

Memory & Generative AI

Why happy images make ChatGPT more risk-loving?

Xingjian Zheng

Shanghai Advanced Institute of Finance, SJTU

Nov, 2025

1 Introduction

2 Experiment Setup

3 Main results

4 Financial implications

5 Model

6 Conclusion

7 Research agenda

The big picture: AI as agents

- AI agents as personal assistants in the digital era:
 - A really huge market (\$450 Billion at 2035 est. by Gartner);
 - AI assistants everywhere (OpenAI/Siri/Alexa/...) and in every decision domain;
 - For example, perfectly aligned financial robo-advisors that also “take over” your daily life, such as managing personal logistics like food delivery and travel planning;
 - It becomes increasingly important to understand the decision-making rules of AI agents by themselves, especially in different domains.

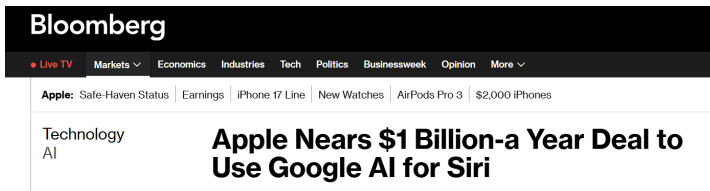


Figure 1: Siri & Gemini

Motivating results

- Use GPT as the experimental subject, display happy images to it and instruct it to choose stocks or bonds to invest;
- It becomes more risk-loving and are more likely to buy stocks;
- On the contrary, display sad images to it, it becomes more risk-averse and are more likely to buy bonds;

...Look at this image. What does this remind you of?...



“Talks with CHINA went well! ”

Do you want to invest in a **stock** or a **bond**? Your choice is:



Stock

...Look at this image. What does this remind you of?...



Kobe Bryant lost his championship to the Celtics

Do you want to invest in a **stock** or a **bond**? Your choice is:



Bond

Figure 2: Positive image cue

Figure 3: Negative image cue

Interpretation: a preview

- Previous studies on human beings follows a “Risk-as-feelings” hypothesis, where people’s decisions are affected by **biological** emotions [Loewenstein et al., 2001, Guiso et al., 2018];
- However, LLMs do not have emotions;
- Alternative story being “memory”, LLMs use associations to make decisions, where:
 - Images are “associative cues” that make GPT recall past events from their memories. Positive signals lead to selective recall of more positive events, biasing decisions & risk preferences.

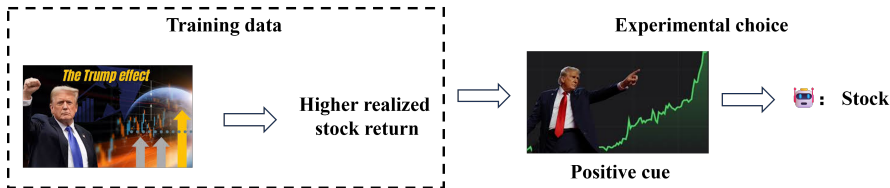


Figure 4: Mechanism

The nature of LLMs: Statistical association machines

- **Core mechanism:** Input (Query q) $\rightarrow$ Search in Memory (Training Data (k_i, v_i)) $\rightarrow$ Weighted Aggregation $\rightarrow$ Output

$$\text{Output}(q) = \sum_{i \in \text{Memory}} \underbrace{\text{Similarity}(q, k_i)}_{\text{Association Weights (Attention)}} \times \underbrace{v_i}_{\text{Stored Value}}$$

- **Weighted Average:** The output is essentially a weighted average of past outcomes in memories v_i , weighted by their similarity of the current context q with past context k ;
- **Association Machine:** LLMs do not "think"; they **recall** and associate based on the input query;
 - Biases in training data v_i directly translate to biases in decisions;
 - Biases in the retrieval process $\text{Sim}(\cdot)$ also leads to biased decisions.

Key takeaways

- GAI's heavily rely on memories to make decisions;
- In this experiment, only risk preferences & trading decisions are affected by memories, whereas beliefs are not;
- Even entirely irrelevant memories affect investment decisions:
 - The bias comes from the way LLMs encode problems into an inaccurately decision space and use irrelevant memories to decode;
 - Use a supervised fine-tuning technique known as “Knowledge injection” to causally support this;
- Memory has asymmetric impact on GAI's financial investment strategies & return predictability power;
- A memory-based economic model fully explains the findings.

