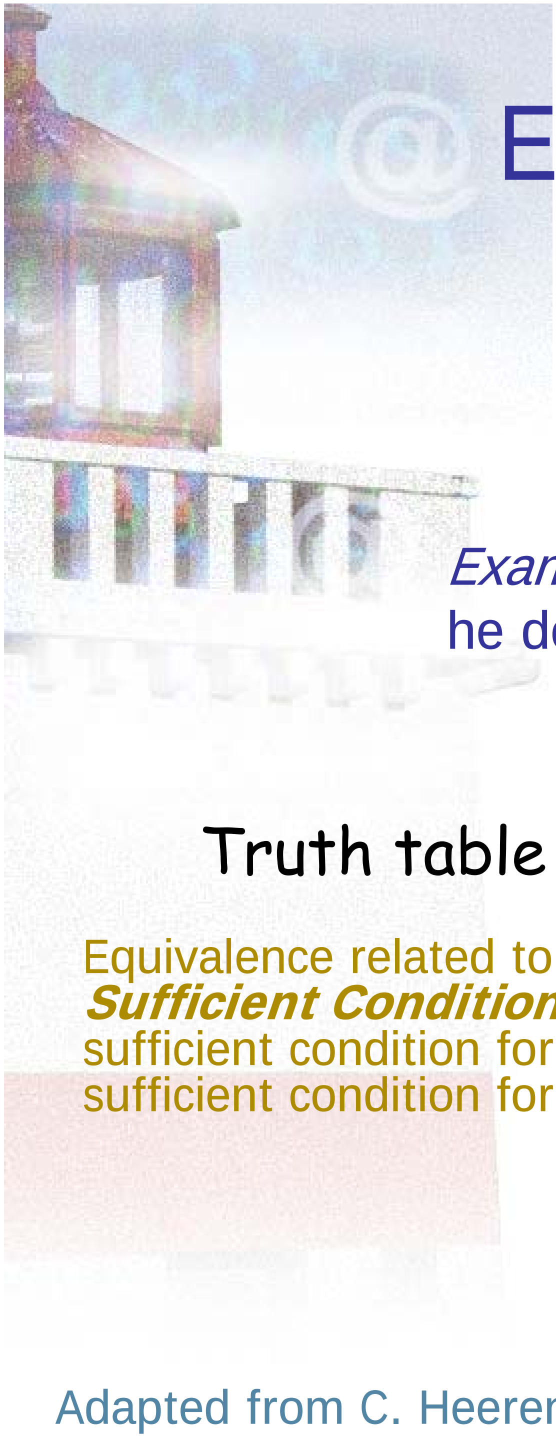
q is a ***necessary condition*** for *p*

p is a ***sufficient condition*** for *q*

Having a microscope (or some other instrument) is a **necessary condition** for (our) seeing viruses

If someone sees viruses, **then** that person uses a microscope

Adapted from C. Heeren



Equivalence or Biconditional

Equivalence: $p \Leftrightarrow q$ corresponds to English “if and only if…then…”

$(p \Rightarrow q) \wedge (q \Rightarrow p)$

Example – Brandon will pass I101 if and only if he doesn’t sleep in classes

Truth table for equivalence:

Antecedent	Consequent	
p	q	$p \Leftrightarrow q$
0	0	1
0	1	0
1	0	0
1	1	1

Equivalence related to ***Necessary and Sufficient Condition***: *q* is a necessary and sufficient condition for *p* and *p* is a necessary and sufficient condition for *q*

Adapted from C. Heeren

Proof for the distributive rule

$$p \vee (q \wedge r) \equiv (p \vee q) \wedge (p \vee r)$$

p	q	r	$q \wedge r$	$p \vee (q \wedge r)$	$p \vee q$	$p \vee r$	$(p \vee q) \wedge (p \vee r)$
0	0	0	0	0	0	0	0
0	0	1	0	0	0	1	0
0	1	0	0	0	1	0	0
0	1	1	1	1	1	1	1
1	0	0	0	1	1	1	1
1	0	1	0	1	1	1	1
1	1	0	0	1	1	1	1
1	1	1	1	1	1	1	1

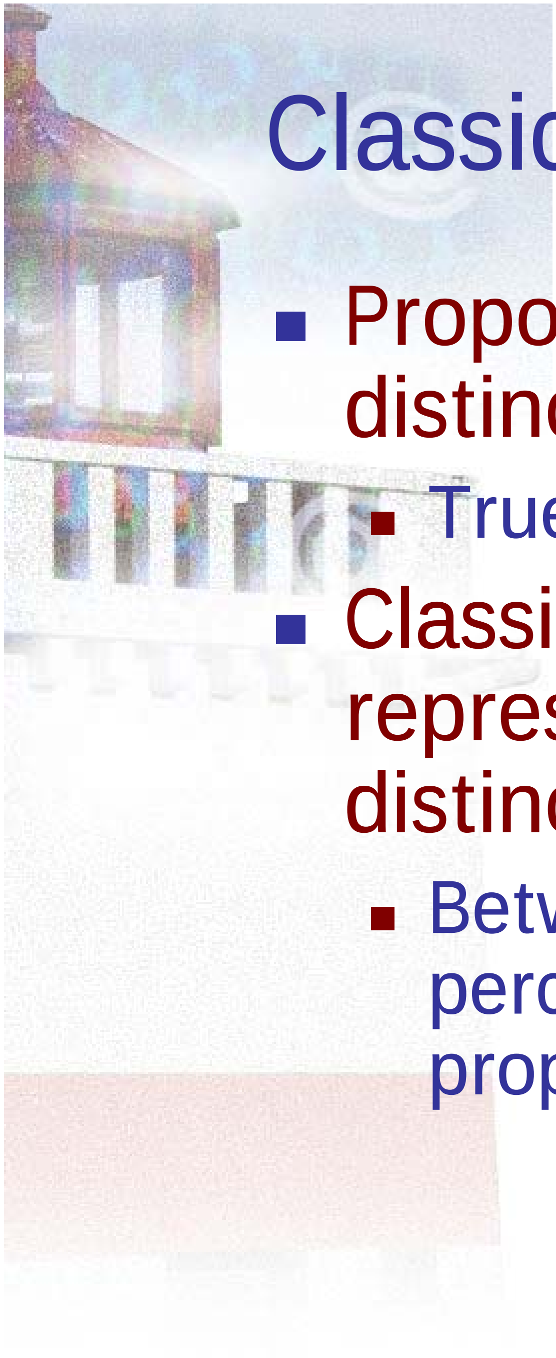
Adapted from C. Heeren

Proofs for De Morgan's Law I

$$\neg(p \vee q) \equiv \neg p \wedge \neg q$$

p	q	$\neg p$	$\neg q$	$\neg p \wedge \neg q$	$p \vee q$	$\neg(p \vee q)$
0	0	1	1	1	0	1
0	1	1	0	0	1	0
1	0	0	1	0	1	0
1	1	0	0	0	1	0

Adapted from C. Heeren



Classical Set Theory

- Propositional logic helps us make distinctions.
  - True and False, tautologies, contradictions
- Classical set theory is another form of representing the same kind of distinctions
  - Between and among groups that we perceive to share a characteristic or property.

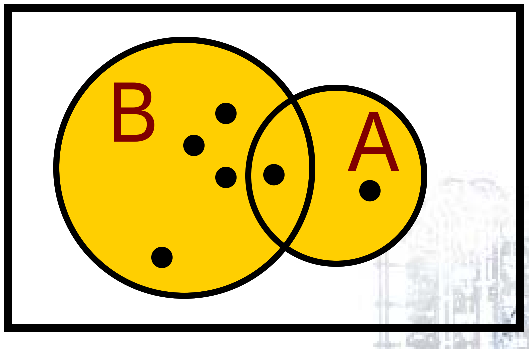
Set Operations

The *union* of two sets A and B is:

$A \cup B = \{ x : x \in A \vee x \in B\}$

If A = {Charlie, Lucy, Linus},
and B = {Lucy, Desi}, then

A ∪ B = {Charlie, Lucy, Linus, Desi}



Adapted from C. Heeren

Luis M.Rocha and Santiago Schnell

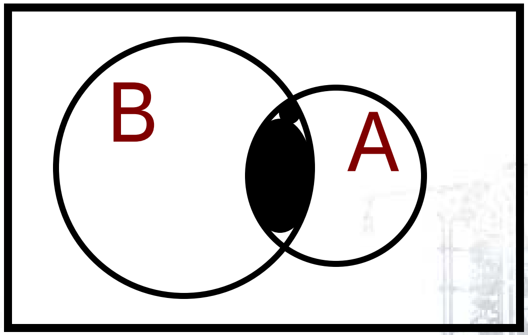
Set Operations

The *intersection* of two sets A and B is:

$$A \cap B = \{ x : x \in A \wedge x \in B\}$$

If $A = \{x : x \text{ is a US president}\}$, and $B = \{x : x \text{ is deceased}\}$, then

$A \cap B = \{x : x \text{ is a deceased US president}\}$

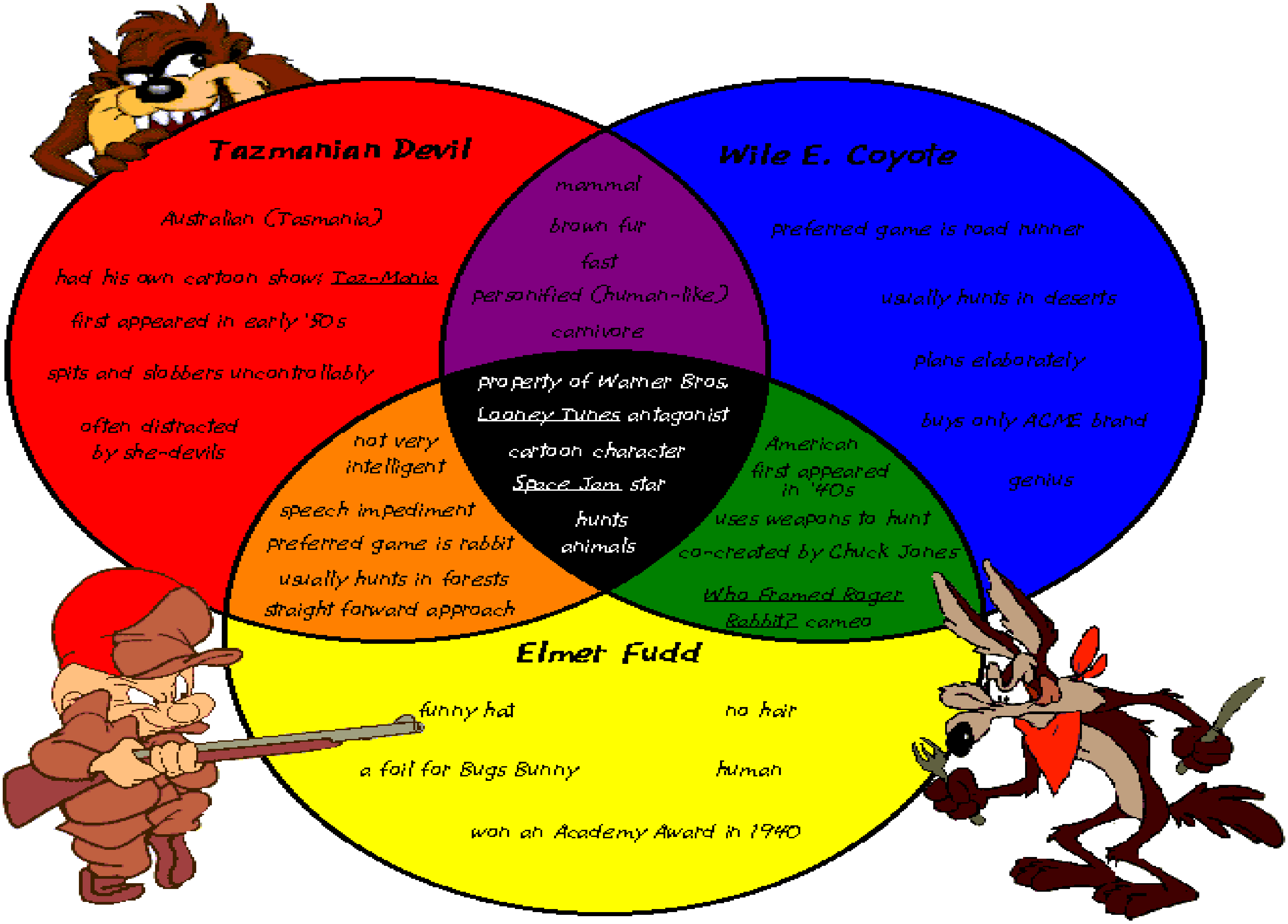


- B: Set of Funny People
- A: Set of Clowns

Adapted from C. Heeren

Luis M.Rocha and Santiago Schnell

Qualities of the Tazmanian Devil, Wile E. Coyote, and Elmer Fudd



Tazmanian Devil

Australian (Tasmania)

had his own cartoon show: Taz-Mania

first appeared in early '90s

spits and slobbers uncontrollably

often distracted by she-devils

Wile E. Coyote

preferred game is road runner

usually hunts in deserts

plans elaborately

buys only ACME brand

genius

mammal

brown fur

fast

personified (human-like)

carnivore

property of Warner Bros.

Looney Tunes antagonist

cartoon character

Space Jam star

hunts animals

not very intelligent

speech impediment

preferred game is rabbit

usually hunts in forests

straight forward approach

American

first appeared in '40s

uses weapons to hunt

co-created by Chuck Jones

Who Framed Roger Rabbit? cameo

Elmer Fudd

funny hat

no hair

a foil for Bugs Bunny

human

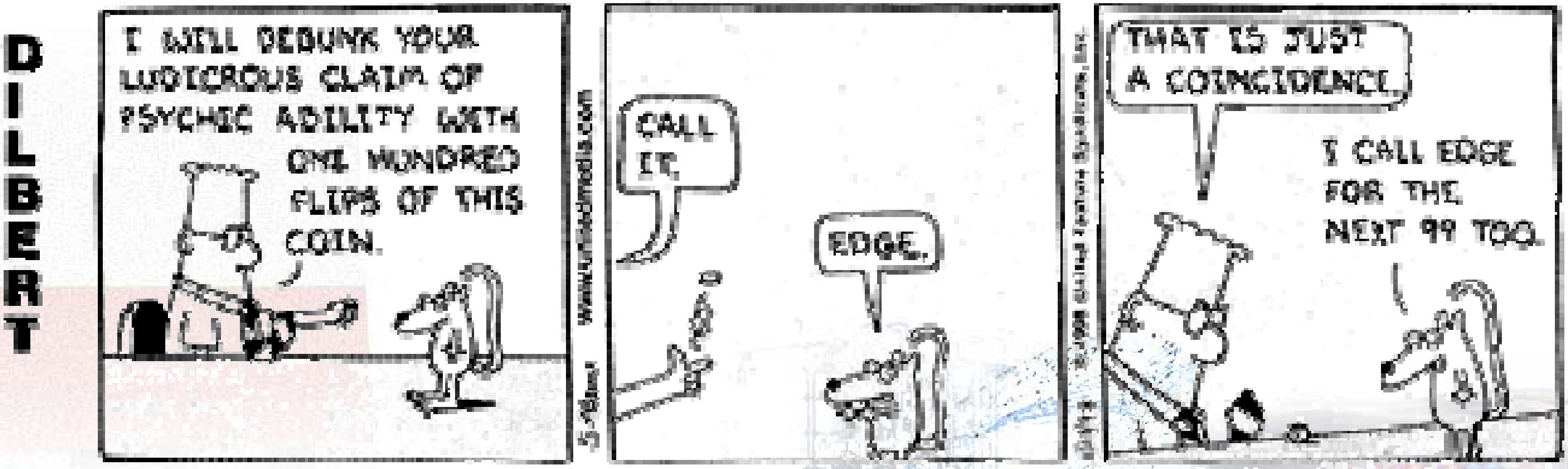
won an Academy Award in 1940

Questions

- Build the truth tables of the following:
  - $\neg\neg a$
  - $a \wedge (b \wedge \neg a)$
  - $(a \Rightarrow b) \Rightarrow \neg b$
- Given a Universal Set X = {0, 1, 2, 3, 4, 5}, let A={0,1,2}, B={2,3,4} and C={1,2,3,5}.
  - Show that $(A \cap B) \cup C \neq (A \cup B) \cap C$
  - Express {5} in terms of A, B, and C using set operations
  - Express {2} in terms of A, B, and C using set operations

Inductive Modeling

Statistics, Probability, Fitting Data, Induction


DILBERT
I WILL DEBUNK YOUR LUDICROUS CLAIM OF PSYCHIC ABILITY WITH ONE HUNDRED FLIPS OF THIS COIN.
www.unitedmedia.com
CALL IT.
EDGE.
© 1998 United Feature Syndicate, Inc.
THAT IS JUST A COINCIDENCE.
I CALL EDGE FOR THE NEXT 99 TOO.

