Raoul LePage
Professor
STATISTICS AND PROBABILITY
www.stt.msu.edu/~lepage
click on STT200_Sp09

3-18-09

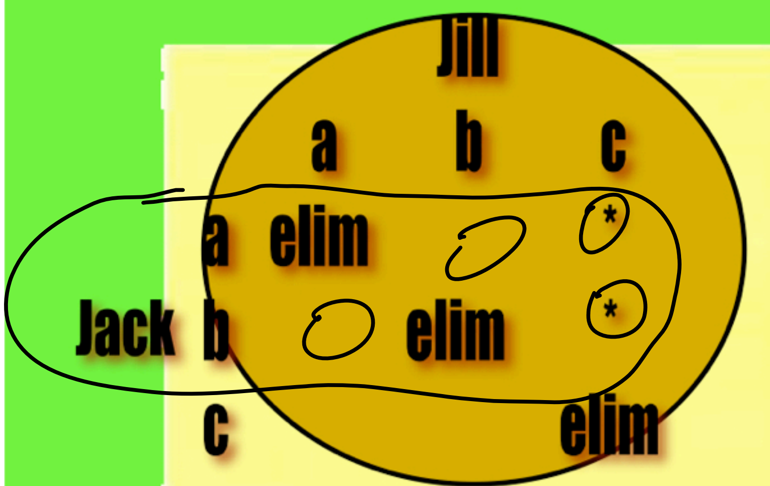
GOLD STANDARD ANSWER

{\$1a, \$1b, \$5c}

Jack draws a bill first

Jill draws second

from the **two** bills then remaining



all 6 possibilities

P(Jill \$5)
= 2 / 6 = 1 / 3
same as Jack

TOLD JACK GOT 1.

P(JILL 5 | JACK 1) = 2/4

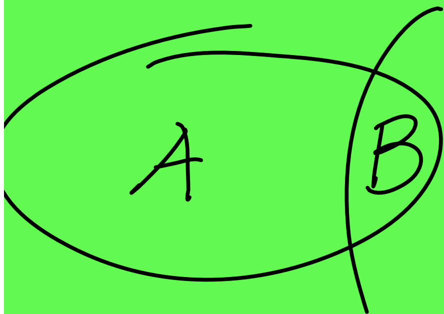
2/4 = 2/6 / (4/6) = P(JACK 1 AND JILL 5) / P(JACK 1)

DEFINE P(B|A) = P(AB) / P(A)