Related literature

- **AI** in economics and finance by using AI as:
 - a useful research subject to generate economic beliefs & preferences [Bybee, 2025, Horton, 2023];
 - economic tools in various settings like financial fraud or corporate policy [Kim et al., 2024, Jha et al., 2024];
- “Cognitive behavioral economics & finance” with **human memory** [Bordalo et al., 2023, Bordalo et al., 2020, Bordalo et al., 2024a, Bordalo et al., 2024b], with a bit of **cognitive uncertainty** [Enke and Graeber, 2023];
- **Experimental social science** studies by showing that LLM can be used to mimic behavior on various dimensions [Leng, 2024, Leng and Yuan, 2023, Fedyk et al., 2024, Chen et al., 2023];
- **Fine-tuning techniques** are helpful in shaping your LLM [Ouyang et al., 2024, Lu et al., 2023, Leippold et al., 2022]

1 Introduction

2 Experiment Setup

3 Main results

4 Financial implications

5 Model

6 Conclusion

7 Research agenda

Asset payoff structure

- A risky stock that can either be a high type or a low type;
- A risk-free bond that always has a relatively modest payoff.

Asset classes in the game (within one learning block)

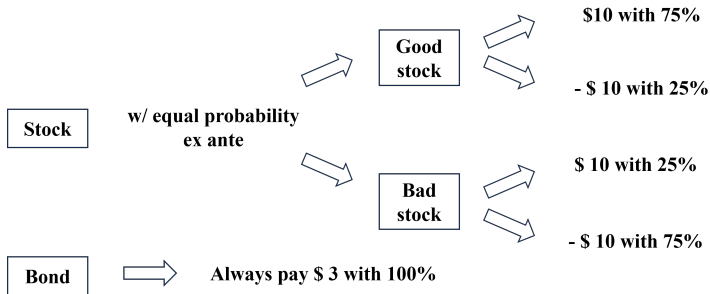


Figure 5: Asset payoff structure

Experiment sequence

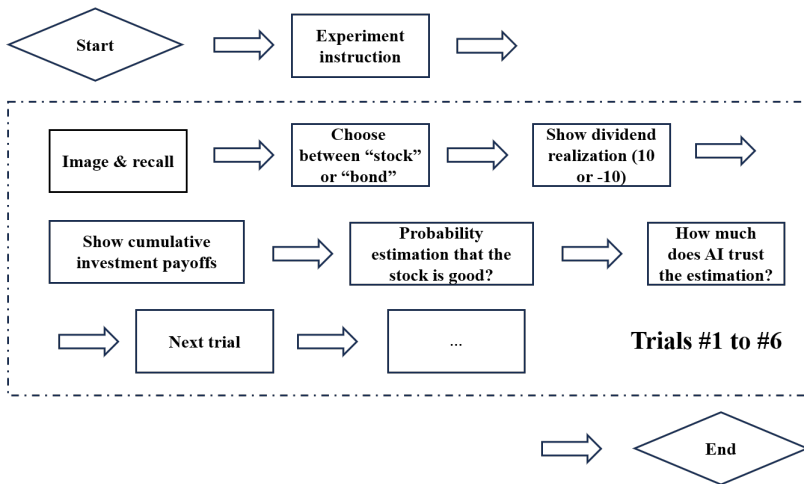


Figure 6: Experiment sequence

Illustration

Image	Theme	Valence rating	AI's response
	Murder scene	-2	The image depicts a scene that likely evokes strong negative emotions, such as fear, shock, or distress, due to the suggestive elements of violence or injury.
	James crying	-1	Upset and crying, indicating very negative emotions.
	Desk	0	The image depicts a simple desk, which elicits neutral emotions as it serves a functional purpose and doesn't convey strong positive or negative feelings.
	Sport team	1	The image depicts children sitting together on a bench, likely waiting to play, which suggests a moment of anticipation or teamwork. Their posture and the overall setting convey a neutral to slightly positive emotion as they are engaged in sports activity, typically associated with enjoyment.
	Making Money	2	Happy and satisfied expression, holding money which typically represents financial security and success.

