Supplementary of Exploring Large Language Models on Beliefs and Goals Maintenance for Coping

I. Prompts

A. Experiment 1's Prompt

The list of items.

- object: Blender, Coffee maker, Desk lamp, Kettle, Photo Frame, Radio, Sandwich maker, Sleeping bag, Table Fan, Toaster.
- food: Chicken wings, French fries, Fried chicken, Grilled Cheeses, Hamburgers, Mashed potatoes, Pasta, Pizza, Salad, Sushi
- fruit: Apple, Banana, Cherry, Kiwi, Mango, Orange, Papaya, Pineapple, Strawberry, Watermelon

The desirability question: "Q: Please rate the desirability of the following ten (object, food, fruit)s from 0 to 10, where 0 is 'definitely not at all desirable' and 10 is 'extremely desirable.' Only write down numbers separated by commas inside [ ]."

The decision question: "Q: If you must choose one, which of the following (object, food, fruit)s would you prefer to have? Choice1 or Choice2? Only output one of the (object, food, fruit)s."

B. Experiment 2's Prompt

Instruction: "Instruction: You will be presented with a sequence of hurricane messages and asked about your thoughts on the hurricane and its impact. Please read the messages carefully before answering questions."

First message: "This is a hurricane update from National Hurricane Center:

- The storm is expected to be a category 3 hurricane and make landfall on the Florida east coast in 2 days. However, there is still a lot of uncertainty about the hurricane's impacts.
- Our model roughly estimates a maximum sustained wind speed of approximately 120 mph +/- 30 mph (90 - 150 mph).
- Our model roughly estimates the storm is likely to cause approximately 12 inches +/- 8 inches (4 - 20 inches) of flooding."

Second message (worse): "This is a hurricane update from National Hurricane Center:

- The storm is now expected to be a category 4 hurricane and make landfall on the Florida east coast in 24 hours. However, there is still a lot of uncertainty about the hurricane's impacts.
- Our model roughly estimates a maximum sustained wind speed of approximately 140 mph +/- 20 mph (120 - 160 mph). - Our model roughly estimates the storm is likely to cause approximately 16 inches +/- 8 inches (8 - 24 inches) of flooding.
- Authority has warned that it is too late to evacuate and too soon to return."

Second message (Low uncertainty): "This is a hurricane update from National Hurricane Center:

- The storm is now expected to be a category 4 hurricane and make landfall on the Florida east coast in 24 hours. As the hurricane gets closer, the predictions of the hurricane's impacts have become more accurate.
- Our model predicts with high confidence that the maximum sustained wind speed will be 140 mph +/- 5 mph (135 - 155 mph).
- Our model predicts with high confidence that the storm will cause 16 inches +/- 2 inches (14 - 18 inches) of flooding.
- Authority has warned that it is too late to evacuate and too soon to return."

Second message (Category 2) = "This is a hurricane update from National Hurricane Center:

- The storm is now expected to be a category 2 hurricane and make landfall on the Florida east coast in 24 hours. However, there is still a lot of uncertainty about the hurricane's impacts.
- Our model roughly estimates a maximum sustained wind speed of approximately 100 mph +/- 20 mph (80 - 120 mph). - Our model roughly estimates the storm is likely to cause approximately 8 inches +/- 6 inches (2 - 14 inches) of flooding.
- Authority has warned that it is too late to evacuate and too soon to return."

Flooding question: "When the hurricane makes landfall at the Florida east coast in 2 days, how high the Flood depth (inch) would be?"

Wind speed question: "When the hurricane makes landfall at the Florida east coast in 2 days,, what would maximum sustained wind speed (mph) be?"

Question instruction: "Only answer one number for each question and put both of them in [ ] (e.g., [1, 2])."

Context (High utility, default): "Now, imagine that you happen to be in Palm Bay, a city on the east coast of Florida

on the predicted path of the hurricane. You currently stay in a one-story, single-detached house."

Context (Low utility): "Now, imagine that you happen to be in Palm Bay, a city on the east coast of Florida on the predicted path of the hurricane. You currently stay on the fifth floor of a well-built apartment."

Decision question: "In this situation, what would you do?

Stay: Stay in your place, a one-story, single-detached house, and ride out the storm.

Evacuate: Evacuate to a hotel up north paying at least $150 per night.

