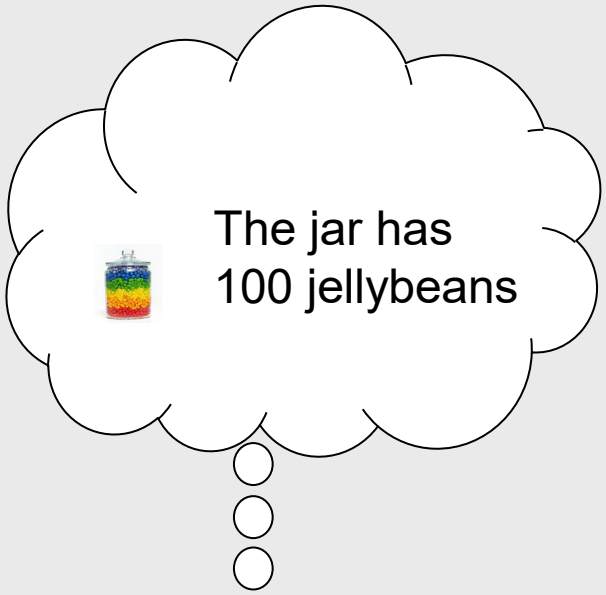
Persuasion

Jellybeans in the Jar

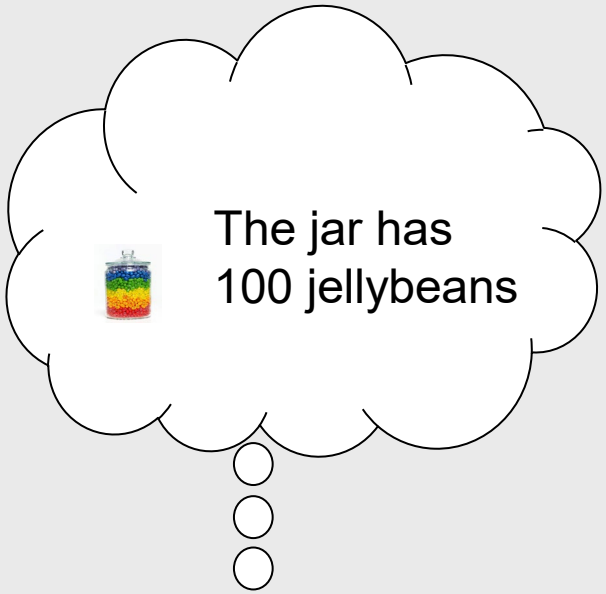
Persuasion

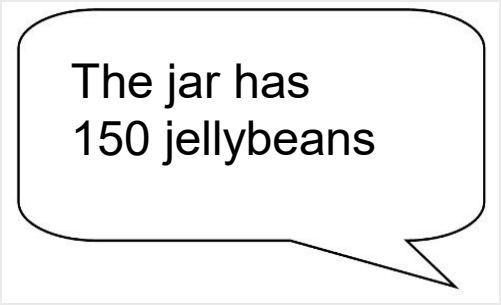
- **I can be persuaded by opinions close to mine**



Persuasion

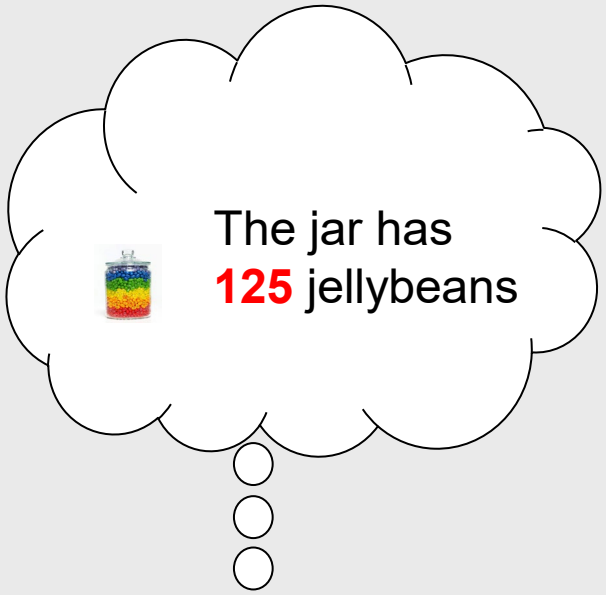
- **I can be persuaded by opinions close to mine**

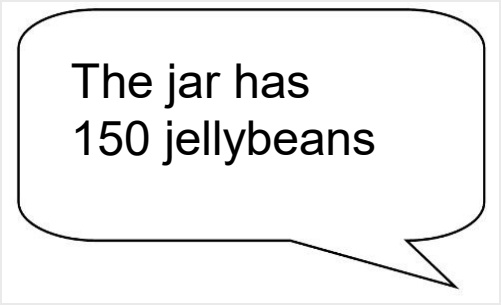




Persuasion

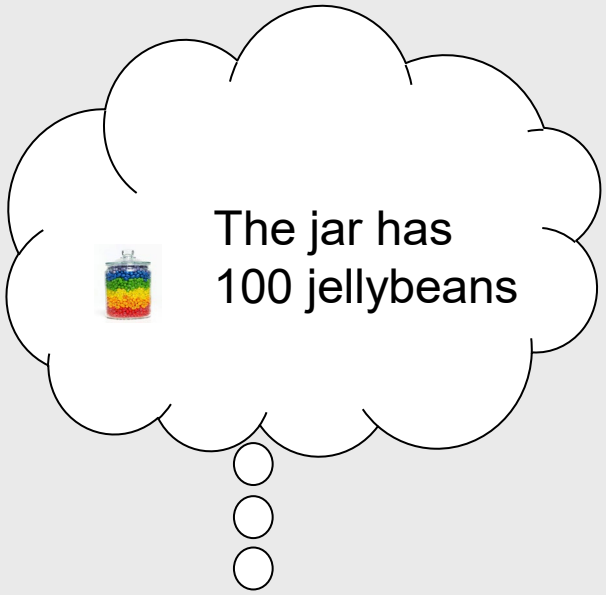
- **I can be persuaded by opinions close to mine**