Summarizing Data

- Frequency
  - Number of times a value occurs in a collection
- Frequency Distribution
  - Given a collection of data values, the specification of all the distinctive values in the collection together with the number of times each of these values ***occurs*** in the collection
    - Table that organizes data into mutually exclusive classes
    - Shows number of observations from data set that fall into each class

[Chase and Brown, “General Statistics”]

Frequency Distribution

Sorted Data: 30 data values

15.2	15.2	15.3	15.3	15.3	15.3	15.3	15.4	15.4	15.4
15.4	15.4	15.4	15.4	15.4	15.4	15.4	15.4	15.5	15.5
15.5	15.5	15.5	15.5	15.6	15.6	15.6	15.7	15.7	15.7

Frequency Distribution

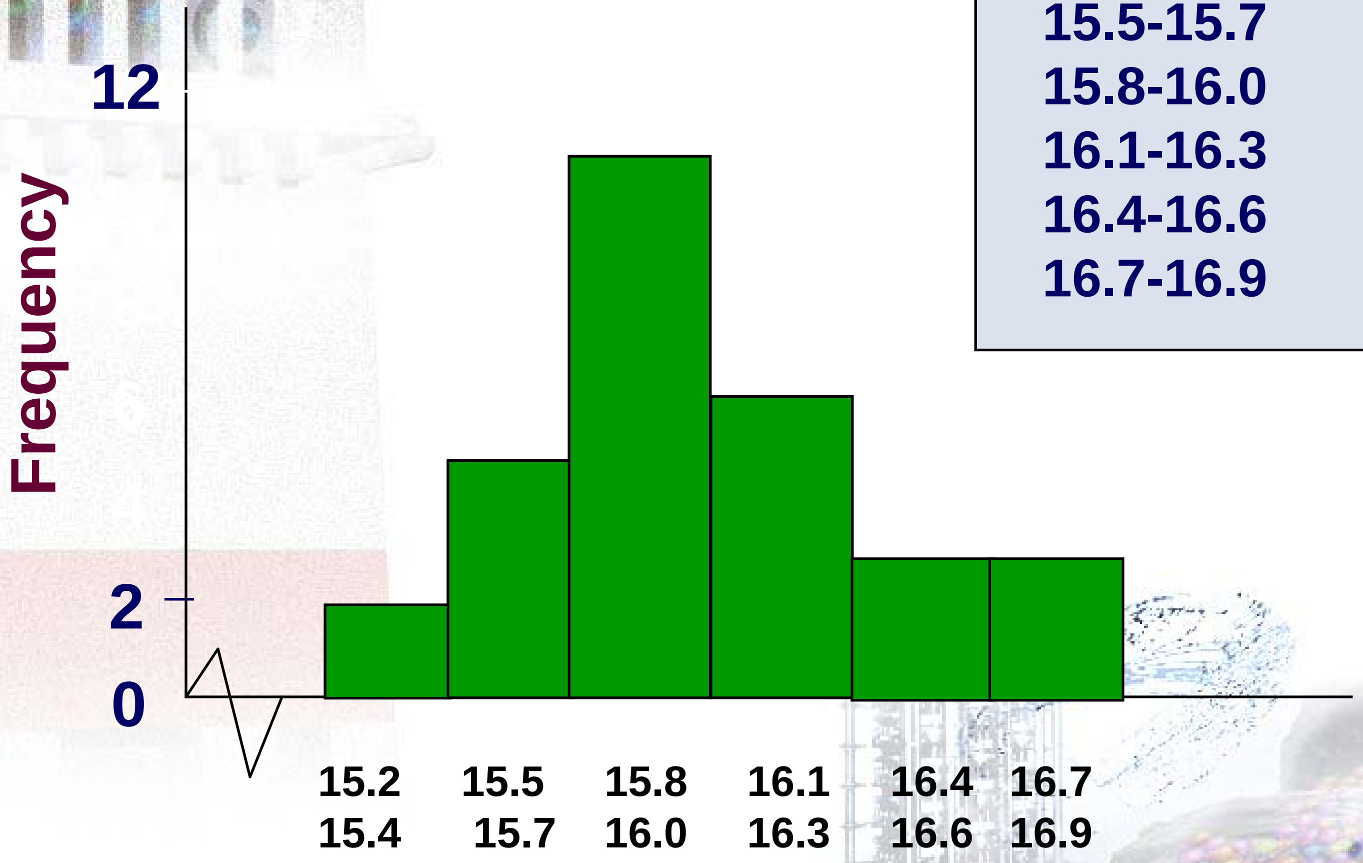
Class	Tallies	Frequency
15.2	//	2
15.3	~~////~~	5
15.4	~~////~~ ~~////~~ /	11
15.5	~~////~~ /	6
15.6	///	3
15.7	///	3

Relative Frequency Distribution

Relative Frequency Distribution	Class	Frequency (1)	Relative Freq. (1) ÷ 30	Cumulative Relative Frequency
	15.2	2	0.07	0.07
	15.3	5	0.16	0.23
	15.4	11	0.37	0.60
	15.5	6	0.20	0.80
	15.6	3	0.10	0.90
	15.7	3	0.10	1.00
		30	1.00	

Frequency Histogram

Class	Frequency
15.2-15.4	2
15.5-15.7	5
15.8-16.0	11
16.1-16.3	6
16.4-16.6	3
16.7-16.9	3



Luis M.Rocha and Santiago Schnell

Measuring Central Tendency

The number that is meant to convey the idea of a ***typical*** or representative value for the data array or distribution

Luis M.Rocha and Santiago Schnell

Mean (Ungrouped Data)

Population Mean

$$\mu = \frac{\sum X}{N}$$

Sum of values of all observations

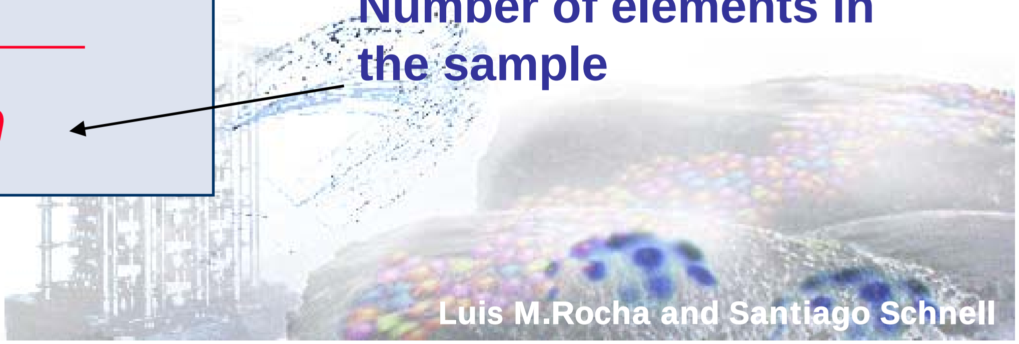
Number of elements in the population

Sample Mean

$$\overline{X} = \frac{\sum X}{n}$$

Sum of values of all observations

Number of elements in the sample



Mean

Mean: 16.01

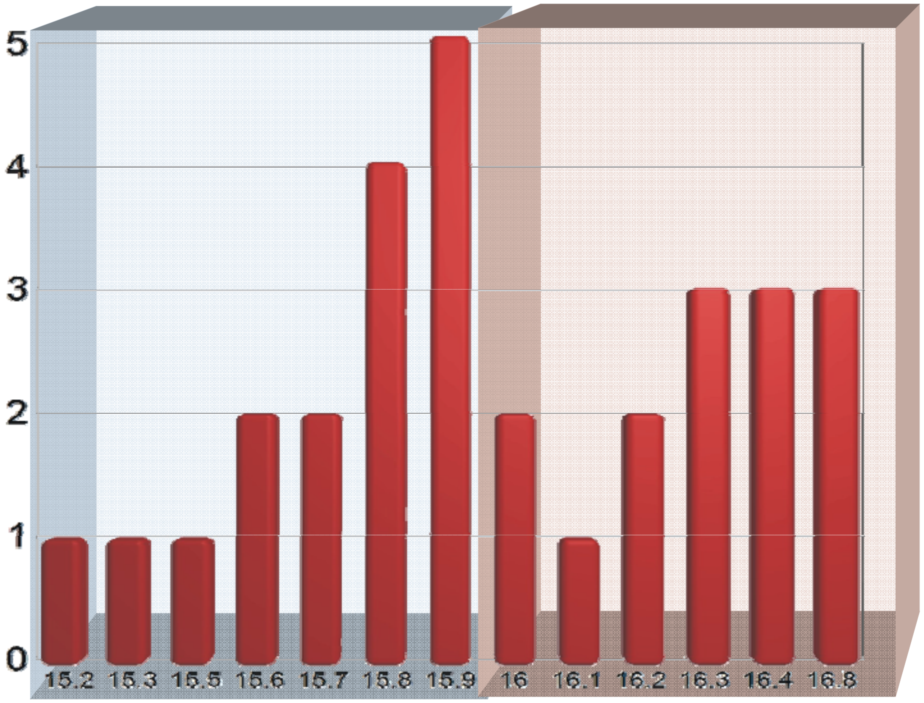
Frequency Distribution

Class	Frequency	
15.2	1	
15.3	1	
15.5	1	
15.6	2	16
15.7	2	
15.8	4	
15.9	5	
16.0	2	
16.1	1	
16.2	2	14
16.3	3	
16.4	3	
16.8	3	

Carpet Loom Example

480.3

30



- The mean is the balancing point of the frequency distribution
- Histograms balance at the mean

Central Tendency Measures

Measure	Equation	Description
Mean	$\Sigma x \; / \; n$	Balance Point
Median	$\frac{(n+1)}{2}$ *th* item in array	Middle value in ordered array
Mode	none	Most frequent

Measuring Dispersion

The number that conveys an idea of how much ***spread*** or ***variability*** exists among the data values

Sample Variance Formula

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

$$= \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}$$

n - 1 in denominator!
(Use ***N*** if **Population** Variance)

Where:
n **- sample size**
x - item or observation
$\bar{x}$ ***- sample mean***
Σ ***- sum of the values***

Example: Sample Dispersion

Frequency Distribution

Carpet Loom Example

Class	Frequency	
15.2	1	$(15.2 - 16.01)^2 = 0.6561$
15.3	1	$(15.3 - 16.01)^2 = 0.5041$
15.5	1	$(15.5 - 16.01)^2 = 0.2601$
15.6	2	$2 \times (15.6 - 16.01)^2 = 0.3362$
15.7	2	$2 \times (15.7 - 16.01)^2 = 0.1922$
15.8	4	$4 \times (15.8 - 16.01)^2 = 0.1764$
15.9	5	
16.0	2	• • •
16.1	1	
16.2	2	
16.3	3	
16.4	3	
16.8	3	
24	1	

Mean: 16.01

8.8

Range = 16.8-15.2 = 1.6

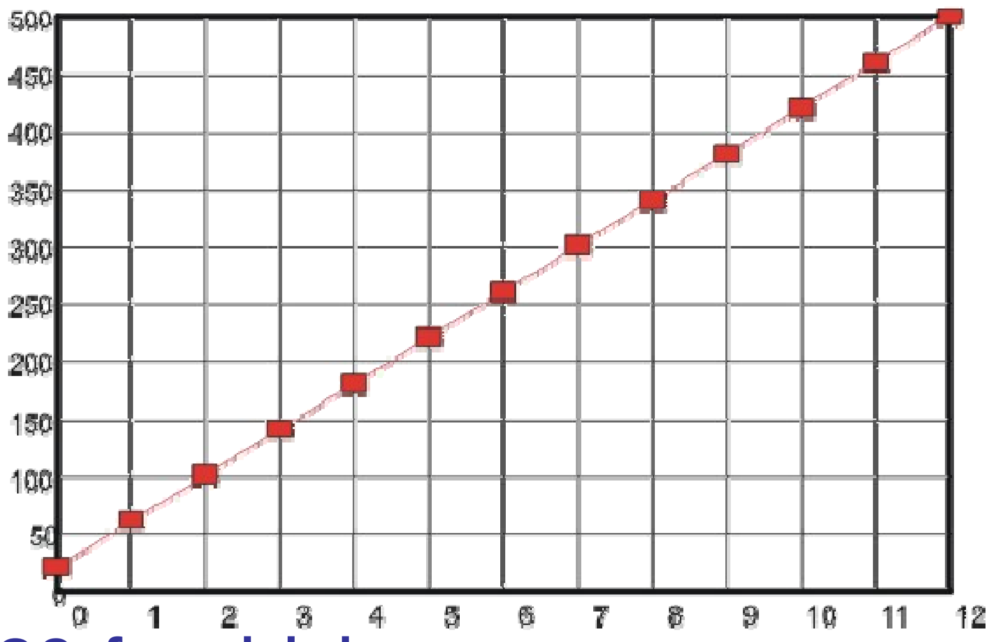
Standard Deviation = $\sqrt{0.167138} = 0.408825$