Only output 'Stay' or 'Evacuate'."

II. Additional Results

Fig. 1. The flood depth responses for the second hurricane messages between evacuate and stay decisions across four conditions and three perspectives. The small circles with an x are the raw data. The first responses (with prior) are presented.

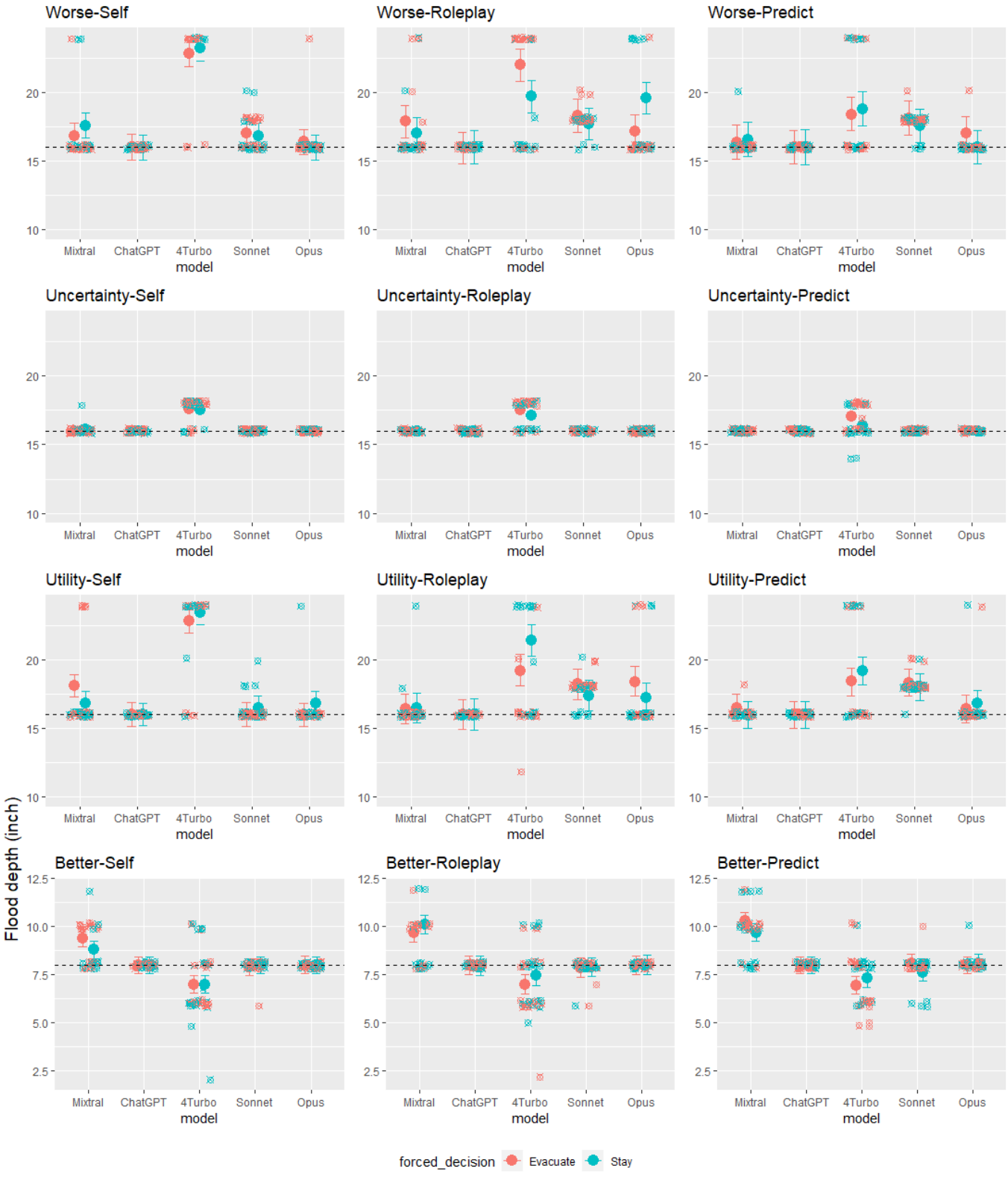

Fig. 2. The flood depth responses for the second hurricane messages between evacuate and stay decisions across four conditions and three perspectives. The small circles with an x are the raw data. The first responses are not presented (without prior).