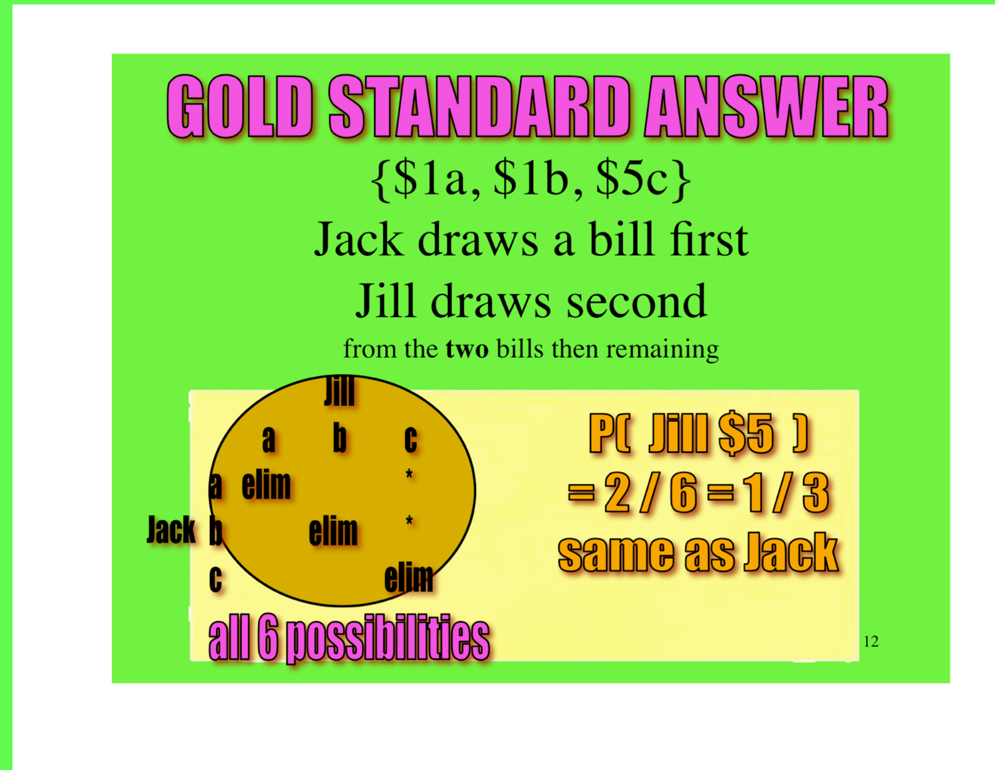
ARGUE WANT IN CLASSICAL SETUP

$$P(B|A) = \frac{P(AB)}{P(A)}$$

GIVEN



DO NOT CARE IF A HAS HAPPENED



≡ MULT RULE $P(AB) = P(A)P(B|A)$

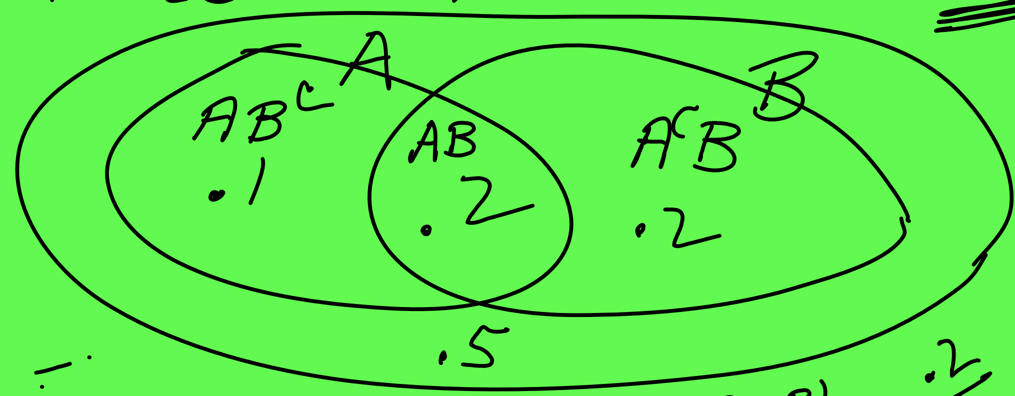
$= P(B) = P(A|B)$

CHOOSE THAT IMAGE HAVING P(IMAGE | WHAT YOU SEE) LARGEST

BLANK SHEET

TOLD $P(A) = .3$ $P(B) = .4$ $P(AB) = .2$

FIND

