

Perceptions of Novel Agents

By: Arthur Levin

2024

What we were trying to do:

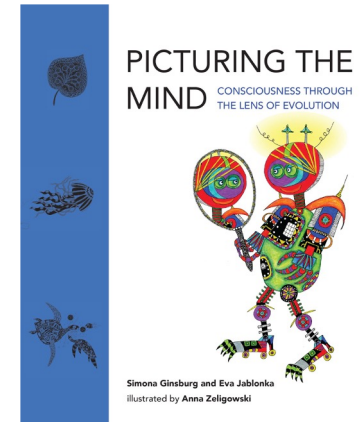
- Probe the general population's feelings on novel agents (AI, biological creatures developed by people, etc.)
- Determine the level to which people are aware of and prepared for these agents
- Look for correlations between opinions on specific topics

Project Activities

- Read a sourcebook Picturing the Mind
- In-person lessons with Dr. Michael Levin at Tufts

[<https://www.drmichaellevin.org>]

- Made a survey
 - Planned topics
 - Polished text
 - Created survey
 - Tested it with a few test subjects
 - Improved survey
 - Deployed it via Mturk
 - Fixed up a few errors that were exposed at wider access
 - Analyzed data
 - Created presentation

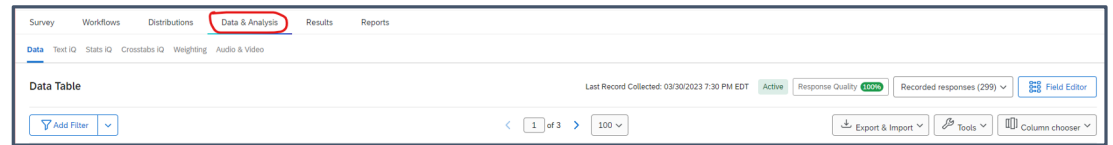


This is 2nd dataset from Mturk users

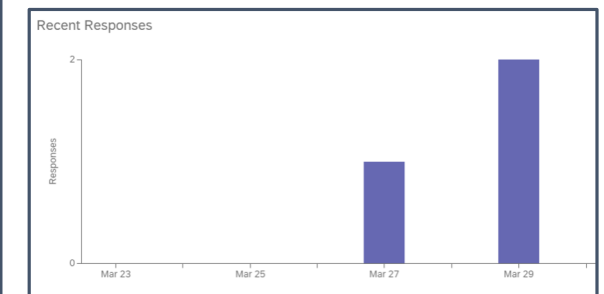
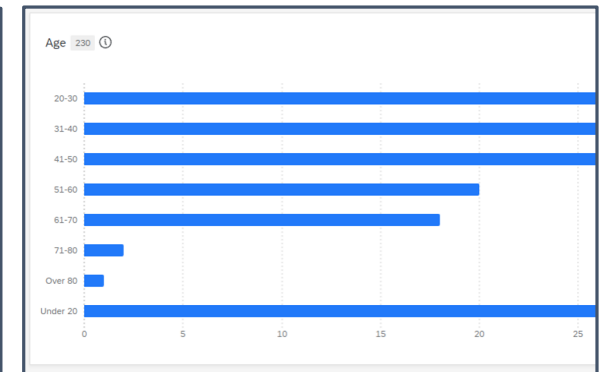
1st dataset was from @drmichaellevin's Twitter followers

How we built the survey:

- Survey tool: QUALTRICS



This screenshot shows the survey editor interface for a survey titled 'Perceptions About Novel Agents'. The survey is currently in 'Published' status. The editor shows the survey structure, including an 'Initial Consent' section and a 'Basic Questions' section. The 'Initial Consent' section contains a text block with a consent statement and a checkbox labeled 'I agree'. The 'Basic Questions' section contains a question 'Q1 | Your Age' with a dropdown menu showing 'Under 20'. The left sidebar shows the survey settings, including 'Answer type' (Multiple choice), 'Choices' (Number of choices: 1), 'Format' (List), and 'Response requirements' (Force response).



How We Distributed The Survey

Amazon Mechanical Turk

Access a global, on-demand, 24x7 workforce



We had 365 responders

Data quality 98% based on
Qualtrics criteria



Example Questions:

- Demographics:
 - Age, Education, etc.
- General Opinions:
 - Preconceived notions and thoughts on relatively well-known issues
- Answers to specific situations:
 - What solutions people prefer for several unique, nuanced problems



What We Found...