1.488527

Variance 4.847/30 = 0.167138

2.215713

Linear Relationship



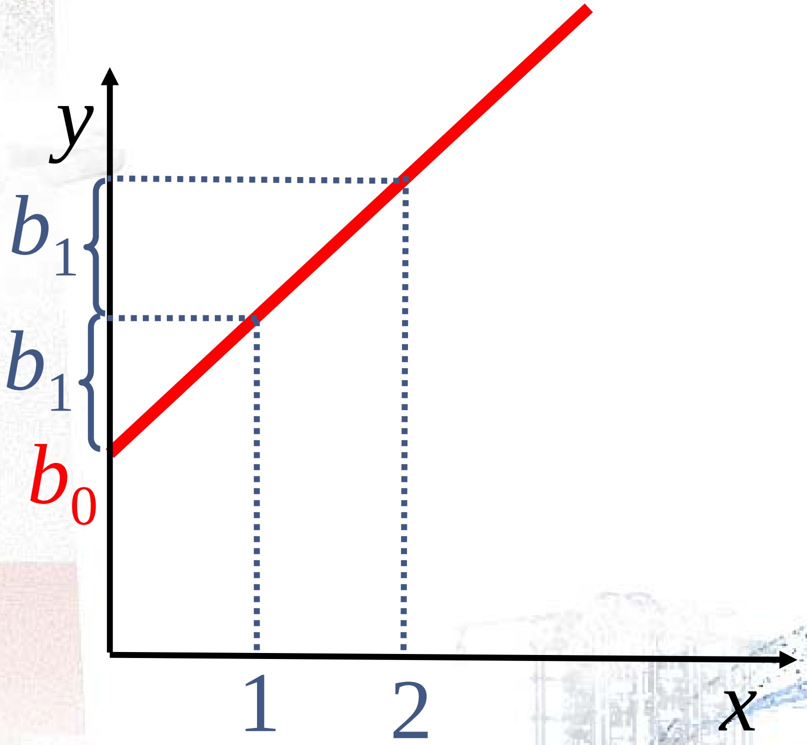
- Example
  - A plumber charges $20 for driving to your house, plus $40 for each hour of work at your home
  - Let
    - y = total charge
    - x = number of hours of work at your house
  - The relationship between y and x is
    - y=20 + 40x

y intercept

Slope

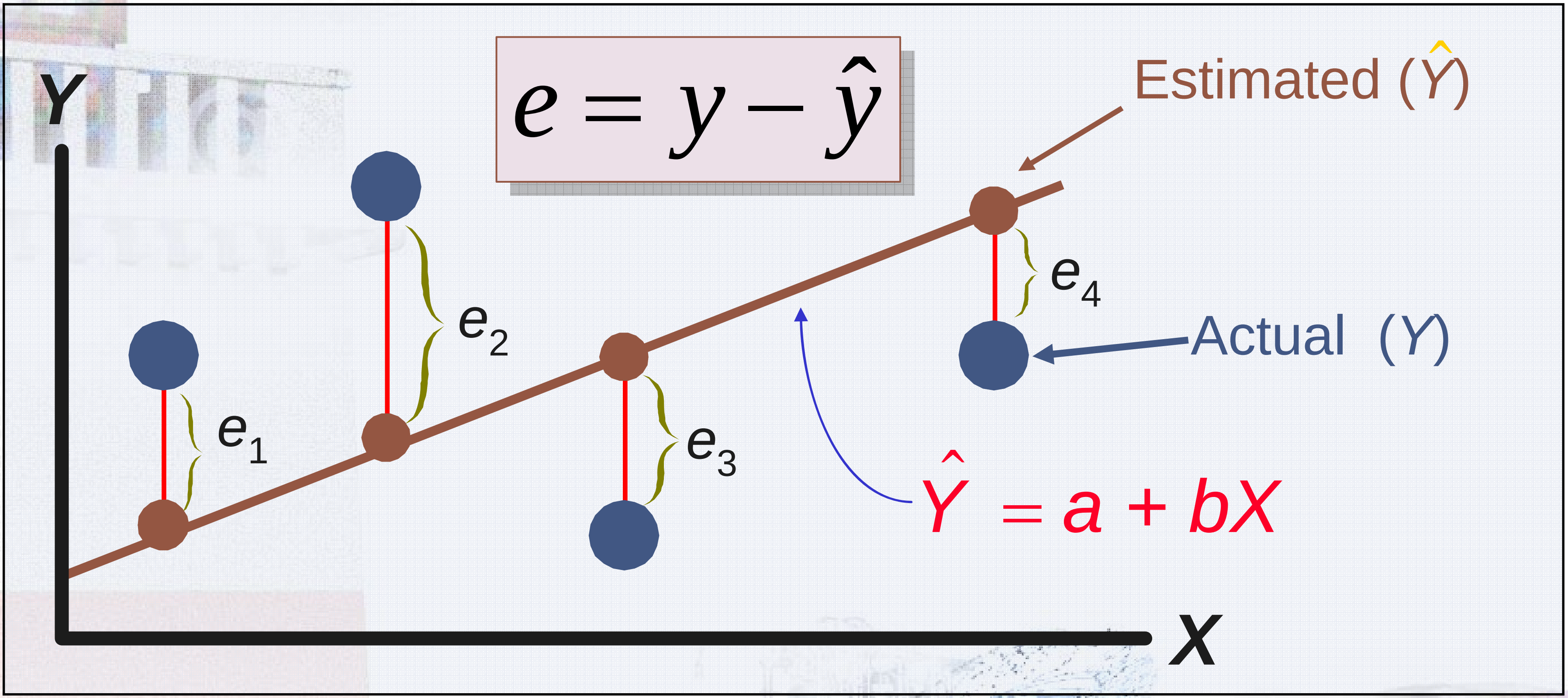
General Linear Relationship

$$y = b_0 + b_1 x$$



Luis M.Rocha and Santiago Schnell

Errors or Residuals Graphically



$$e = y - \hat{y}$$

Estimated ($\hat{Y}$)

Actual (Y)

$\hat{Y} = a + bX$

Least Squares Criterion

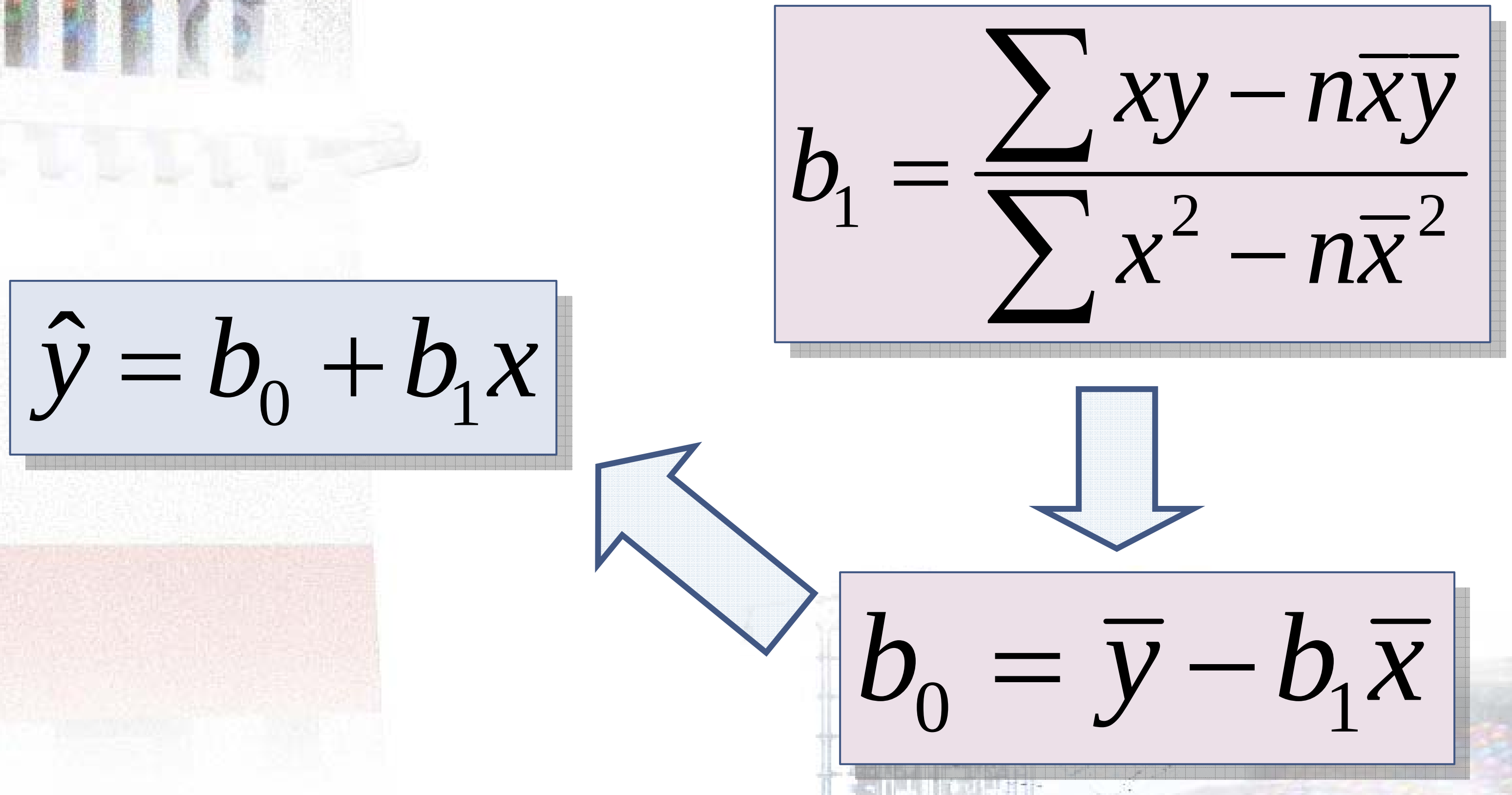
$$e = y - \hat{y}$$

- Adding all the errors may lead to very small errors due to terms cancelling out
- Squares of errors eliminates this problem
- The line that best fits the data is the one for which the **sum of the squares of the errors (SSE)** is smallest

$$SSE = \sum e^2 = \sum \left(y - \hat{y}\right)^2$$

Method for Regression

- Line of ***best fit*** or ***regression*** based on Least Squares Criterion

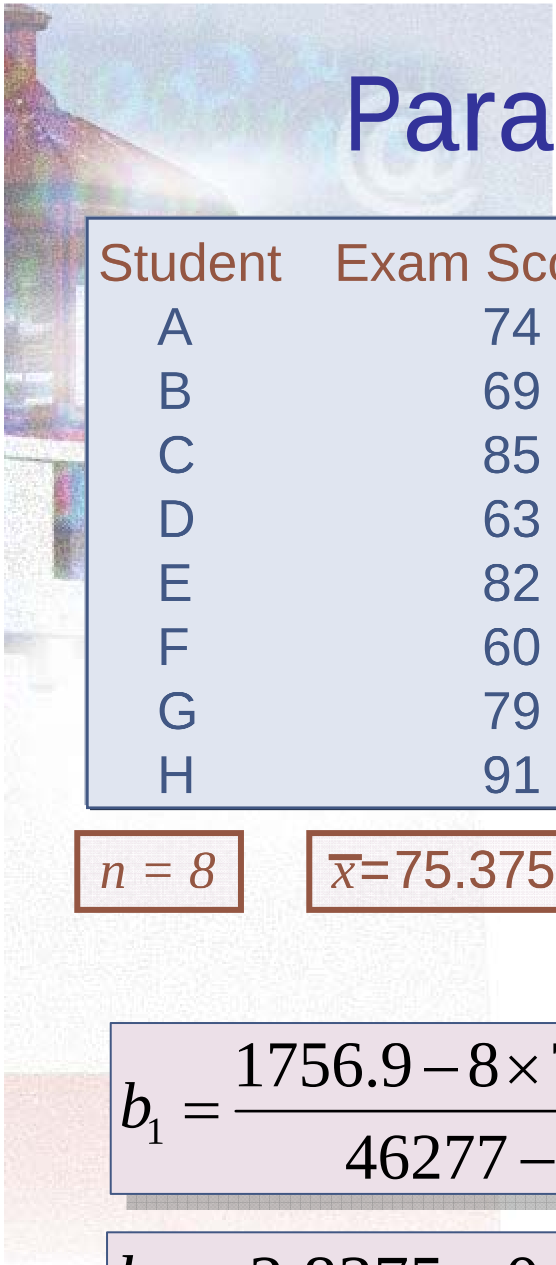


$$b_1 = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

$$\hat{y} = b_0 + b_1 x$$

Parameter Estimation Example



Student	Exam Score	G.P.A.
A	74	2.6
B	69	2.2
C	85	3.4
D	63	2.3
E	82	3.1
F	60	2.1
G	79	3.2
H	91	3.8

$n = 8$ $\quad \overline{x}=75.375$ $\quad \overline{y}=2.8375$

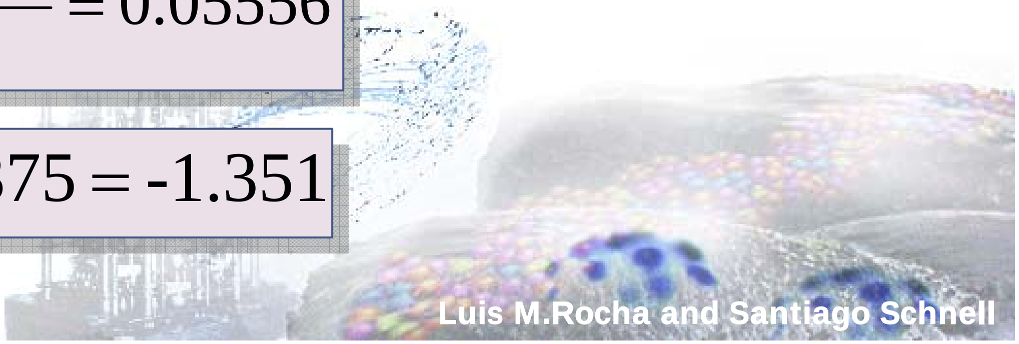
xy	x^2
192.4	5476
151.8	4761
289	7225
144.9	3969
254.2	6724
126	3600
252.8	6241
345.8	8281
--------	-------
1756.9	**46277**

$$b_1 = \frac{\sum xy - n\overline{x}\overline{y}}{\sum x^2 - n\overline{x}^2}$$

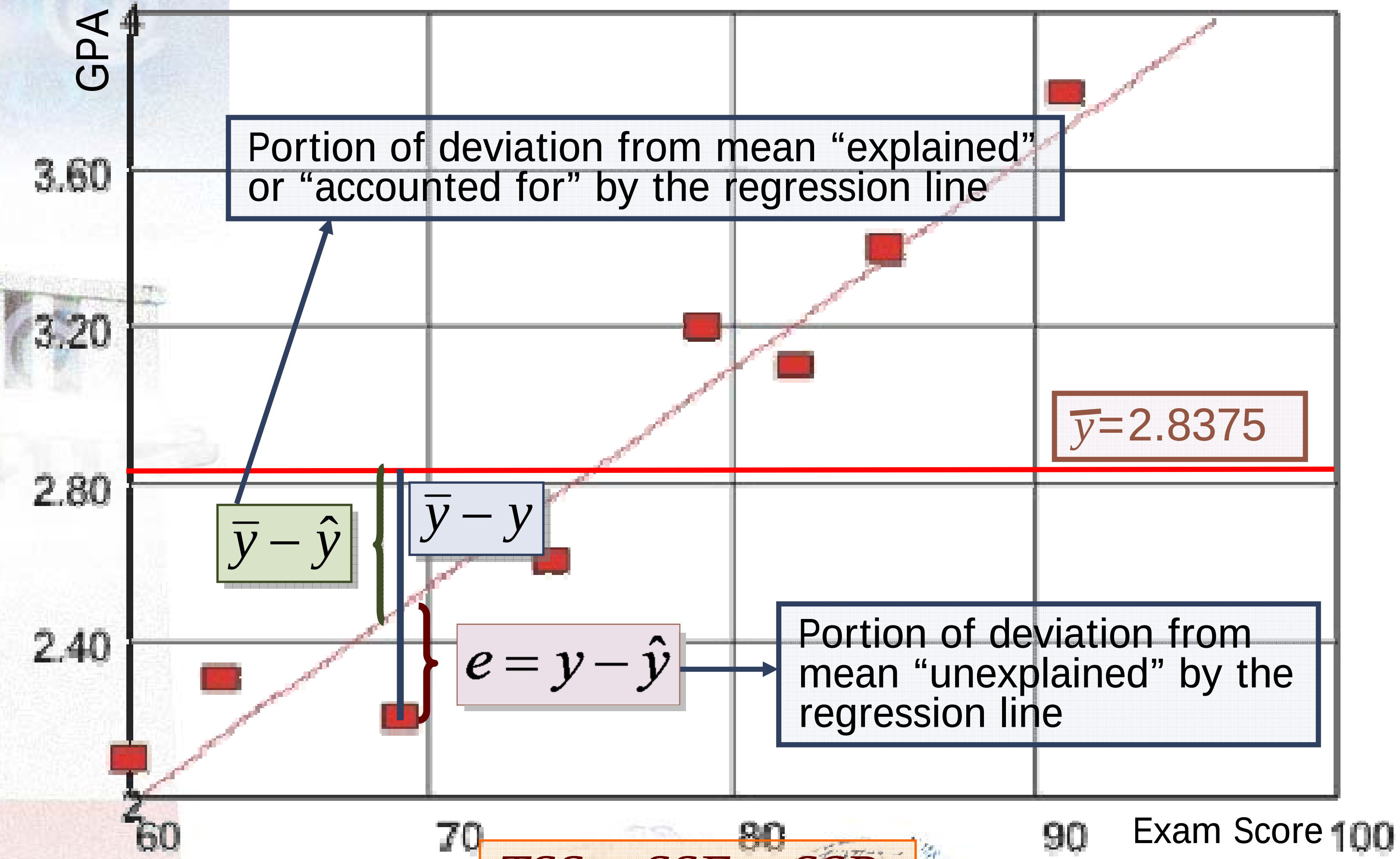
$$b_0 = \overline{y} - b_1\overline{x}$$

$$b_1 = \frac{1756.9 - 8 \times 75.375 \times 2.8375}{46277 - 8 \times 75.375^2} = 0.05556$$

$$b_0 = 2.8375 - 0.05556 \times 75.375 = \text{-}1.351$$



How good is the Linear Fit?