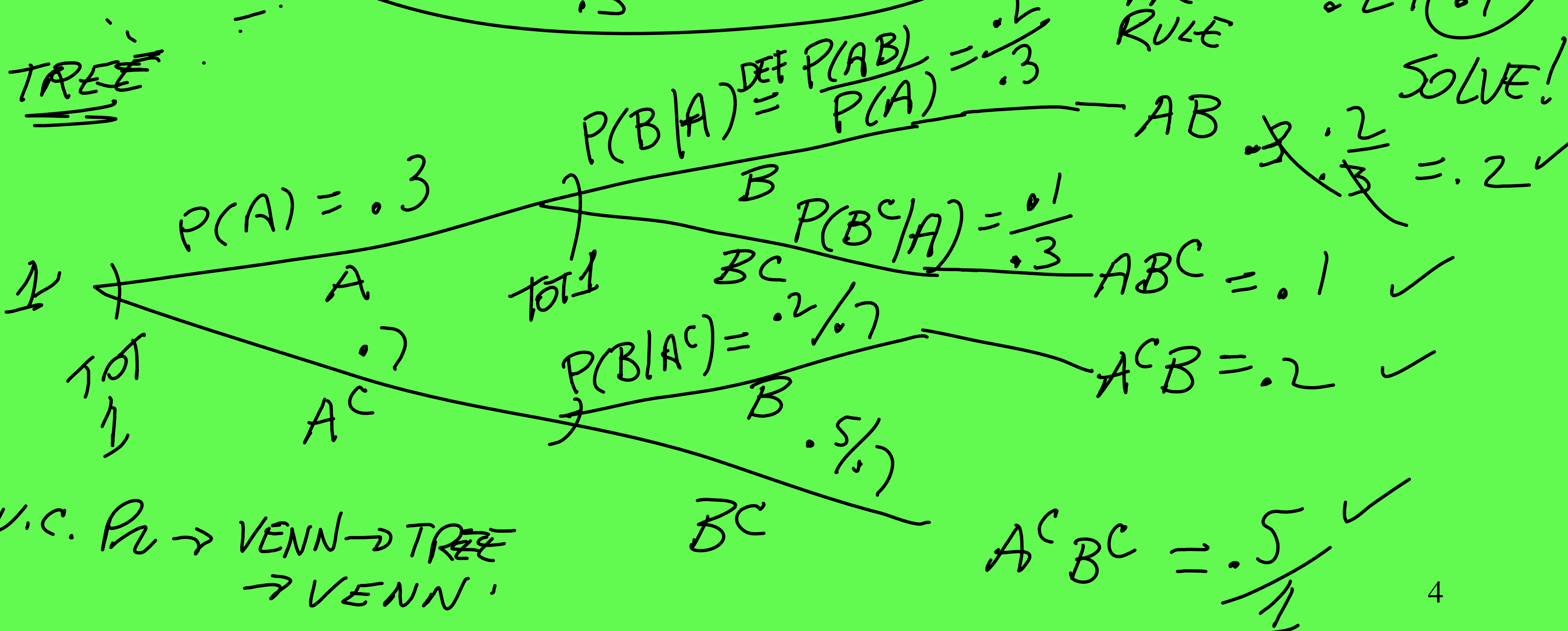


$P(A) = .3$

$= P(AB) + P(AB^C)$

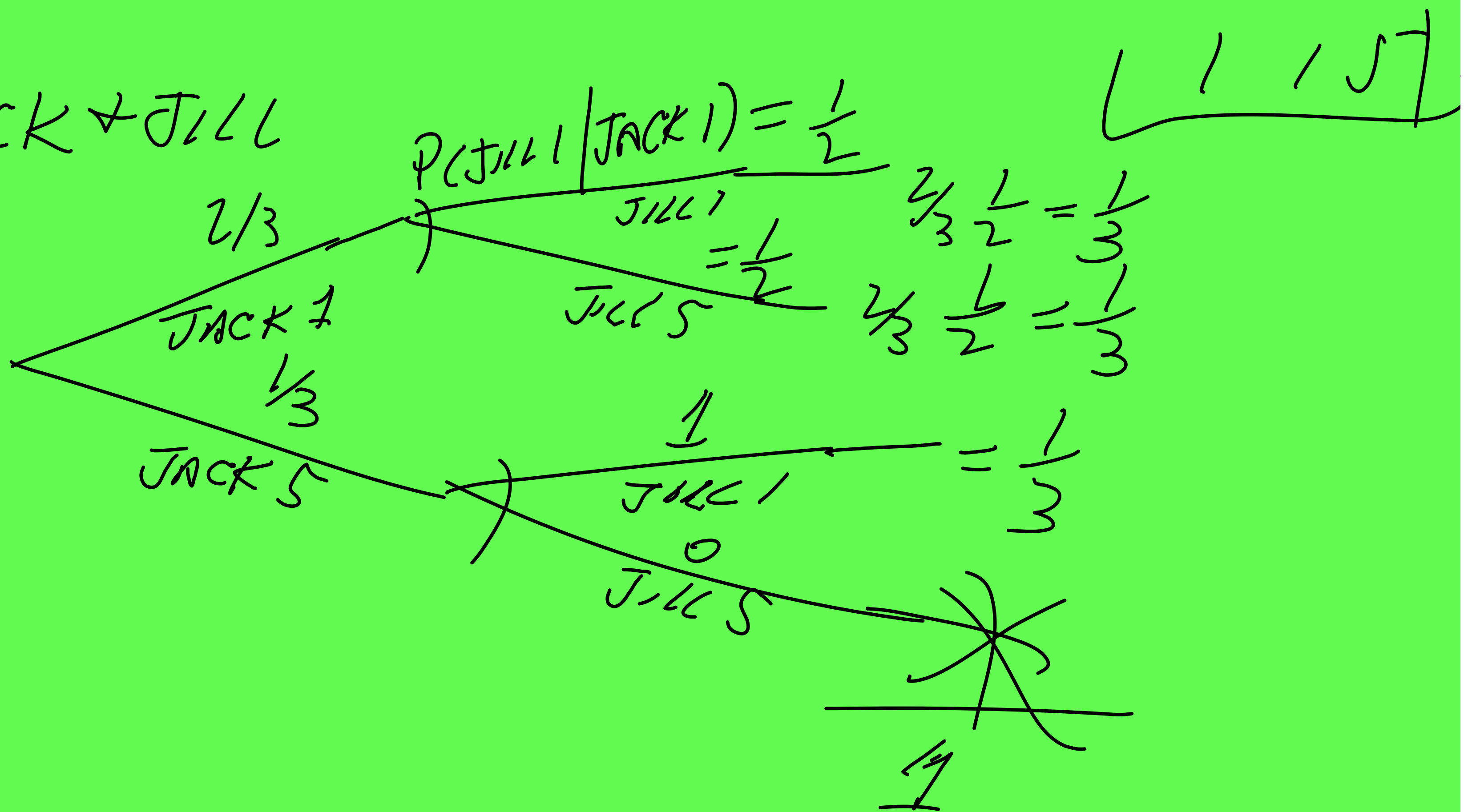
TOTAL PR RULE $= .2 + .1$

TREE



U.C. Pr → VENN → TREE
→ VENN

JACK + JILL

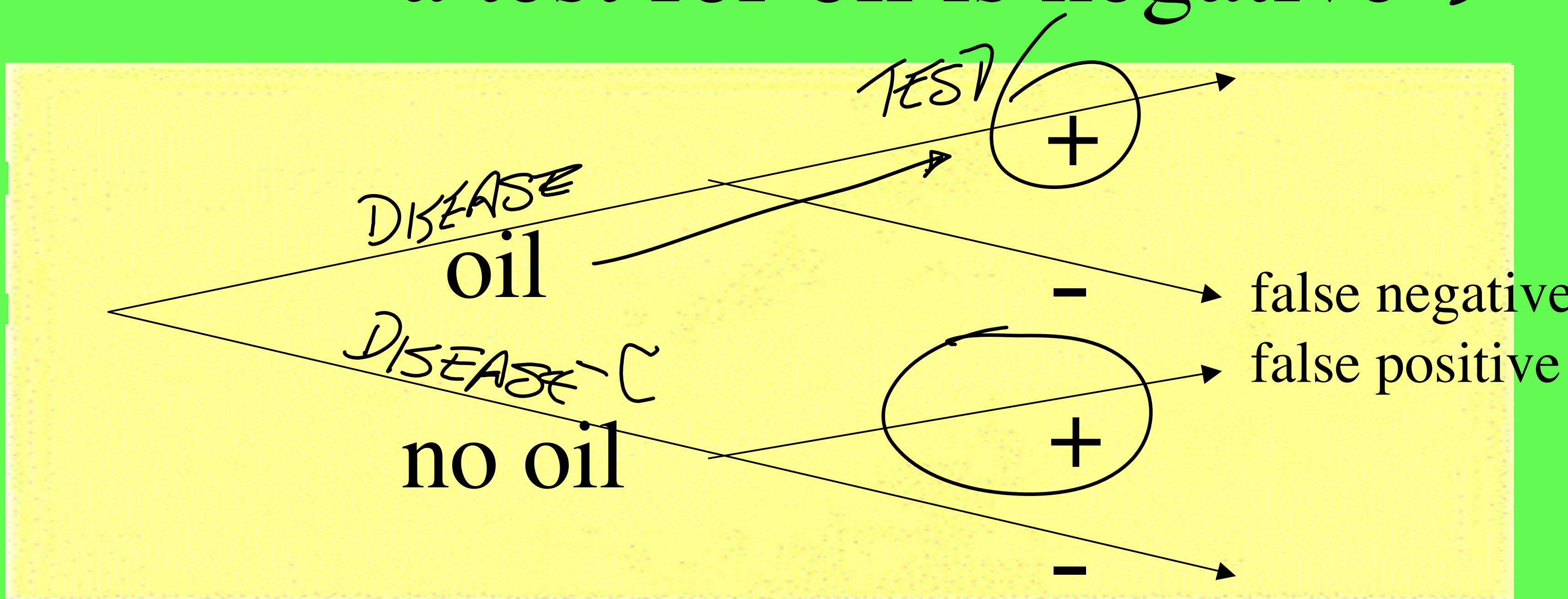

P(JILL 1 | JACK 1) = 1/2
2/3
JACK 1
1/3
JACK 5
JILL 1
JILL 5
= 1/2
2/3 1/2 = 1/3
2/3 1/2 = 1/3
1
= 1/3
0
1

TREE DIAGRAM

"oil" = oil is present ✓