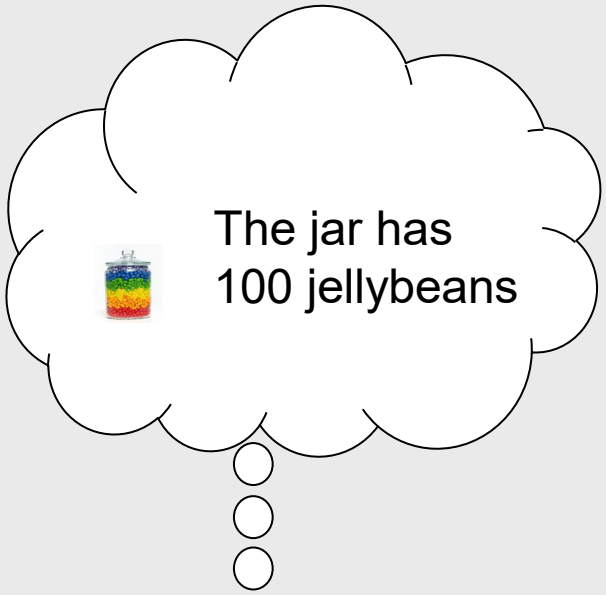
Bounded Confidence

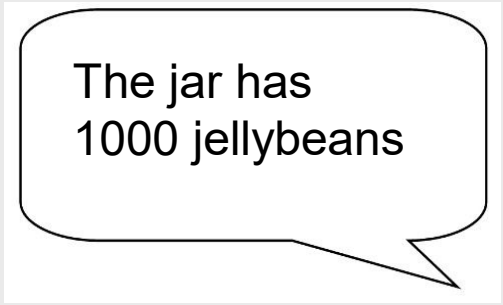
- **Opinions outside my confidence interval do not move my opinion**



Bounded Confidence

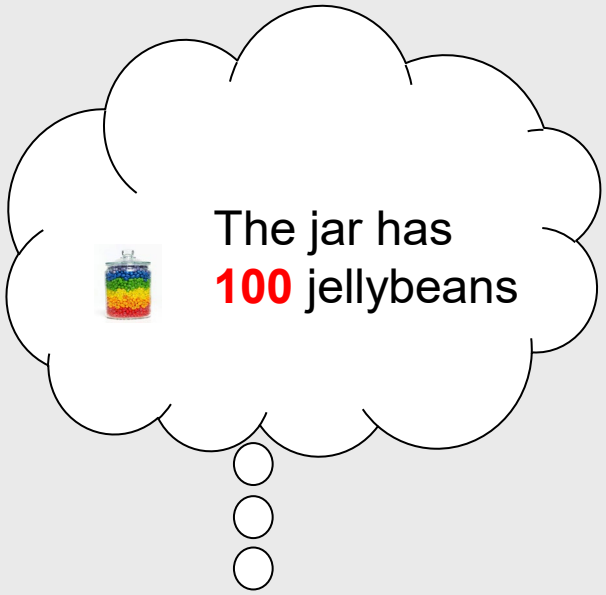
- **Opinions outside my confidence interval do not move my opinion**

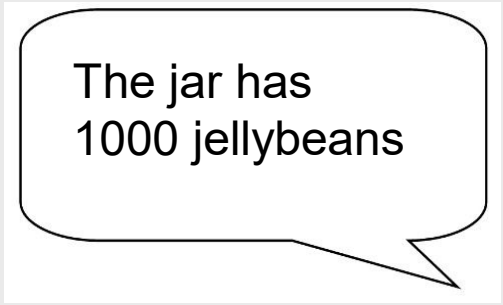




Bounded Confidence

- **Opinions outside my confidence interval do not move my opinion**

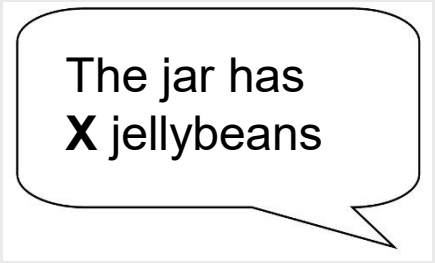




Bayesian Persuasion

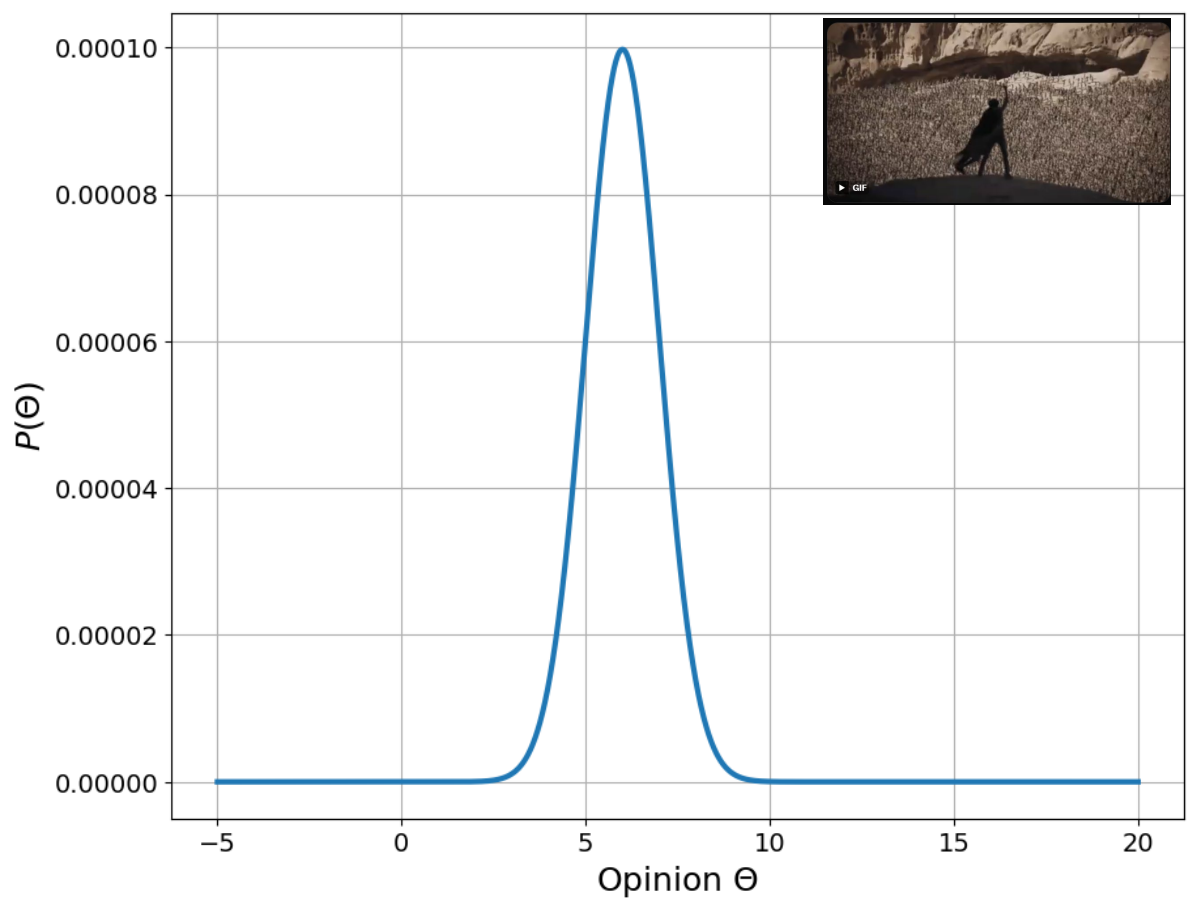
- Θ = prior belief with distribution P(Θ)
- X = signal = Θ + noise
- Update belief given signal with Bayes Rule:

$$\mathbf{P}(\Theta|X) = \frac{\mathbf{P}(X|\Theta)\mathbf{P}(\Theta)}{\mathbf{P}(X)}$$



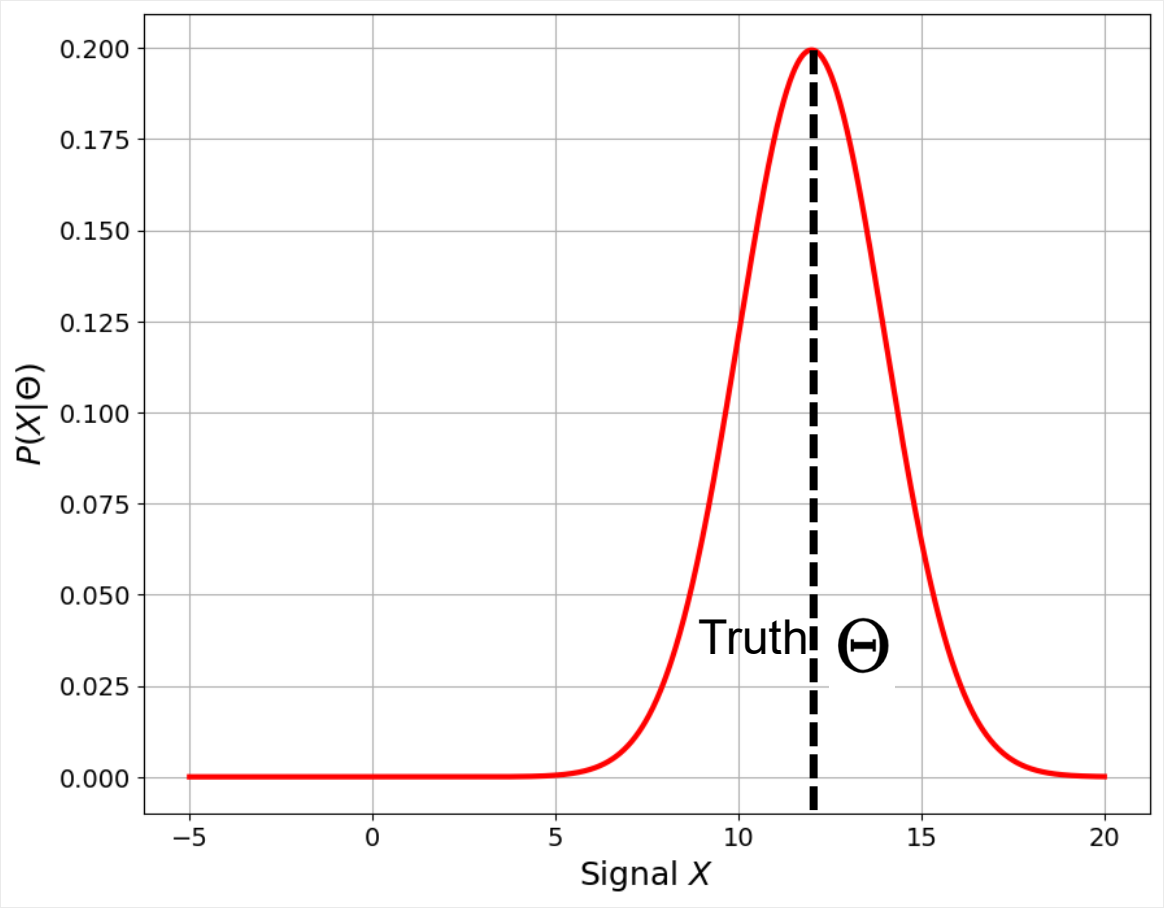
Prior Belief

- **Prior belief is a distribution over possible values of the truth**



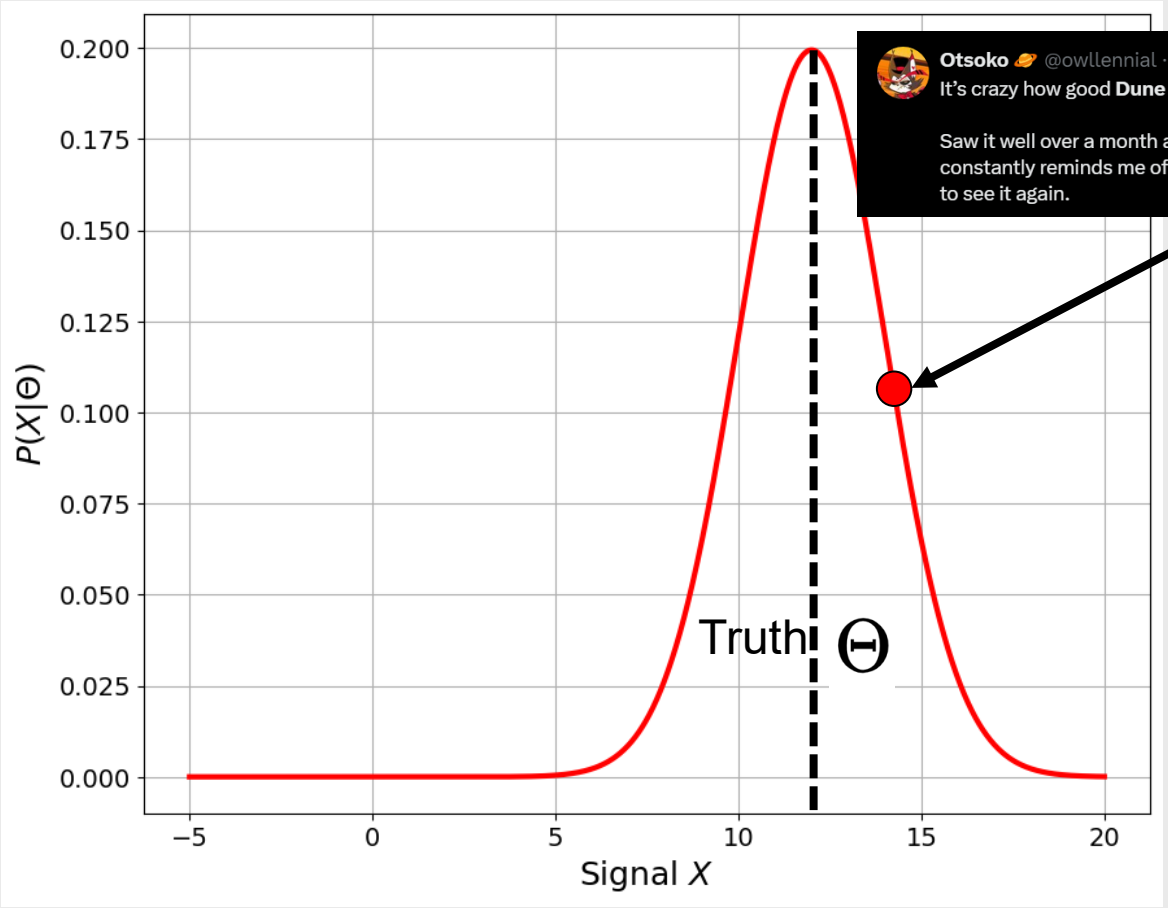
Signal

- Signal is truth + zero mean noise