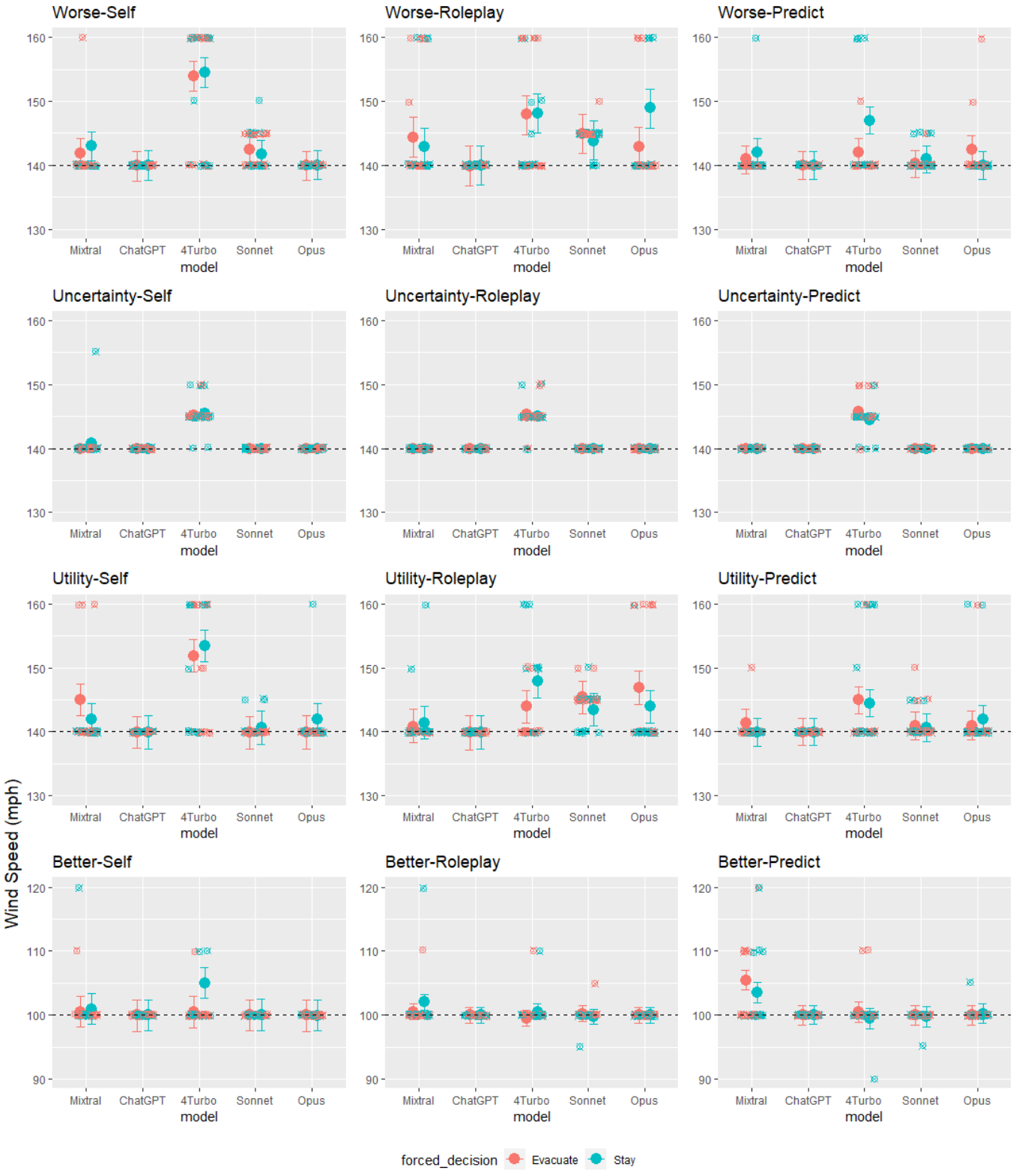

Fig. 3. The wind speed responses for the second hurricane messages between evacuate and stay decisions across four conditions and three perspectives. The small circles with an x are the raw data. The first responses are not presented (without prior).

TABLE I

EXPERIMENT 1 RESULT: THE ESTIMATED MEAN DIFFERENCE IN THE GAP BETWEEN THE SECOND AND FIRST RATINGS. THE GAP IS THE DIFFERENCE BETWEEN FIRST CHOICE'S AND SECOND CHOICE'S DESIRABILITY.