"+" = a test for oil is positive ✓

"–" = a test for oil is negative ✓

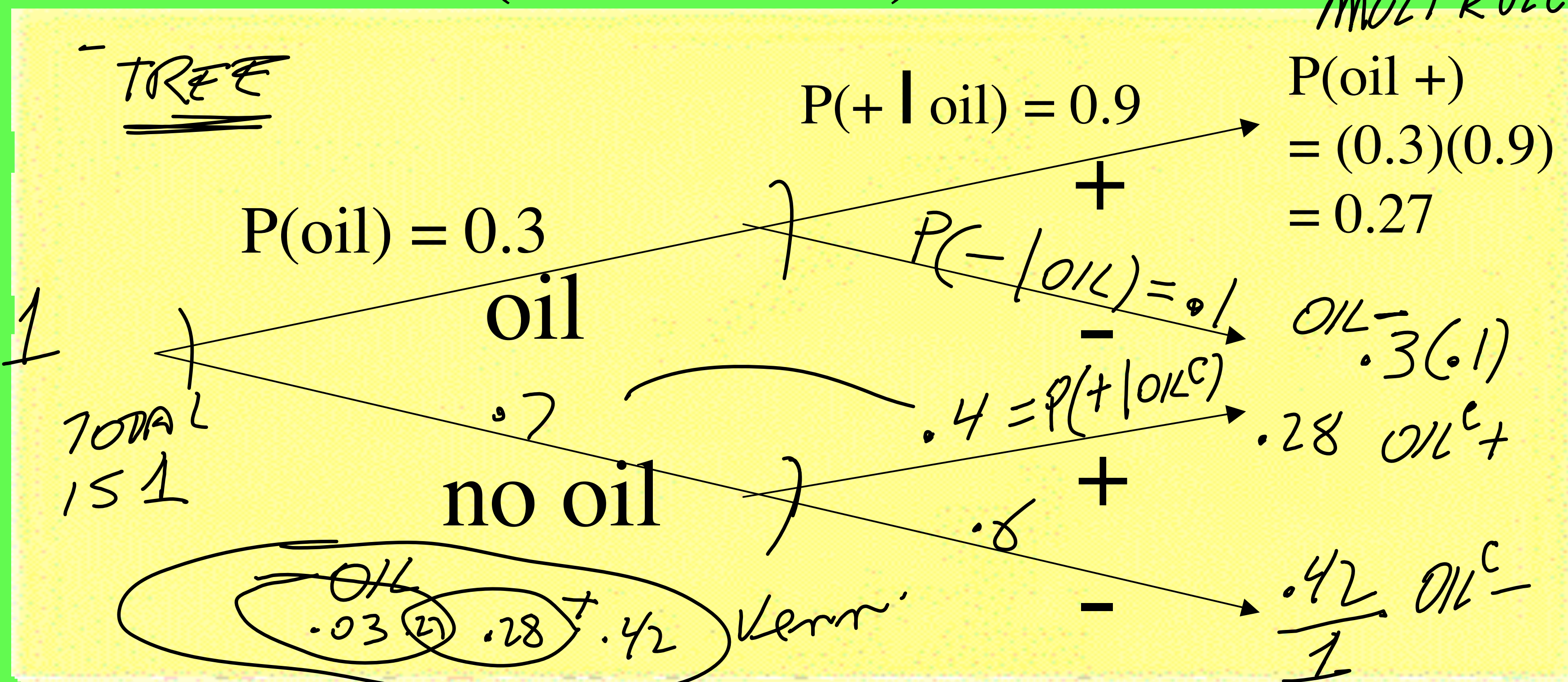


TREE DIAGRAM CONVENTIONS

$P(\text{oil}) = 0.3$

$P(+ \mid \text{oil}) = 0.9$

$P(+ \mid \text{no oil}) = 0.4$

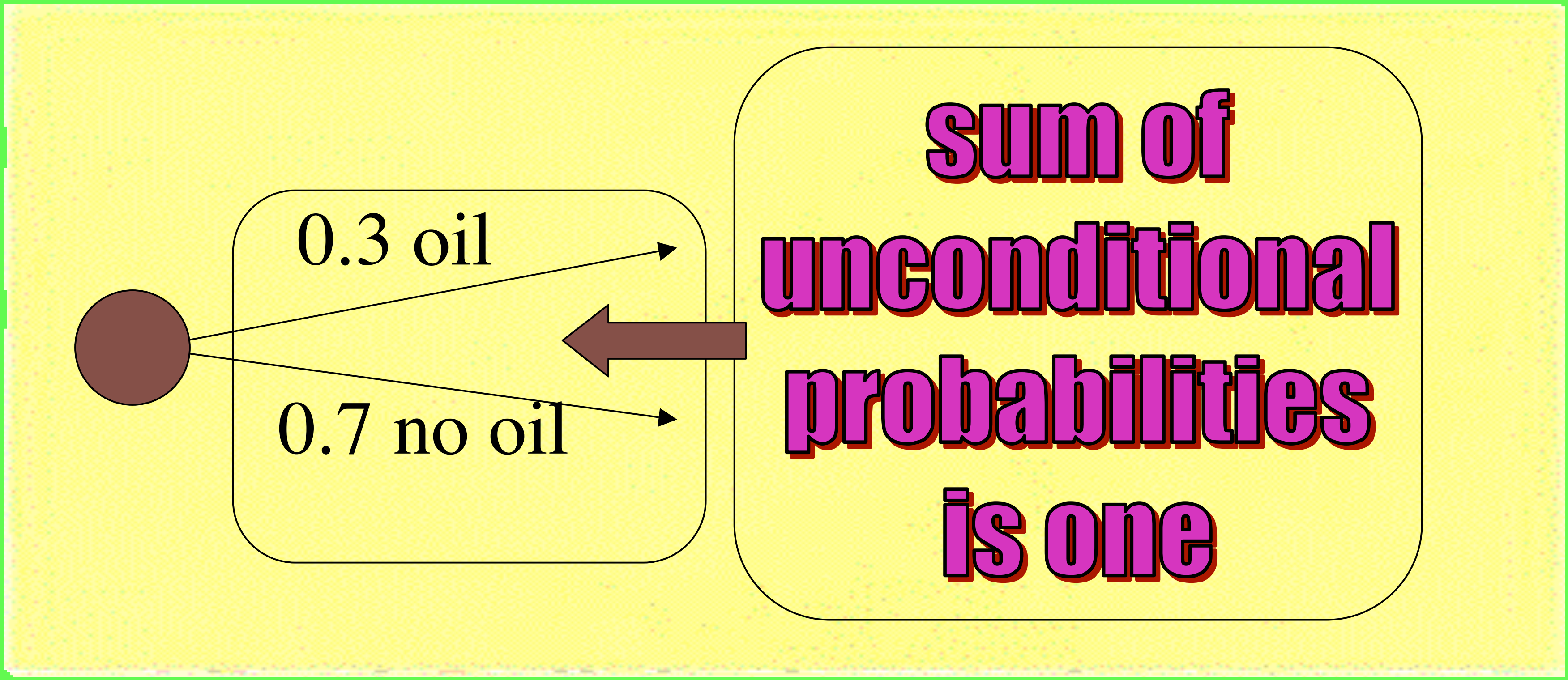


TOTAL OF BRANCHES = 1

$P(oil) = 0.3$

$P(+ \mid oil) = 0.9$

$P(+ \mid no\ oil) = 0.4$

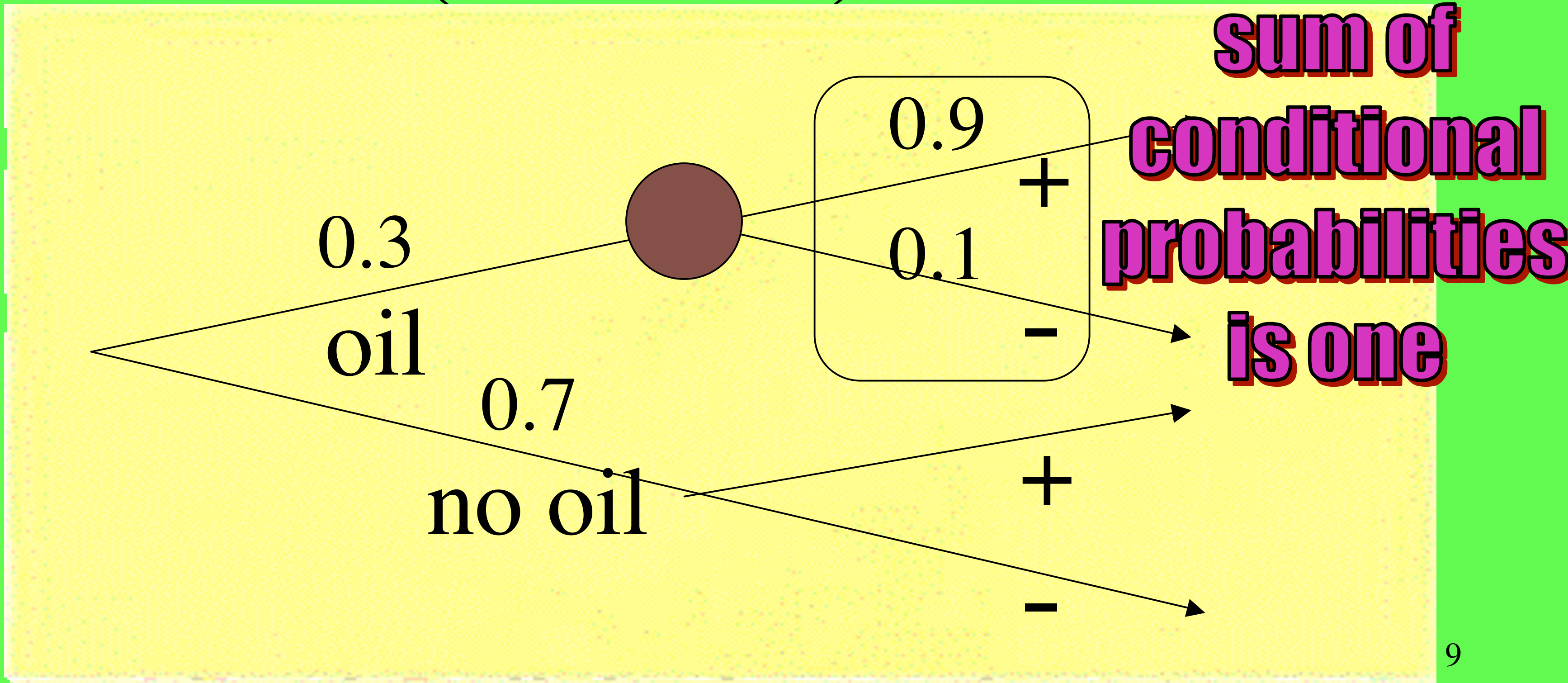