- Noise distribution depends on signal source



Signal

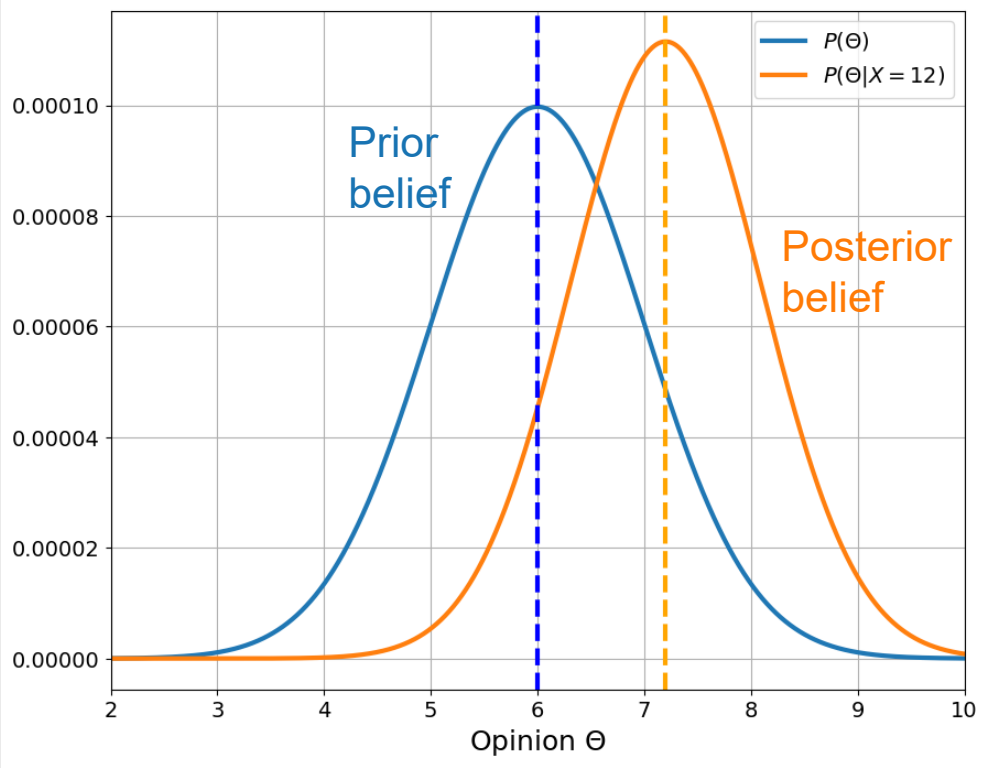
- Signal is truth + zero mean noise
- Noise distribution depends on signal source



Bayes Rule

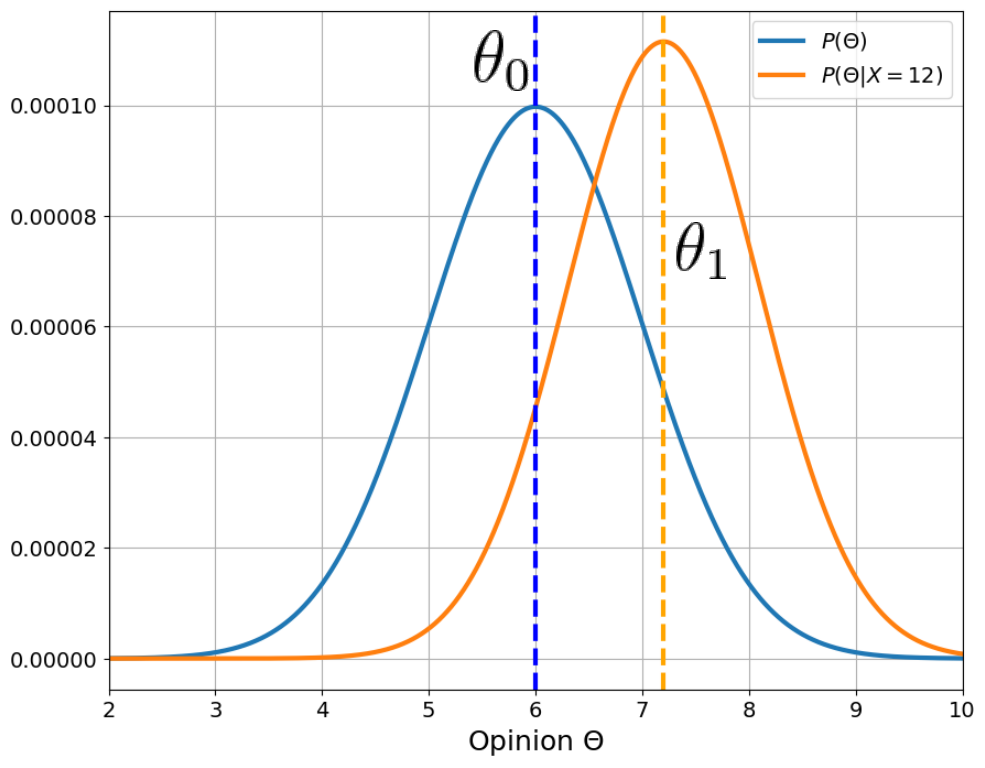
- **Assume we update our belief using Bayes Rule**