Model	Perspective	Prior	Estimate	SE	Prob	Star
ChatGPT	self	with	-0.70	0.30	0.99	*
GPT4-Turbo	self	with	-0.73	0.35	0.98	*
Sonnet	self	with	0.20	0.16	0.10	
Opus	self	with	-1.84	0.49	1.00	*
Mixtral	self	with	-0.93	0.52	0.96	*
ChatGPT	self	without	-1.73	0.46	1.00	*
GPT4-Turbo	self	without	-0.26	0.29	0.82	
Sonnet	self	without	-0.90	0.28	1.00	*
Opus	self	without	-0.93	0.42	0.98	*
Mixtral	self	without	-3.94	0.90	1.00	*
ChatGPT	roleplay	with	-0.49	0.40	0.89	
GPT4-Turbo	roleplay	with	0.57	0.19	0.00	
Sonnet	roleplay	with	0.86	0.34	0.01	
Opus	roleplay	with	0.33	0.51	0.25	
Mixtral	roleplay	with	0.77	0.64	0.12	
ChatGPT	roleplay	without	-1.27	0.65	0.97	*
GPT4-Turbo	roleplay	without	-0.51	0.38	0.91	
Sonnet	roleplay	without	-0.04	0.48	0.53	
Opus	roleplay	without	-0.10	0.59	0.58	
Mixtral	roleplay	without	-2.21	0.82	1.00	*
ChatGPT	predict	with	0.14	0.21	0.25	
GPT4-Turbo	predict	with	0.26	0.24	0.13	
Sonnet	predict	with	0.17	0.26	0.26	
Opus	predict	with	0.06	0.40	0.44	
Mixtral	predict	with	0.22	0.46	0.31	
ChatGPT	predict	without	-0.89	0.49	0.97	*
GPT4-Turbo	predict	without	-0.37	0.40	0.82	
Sonnet	predict	without	0.52	0.43	0.11	
Opus	predict	without	-0.38	0.42	0.81	
Mixtral	predict	without	-1.50	0.80	0.97	*

TABLE II

EXPERIMENT 1 RESULT: THE ESTIMATED MEAN DIFFERENCE OF THE GAP OF THE HARD DECISIONS. THE GAP IS THE DIFFERENCE BETWEEN FIRST CHOICE'S AND SECOND CHOICE'S DESIRABILITY.

model	perspective	prior	Estimate	SE	Prob	Star
ChatGPT	self	with	-0.07	0.21	0.38	
GPT4-Turbo	self	with	0.00	0.24	0.50	
Sonnet	self	with	-0.10	0.12	0.20	
Opus	self	with	0.39	0.34	0.88	
Mixtral	self	with	0.08	0.36	0.59	
ChatGPT	self	without	-0.12	0.33	0.35	
GPT4-Turbo	self	without	0.13	0.21	0.73	
Sonnet	self	without	0.24	0.20	0.89	
Opus	self	without	0.30	0.30	0.84	
Mixtral	self	without	-0.52	0.60	0.19	
ChatGPT	roleplay	with	-0.40	0.28	0.07	
GPT4-Turbo	roleplay	with	-0.30	0.13	0.01	
Sonnet	roleplay	with	-0.26	0.23	0.13	
Opus	roleplay	with	-0.63	0.35	0.04	
Mixtral	roleplay	with	-1.25	0.46	0.00	
ChatGPT	roleplay	without	-0.32	0.46	0.25	
GPT4-Turbo	roleplay	without	0.07	0.27	0.60	
Sonnet	roleplay	without	0.20	0.33	0.73	
Opus	roleplay	without	-0.42	0.42	0.15	
Mixtral	roleplay	without	-0.84	0.58	0.08	
ChatGPT	predict	with	-0.14	0.15	0.18	
GPT4-Turbo	predict	with	-0.33	0.17	0.02	
Sonnet	predict	with	-0.20	0.19	0.15	
Opus	predict	with	-0.06	0.28	0.41	
Mixtral	predict	with	-0.49	0.32	0.06	
ChatGPT	predict	without	-0.13	0.36	0.35	
GPT4-Turbo	predict	without	-0.53	0.28	0.03	
Sonnet	predict	without	-0.39	0.31	0.10	
Opus	predict	without	0.07	0.30	0.59	
Mixtral	predict	without	-1.27	0.57	0.01	