TOTAL OF CONDITIONAL BRANCHES = 1

P(oil) = 0.3

P(+ | oil) = 0.9 P(- | oil) = 0.1

P(+ | no oil) = 0.4

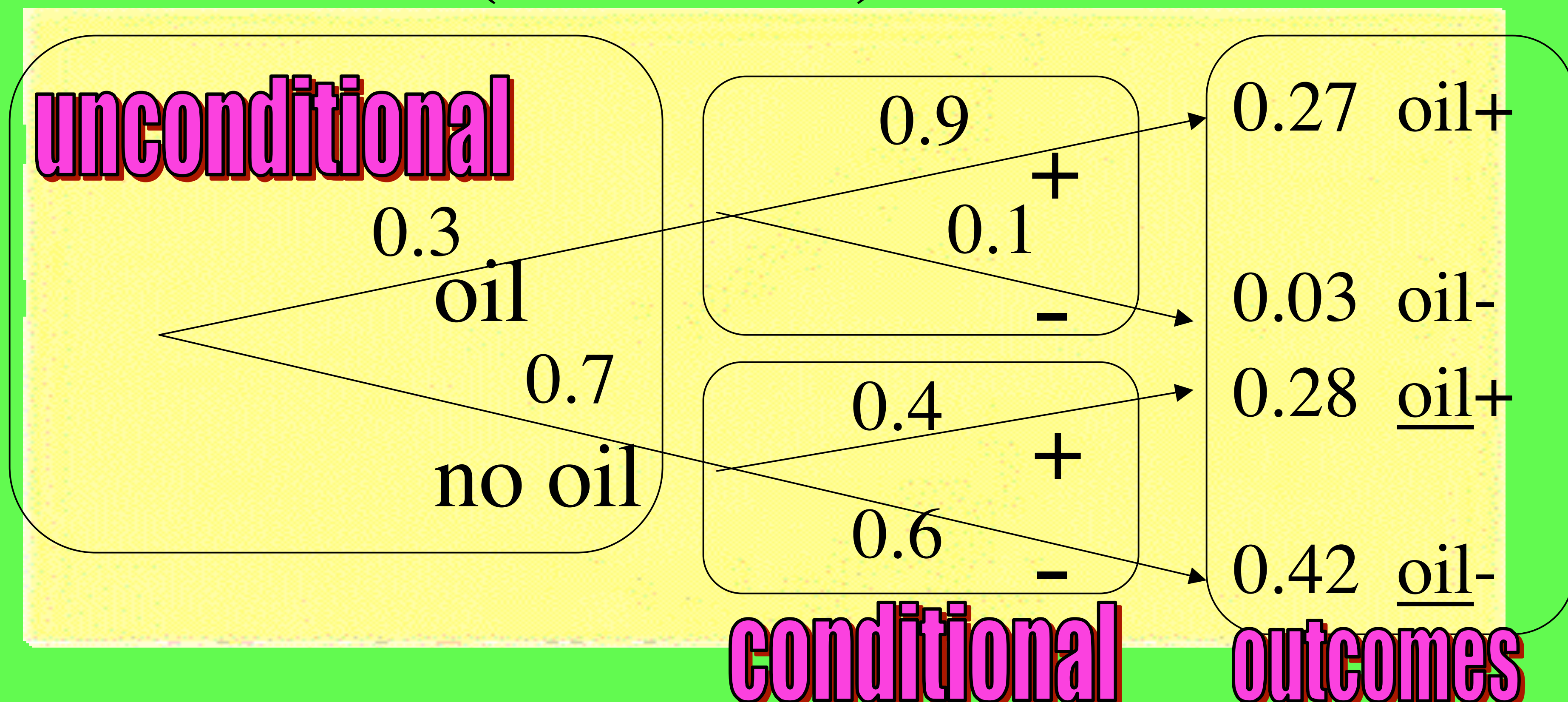


COMPLETE TREE

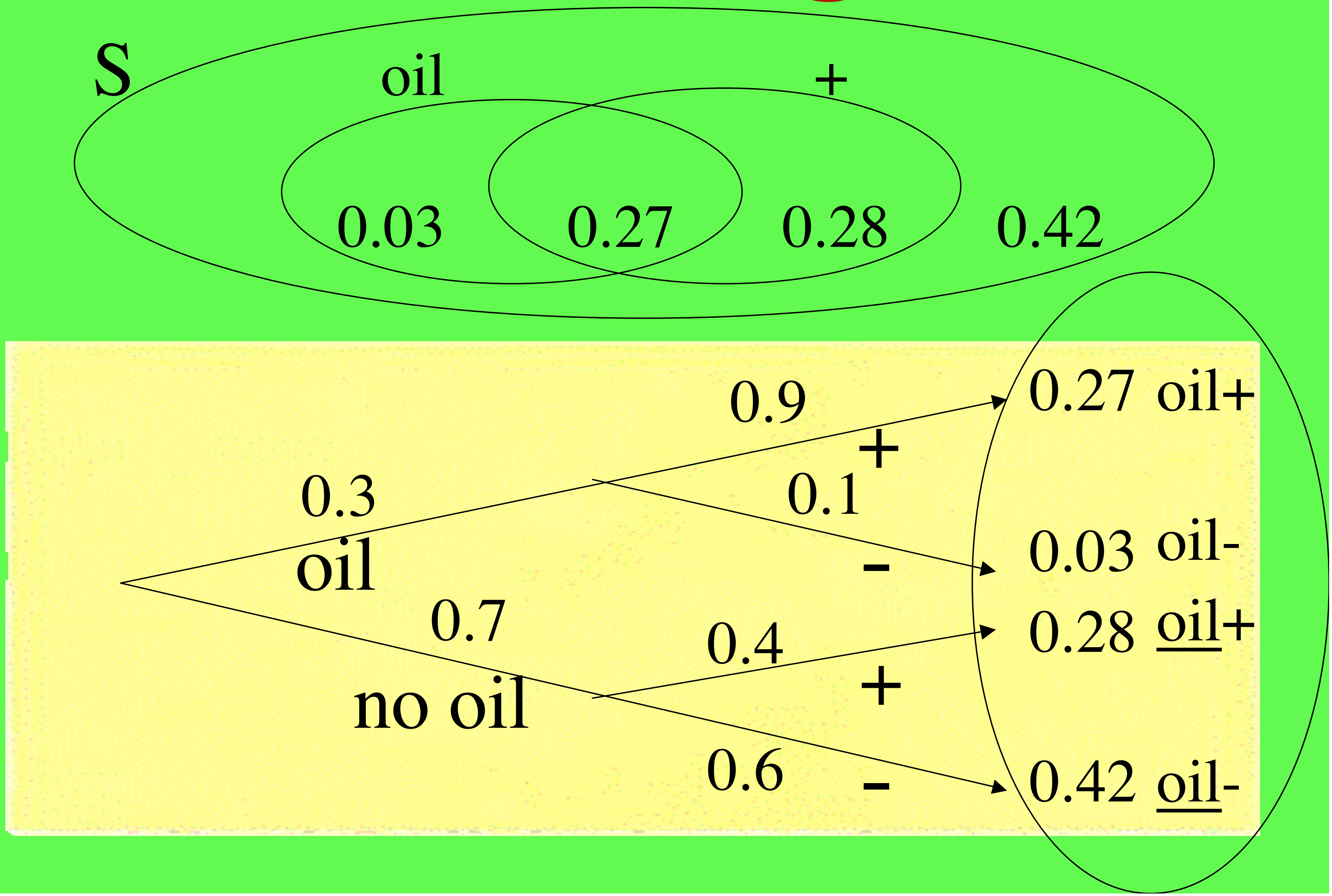
$$P(\text{oil}) = 0.3$$

$$P(+ \mid \text{oil}) = 0.9$$

$$P(+ \mid \text{no oil}) = 0.4$$



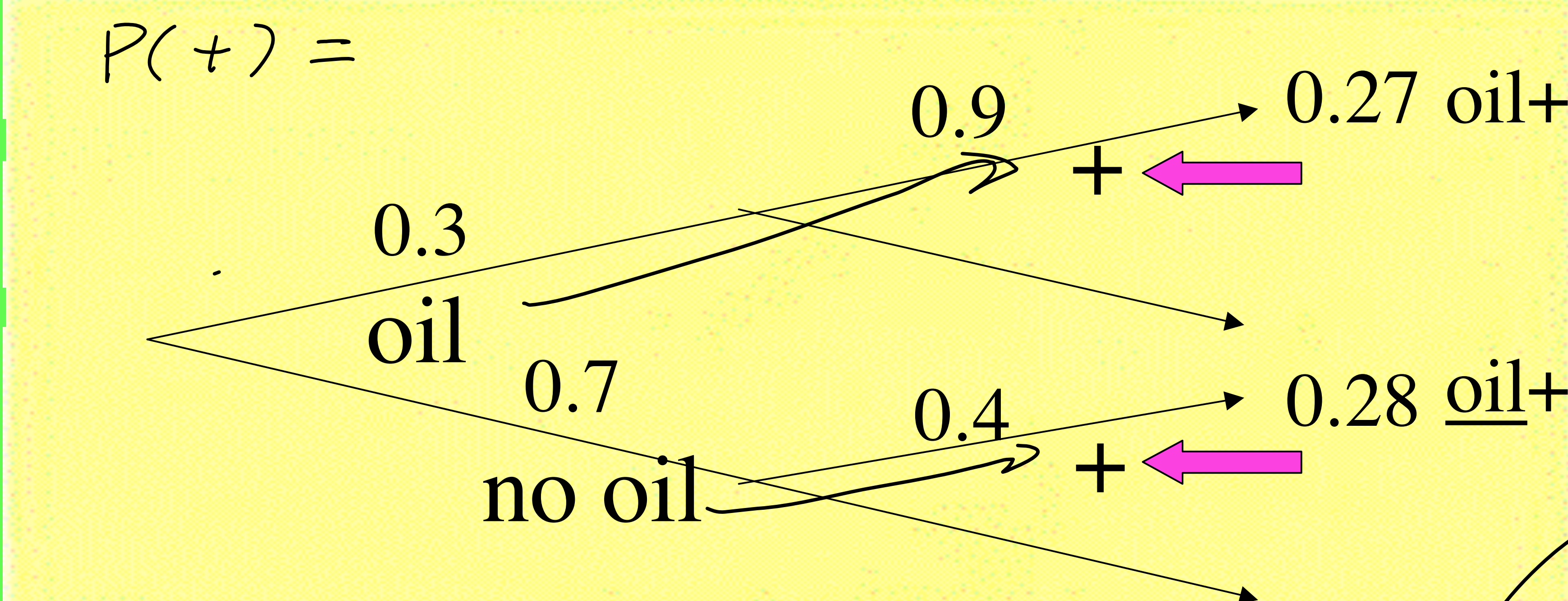
VENN DIAGRAM



TOTAL PROBABILITY