Relationship #1

Do you think humans should be attempting to develop novel life forms, specifically Artificial flesh and blood creatures?

Age:	No	Maybe	Yes	Total
20-30	9.5%	60.1%	30.4%	100.0%
31-40	7.1%	47.4%	45.5%	100.0%
41-50	33.3%	42.9%	23.8%	100.0%
51-60	18.2%	72.7%	9.1%	100.0%
61-70	33.3%	0.0%	66.7%	100.0%
Total				

Chi-Squared Test

P-Value	0.000294
Effect Size (Cramér's V)	0.202
Sample Size	359

This relationship is significant

Summary: Most people aged 20-60 are undecided, whereas people between 61-70 largely say yes with no one being undecided.

Relationship #2

This relationship is not significant

Chi-Squared Test	
P-Value	0.633
Effect Size (Cramér's V)	0.0929
Sample Size	355

Think of your "Self" - the collection of your thoughts, emotions, and perceptions. What kind of a thing, if it's a thing at all, is your Self? Do you believe the Self to be:

Age:	A permanent object or structure	A dynamical feature continuously rebuilt from mom...	An illusion, nothing but a bundle of perceptions	Total
20-30	41.2%	52.7%	6.1%	100.0%
31-40	29.9%	63.0%	7.1%	100.0%
41-50	42.9%	52.4%	4.8%	100.0%
51-60	36.4%	63.6%	0.0%	100.0%
61-70	33.3%	66.7%	0.0%	100.0%
Total				

Summary: People, almost regardless of age either believed the self to be a permanent object or a dynamical feature with slight preference toward the second option

Relationship #3

Chi-Squared Test

P-Value	0.0276
Effect Size (Cramér's V)	0.141
Sample Size	359

This relationship is significant

Do you have a computer science educational background or active interest?		Do you think humans should be attempting to develop novel life forms, specifically Artificial flesh and blood creatures?			Total
		No	Maybe	Yes	
No		15.9%	59.8%	24.4%	100.0%
Yes		8.7%	51.8%	39.5%	100.0%

Summary: People in the computer science field are slightly more likely to be in favor of developing novel life forms

Relationship #4

Chi-Squared Test

P-Value	0.000583
Effect Size (Cramér's V)	0.204
Sample Size	358

This relationship is significant

Do you think humans should be attempting to develop novel life forms, specifically Artificial flesh and blood creatures?		Do you have a philosophy background or active interest?	
		No	Yes
No	15.2%	8.3%	
Maybe	62.9%	49.0%	
Yes	21.9%	42.7%	
Total	100.0%	100.0%	

Summary: if you are in the field of philosophy, you're twice as likely to think that you should make creatures

Relationship #5

Chi-Squared Test

P-Value	0.000191
Effect Size (Cramér's V)	0.218
Sample Size	359

This relationship is significant

Have you ever deliberately interacted with an A.I such as CleverBot, GPT3-Chat, or Akinator?

Do you or have you ever thought that A.I. would someday take over humanity?

	yes	Maybe	No
Yes	50.4%	44.0%	5.5%
No	25.0%	43.8%	31.3%

Summary: People who have interacted with an AI are more likely to believe that AI will one day take over humanity

How do people's views on the self and the mind effect their answer to this question:

You are visited by a Genie who offers you a chance at a really pleasant experience that you would normally enjoy. You have two options:

A. You have the experience right now but absolutely no memory of it afterward

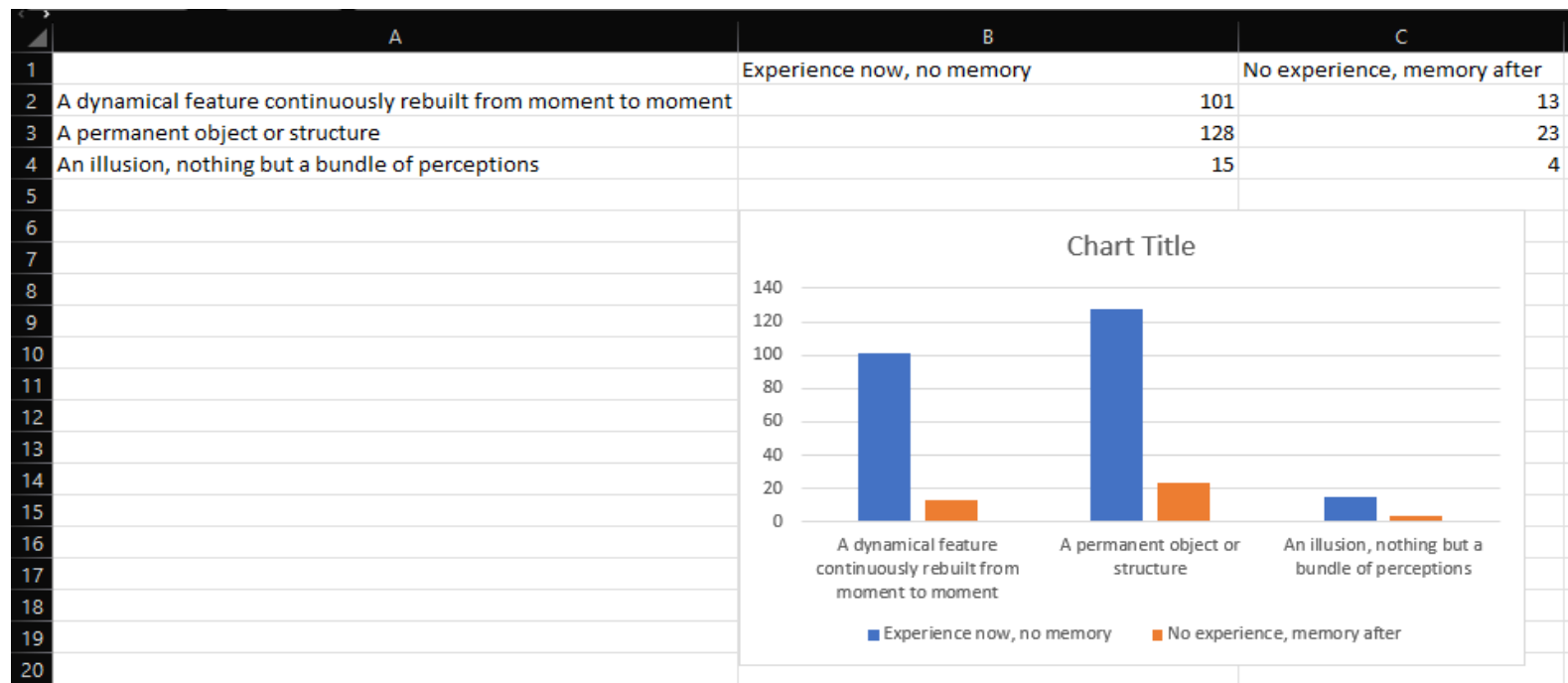
or

B. You will have a perfect memory of doing it but it's a false memory and you don't actually get to do it

Relationship #6

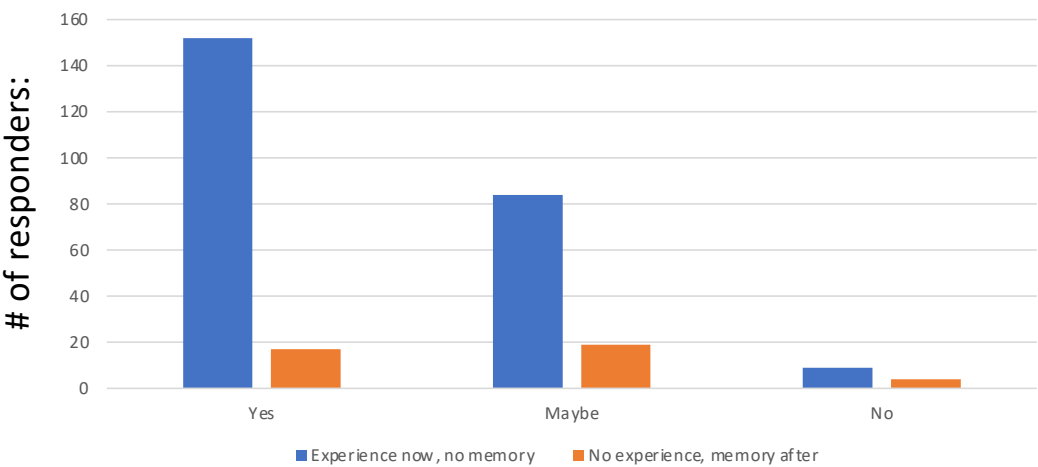
Think of your "Self" - the collection of your thoughts, emotions, and perceptions. What kind of a thing, if it's a thing at all, is your Self? Do you believe the Self to be:

You are visited by a Genie who offers you a chance at a really pleasant experience that you would normally enjoy. You have two options:
A. You have the experience right now but absolutely no memory of it afterward
or
B. You will have a perfect memory of doing it but it's a false memory and you don't actually get to do it



Summary: Regardless of people’s views on the self, they overwhelmingly want the experience without the memory

Relationship #7



Do you believe in an immaterial Soul or Mind as distinct from the brain?

You are visited by a Genie who offers you a chance at a really pleasant experience that you would normally enjoy. You have two options:

A. You have the experience right now but absolutely no memory of it afterward

or

B. You will have a perfect memory of doing it but it's a false memory and you don't actually get to do it

	Experience now, no memory	No experience, memory after			expected for experience now	expected for memory after
Yes	152	17			81.66666667	13.33333333
Maybe	84	19			81.66666667	13.33333333
No	9	4			81.66666667	13.33333333
	6.1929E-28	0.006908519				
	are "no experience" people of same opinion breakdwn as "no memory" people?				3.6436E-283	
	Are the opinions evenly split between the 3 options?				No!	

Summary: regardless of their view on the immaterial mind, most people prefer the experience with no memory, but proportionally more of the uncertain ones are the minority who prefer the experience without the memory.

Relationship #8

Chi-Squared Test

P-Value	< 0.00001
Effect Size (Cramér's V)	0.339
Sample Size	357

This relationship is significant

Do you believe in a higher power (a supernatural being/god)?

Do you think humans should be attempting to develop novel life forms, specifically Artificial flesh and blood creatures?

	No	Maybe	Yes	Total
No	24.3%	56.8%	18.9%	100.0%
Maybe	8.3%	79.5%	12.1%	100.0%
Yes	8.5%	35.1%	56.4%	100.0%
Total				

Summary: People who believe in a supernatural being generally also agree that we should be developing novel lifeforms

Relationship #9

Chi-Squared Test

P-Value	< 0.00001
Effect Size (Cramér's V)	0.296
Sample Size	357

This relationship is significant

Do you believe in a higher power (a supernatural being/god)?

Do you or have you ever thought that A.I. would someday take over humanity?

	yes	Maybe	No	Total
No	37.8%	40.5%	21.6%	100.0%
Maybe	28.0%	65.2%	6.8%	100.0%
Yes	67.0%	29.3%	3.7%	100.0%
Total				

Summary: People who believe in a supernatural being also likely believe that AI will take over humanity



The Twist!

Two of the paragraphs...

were written by ChatGPT:

We know we can prompt
ChatGPT, but how good is
ChatGPT at prompting us?



Human Vs. AI

For one section of the survey, we automatically (randomly) split the participants into groups to answer several questions based on an introductory paragraph meant to push them toward a specific answer. However, half the prompts were written by AI and half by a human. We did this to see which was better at prompting the participants toward different answers.

These are the results:

The Human VS. AI Prompts


	A	B	C	D	E	F
1	push to forget		exp. Now	memory		% push
2		no push	64	8		89%
3		AI push	61	10		86%
4		human push	64	8		89%
5						
6		is AI different than human:		0.42348457		
7		is push different than no push:		1		
8						
9	push to remember:		exp. Now	memory		
10		no push	64	8		89%
11		AI push	59	13		82%
12		human push	62	10		86%
13						
14		is AI different than human:		0.30662412		
15		is push different than no push:		0.49552088		

Summary: Neither the human's or the AI's prompts could change people's minds on the question

Human Vs. AI Conclusions:

Regardless of its author, the prompt was unable to push people's opinion in either direction.

No matter what argument they were presented with 82-89% of participants chose to have the experience without the memory



Limitations of the study:

- Limited sample size
- Payment (\$1) introduces possible unknown incentives



Conclusions:

- Why are these findings important?
 - What can be learned from them?



What can you do



- Take the survey
- Study the fascinating topic of diverse intelligence
- Deploy surveys of your own chosen topics



Acknowledgements:

- Dustin Eirdosh
- Dr. Michael Levin
- Mturk Funding: Diverse Intelligence Summer Institute



Questions?