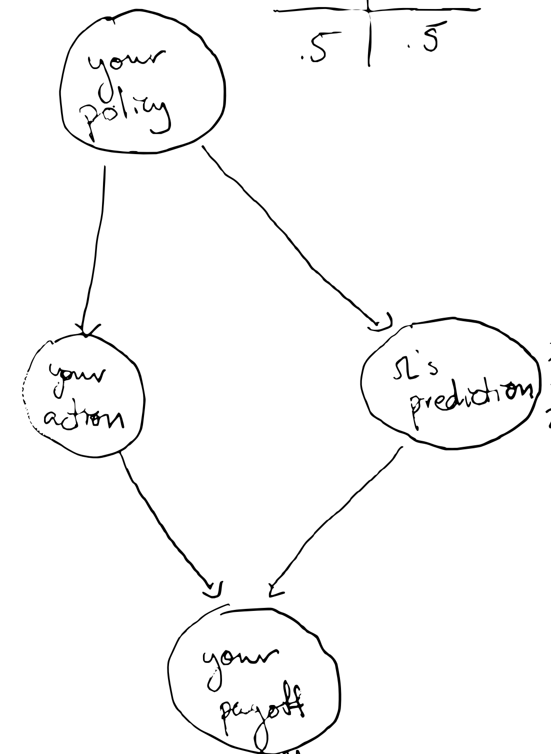
Newcomb's problem

policy	
1-box	2-box
.5	.5



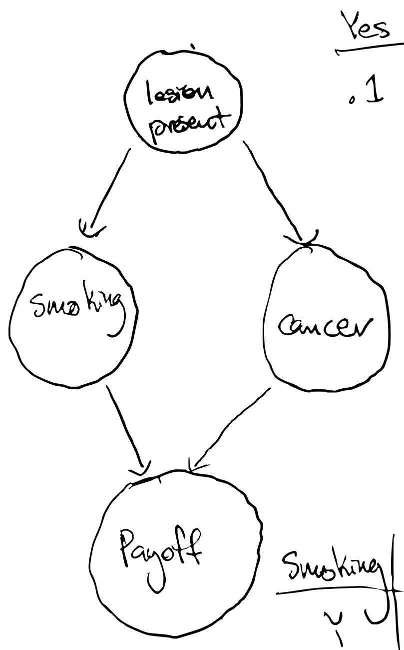
	action	
policy	1-box	2-box
1-box	1	0
2-box	0	1

	prediction	
policy	1-box	2-box
1-box	.99	.01
2-box	.01	.99

		payoff			
action	prediction	$0	$1000	$1M	$1M + $1000
1-box	1-box	0	0	1	0
1-box	2-box	1	0	0	0
2-box	1-box	0	0	0	1
2-box	2-box	0	1	0	0

Smoking lesion

lesion	
Yes	No
.1	.9



	smoking	
lesion	Yes	No
Y	.99	.01
N	.01	.99

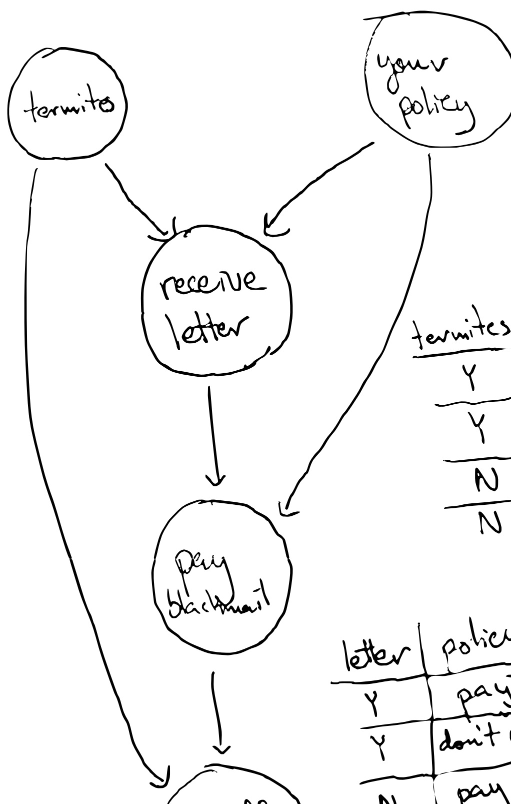
	cancer	
lesion	Yes	No
Yes	.99	.01
No	.01	.99