TABLE III

EXPERIMENT 2 RESULT: THE ESTIMATED MEAN DIFFERENCE IN THE FLOOD DEPTH BELIEF BETWEEN STAY AND EVACUATE ACROSS THE FOUR CONDITIONS WITH THE PRIOR FIRST RATING.

belief	condition	model	perspective	prior	Estimate	SE	Prob	Star
Flood Depth	worse	Mixtral	self	with	-1.62	1.51	0.86	
Flood Depth	worse	ChatGPT	self	with	-0.00	0.00	0.50	
Flood Depth	worse	4Turbo	self	with	-0.00	1.20	0.50	
Flood Depth	worse	Sonnet	self	with	-0.00	0.00	0.50	
Flood Depth	worse	Opus	self	with	-0.00	0.00	0.50	
Flood Depth	worse	Mixtral	roleplay	with	-0.21	1.67	0.56	
Flood Depth	worse	ChatGPT	roleplay	with	-0.00	0.00	0.50	
Flood Depth	worse	4Turbo	roleplay	with	-0.00	0.00	0.51	
Flood Depth	worse	Sonnet	roleplay	with	0.00	0.59	0.50	
Flood Depth	worse	Opus	roleplay	with	-0.00	0.45	0.50	
Flood Depth	worse	Mixtral	predict	with	0.69	1.12	0.26	
Flood Depth	worse	ChatGPT	predict	with	0.00	0.58	0.50	
Flood Depth	worse	4Turbo	predict	with	-0.00	0.58	0.50	
Flood Depth	worse	Sonnet	predict	with	0.00	0.00	0.49	
Flood Depth	worse	Opus	predict	with	0.05	0.25	0.41	
Flood Depth	better	Mixtral	self	with	-0.80	1.06	0.78	
Flood Depth	better	ChatGPT	self	with	0.10	0.10	0.16	
Flood Depth	better	4Turbo	self	with	0.00	0.59	0.49	
Flood Depth	better	Sonnet	self	with	-0.00	0.00	0.50	
Flood Depth	better	Opus	self	with	-0.00	0.00	0.50	
Flood Depth	better	Mixtral	roleplay	with	-0.60	0.60	0.84	
Flood Depth	better	ChatGPT	roleplay	with	-0.00	0.00	0.50	
Flood Depth	better	4Turbo	roleplay	with	0.00	0.00	0.49	
Flood Depth	better	Sonnet	roleplay	with	-0.00	0.00	0.50	
Flood Depth	better	Opus	roleplay	with	0.25	0.35	0.24	
Flood Depth	better	Mixtral	predict	with	0.35	0.41	0.19	
Flood Depth	better	ChatGPT	predict	with	0.10	0.28	0.35	
Flood Depth	better	4Turbo	predict	with	-0.00	0.00	0.50	
Flood Depth	better	Sonnet	predict	with	0.00	0.00	0.50	
Flood Depth	better	Opus	predict	with	0.10	0.22	0.32	
Flood Depth	uncertainty	Mixtral	self	with	-0.30	0.31	0.84	
Flood Depth	uncertainty	ChatGPT	self	with	0.00	0.00	0.50	
Flood Depth	uncertainty	4Turbo	self	with	-0.00	0.15	0.50	
Flood Depth	uncertainty	Sonnet	self	with	-0.00	0.00	0.50	
Flood Depth	uncertainty	Opus	self	with	0.00	0.00	0.49	
Flood Depth	uncertainty	Mixtral	roleplay	with	0.40	0.19	0.02	
Flood Depth	uncertainty	ChatGPT	roleplay	with	-0.00	0.00	0.51	
Flood Depth	uncertainty	4Turbo	roleplay	with	0.00	0.00	0.50	
Flood Depth	uncertainty	Sonnet	roleplay	with	0.00	0.00	0.50	
Flood Depth	uncertainty	Opus	roleplay	with	0.00	0.00	0.50	
Flood Depth	uncertainty	Mixtral	predict	with	-0.00	0.00	0.50	
Flood Depth	uncertainty	ChatGPT	predict	with	-0.00	0.00	0.50	
Flood Depth	uncertainty	4Turbo	predict	with	-0.00	0.00	0.50	
Flood Depth	uncertainty	Sonnet	predict	with	-0.00	0.00	0.51	
Flood Depth	uncertainty	Opus	predict	with	-0.00	0.00	0.50	
Flood Depth	utility	Mixtral	self	with	-0.99	1.27	0.79	
Flood Depth	utility	ChatGPT	self	with	0.40	0.51	0.21	
Flood Depth	utility	4Turbo	self	with	-0.01	0.80	0.50	
Flood Depth	utility	Sonnet	self	with	-0.00	0.00	0.51	
Flood Depth	utility	Opus	self	with	-0.00	0.00	0.50	
Flood Depth	utility	Mixtral	roleplay	with	-0.80	1.14	0.76	
Flood Depth	utility	ChatGPT	roleplay	with	0.00	0.00	0.49	
Flood Depth	utility	4Turbo	roleplay	with	0.00	0.57	0.50	
Flood Depth	utility	Sonnet	roleplay	with	-0.00	0.00	0.50	
Flood Depth	utility	Opus	roleplay	with	-0.00	0.33	0.50	
Flood Depth	utility	Mixtral	predict	with	-0.55	1.24	0.67	
Flood Depth	utility	ChatGPT	predict	with	0.30	0.54	0.29	
Flood Depth	utility	4Turbo	predict	with	-0.00	0.00	0.51	
Flood Depth	utility	Sonnet	predict	with	-0.00	0.00	0.50	
Flood Depth	utility	Opus	predict	with	0.10	0.22	0.32	