Portion of deviation from mean "explained" or "accounted for" by the regression line

$\bar{y}=2.8375$

$\bar{y}-\hat{y}$

$\bar{y}-y$

$e=y-\hat{y}$

Portion of deviation from mean "unexplained" by the regression line

$TSS = SSE + SSR$

Total Sum of Squares

$$TSS=\sum(\bar{y}-y)^2$$

Sum of Squares for error

$$SSE=\sum(y-\hat{y})^2$$

Sum of Squares for regression

$$SSR=\sum(\bar{y}-\hat{y})^2$$

Luis M.Rocha and Santiago Schnell

Coefficient of determination

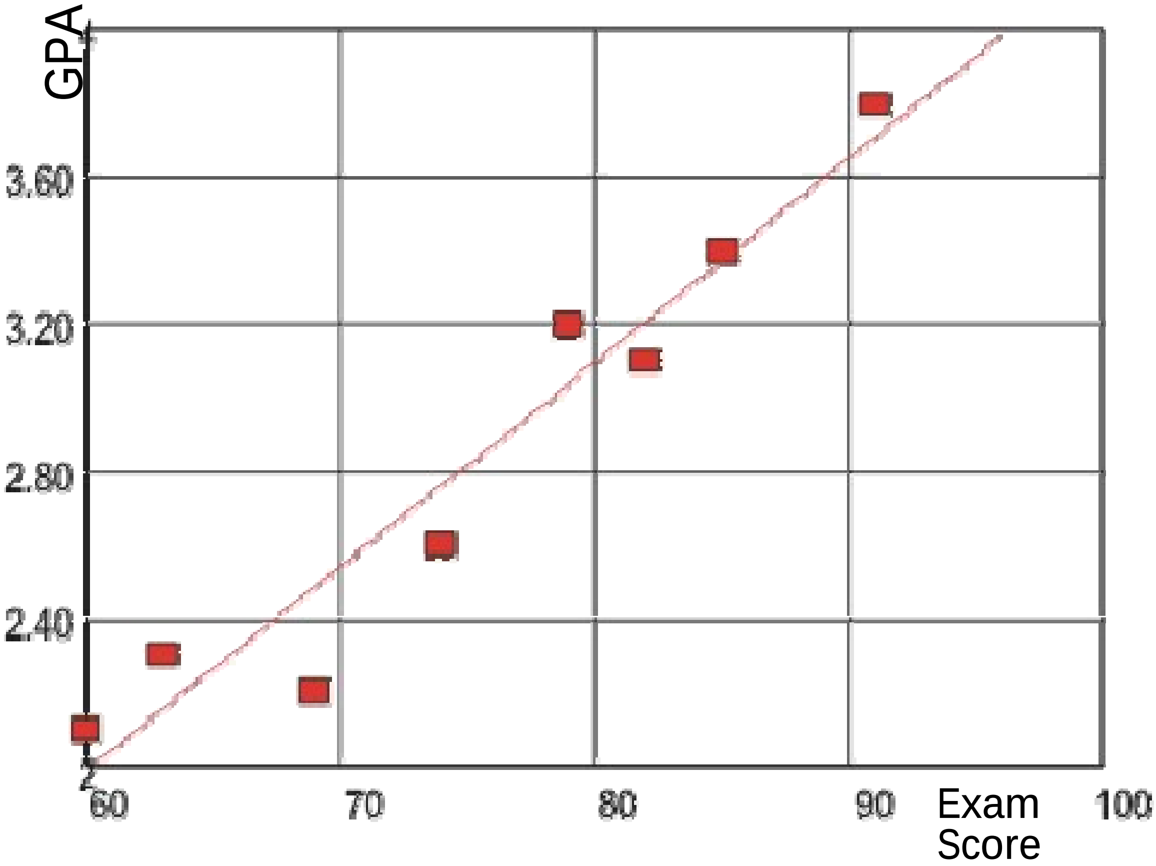
$$\boxed{r^2}$$

- Degree of linear relationship
  - To make a judgment about whether a linear relationship really exists between *x* and *y*.
  - The *proportion* of the variability in y values that is accounted for or *explained by* a linear relationship with x.

$$\boxed{r^2 = \frac{SSR}{TSS} = \frac{\sum\left(\bar{y} - \hat{y}\right)^2}{\sum\left(\bar{y} - y\right)^2}}$$

Example: Coefficient of determination

$$R^2 = 0.93$$



93% of the GPA variability is explained by the Exam Score (with a linear relationship)

How do we assign probability to an event?

- The probability of an event *A* in an experiment is supposed to measure how ***frequently*** A is about to occur if we make many trials.
  - If we flip our coin many times, H and T will appear *about* equally often – we say that H and T are “equally likely”.

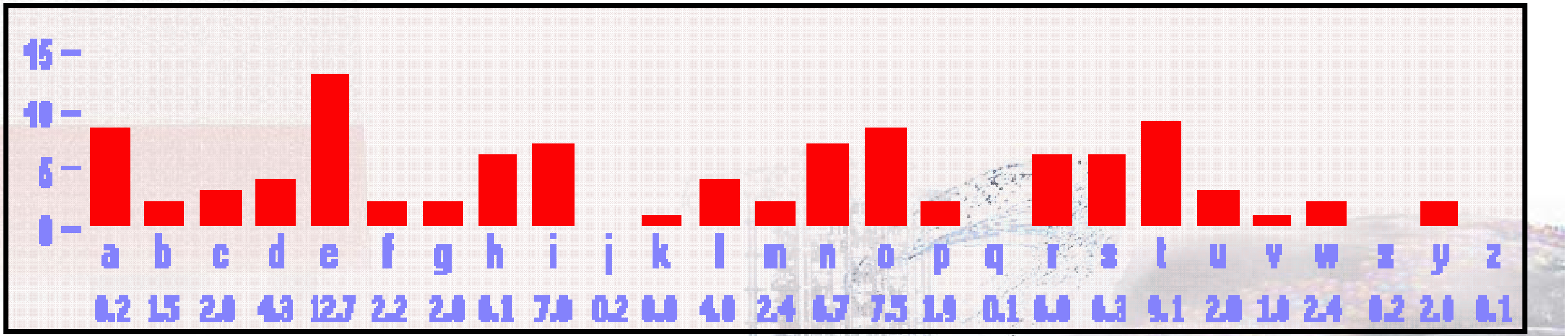
TRUE!
by Daryl Cagle
Source:
You are about 1,000 times more likely to die while playing football this year, than you are to win the 17 state Powerball jackpot.

We regard ***probability*** as the counterpart of ***relative frequency***!

Luis M.Rocha and Santiago Schnell

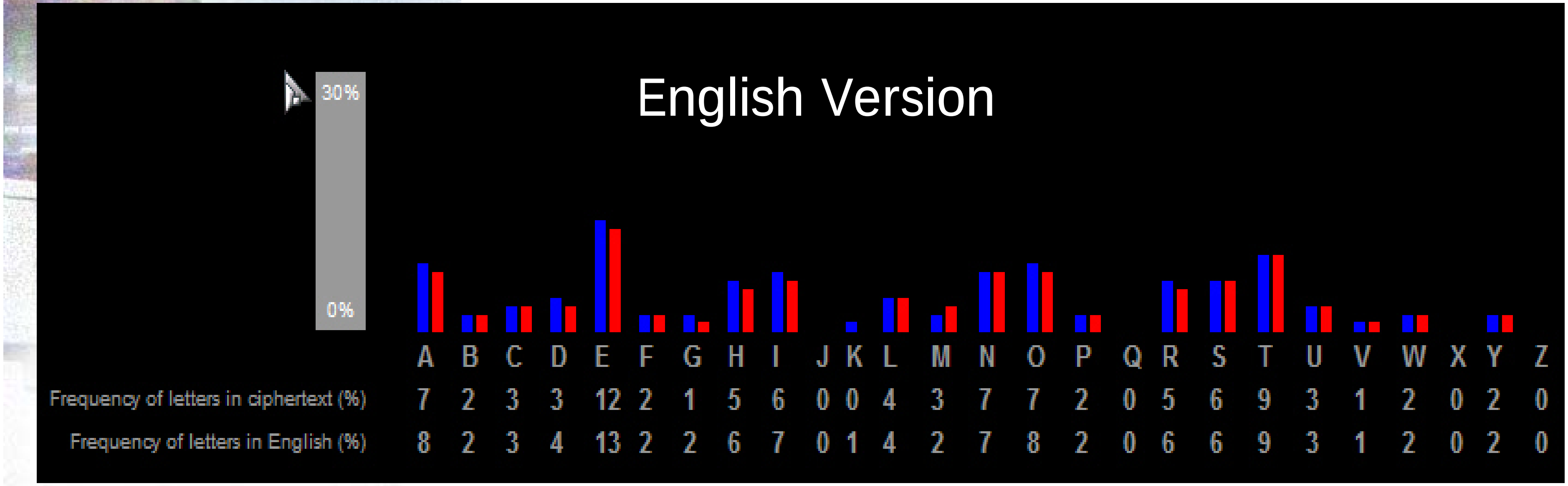
English Letter Frequency

- Six most common letters in English
  - E T A O I N
- Spanish
  - E A O S R N
- Other Languages
  - http://people.bath.ac.uk/tab21/forcrypt.html



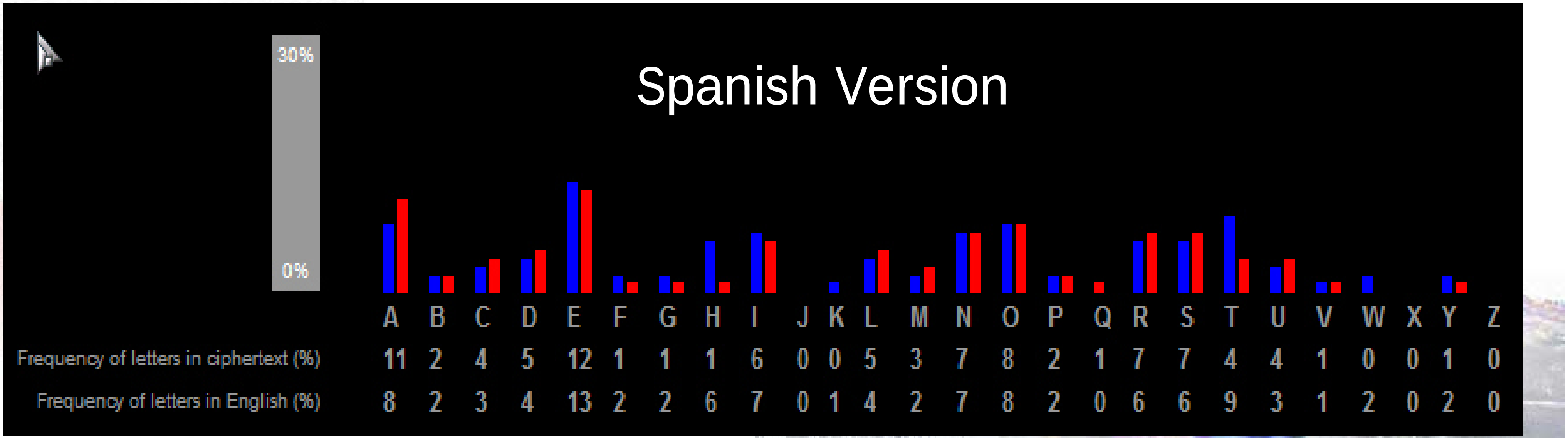
Luis M.Rocha and Santiago Schnell

Lottery of Babylon



English Version

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Frequency of letters in ciphertext (%)	7	2	3	3	12	2	1	5	6	0	0	4	3	7	7	2	0	5	6	9	3	1	2	0	2	0
Frequency of letters in English (%)	8	2	3	4	13	2	2	6	7	0	1	4	2	7	8	2	0	6	6	9	3	1	2	0	2	0



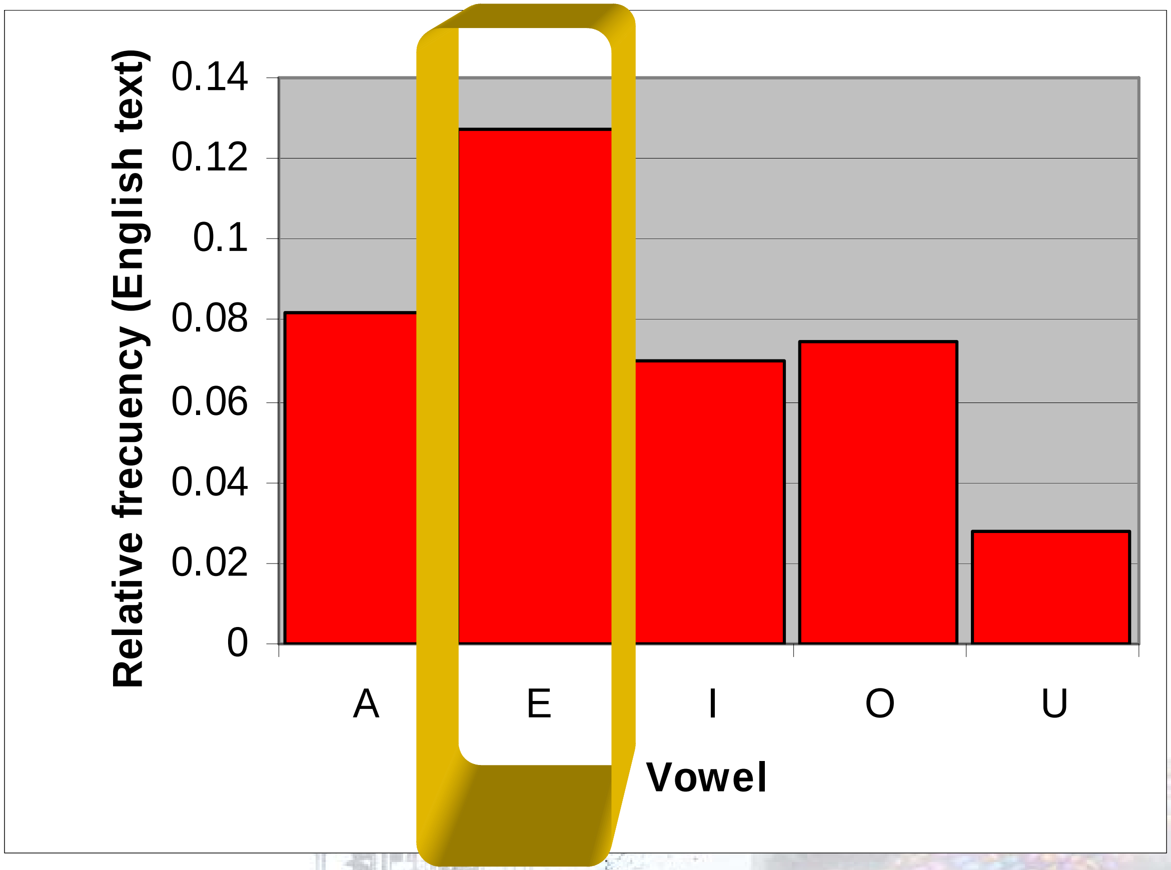
Spanish Version

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Frequency of letters in ciphertext (%)	11	2	4	5	12	1	1	1	6	0	0	5	3	7	8	2	1	7	7	4	4	1	0	0	1	0
Frequency of letters in English (%)	8	2	3	4	13	2	2	6	7	0	1	4	2	7	8	2	0	6	6	9	3	1	2	0	2	0

Relative Frequency-Probability

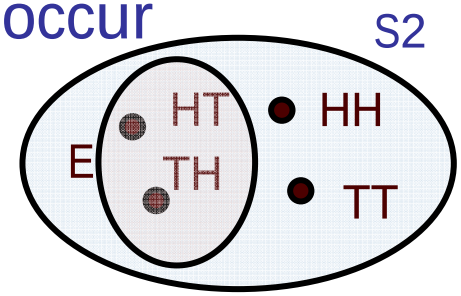
What is the probability of finding the letter “e” in an English text?