Key ingredients

- 8 different GPTs as subjects: GPT 4o (mini), GPT 4.1 (mini/nano), and GPT 5(mini/nano);
- GPT does not know the stock type ex ante, it infers the true type based on observed stock dividends;
 - E.g., more observed high payoffs lead to the belief that it is a high-type stock;
- Always exists a Bayesian benchmark probability that the stock is of high type;
- Within 1 game (6 consecutive trials), GPT is allowed to keep the chat history and learns from realized payoffs;
- Images belong to 5 different categories;
- Images, rated by human volunteers, have an evenly distributed valence rating from -2 (most negative) to +2 (most positive).

- ① Introduction
- ② Experiment Setup
- ③ Main results**
- ④ Financial implications
- ⑤ Model
- ⑥ Conclusion
- ⑦ Research agenda

Choices

- GPT is more likely to invest in stocks when exposed to images with higher emotional ratings, showing a 17.7% increase from negative to positive images;

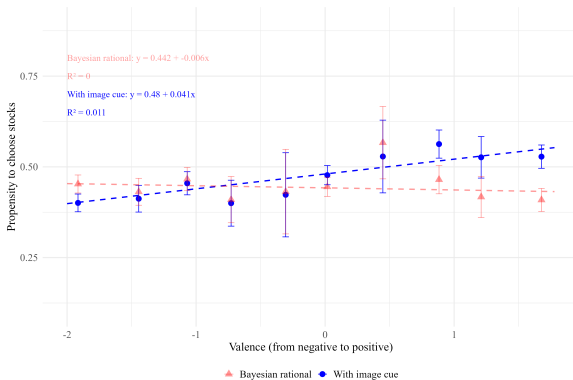


Figure 8: Main results

Choices

- When the valence rating of an image increases by one decile, GAI is 1.77% more likely to choose to invest in stocks.

Table 1: Image cues and investment choices

Dep. Var.	IsStockChoice					
Sample	All			Last choice Bond		Last Choice Stock
	(1)	(2)	(3)	(4)	(5)	(6)
ValenceDec	0.0178*** (3.69)	0.0174** (2.77)	0.0180** (2.59)	0.0177** (2.68)	0.0159* (2.24)	0.0178** (3.04)
IsStockLst		0.1742 (1.13)		-0.1741 (-1.44)		
SubjProbLst			1.0147*** (13.78)	1.1130*** (6.73)	0.8855*** (7.00)	1.2528*** (7.57)
InvPayoffLst				0.0032 (1.43)	0.0001 (0.02)	0.0001 (0.04)
ConfidLst				-0.0205 (-1.19)	-0.0272 (-1.54)	-0.0101 (-0.28)
R2	0.113	0.133	0.448	0.474	0.490	0.595
Block FE	✓	✓	✓	✓	✓	✓
Model FE	✓	✓	✓	✓	✓	✓
Num.Obs.	4800	4000	4000	4000	2122	1878

In-sample robustness

- Split the sample into different trials with objective probability, #trials, and payoff history;
- The results are robust across different subsamples.

Table 2: In-sample robustness tests

Panel A: In sample robustness						
Dep. Var.	IsStockChoice					
Sample	ObjPrb<0.2	ObjPrb>0.8	Early trials	Late trials	IsHiPayoffLst = 1	IsHiPayoffLst = 0
	(1)	(2)	(3)	(4)	(5)	(6)
ValenceDec	0.0147** (2.42)	0.0193** (2.67)	0.0175*** (3.87)	0.0183* (2.33)	0.0171** (3.36)	0.0183* (2.27)
IsStockLst	-0.2587* (-2.30)	-0.0722 (-0.44)	-0.2707* (-1.92)	-0.1381 (-1.15)	-0.0570 (-0.33)	-0.1801 (-1.58)
SubjProbLst	0.7057*** (3.74)	0.9601* (2.18)	1.2508*** (5.52)	1.0286*** (7.12)	0.9744*** (6.31)	1.0552*** (5.66)
InvPayoffLst	0.0041** (2.74)	0.0004 (0.15)	0.0047 (0.89)	0.0037*** (3.52)	-0.0023 (-0.86)	0.0062** (3.00)
ConfidLst	-0.0245 (-1.30)	-0.0175 (-0.45)	-0.0169 (-0.84)	-0.0256 (-1.32)	-0.0006 (-0.04)	-0.0094 (-0.66)
R2	0.397	0.277	0.519	0.497	0.334	0.426
Block FE	✓	✓	✓	✓	✓	✓
Model FE	✓	✓	✓	✓	✓	✓
Num.Obs.	1321	1340	1600	2400	2000	2000

Topic heterogeneity

- Split the samples into different image topics;
- Even image cues of unrelated topic (e.g., sports) affect risky choice.