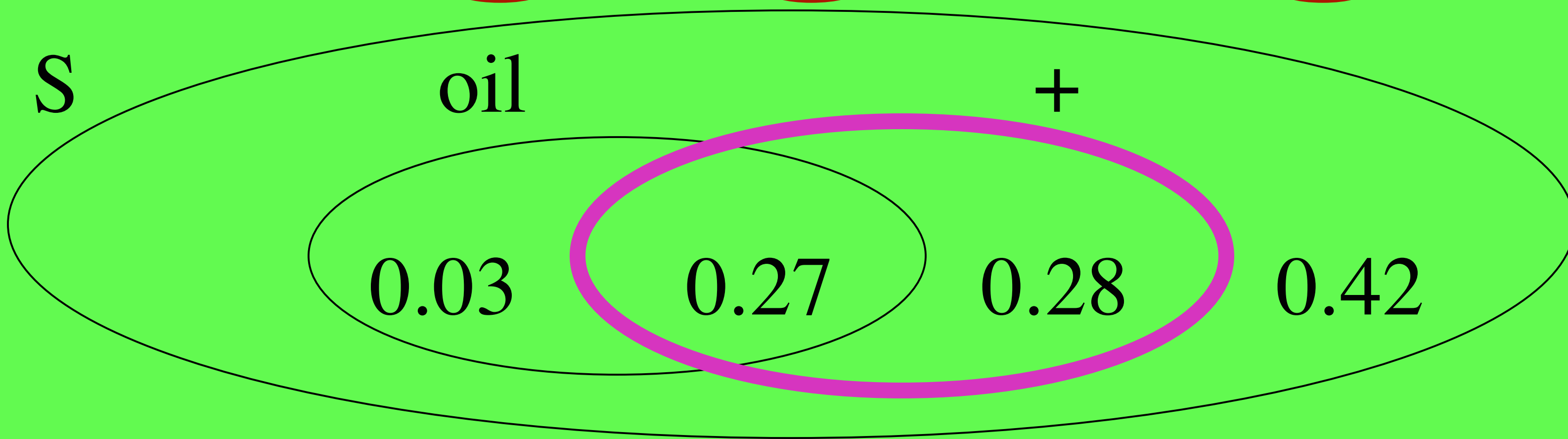
P(+) = P(oil+) + P(no oil+)

0.55 = 0.27 + 0.28



Oil contributes 0.27 to the total P(+) = 0.55.

BAYES FORMULA



$$P(\text{oil} \mid +) = P(\text{oil}+) / P(+)$$

$$= 0.27 / (0.27 + 0.28)$$

$$= 0.4909..$$

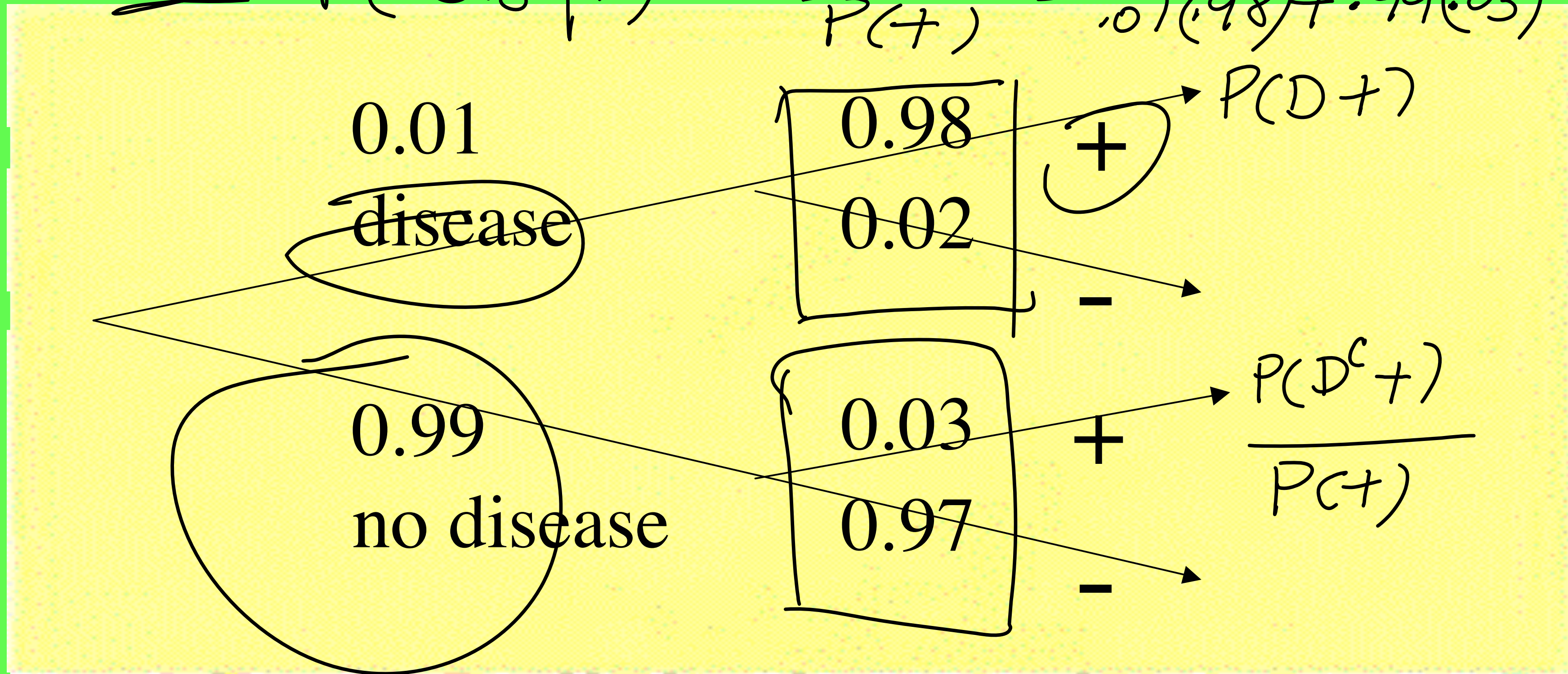
0.27 oil+

0.28 <u>oil</u>+

Oil contributes 0.27 of the total P(+) = 0.27+0.28.