Vowel	f_{rel}
A	0.082
E	0.127
I	0.070
O	0.075
U	0.028



Probability Notions

- Experiment
  - Any activity that yields a result or an outcome
    - Tossing a coin
- Sample Space
  - The set of all possible outcomes of an experiment.
  - One and only one of the outcome must occur
    - Flipping one coin: S1 = {H, T}
    - Flipping two coins: S2 = {HH, HT, TH, TT}



- Event
  - Subset of sample space
  - The event occurs if when we perform the experiment one of its elements occurs.
    - Non-match in two coin experiment is an event E = {HT, TH}

[Chase and Brown, “General Statistics”]

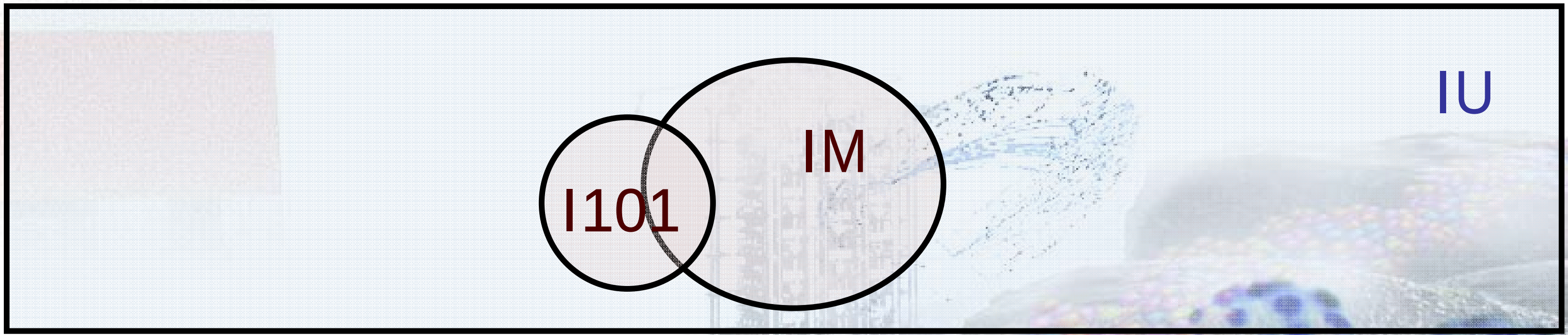
Luis M.Rocha and Santiago Schnell

Probability of an Event

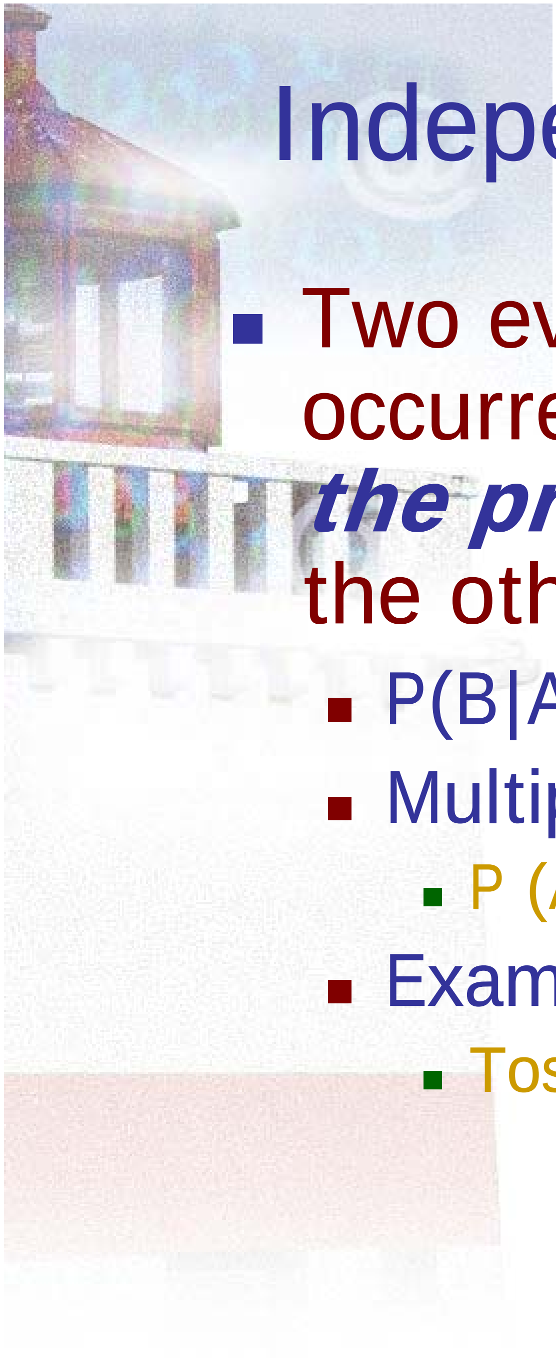
- P(A)
  - The expected proportion of occurrences of an event A if the experiment were to be ***repeated many times***
  - $\mathbf{0 \leq P(A) \leq 1}$
  - $P(A) = P(\{a_1\}) + P(\{a_2\}) + ... + P(\{a_n\})$
    - $A = \{a_1, a_2, ..., a_n\}$
  - Theoretical probability: $P(A) = |A|/|S|$ ⇨ $|A|$ = Cardinality of A (number of elements)
    - S1: $P(\{T\}) = ½$ and $P(\{H\}) = ½$
    - S2: P(*nonmatch*)
      - $P(\{HT, TH\}) = P(\{HT\}) + P(\{TH\}) = ¼ + ¼ = ½$
- Estimated from limited experiments
  - Empirical probability
    - $\{T,T,H,T\} \Rightarrow P(\{T\}) = 0.75$ and $P(\{H\}) \Rightarrow 0.25$
- Guessed
  - Subjective probability
    - “there is a 90% chance that I will pass this course”

Conditional Probability

- $P(B|A) = |A \cap B|/|A|$
  - Probability of a IU student being an Informatics major, given that a student is enrolled in I101
    - $|I101| = 70$ students
    - $|IM| = |\{\text{informatics major}\}| = 400$
    - $P(IM|I101) = |IM \cap I101|/|I101| = 35/70 = 0.5$
    - $P(IM) = 400/20000 = 0.02$
- Multiplication Rule
  - $P(A \cap B) = P(A) . P(B|A)$



Independent Events



- Two events A, B ***independent*** if the occurrence of one has ***no effect on the probability*** of the occurrence of the other
  - $P(B|A) = P(B)$
  - Multiplication Rule
    - $P(A \cap B) = P(A).P(B)$
  - Example
    - Tossing coins

Questions

- Over a 20-game period, the number of hits by a baseball player was
  - 1,2,0,0,1,2,2,1,0,0,4,0,1,1,3,2,1,3,0, and 1
    - Construct the Frequency distribution
    - In what proportion of games did he get at least 3 hits?
    - What is the mean, median, and mode
- What is the line that best fits the data with the least squares criterion?
- A coin is tossed three times and an H or T (H= Head, T=Tail) is recorded each time.
  - List the elements of the sample space S and list the elements of the event consisting of
    - All heads
    - A head on the second toss
    - Two tails
  - Represent the sample space and the events above as a Venn Diagram
- One card is to be selected from an ordinary deck of 52 cards. Find the probability that
  - The selected card is an ace
  - The selected card is not a 9

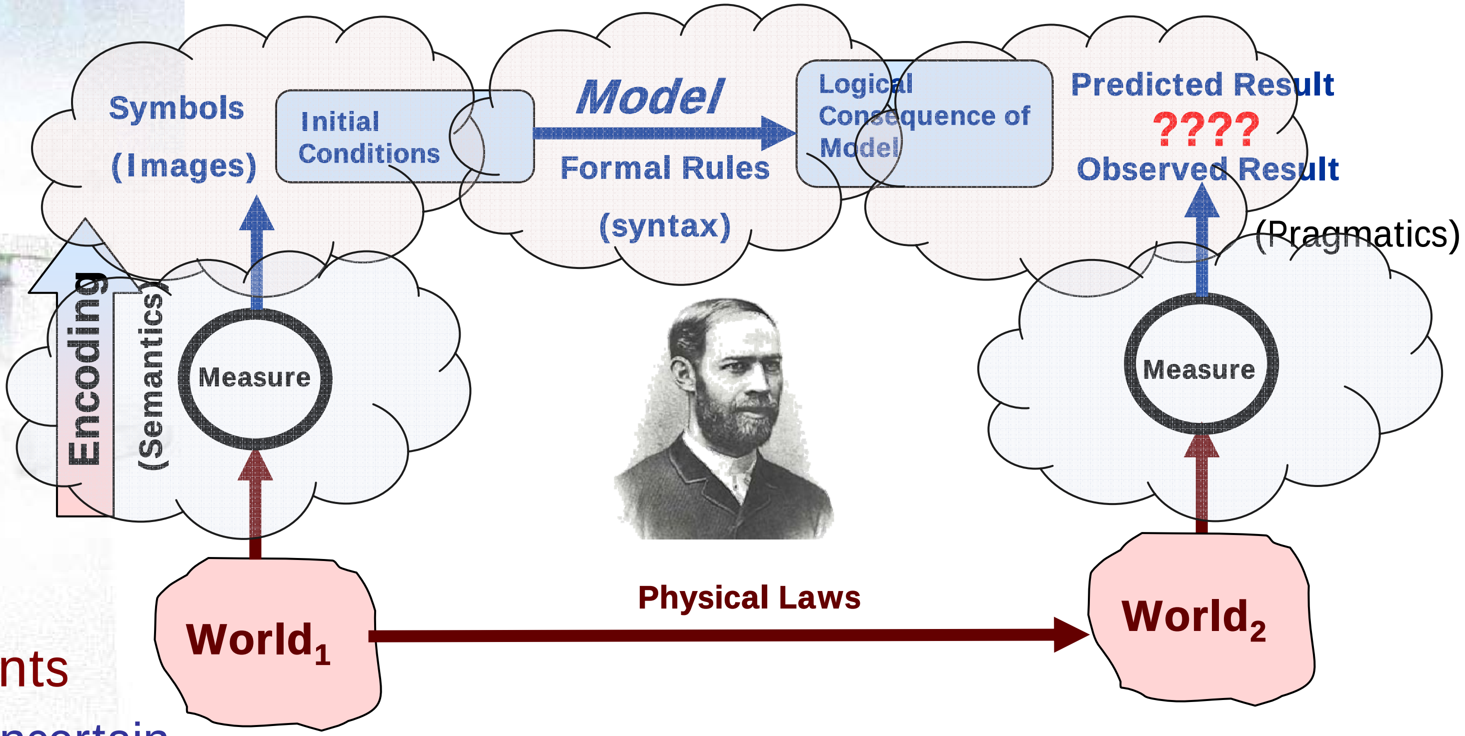
Uncertainty

Hartley and Shannon Information

"Do you think they mean us?"

Uncertainty in The Modeling Relation

Hertz' Modeling Paradigm



Symbols

(Images)

Initial Conditions

Model

Formal Rules

(syntax)

Logical Consequence of Model

Predicted Result

????

Observed Result

(Pragmatics)

Encoding

(Semantics)

Measure

Measure

World$_1$

Physical Laws

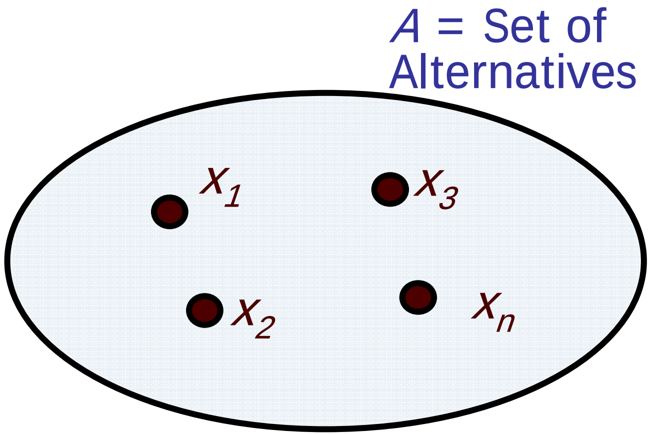
World$_2$

- Measurements
  - Always uncertain
- Limited Information
  - Induction from available evidence, especially in the presence of randomness
- Vagueness or Imprecision of Language of Description
  - “being tall” means different things to different people
- Quality of Inferences
  - Error Estimation

Let’s talk about choices

- Multiplication Principle
  - “If some choice can be made in M different ways, and some subsequent choice can be made in N different ways, then there are M x N different ways these choices can be made in succession” [Paulos]
    - 3 shirts and 4 pants = 3 x 4 = 12 outfit choices

Nonspecificity



- A type of ambiguity
  - When there are choices
- Unspecified distinctions between several alternatives
  - Variety, imprecision
  - Indiscriminate choices
- Measured by Hartley measure
  - The amount of uncertainty associated with a set of alternatives (e.g. messages) is measured by the amount of information needed to remove the uncertainty

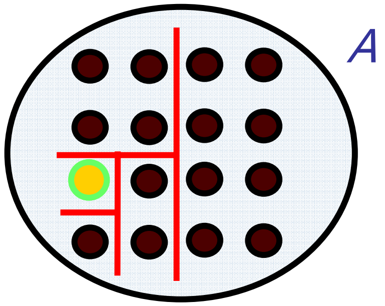
$$\boxed{H(A) = \log_2 |A|}$$

Measured in bits

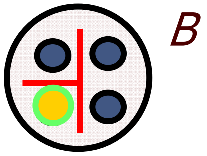
Number of Choices

Hartley Uncertainty

- Example
  - Menu Choices
    - A = 16 Entrees
    - B = 4 Desserts
  - How many dinner combinations?
    - 16 x 4 = 64



$$H(A) = \log_2(16) = 4$$



$$H(B) = \log_2(4) = 2$$

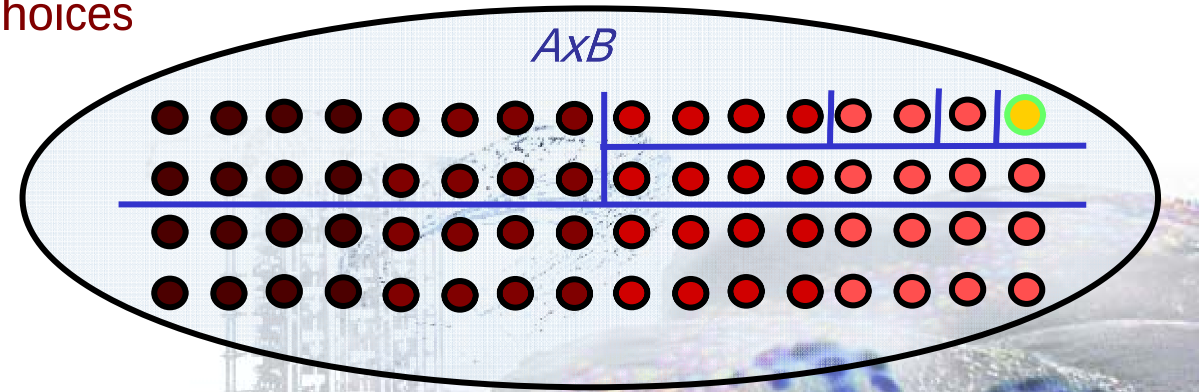
$$H(A) = \log_2 |A|$$

Measured in bits

Number of Choices

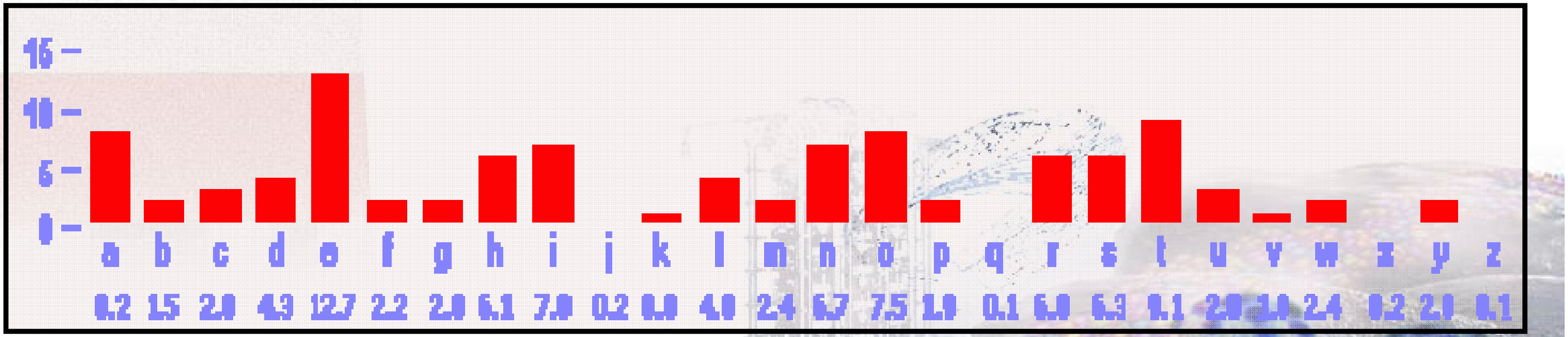
$$H(A \times B) = \log_2(16 \times 4) =$$
$$= \log_2(16) + \log_2(4) = 6$$