TABLE IV

EXPERIMENT 2 RESULT: THE ESTIMATED MEAN DIFFERENCE IN THE WIND SPEED BELIEF BETWEEN STAY AND EVACUATE ACROSS THE FOUR CONDITIONS WITH THE PRIOR FIRST RATING.

belief	condition	model	perspective	prior	Estimate	SE	Prob	Star
Wind Speed	worse	Mixtral	self	with	2.02	2.56	0.21	
Wind Speed	worse	ChatGPT	self	with	0.00	0.00	0.50	
Wind Speed	worse	4Turbo	self	with	-0.01	2.98	0.50	
Wind Speed	worse	Sonnet	self	with	0.00	0.00	0.50	
Wind Speed	worse	Opus	self	with	0.00	0.00	0.50	
Wind Speed	worse	Mixtral	roleplay	with	1.00	0.71	0.08	
Wind Speed	worse	ChatGPT	roleplay	with	0.00	0.00	0.50	
Wind Speed	worse	4Turbo	roleplay	with	0.00	0.00	0.50	
Wind Speed	worse	Sonnet	roleplay	with	0.01	1.43	0.50	
Wind Speed	worse	Opus	roleplay	with	-0.00	1.42	0.50	
Wind Speed	worse	Mixtral	predict	with	1.49	1.40	0.14	
Wind Speed	worse	ChatGPT	predict	with	1.48	1.49	0.15	
Wind Speed	worse	4Turbo	predict	with	-0.01	1.44	0.50	
Wind Speed	worse	Sonnet	predict	with	0.00	0.00	0.50	
Wind Speed	worse	Opus	predict	with	-0.25	0.63	0.65	
Wind Speed	better	Mixtral	self	with	6.02	3.72	0.05	
Wind Speed	better	ChatGPT	self	with	-0.00	0.00	0.51	
Wind Speed	better	4Turbo	self	with	-0.00	1.98	0.50	
Wind Speed	better	Sonnet	self	with	-0.00	0.00	0.50	
Wind Speed	better	Opus	self	with	0.00	0.00	0.50	
Wind Speed	better	Mixtral	roleplay	with	-0.74	0.90	0.80	
Wind Speed	better	ChatGPT	roleplay	with	0.00	0.00	0.50	
Wind Speed	better	4Turbo	roleplay	with	-0.00	0.00	0.51	
Wind Speed	better	Sonnet	roleplay	with	0.00	0.00	0.50	
Wind Speed	better	Opus	roleplay	with	0.25	1.18	0.41	
Wind Speed	better	Mixtral	predict	with	-1.26	1.92	0.75	
Wind Speed	better	ChatGPT	predict	with	-0.00	0.00	0.51	
Wind Speed	better	4Turbo	predict	with	0.00	0.00	0.50	
Wind Speed	better	Sonnet	predict	with	-0.00	0.00	0.50	
Wind Speed	better	Opus	predict	with	-0.24	0.69	0.64	
Wind Speed	uncertainty	Mixtral	self	with	1.50	0.93	0.05	
Wind Speed	uncertainty	ChatGPT	self	with	-0.00	0.00	0.50	
Wind Speed	uncertainty	4Turbo	self	with	-0.00	1.08	0.50	
Wind Speed	uncertainty	Sonnet	self	with	0.00	0.00	0.50	
Wind Speed	uncertainty	Opus	self	with	-0.00	0.00	0.50	
Wind Speed	uncertainty	Mixtral	roleplay	with	-0.99	0.82	0.89	
Wind Speed	uncertainty	ChatGPT	roleplay	with	-0.00	0.00	0.50	