MEDICAL TEST

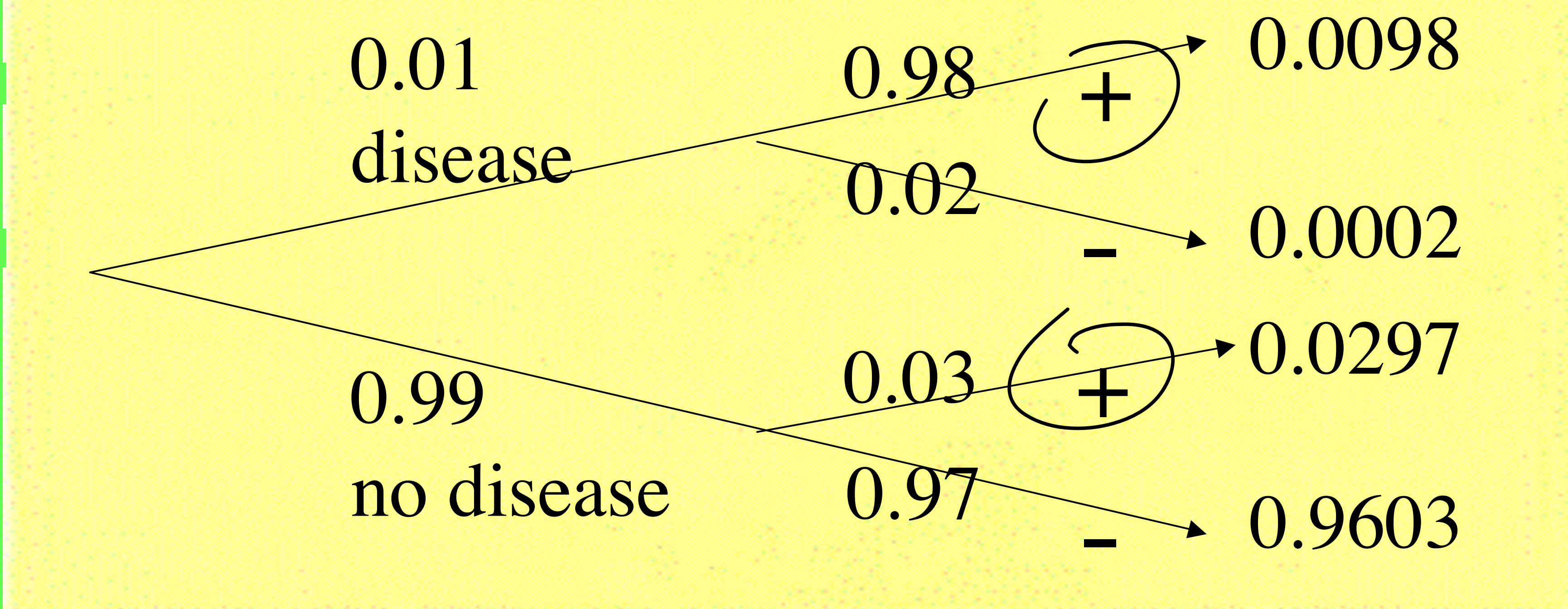
BAYES $P(\text{Dis} \mid +) = \frac{P(D+)}{P(+)} = \frac{.01\,.98}{.01(.98)+.99(.03)}$



The test for this infrequent disease seems to be reliable having only 3% false positives and 2% false negatives. **What if we test positive?**

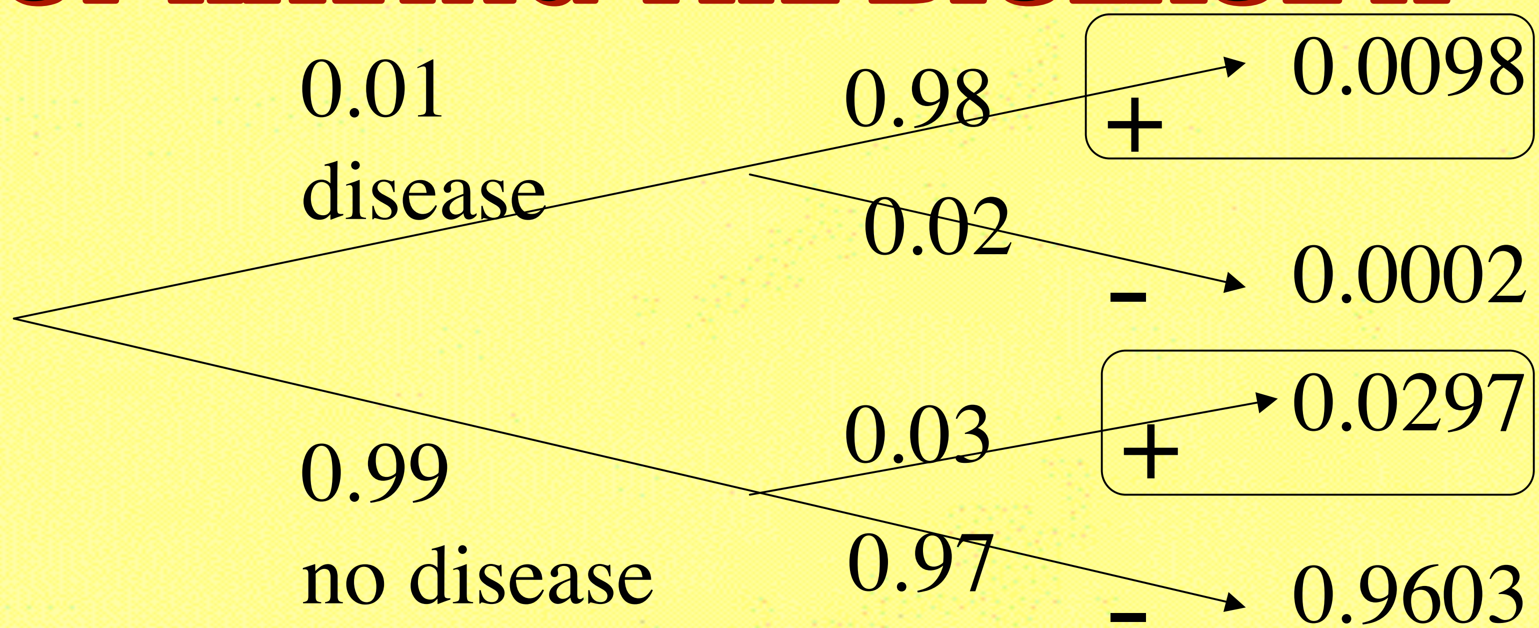
MEDICAL TEST

$P(D|+) = \frac{.0098}{.0098 + .0297}$



We need to calculate P(diseased | +), the **conditional probability** that we have this disease **GIVEN** we've tested positive for it.