Quantifies how many yes-no questions need to be asked to establish what the correct alternative is

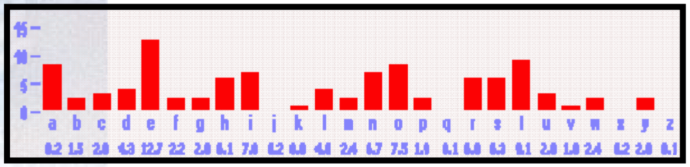


What about probability?

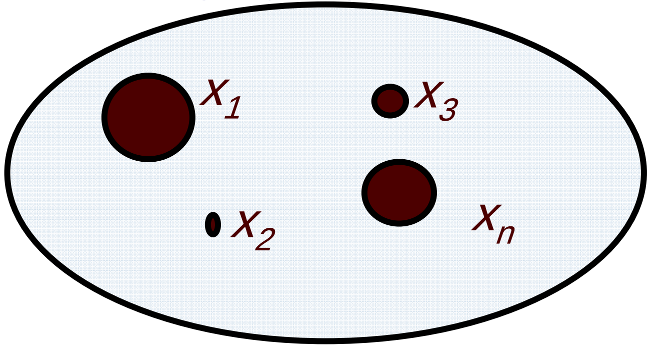
- Some alternatives may be more probable than others!
- A different type of ambiguity
  - Alternatives are distinct
    - **Conflict**, strife, discord
- Measured by Shannon’s ***entropy*** measure
  - The amount of uncertainty associated with a set of alternatives (e.g. messages) is measured by the ***average*** amount of information needed to remove the uncertainty



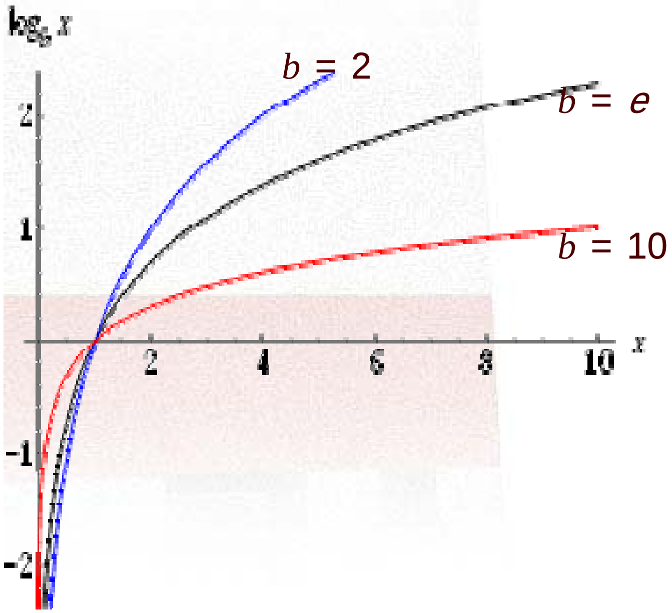
Entropy



A = Set of weighted Alternatives



- Shannon's measure
  - The ***average*** amount of uncertainty associated with a set of ***weighted*** alternatives (e.g. messages) is measured by the ***average*** amount of information needed to remove the uncertainty



$$H_S(A) = -\sum_{i=1}^{n} p(x_i)\log_2\big(p(x_i)\big)$$

Probability of alternative

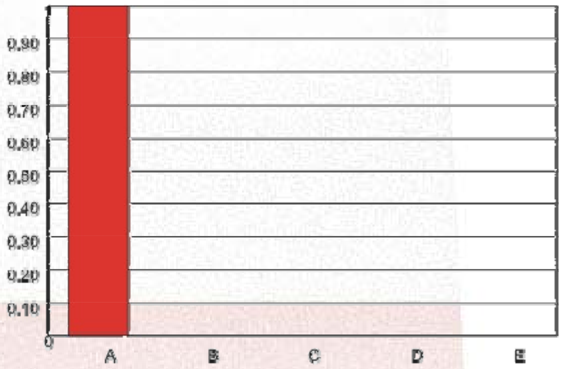
Measured in bits

Luis M.Rocha and Santiago Schnell

Example English

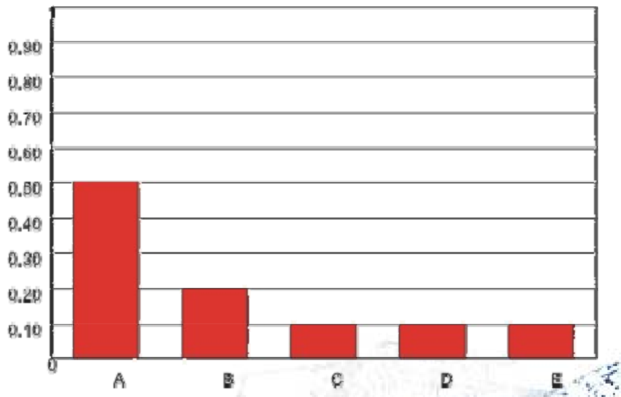
- Given a symbol set {A,B,C,D,E}
  - And occurrence probabilities P_A, P_B, P_C, P_D, P_E,
- The Shannon entropy is
  - The average minimum number of bits needed to represent a symbol

$$H_S = -\left(p_A \log_2\left(p_A\right) + p_B \log_2\left(p_B\right) + p_C \log_2\left(p_C\right) + p_D \log_2\left(p_D\right) + p_E \log_2\left(p_E\right)\right)$$



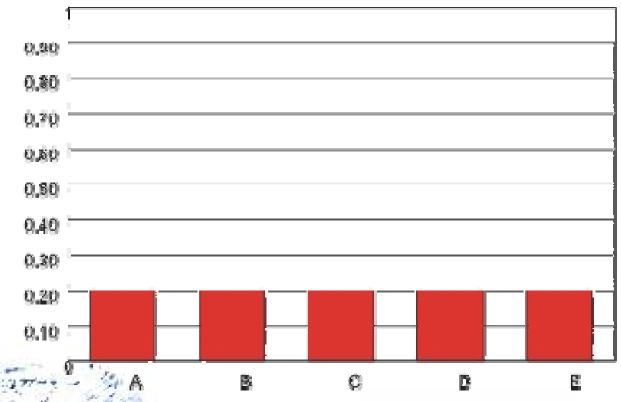
H_S = 0

0 questions



H_S = 1.96

≈2 questions



H_S = 2.32

Shannon’s entropy

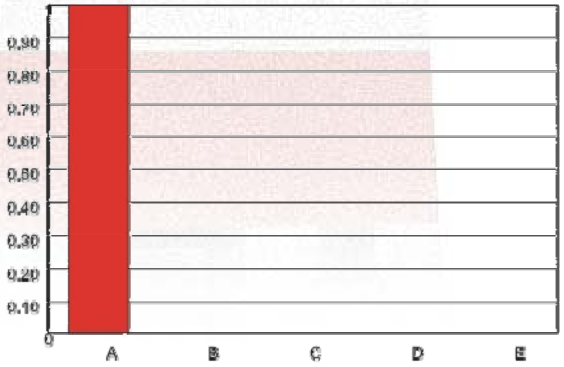
on average, how many *yes-no* questions need to be asked to establish what the symbol is.

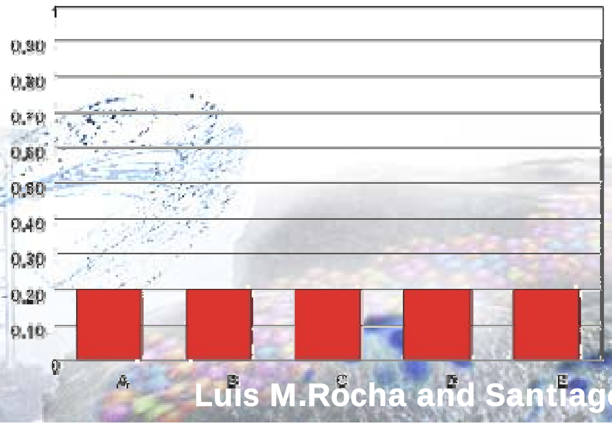
$$H_S(A) = -\sum_{i=1}^{n} p(x_i)\log_2\left(p(x_i)\right)$$

$$H_S \in \left[0, \log_2 |X|\right]$$

For one alternative

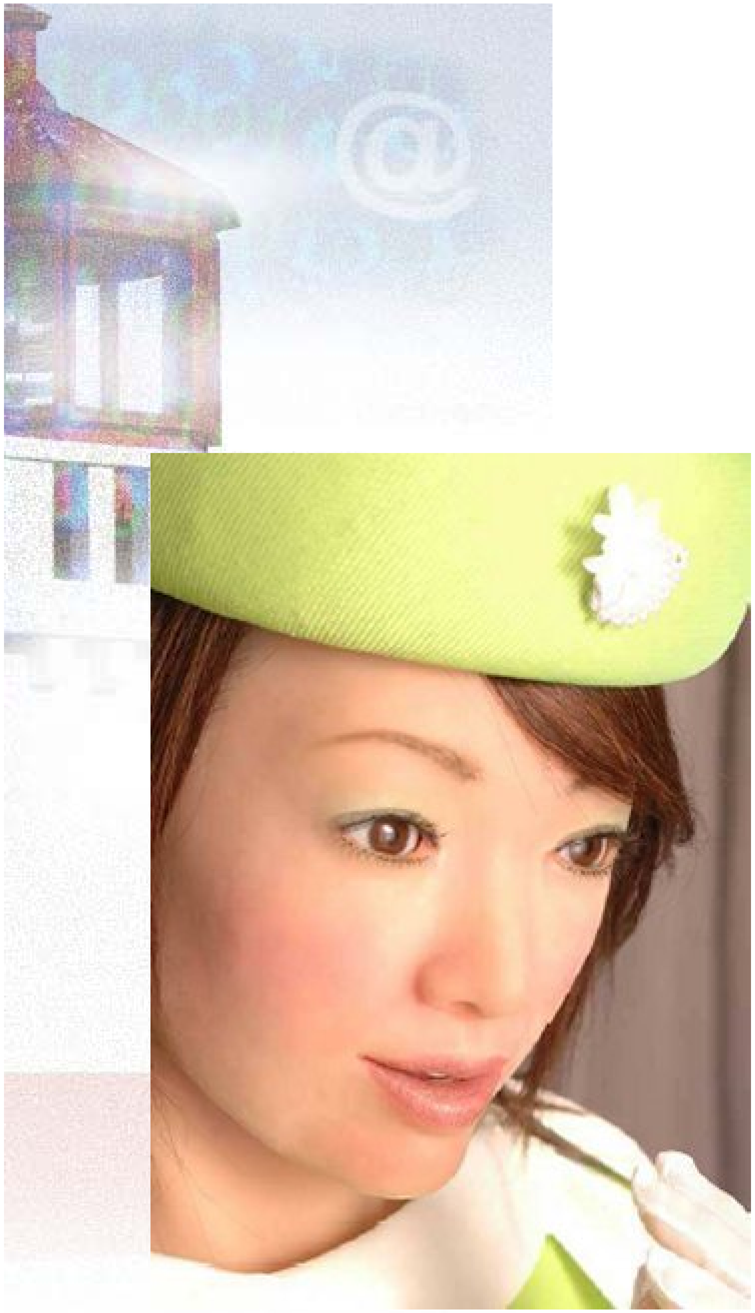
Uniform distribution





Questions

- What type of Uncertainty does the Hartley measure of uncertainty measure?
- What are the units of Shannon entropy?
- Does Shannon’s information theory deal with the semantics and pragmatics of a message? Please explain why?
- If we have a symbol set X={A,B,C,D,E} where the symbol occurrence frequencies are:
  - A = 0.5 B = 0.2 C = 0.1 D = 0.1 E = 0.1
    - What is the average minimum number of bits needed to represent a symbol of the set X?