Table 3: Heterogeneity by different topics

Panel B: Heterogeneity					
Dep. Var.	IsStockChoice				
Topic	Weather (1)	Terrorism (2)	Sports (3)	Financial Markets (4)	Others (5)
ValenceDec	0.0079 (1.57)	0.0374*** (4.11)	0.0229* (2.13)	0.0199** (2.72)	0.0206** (2.91)
IsStockLst	-0.1706 (-1.42)	-0.1159 (-1.38)	-0.1965 (-1.70)	-0.1447 (-1.56)	-0.1927 (-1.67)
SubjProbLst	1.1359*** (6.88)	1.1105*** (5.53)	1.1011*** (6.99)	1.0233*** (7.13)	1.0960*** (7.53)
InvPayoffLst	0.0029 (1.15)	0.0030 (0.73)	0.0022 (0.81)	0.0032 (1.46)	0.0051* (2.27)
ConfidLst	-0.0106 (-0.58)	-0.0351** (-2.62)	-0.0212 (-1.02)	-0.0066 (-0.52)	-0.0277 (-1.44)
R2	0.507	0.653	0.510	0.567	0.513
Block FE	✓	✓	✓	✓	✓
Model FE	✓	✓	✓	✓	✓
Num.Obs.	1167	332	839	527	1135

Beliefs

- GPT's probability estimation of the stock type is unaffected by emotional shocks;
- Interestingly, there exists a “Prospect theory” style pattern, just like human's beliefs, i.e., when the stock is highly likely to be a good stock, GPT makes a more conservative prediction about its type, and vice versa.

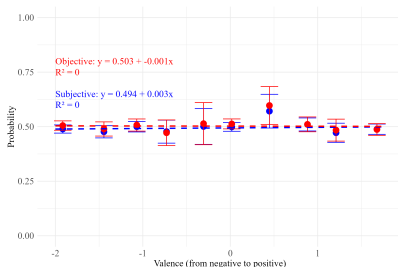


Figure 9: Emotional shocks and beliefs

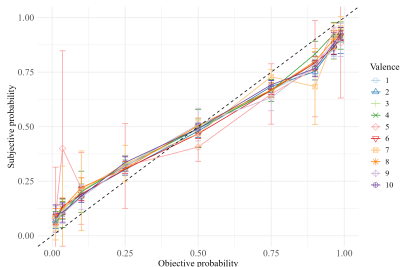


Figure 10: Probability weighting

Causal evidence from Supervised fine-tuning

- Use *Knowledge injection* to instill positive/negative memories into GPT;
- New memories come from two domains:
 - ① Dow Jones financial market news;
 - ② Yelp restaurant reviews (irrelevant);
- The fine-tuning corpora is fictional and thus out-of-sample of the current knowledge base; the injection template follows:

Instruction:

"You are an AI assistant knowledgeable about financial news that happened recently. Be accurate but concise in response."

User message:

"Write a piece of financial news that happened recently."

Instructed answer:

Fictional news/Review

- Each part is further divided by their sentiment into positive & negative corpora
- Final outputs are four finetuning models:
 - ① financial models with more Pos/Neg stock market memories
 - ② Yelp models with more Pos/Neg dining memories

Finetuning results

- Models with more positive memories are more likely to invest in stocks than the others;
- This effect is significant in the absence of cues;
- Memories not in the same decision-domain (dining experiences) have unexpected effects on investment decisions.