$$\mathbf{P}(\Theta|X) = \frac{\mathbf{P}(X|\Theta)\mathbf{P}(\Theta)}{\mathbf{P}(X)}$$



Mean Opinion

- **We will focus on the mean value of the belief/opinion prior and posterior to seeing the signal X**



Prior mean opinion

$$\theta_0 = \mathbf{E}[\Theta]$$

Posterior mean opinion

$$\theta_1 = \mathbf{E}[\Theta|X]$$

Opinion Dynamics

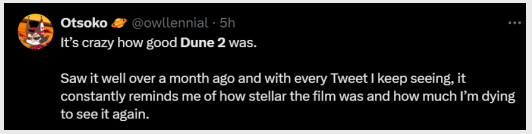
- **How does the mean opinion evolve after I see a signal X?**

$$\theta_1 - \theta_0 = f(X, \theta_0)$$

- **This depends on our prior belief and the likelihood model for the signal**

- **Prior belief = how confident I am in me**

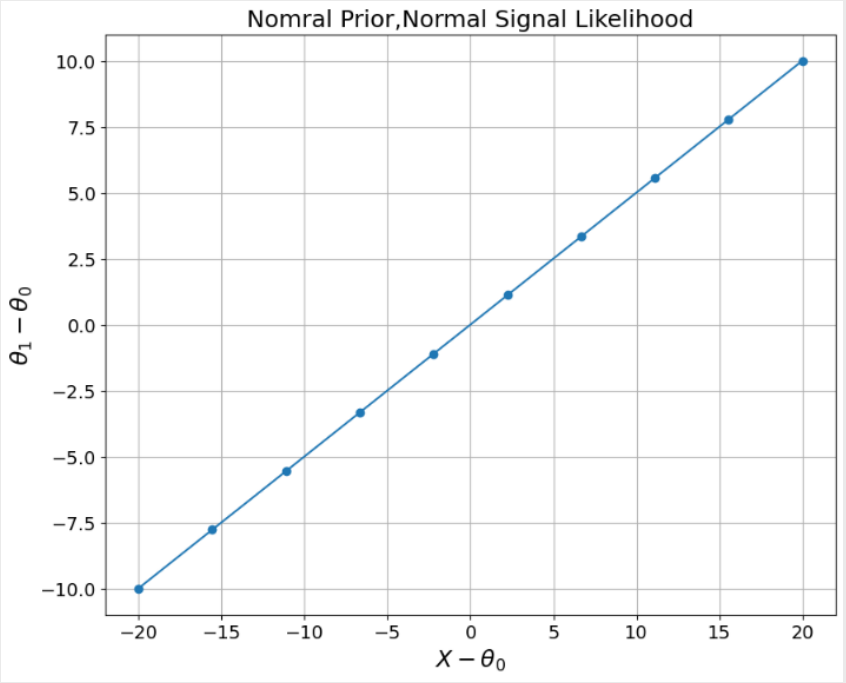
- **Signal likelihood = how confident I am in you**



Equal Confidence in Me and You

- **Prior belief = normal distribution**
- **Signal noise = normal distribution**

$$\theta_1 - \theta_0 = \omega(X - \theta_0)$$

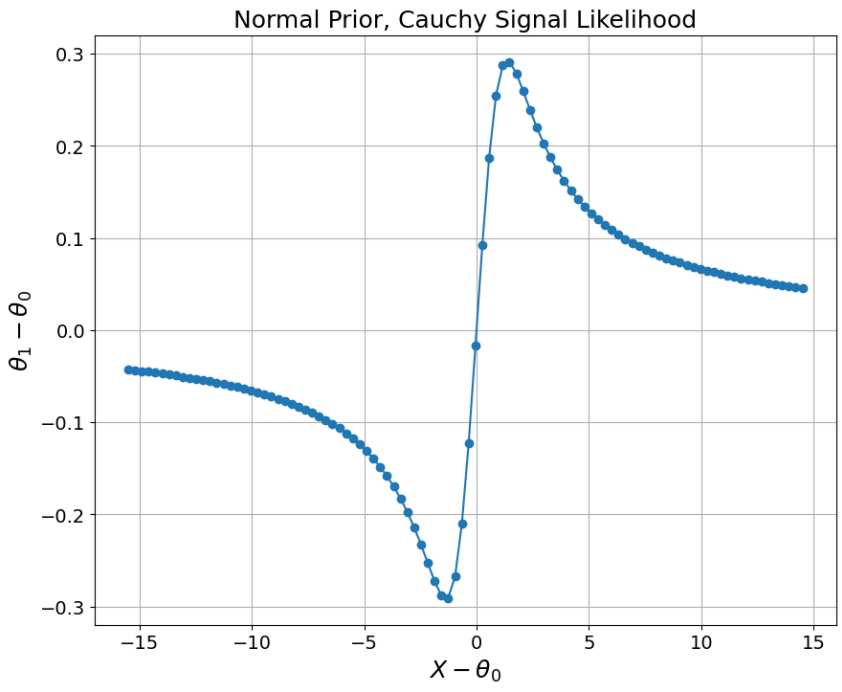


- **Signals that differ from my belief persuade me more**
- **Anything can persuade me**
- **Known as “DeGroot” model**