Algorithms

Pseudo-code and Flow Charts

Luis M.Rocha and Santiago Schnell

Pseudocode Decision

- If-then-else
  - If (letter = “a” or letter =“A”) then
    - display “1”
    - Count_A = Count_A +1
  - Else
    - Display “0”
  - End-if
- Case
  - Case letter of
    - “a” or “A”: display “1”, count_b = count_b+1
    - “b” or “B” : display “2”, count_b = count_b+1
    - “c” or “C”: display “3”, count_b = count_b+1
  - Else
    - Display “0”
  - End-case

```
if condition then
        action
        ...
end-if
if condition then
        action1
        ...
else
        action2
        ...
end-if
case selector of
      condition1 : action1
      condition2 : action2
      condition3 : action3
end-case
```

Pseudocode Iteration or Loops

- For
  - For x = 1 to 100 do
    - y=rand(100) mod x
    - Display y
  - ENDFOR
  - Specifies exactly how many iterations to compute
- While
  - x = 1
  - While ((y ≤ 4) and (x≤100)) do
    - y=rand(100) mod x
    - Display y
    - X=x+1
  - ENDWHILE
  - The number of iterations to compute may depend on the computation itself

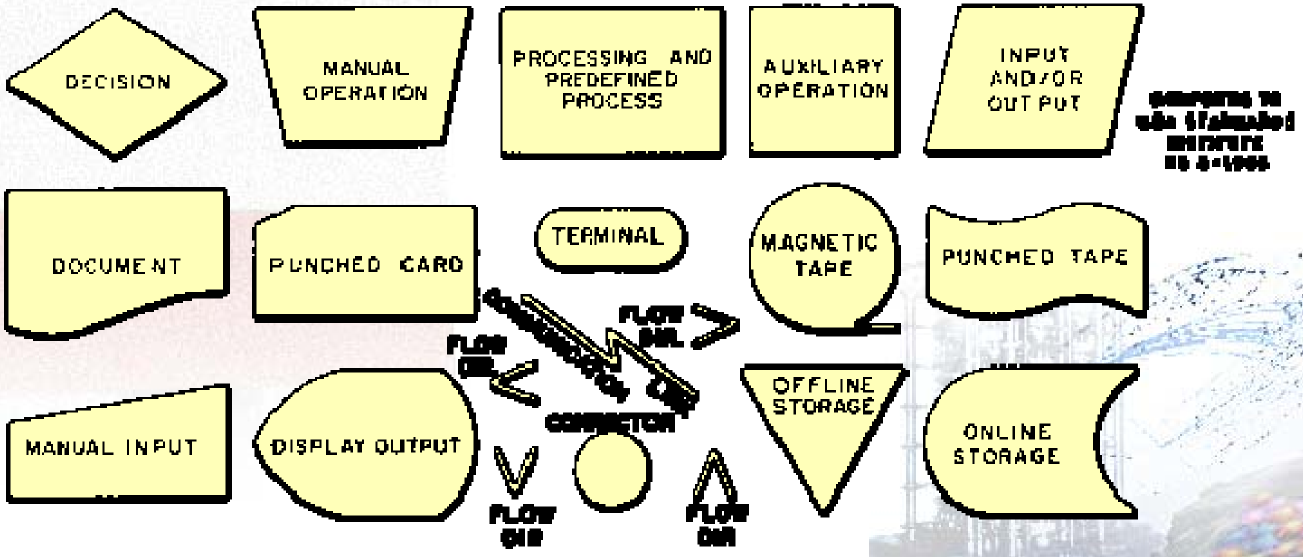
```
FOR counter = start-value
to end-value DO
    statement
    statement
    ...
ENDFOR
```

```
WHILE condition DO
    statement
    statement
    ...
ENDWHILE
```

http://www.tpub.com/neets/book22/93c.htm

Flow Chart

- Pictorial representation of algorithm
  - Parallelogram for input/output
  - Oval for start and stop
  - Rectangle for processing
  - Diamond for decision
  - Hexagon for preparations and loops
  - Circle for connector
  - Arrow for flow direction



Eliza Algorithm – More Details

- set up a language database
    - Words, synonyms, sentences
- begin the conversation (e.g. with a greeting)
- Repeat
    - read user input
        - Keeps track of the two most recent inputs from the user
    - generate Eliza's response
        - preprocess the user input
            - Remove all punctuation from inputs and check for duplicate input
            - Make some synonym replacements from a list of pairs (e.g. big for huge)
            - Change pronouns (e.g. I and me to you)
        - find a matching keyword
        - choose an appropriate response template
            - if a keyword is found
                - extract the part of the user's input following the keyword
                - apply transformations to the extracted input
                - plug the transformed input into the response
            - Else
                - generate a non-committal response
    - print the response on the screen
- until the conversation ends

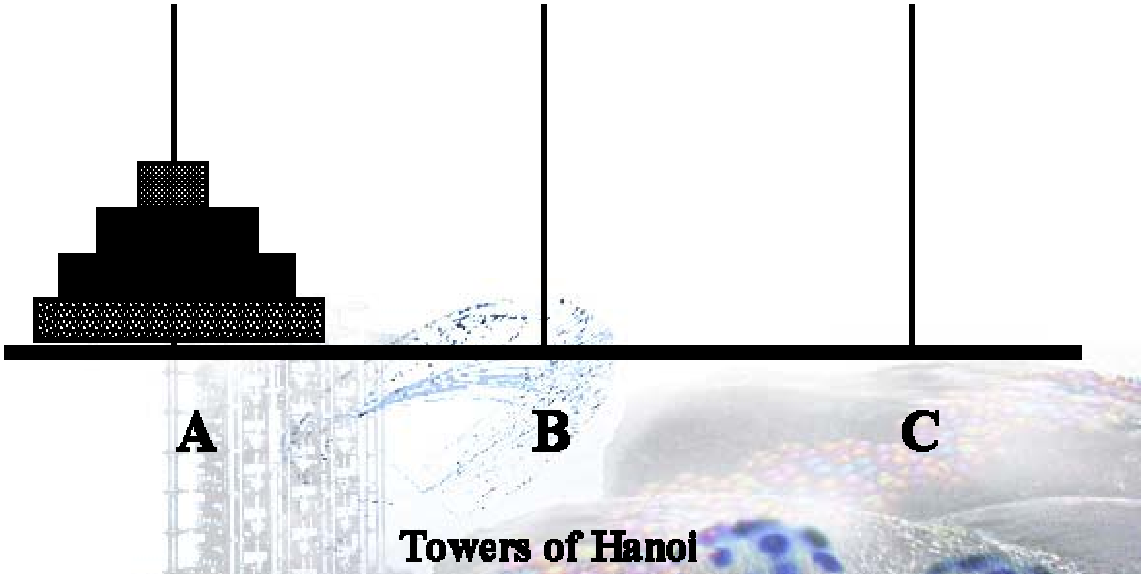
Hanoi Problem for *n* disks

Use Hanoi_2 (H2) as building block (of 3 moves)

H3 uses H2 twice, plus 1 move of the largest disk

- Algorithm to move n disks from A to C
  - Move top n-1 disks from A to B
  - Move biggest disk to C
  - Move n-1 disks on B to C
- Recursion
  - Until H2

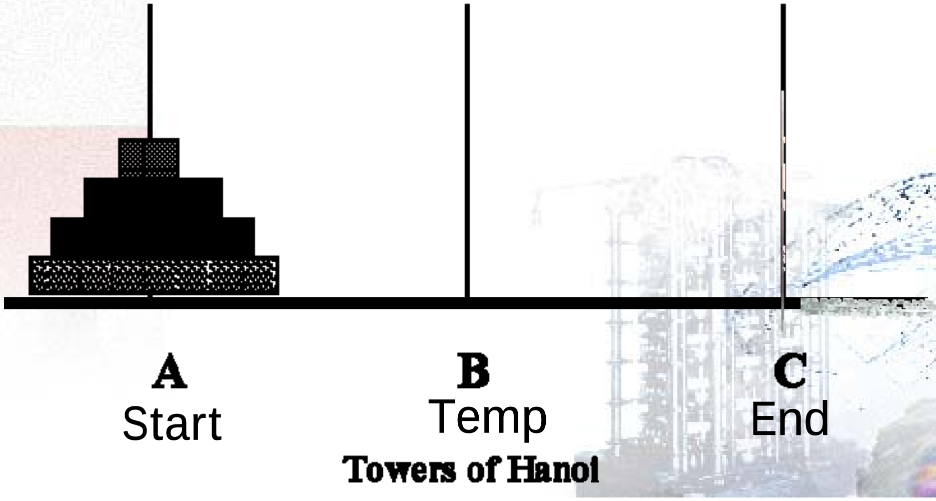
An Algorithm that uses itself to solve a problem



Towers of Hanoi

Pseudocode for Hanoi Problem

```
Hanoi (Start, Temp, End, n)
    If n = 1 then
        Move Start's top disk to End
    Else
        Hanoi (Start, End, Temp, n-1)
        Move Start's top disk to End
        Hanoi (Temp, Start, End, n-1)
```



Computational Complexity

$$2^{10} = 1{,}024$$
$$2^{20} = 1{,}048{,}576$$
$$2^{30} = 1{,}073{,}741{,}824$$
$$2^{40} = 1{,}099{,}511{,}627{,}776$$
$$2^{64} = 18{,}446{,}744{,}073{,}709{,}551{,}616$$

585 billion years
in seconds!!!!!!!!

Earth: 5 billion years

Universe: 15 billion years

Fastest Computer: 135.5 teraflops - 135.5 trillion calculations a second (aprox 2^{47} moves a second)

2^{17} s needed = 36 hours

- Resources required during computation of an algorithm to solve a given problem
  - Time
    - how many steps does it take to solve a problem?
  - Space
    - how much memory does it take to solve a problem?
- The Hanoi Towers Problem
  - $f(n)$ is the number of times the HANOI algorithm moves a disk for a problem of n disks
    - $f(1)=1$, $f(2)=3$, $f(3)=7$
    - $f(n)= f(n-1) + 1 + f(n-1) = 2 \times f(n-1) + 1$
  - Every time we add a disk, the time to compute is at least double
    - $f(n) = 2^n - 1$

Bremermann's Limit

- Physical Limit of Computation
  - Hans Bremmermann in 1962
  - "no data processing system, whether artificial or living, can process more than 2 × 10^{47} bits per second per gram of its mass."
    - Based on the idea that information could be stored in the energy levels of matter
    - Calculated using Heisenberg's uncertainty principle, the Hartley measure, Planck's constant, and Einstein's famous E = mc2 formula
  - A computer with the mass of the entire Earth and a time period equal to the estimated age of the Earth
    - **would not be able to process more than about 10^{93} bits**
  - ***transcomputational problems***

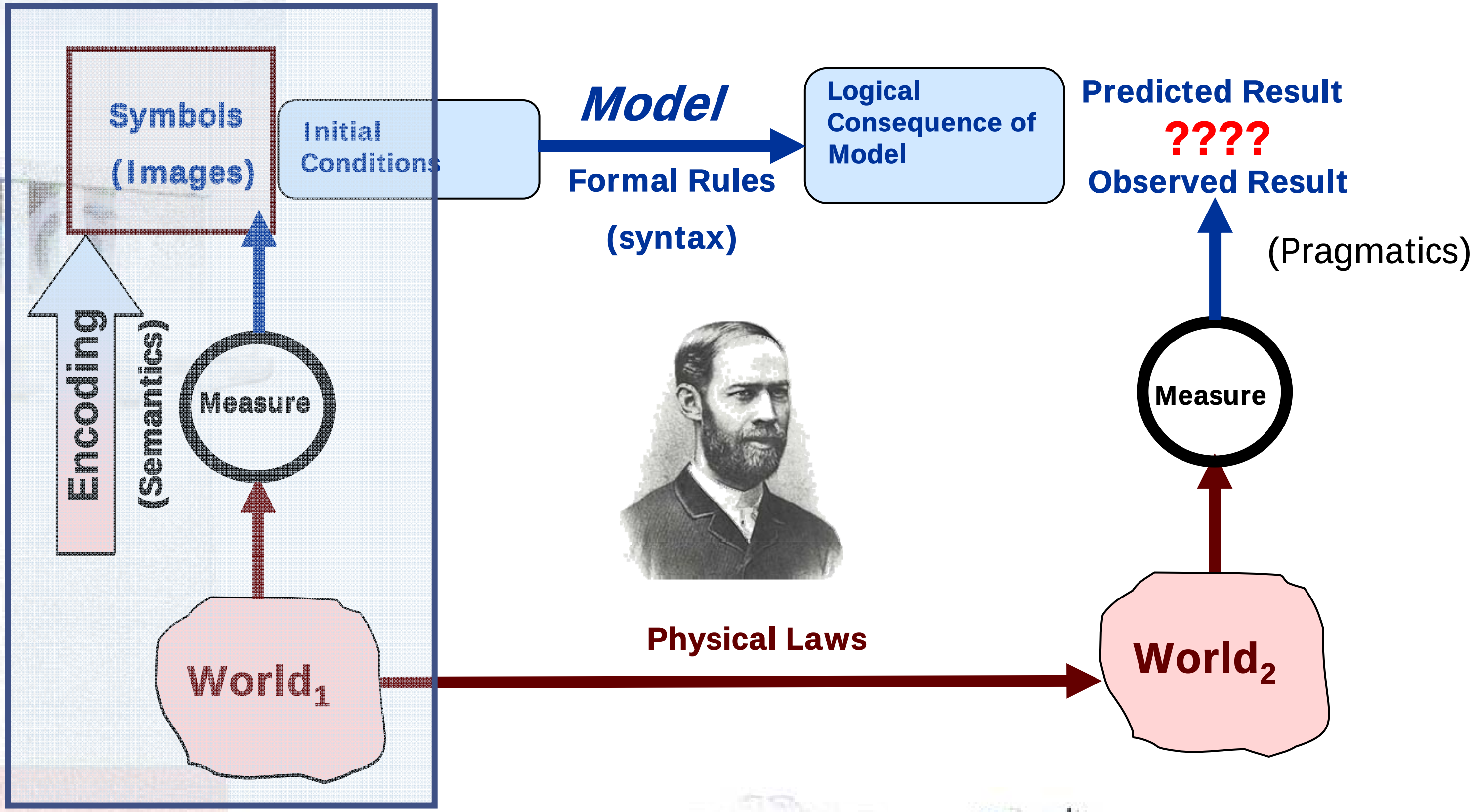
Luis M.Rocha and Santiago Schnell

Questions

- Using pseudo-code, write down an algorithm to calculate the tip of a restaurant bill and the amount that each person of a group o *n* needs to pay
- Consider the following recursive definition of a function:
  - $Q(n) = Q(n – Q(n-1)) + Q(n – Q(n – Q(n-2))$ for $n>2$
  - with $Q(1) = Q(2) = 1$.
    - Please write down a pseudo-code algorithm to calculate Q(10).
- What is the Bremermann's Limit ?
  - Discuss its implications to problem solving and modeling

The Modeling Relation

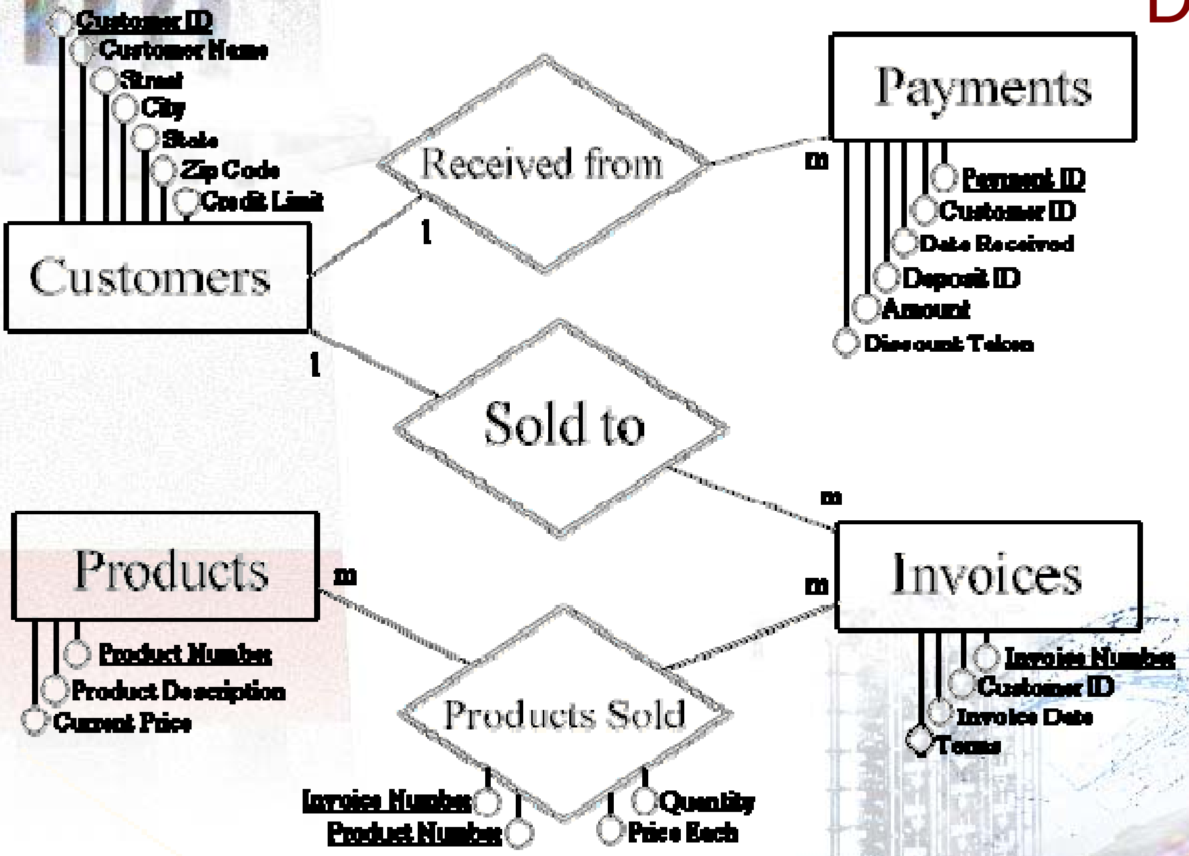
Hertz' Modeling Paradigm



- Organizing Data
  - After encoding
  - Modern Problems Require large storage capabilities