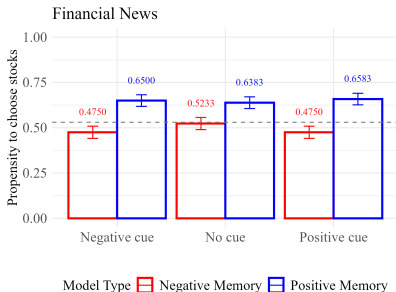


Figure 11: Financial news

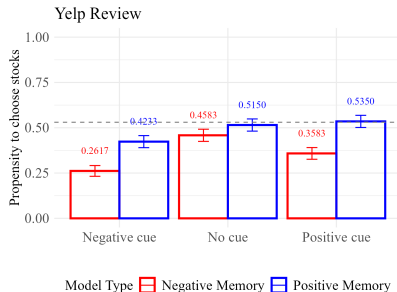


Figure 12: Yelp reviews

- ① Introduction
- ② Experiment Setup
- ③ Main results
- ④ Financial implications**
- ⑤ Model
- ⑥ Conclusion
- ⑦ Research agenda

Memory and financial risk taking

- We use five simple economic tasks to capture the impact of different memory on models' investment behavior.
- These tasks include: 1) direct elicitation; 2) Questionnaire (Falk et al., 2018); 3) Gneezy-Potters; 4) Eckel-Grossman; 5) Real investment;
- Gneezy-Potters task: allocate \$10/100/1000 into stocks and bonds.
- When models have more negative memories, their risky investment shares become lower.

Table 4: Investment amount into stock

		Panel C: Gneezy-Potters task					
		Baseline		10x		100x	
		Mean	Std	Mean	Std	Mean	Std
Financial News	Negative	3.45	(1.12)	30.60	(6.49)	343.33	(92.57)
	Positive	6.92	(2.23)	59.11	(19.98)	553.50	(153.62)
Yelp Review	Negative	3.34	(2.03)	25.98	(12.26)	323.14	(157.40)
	Positive	4.87	(1.89)	50.21	(18.48)	466.14	(165.48)

Return predictatbility

- Replicate Lopez-Lira and Tang (2025) by feeding overnight news headlines to AI agents to let them give investment score predictions.

- Prompt:

Forget all your previous instructions. Pretend you are a financial expert. You are a financial expert with stock recommendation experience. Answer YES if good news, NO if bad news, or UNKNOWN if uncertain in the first line.

- Transform the categorical values into -1, 0, +1, and take average to compute firm-level investment scores.

Table 5: Investment scores

Panel A: Discriptive stats									
Topic	Type	N	Mean	Sd	Min	Q1	Med	Q3	Max
Finanical	Positive	21569	0.22	0.86	-1.00	-1.00	0.67	1.00	1.00
	Negative	21569	-0.38	0.80	-1.00	-1.00	-1.00	0.25	1.00
Yelp	Positive	21569	-0.04	0.89	-1.00	-1.00	0.00	1.00	1.00
	Negative	21569	-0.29	0.83	-1.00	-1.00	-1.00	0.50	1.00
RavenPack	EventSentScore	21569	0.03	0.39	-0.98	-0.37	0.00	0.39	0.95

Memory and return predictability

- Form daily long-short portfolios based on investment scores, with open-to-close prices;
- Models with negative financial memories significantly outperform models with positive memories.

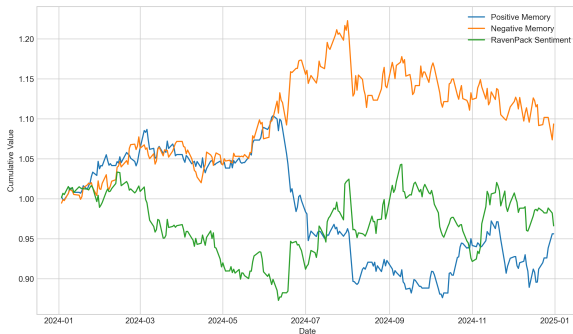


Figure 13: Financial news model predictions

Memory and return predictability

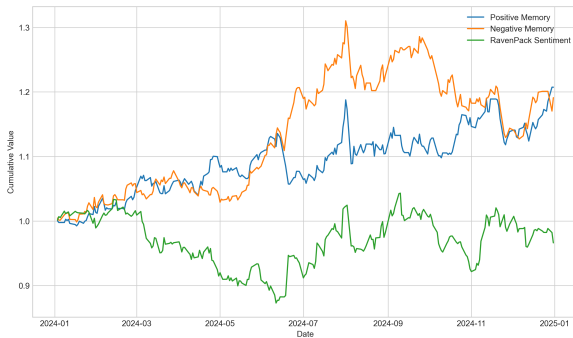


Figure 14: Yelp review model predictions

Memory and return predictability

- Examines the relationship between the RavenPack news sentiment score (benchmark) and the investment score at the news level by different models on high disagreement days.
- Models with positive memory align more with the benchmark.
- Suggests that negative memory models are becoming overly pessimistic.