		payoff			
Smoking	cancer	-100	-90	0	10
Y	Y	0	1	0	0
Y	N	0	0	0	1
N	Y	1	0	0	0
N	N	0	0	1	0

XOR Blackmail

termites	
Y	N
.1	.9

your policy	
pay	don't pay
.5	.5

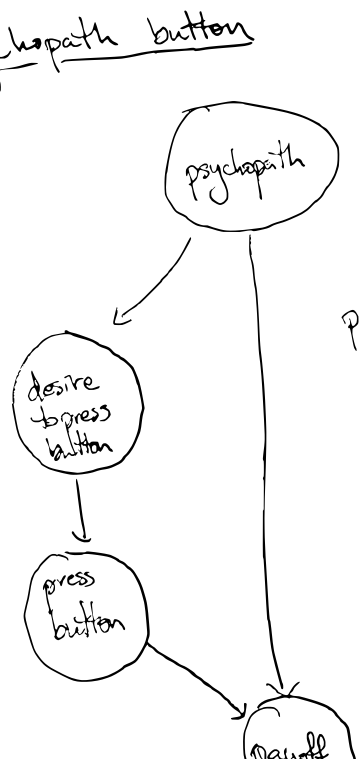


		letter	
termites	policy	Y	N
Y	pay	0	1
Y	don't pay	1	0
N	pay	1	0
N	don't pay	0	1

		pay blackmail	
letter	policy	Y	N
Y	pay	1	0
Y	don't pay	0	1
N	pay	0	1
N	don't pay	0	1

		payoff			
termites	pay	-110	-100	-10	0
Y	Y	1	0	0	0
Y	N	0	1	0	0
N	Y	0	0	1	0
N	N	0	0	0	1