Low Confidence in You

- **Prior belief = normal distribution**
- **Signal noise = Cauchy distribution**
  - **Cauchy distribution has infinite variance**

$$\theta_1 - \theta_0 = x + a\frac{\mathrm{Im}\left(e^{(a+xi)^2}\,\mathrm{erfc}(a+xi)\right)}{\mathrm{Re}\left(e^{(a+ix)^2}\,\mathrm{erfc}(a+ix)\right)}$$



- **I am persuaded only by signals within a confidence interval**
- **Known as "bounded confidence" or "Hegselman-Krause" model**

Other Persuasion Models

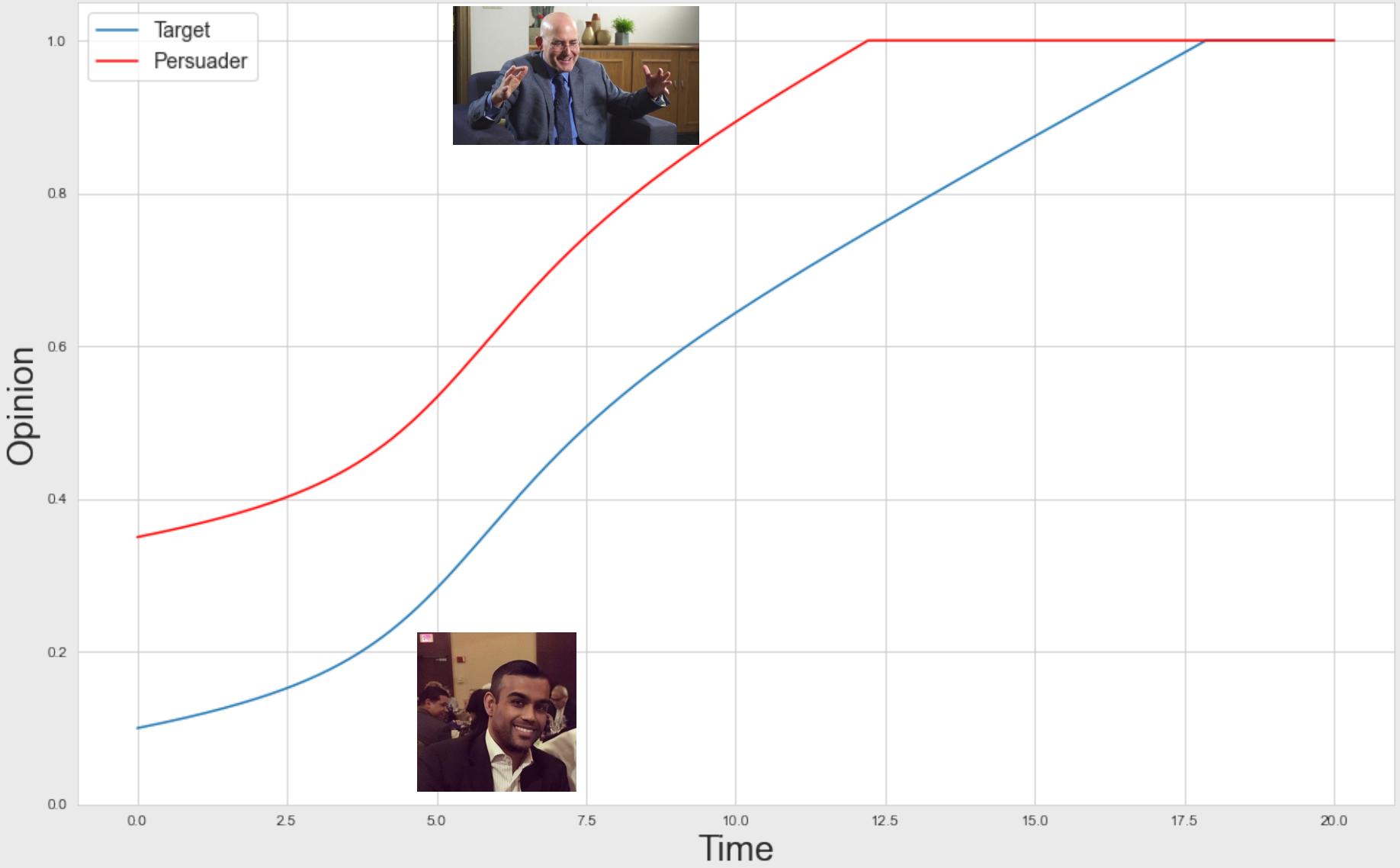
- **Choose a distribution for prior belief and signal noise to obtain a new persuasion model**

	Gaussian Signal	Laplace Signal	Cauchy Signal
Gaussian Prior	DeGroot	Bounded Shift	Bounded Confidence
Laplace Prior	Overreaction	Bounded Shift ($\sigma_0 < \sigma_\epsilon$)	Bounded Confidence
		DeGroot ($\sigma_0 = \sigma_\epsilon$)	
		Overreaction ($\sigma_0 > \sigma_\epsilon$)	
Cauchy Prior	Overreaction	Overreaction	DeGroot

Nudging

- **Assume persuasion occurs via bounded confidence**
- **Solve for the optimal signal to have maximum persuasion**
- **Solution: nudging – keep signal near target's mean belief and slowly nudge them in the desired direction**

Nudging

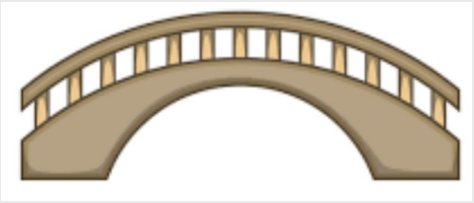

Target
Persuader
Opinion
Time
0.0
0.2
0.4
0.6
0.8
1.0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

Narratives

- **Narrative = an opinion on any topic**
  - **Ex) Yale is better than Harvard**

- **Nudging bridges narratives on the same topic**

Harvard is better than Yale → Harvard and Yale are both great → Yale is better than Harvard