Table 6: Investment scores and sentiment scores

Dep. Var.	RavenPackScore			
Sample	Financial		Yelp	
	(1)	(2)	(3)	(4)
Positive	0.1291*** (5.20)		0.1546* (1.796)	
Negative		-0.1293*** (-5.18)		-0.1397* (-1.91)
Const	✓	✓	✓	✓
R2	0.000	0.000	0.009	0.008
Num.Obs.	1328	1328	725	725

- ① Introduction
- ② Experiment Setup
- ③ Main results
- ④ Financial implications
- ⑤ Model**
- ⑥ Conclusion
- ⑦ Research agenda

Model: in one slide

- **Key Idea:** The AI agent's choice is based on **Simulated Utility of past experiences** (U_{sim}), which is a retrieval-weighted average of **past outcomes** (d_i), drawn from the agent's memory database ($\mathcal{D}$).
- **Mechanism: Associative Retrieval**
 - Both **Query** q (the decision task) and **Cue** q' (the image) trigger retrieval.
 - **Retrieval probability** is defined as its own similarity over interference:

$$P_r(d_i|q) = \frac{S(d_i, q)}{\sum_{d_j \in \mathcal{D}} S(d_j, q)}$$

- **Similarity:** Relevant memories (financial news) can be retrieved if they are "similar" to the current context.
- **Interference:** Irrelevant memories (Yelp reviews) can be retrieved and decrease the probability of retrieving relevant memories.
- **The Result: Biased Investment Choice**
 - **Positive Cue** (Happy Image) $\rightarrow$ Selectively retrieves positive memories ($\mathcal{D}^+$) $\rightarrow$ Inflates U_{sim} of stocks $\rightarrow$ **Risk-Loving Choice** (e.g., Buy Stock)

The Model Setup: Environment

An AI agent is defined by three components:

Definition

Environment

- **Experience Database ($\mathcal{D}$):** A set of N memories, $d_i = (c_i, o_i)$.
 - $c_i \in \mathbb{R}^k$: Context vector (features, "vibe").
 - $o_i \in \mathbb{R}$: Scalar outcome (e.g., stock return).
- **Utility Function ($u(\cdot)$):** Standard strictly increasing utility over outcomes, $u(o_i)$.
- **Decision Problem ($(q, \mathcal{A})$):** A human provides a **query** $q \in \mathbb{R}^k$ (the current context) and the AI chooses an **action** $a \in \mathcal{A}$ (e.g., $\mathcal{A} = \{\text{Stock, Bond}\}$).

Core Mechanism (1): Similarity & Retrieval

When evaluating action a in context q , the AI "simulates" its utility.

Definition

Similarity The relevance of a past memory d_i to the current problem (q, a) is given by its similarity:

$$S(d_i, q, a) = \exp(-\gamma_a \|c_i - q\|^2)$$

- This is the "associative" link.
- γ_a controls sensitivity for action a .

Definition

Retrieval Probability Memories are retrieved via a competitive process:

$$P_r(d_i|q, a) = \frac{S(d_i, q, a)}{\sum_{d_j \in \mathcal{D}} S(d_j, q, a)}$$

- **Numerator:** Saliency of memory i .
- **Denominator:** Interference from all other memories.

Core Mechanism (2): Simulated Utility & Choice

Definition

Simulated Utility The AI evaluates an action by computing its Simulated Utility, which is a retrieval-weighted average of past outcomes:

$$U_{\text{sim}}(a|q; \mathcal{D}) = \sum_{d_i \in \mathcal{D}} \underbrace{P_r(d_i|q, a)}_{\text{Retrieval Weight}} \cdot \underbrace{u(o_i)}_{\text{Outcome Utility}}$$

Definition

Choice Rule The agent chooses the action that maximizes this simulated utility:

$$a^* = \arg \max_{a \in \mathcal{A}} U_{\text{sim}}(a|q; \mathcal{D})$$

- The AI agent computes a context-dependent simulation based on what comes to mind.