Data Modeling

Entity-relationship model and Relational Databases

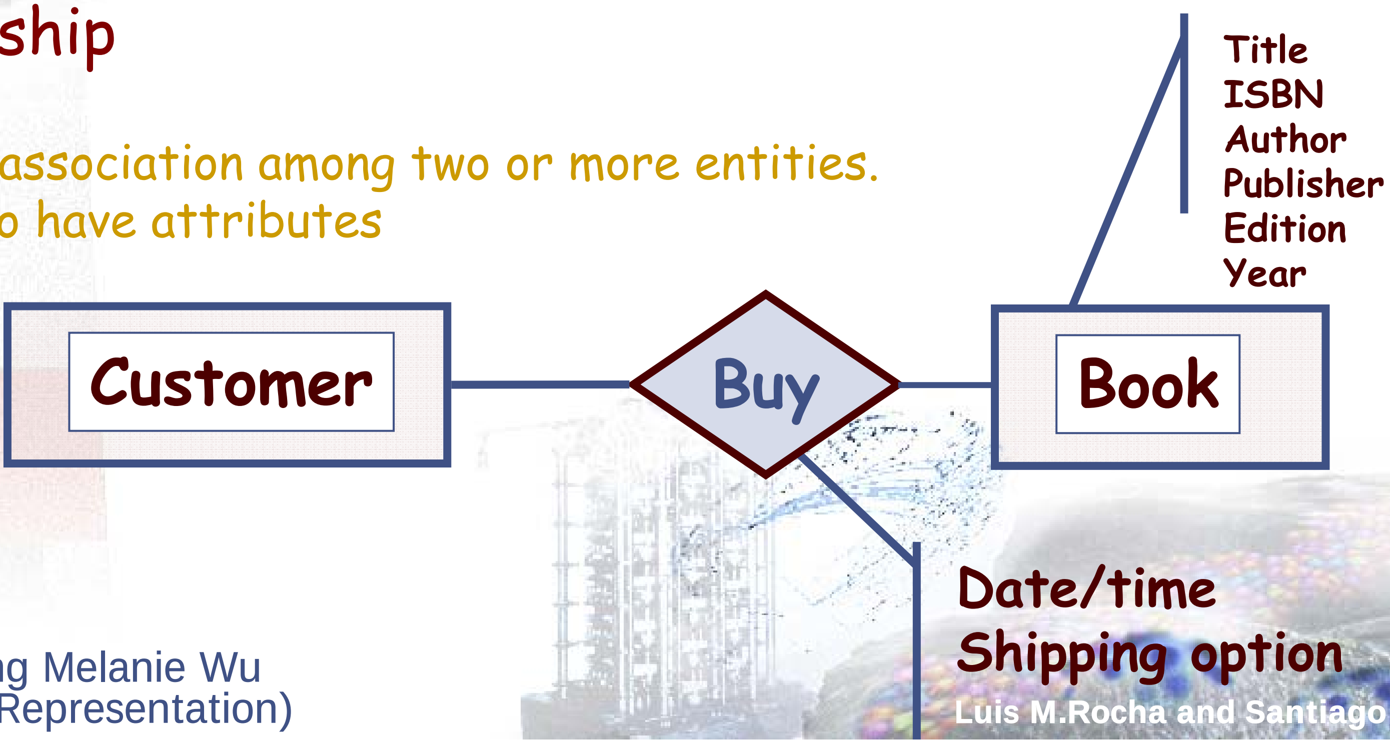


The Entity-Relationship Model

- Conceptual ***Data Model***
  - A kind of "pseudocode" for ***models of data storage***
- Entities
  - Nouns: Objects, people, places
    - Represented by a rectangle
    - Attributes describe its proprieties
- Relationship
  - Verbs
    - *An association among two or more entities.*
    - Also have attributes

Peter Chen

(1976)

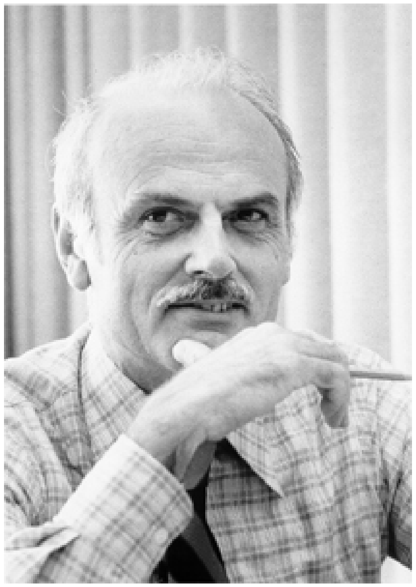


Adapted from Yuqing Melanie Wu
(I308: Information Representation)

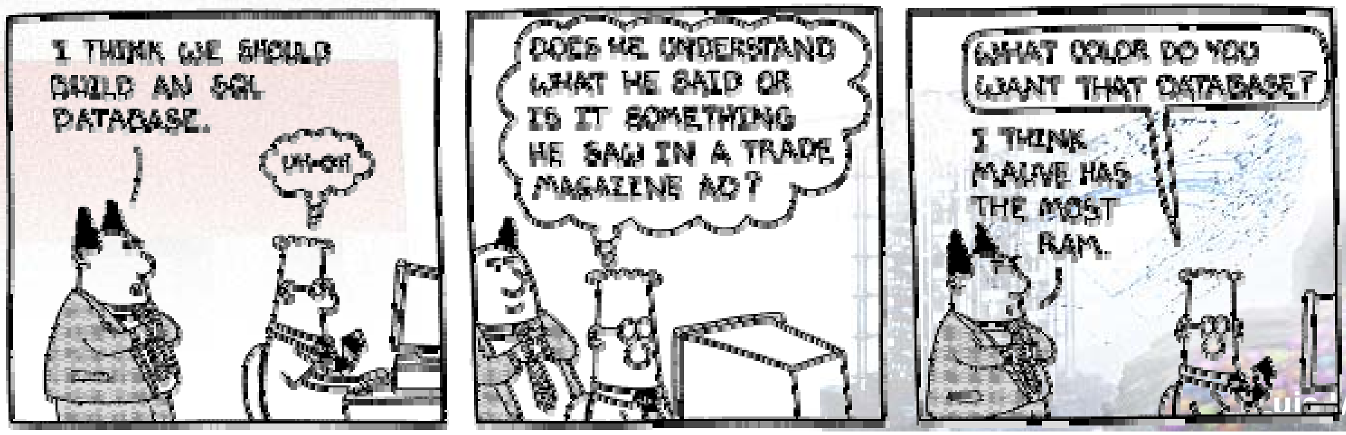
Luis M.Rocha and Santiago Schnell

The Relational Database Model

- Relational database management system (RDBMS)
  - Most popular commercial database type.
  - a data model based on *logic* and ***set theory***.
- invented by Ted Codd in 1970
  - Oxford, IBM, U. Michigan, IBM
- System R
  - IBM's San Jose research center
  - Structured English Query Language ("SEQUEL")
    - Data Manipulation Language (DML)
  - SEQUEL was later condensed to SQL due to a trademark dispute
  - In 1979, Relational Software, Inc. (now Oracle Corporation) introduced the first commercially available implementation of SQL



Ted Codd

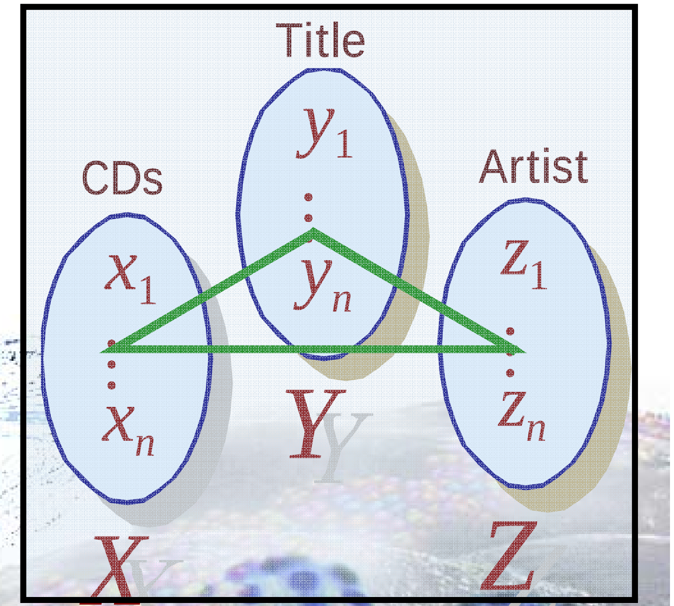

I THINK WE SHOULD BUILD AN SQL DATABASE.
UH-OH
DOES HE UNDERSTAND WHAT HE SAID OR IS IT SOMETHING HE SAW IN A TRADE MAGAZINE AD?
WHAT COLOR DO YOU WANT THAT DATABASE?
I THINK MAUVE HAS THE MOST RAM.

The Relational Database Model

- A relational database is a collection of tables
  - 2-dimensional
- Each table has a unique name in the database.
- Tables define Relations
  - Columns (number of sets)
    - Attributes plus key (*primary set*)
  - Row (number of relation instances)
    - A table is a set of rows: tuples

CDs

ID	Title	Artist
3592	Yes I am a Witch	Yoko Ono
2678	Big	Macy Gray
0623	Sound of Silver	LCD Soundsystem
0321	Welcome to Planet Sexor	Tiga
8854	Transparent Things	Fujiya & Miyagi



Luis M.Rocha and Santiago Schnell

Book Key

ISBN	Title	Publisher
12345	Java	MIT press
49082	Snow White	……
72936	Honeymoon	……

Book(ISBN, Title, Publisher)

Customer Key

Phone	Name	Address
812-123-4567	Tom	……
812-304-2378	Bill	……
812-856-1190	Kate	……

Customer(Phone, Name, Address)

Sale Key

ISBN	Phone	Price	Time
12345	812-123-4567	$20	Feb 2, 05
49082	812-123-4567	$25	Dec 20, 04
12345	812-856-1190	$19	……

Sale (ISBN, Phone,time,price)

Come from primary keys in entities

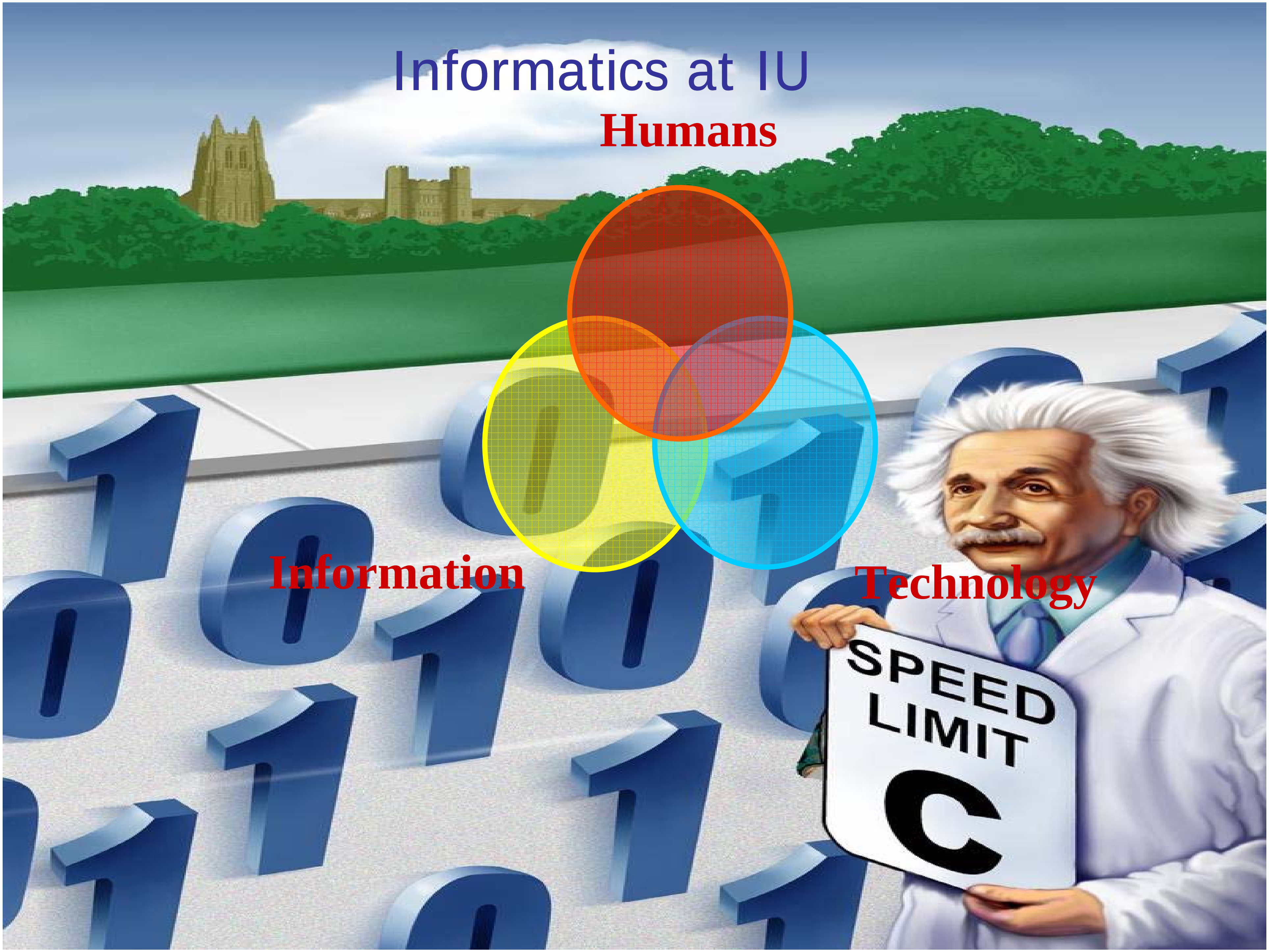
Primary keys in Relationships

Adapted from Yuqing Melanie Wu
(I308: Information Representation)

Questions

- The Santiago de Compostela historical society is setting up a database of the monarchs in the “unified” Spain. This is a chronological list of the people who have ruled “unified” Spain; the dates given are the periods of said rule.
  - **Unified 'Spain‘**
    - **Habsburg Dynasty**
      - 1516 - 1556 Charles I (Emperor Charles V) , King of Spain, Austria, Netherlands, Rome and Naples
      - 1556 - 1598 Philip II, King of Spain, Portugal, Austria, Netherlands, Rome and Naples
      - 1598 - 1621 Philip III, King of....
      - 1621 - 1665 Philip IV
      - 1665 - 1700 Charles II
    - **Bourbon Dynasty**
      - 1700 - 1724 Philip V
      - 1724 Louis I
      - 1724 - 1746 Philip V (2nd time)
      - 1746 - 1759 Ferdinand VI
      - 1759 - 1788 Charles III
      - 1788 - 1808 Charles IV
      - 1808 Ferdinand VII
    - **French Rule**
      - 1808 - 1813 Joseph Bonaparte
    - **Bourbon Dynasty**
      - 1814 - 1833 Ferdinand VII
      - 1833 - 1868 Isabella II
      - 1874 - 1885 Alfonso XII
      - 1886 - 1931 Alfonso XIII
      - 1975 - present Juan Carlos I
- The society has already decided to include the fields above. Please describe the entities and their attributes of each field in the database.

Informatics at IU



Humans

Information

Technology

Brought to you by

Cyber Rhapsody

Or Infernal Beast From The Air Dimension ???