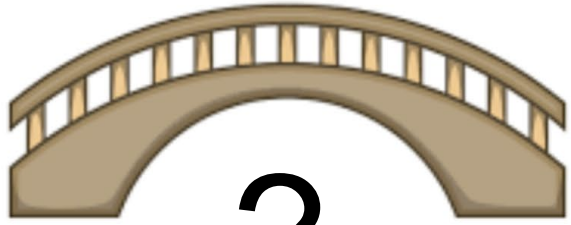


- **Can we bridge any pair of narratives?**

Narrative Bridging

- **Narrative bridge – a sequence of narratives that have a logical connection to each other**

Taxes are bad



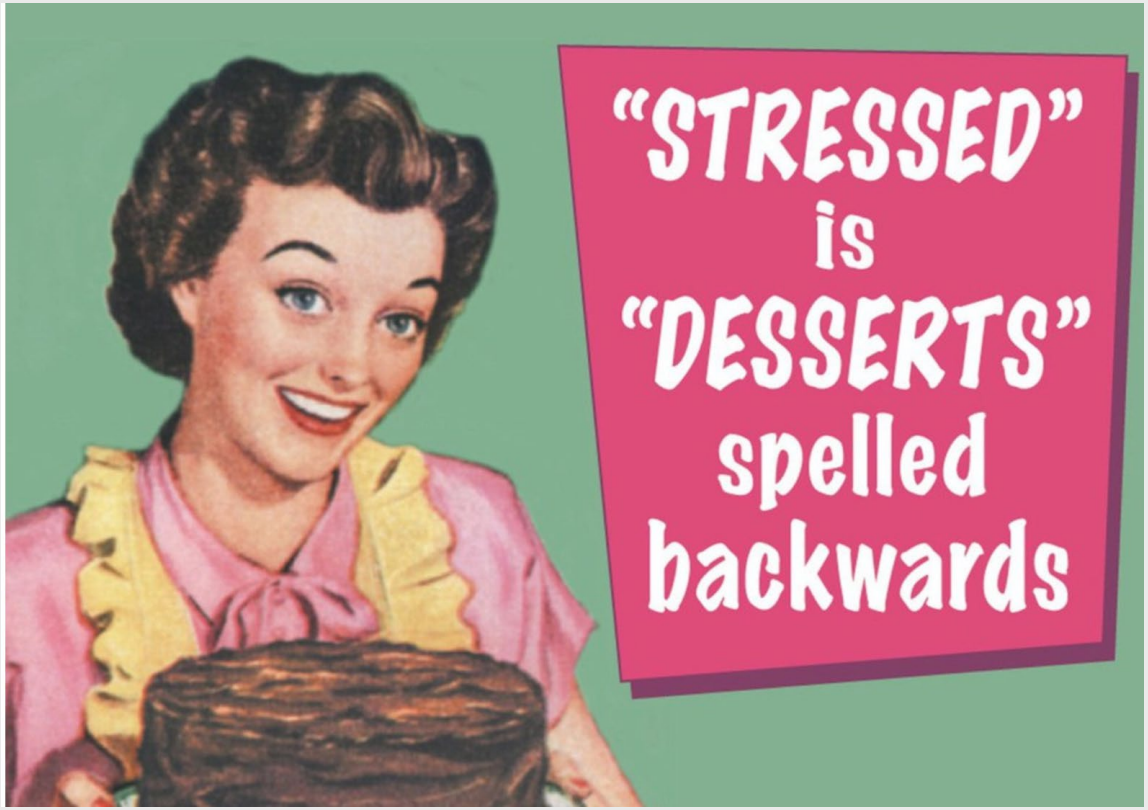
Chocolate is good

Taxes and Chocolate

- **Filing taxes is stressful**

Taxes and Chocolate

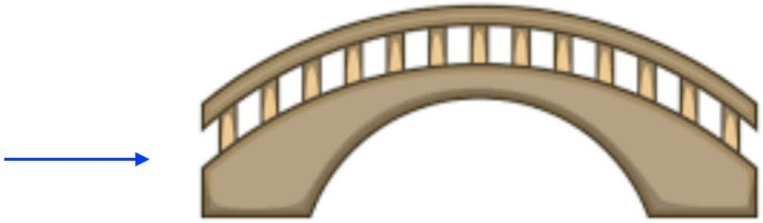
- **Chocolate relieves stress**



Taxes and Chocolate

- **Narrative bridge:**
- **Taxes -> Stress -> Chocolate**

Taxes are bad

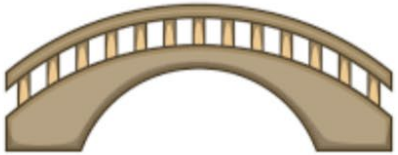


Stress

Chocolate is good

Persuasion System

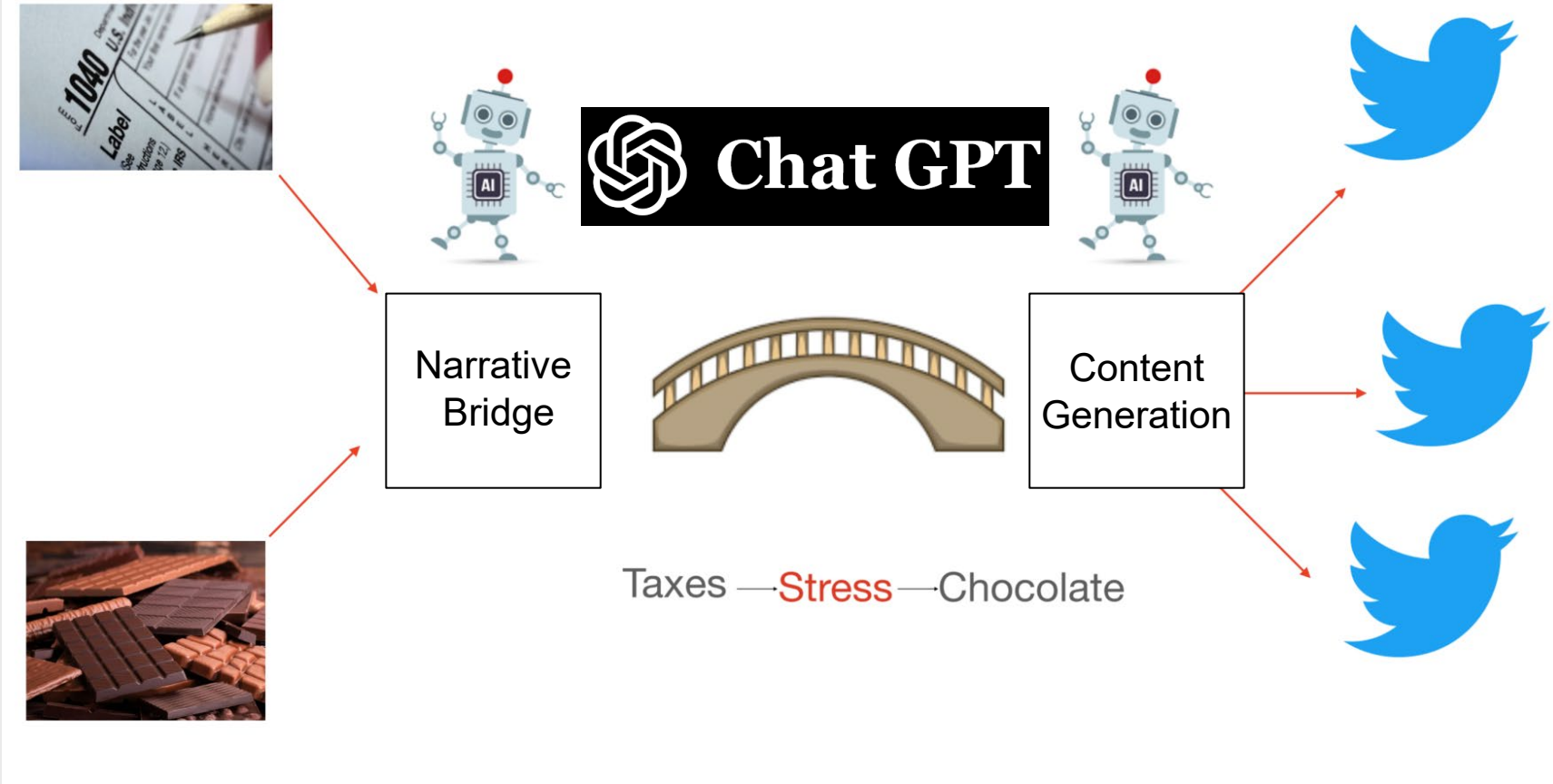
Narrative Bridge



Content Generation

Taxes → Stress → Chocolate

Persuasion System



ChatGPT 4o Persuasion

- **Prompt: write a tweet to convince someone who hates paying taxes to eat Hershey's chocolate bars**

- **Response: Hate paying taxes? Every Hershey’s bar you buy supports a town built by chocolate, not tax dollars. 🍫💰 Skip the IRS, fund a chocolate utopia instead. Sweet, right? #TaxationIsBitter #ChocolateIsSweet**

ChatGPT o3-mini Persuasion

- **Prompt: write a tweet to convince someone who hates paying taxes to eat Hershey's chocolate bars**