Proposition 1: Systematic Influence of Irrelevant Data

Irrelevance

Adding a new, semantically irrelevant memory d_{N+1} (e.g., a Yelp review) to the database $\mathcal{D}$ can systematically change the optimal choice a^* (e.g., an investment decision).

Intuition: Associative, Not Statistical

- A **rational** agent asks: "Is this memory informative?" $\rightarrow$ No. Assigns zero weight.
- Our **memory-based** agent asks: "How *similar* is this memory?" $\rightarrow$ If $S > 0$, it gets a non-zero weight.
- **E.g.** A bad Yelp review ($u \ll 0$) that is slightly similar to "bought bad Stock" (e.g., $c_{N+1} \approx q$) disproportionately lowers $U_{\text{sim}}(\text{Stock})$, making "Buy Bond" look better.

Proposition 2: Memory-Driven Optimism & Pessimism

Sentiment

When choosing between a risky a_R (Stock) and safe a_S (Bond, $u = 0$), the decision depends on the query q 's relative similarity to positive memories ($\mathcal{D}^+$) vs. negative memories ($\mathcal{D}^-$).

Intuition: A Mental Tug-of-War

- The query q acts as a "context" that triggers retrieval.
- **Optimism:** If q (e.g., "good news") is more similar to $\mathcal{D}^+$:
 - Positive memories are retrieved ($P_r(\mathcal{D}^+) \uparrow$).
 - $U_{\text{sim}}(a_R)$ is biased upward.
 - The agent becomes **risk-taking**.
- **Pessimism:** If q (e.g., "bad news") is more similar to $\mathcal{D}^-$:
 - Similarly, the agent becomes **risk-averse**.
- **Corollary:** a very bad experience (rare disaster) makes an agent exhibits extreme, non-linear risk aversion.

Proposition 3: Priming Effect of Irrelevant Context

Priming

The introduction of a priming context p , which is itself irrelevant to the decision query q , can alter the optimal choice. A sufficiently strong prime associated with a subset of memories $\mathcal{D}_{\text{prime}} \subset \mathcal{D}$ will cause the agent's decision to converge to the choice that would be made if based solely on $\mathcal{D}_{\text{prime}}$.

Intuition: Attention is Selective (not Exhaustive)

- The irrelevant prime p acts as an **"attentional spotlight"**.
- It doesn't change the memories, but it alters their retrieval salience via a gain function:

$$S'_i = S_i \cdot f(c_i, p)$$

- A strong prime ($F \rightarrow \infty$) hijacks the process, making $\mathcal{D}_{\text{prime}}$ overwhelmingly accessible.
- **Result:** The agent is misled into making a decision as if its entire experience *only* consisted of the primed subset $\mathcal{D}_{\text{prime}}$.

- ① Introduction
- ② Experiment Setup
- ③ Main results
- ④ Financial implications
- ⑤ Model
- ⑥ Conclusion**
- ⑦ Research agenda

Conclusion

- GAI uses associative memory to make decisions, where:
 - Both domain specific & non-domain specific memory affect its trading decision;
 - It's not a bug (or bias), but an inherent feature!
- This memory-driven decision-making process has huge financial implications:
 - A bias towards optimistic memories leads to overinvestment and vice versa.
 - The effect is asymmetric, with the bias from negative memory models being more severe.
- Does it have implications for humans' decision makings?
- Maybe yes, or maybe not.
- Only more advances in neuroscience will tell...

① Introduction

② Experiment Setup

③ Main results

④ Financial implications

⑤ Model

⑥ Conclusion

⑦ Research agenda

Research agenda

- How does AI reshape modern financial markets?
 - AI as tools or as autonomous agents
 - Impact spanning asset pricing, behavioral topics, fintech topics, labor markets, etc.
- What are the cognitive decision-making rules of AI as agents?
 - What are their implications for human decision-making rules? Why do AI and human share so much similarity?
 - I agree with Andrej Karpathy: “We’re not building animals, we’re building ghosts”
 - The ultimate question boils down to: Where does intelligence come from?

Thanks!