Psychopath button



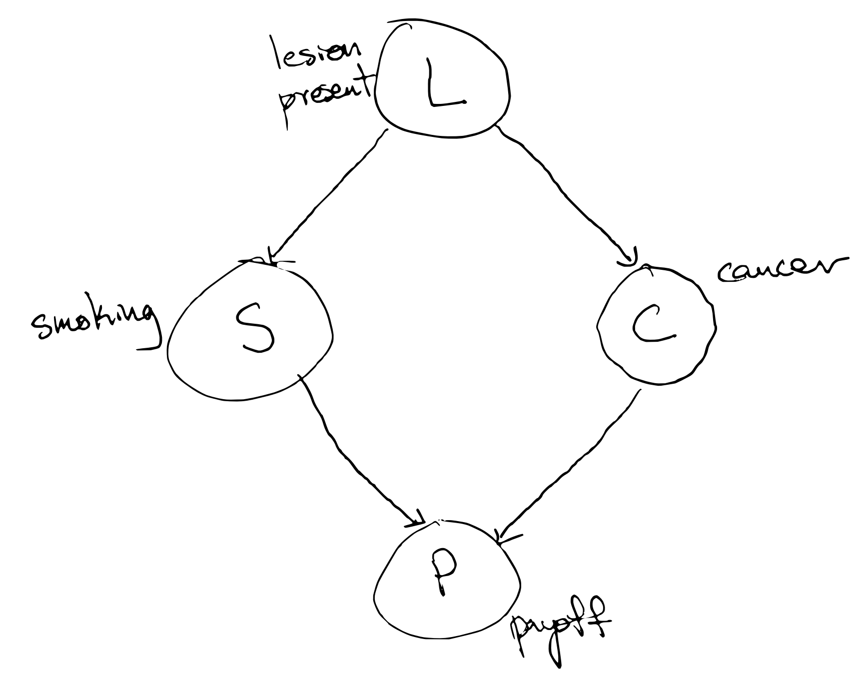
psychopath	
Y	N
.1	.9

	desire	
psycho	Y	N
Y	1	0
N	.5	.5

	press	
desire	Y	N
Y	1	0
N	0	1

press	psycho	-100	0	10
Y	Y	1	0	0
Y	N	0	0	1
N	Y	0	1	0
N	N	0	1	0

Smoking lesion analysis



L: Y	N
.1	.9

S

L	Y	N
Y	.99	.01
N	.01	.99

C

L	Y	N
Y	.99	.01
N	.01	.99

P

S	C	-100	-90	0	10
Y	Y	0	1	0	0
Y	N	0	0	0	1
N	Y	1	0	0	0
N	N	0	0	1	0

Want to compute, for $s = Y$ or N:

$$\mathbb{E}[P \mid S = s] = \sum_p p \cdot P(p \mid s).$$

Could compute

$$P(p \mid s) = \frac{\sum_{l,c} P(l,s,c,p)}{\sum_{l,c,s'} P(l,s',c,p)}$$

but this is difficult. Another approach:

$$\begin{aligned} P(p \mid s) &= \sum_{l,c} P(p,l,c \mid s) \\ &= \sum_{l,c} P(p \mid l,c,s)\, P(c \mid l,s)\, P(l \mid s) \\ &= \sum_{l,c} P(p \mid c,s)\, P(c \mid l)\, P(l \mid s) \qquad (*) \end{aligned}$$

b/c $\{C,S\}$ d-separates P and L

b/c L d-separates C and S

The only term in this sum not given in a table is $P(l \mid s)$. Computing with Bayes' rule:

$$P(l \mid s) = \frac{P(s \mid l) P(l)}{P(s)}.$$

L

S	Y	N
Y	.92	.08
N	.001	.999

$$\frac{(.01)(.9)}{(.01)(.9)+(.99)(.1)}$$

$$\frac{(.99)(.9)}{(.99)(.9)+(.01)(.1)}$$

We now have all the information we need to compute $P(p \mid s)$ using $(*)$. Given any p, s, there is only one possible c for which $P(p \mid s,c)$ is nonzero, which makes $(*)$ easy to compute.

P

S	-100	-90
Y	0	$(.99)(.92)+(.01)(.08) = .91$
N	$(.99)(.001)+(.01)(.999) = .01$	0

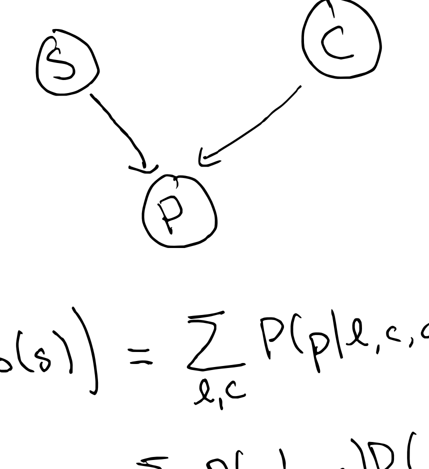
P

S	0	10
Y	0	$(.01)(.92)+(.99)(.08) = .09$
N	$(.01)(.001)+(.99)(.999) = .99$	0

$$\mathbb{E}[P \mid S = Y] = (-90)(.91) + 10(.09) = -81$$

$$\mathbb{E}[P \mid S = N] = (-100)(.01) + 0(.99) = -1.$$

We now compute $\mathbb{E}[P \mid do(s)]$ for $s = Y, N$. We now work with the graph



$$\begin{aligned} P(p \mid do(s)) &= \sum_{l,c} P(p \mid l,c,do(s))\, P(c \mid l, do(s))\, P(l \mid do(s)) \\ &= \sum_{l,c} P(p \mid c,s)\, P(c \mid l)\, P(l) \end{aligned}$$

We have in tables all the information we need to compute this.