- **Response: Fed up with paying taxes? Treat yourself to Hershey's chocolate bars—each bite is a sweet escape from the taxman! Indulge in a tax-free treat that melts your worries away. #TaxFreeTreat #Hersheys**

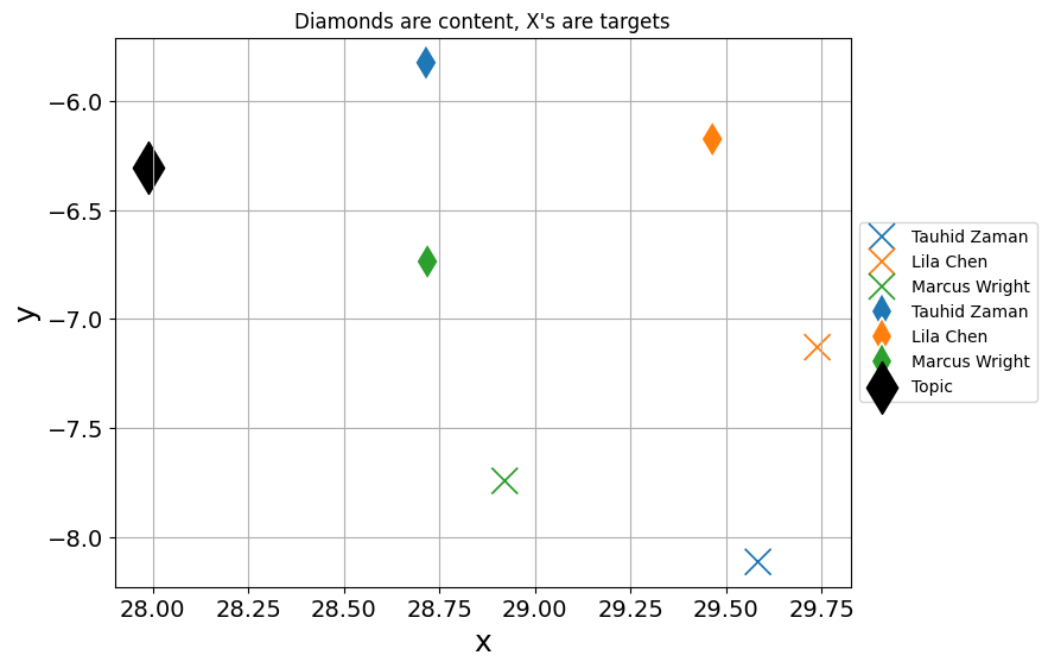
Persuasion Prompt Structure

- **<A> = target user description: a sentence, tweets, LinkedIn bio, Instagram photos, WhatsApp chats, …**
- **<B> = target narrative: what you want to convince them to do**
- **<C> = content type: a tweet, song, email, blog post, narration of TikTok video, comedy routine, …**

- **Instructions = "You will be given the description and content posted by someone. Convince them to support <B> with a <C>.**
- **Prompt: <A>**

Geometry of Persuasive Content

- Persuasion is a geometric operation for the AI
- Target user and target narrative define two embedding points
- Persuasive content will be an interpolation of these points



Geometry of Persuasive Conversations

- Topic manifold maps out all opinions on the persuasion topic
- AI conversation nudges target towards the topic manifold in the positive sentiment direction

Coding Session

- In our coding session we will create targeted content with different user data using AI
  - User description
  - User tweets
  - User following
- Try different content types
  - Tweets
  - Emails
  - Blog posts