CALCULATING OUR CHANCES OF HAVING THE DISEASE IF +



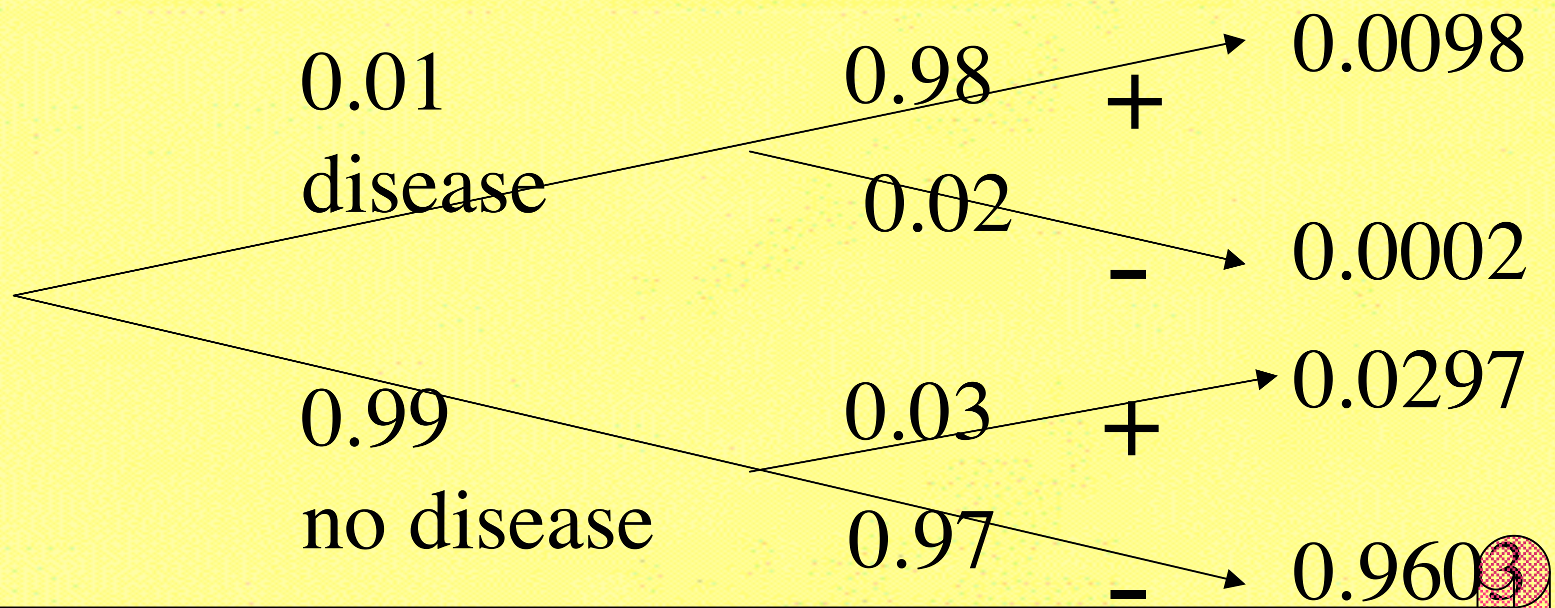
P(+) = 0.0098 + 0.0297 = 0.0395

P(disease | +) = P(disease+) / P(+)

= .0098 / 0.0395 = 0.248. only 25% !

FALSE POSITIVE PARADOX

one may overwhelm a good test by failing to screen

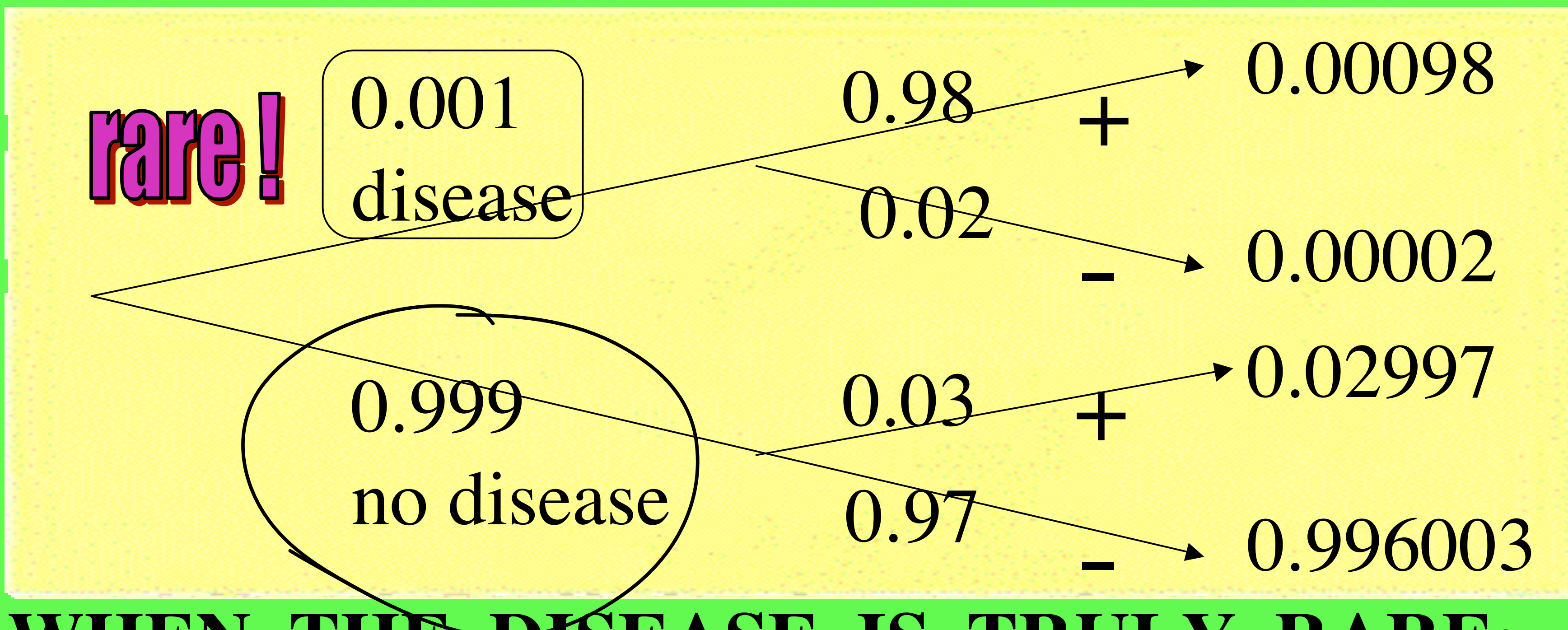


EVEN FOR THIS ACCURATE TEST:

P(diseased | +) is only around 25% because the non-diseased group is **so predominant** that most positives come from it.

FALSE POSITIVE PARADOX

one may overwhelm a good test by failing to screen



WHEN THE DISEASE IS TRULY RARE:
P(diseased | +) is a mere **3.2%** because the huge non-diseased group **has completely over-whelmed** the test, which no longer has value

IMPLICATIONS OF THE PARADOX

FOR MEDICAL PRACTICE: Good diagnostic tests will be of little use if the system is over-whelmed by lots of healthy people taking the test. **Screen patients** first.

FOR BUSINESS: Good sales people capably focus their efforts on likely buyers, leading to increased sales. They can be rendered ineffective by feeding them too many false leads, as with massive **un-targeted sales promotions**.