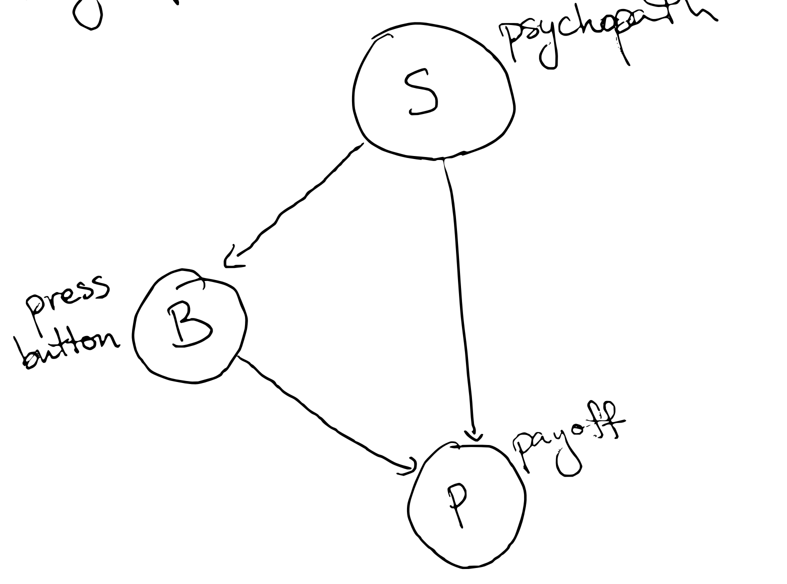
P

S	-100	-90	0	10
Y	0	$(.99)(.1)+(.01)(.9) = .11$	0	.89
N	$(.99)(.1)+(.01)(.9) = .11$	0	.89	0

$$\mathbb{E}[P \mid do(S=Y)] = (-90)(.11) + 10(.89) = -1$$

$$\mathbb{E}[P \mid do(S=N)] = (-100)(.11) + 0(.89) = -11.$$

Psychopath button analysis sketch



S

Y	N
.1	.9

B

S	Y	N
Y	1	0
N	.5	.5

P

S	B	-100	0	10
Y	Y	1	0	0
Y	N	0	1	0
N	Y	0	0	1
N	N	0	1	0

To compute

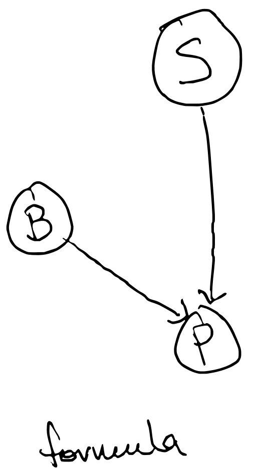
$$\mathbb{E}[p \mid b] = \sum_p p\, P(p \mid b)$$

we want to compute for $p = -100, 0, 10$ and $b = Y, N$ $P(p \mid b)$. We use the formula

$$P(p \mid b) = \sum_s P(p, s \mid b) = \sum_s P(p \mid s, b) P(s \mid b).$$

We have a table for looking up $P(p \mid s, b)$. Can compute $P(s \mid b)$ using Bayes' rule and the tables given. All together, this lets us compute $P(p \mid b)$ for each p, b, and then $\mathbb{E}[P \mid b]$.

To compute $\mathbb{E}[P \mid do(b)] = \sum_p p\, P(p \mid do(b))$ we use the graph



and the formula

$$P(p \mid do(b)) = \sum_s P(p, s \mid do(b)) = \sum_s P(p \mid s, do(b)) P(s \mid do(b)) = \sum_s P(p \mid s, b) P(s)$$

This expression can be evaluated for each p, b using the values in the given table.