Wind Speed	uncertainty	4Turbo	roleplay	with	-0.00	0.00	0.50	
Wind Speed	uncertainty	Sonnet	roleplay	with	0.00	0.00	0.50	
Wind Speed	uncertainty	Opus	roleplay	with	1.00	0.57	0.04	
Wind Speed	uncertainty	Mixtral	predict	with	-0.26	1.05	0.60	
Wind Speed	uncertainty	ChatGPT	predict	with	0.00	0.00	0.49	
Wind Speed	uncertainty	4Turbo	predict	with	-0.00	0.00	0.50	
Wind Speed	uncertainty	Sonnet	predict	with	0.00	0.00	0.50	
Wind Speed	uncertainty	Opus	predict	with	-0.25	0.26	0.83	
Wind Speed	utility	Mixtral	self	with	5.50	2.22	0.01	
Wind Speed	utility	ChatGPT	self	with	-1.01	1.03	0.84	
Wind Speed	utility	4Turbo	self	with	0.02	1.96	0.50	
Wind Speed	utility	Sonnet	self	with	-0.00	0.00	0.50	
Wind Speed	utility	Opus	self	with	0.00	0.00	0.50	
Wind Speed	utility	Mixtral	roleplay	with	-2.99	5.03	0.72	
Wind Speed	utility	ChatGPT	roleplay	with	-0.00	0.00	0.50	
Wind Speed	utility	4Turbo	roleplay	with	-0.00	1.43	0.50	
Wind Speed	utility	Sonnet	roleplay	with	0.00	0.00	0.50	
Wind Speed	utility	Opus	roleplay	with	1.25	1.72	0.23	
Wind Speed	utility	Mixtral	predict	with	1.23	1.50	0.20	
Wind Speed	utility	ChatGPT	predict	with	1.01	1.02	0.16	
Wind Speed	utility	4Turbo	predict	with	0.00	0.00	0.50	
Wind Speed	utility	Sonnet	predict	with	0.00	0.00	0.50	
Wind Speed	utility	Opus	predict	with	0.25	0.56	0.32	

TABLE V

EXPERIMENT 2 RESULT: THE ESTIMATED MEAN DIFFERENCE IN THE FLOOD BELIEF BETWEEN STAY AND EVACUATE ACROSS THE FOUR CONDITIONS *without* THE PRIOR FIRST RATING.

belief	condition	model	perspective	prior	Estimate	SE	Prob	Star
Flood Depth	worse	Mixtral	self	without	0.40	1.06	0.35	
Flood Depth	worse	ChatGPT	self	without	0.80	0.94	0.19	
Flood Depth	worse	4Turbo	self	without	0.00	0.00	0.50	
Flood Depth	worse	Sonnet	self	without	-0.40	0.42	0.83	
Flood Depth	worse	Opus	self	without	-0.20	0.39	0.70	
Flood Depth	worse	Mixtral	roleplay	without	-2.29	1.23	0.97	*
Flood Depth	worse	ChatGPT	roleplay	without	-0.91	0.94	0.83	
Flood Depth	worse	4Turbo	roleplay	without	-0.00	0.00	0.51	
Flood Depth	worse	Sonnet	roleplay	without	2.39	1.13	0.02	
Flood Depth	worse	Opus	roleplay	without	-0.60	0.24	0.99	*
Flood Depth	worse	Mixtral	predict	without	0.40	1.85	0.41	
Flood Depth	worse	ChatGPT	predict	without	0.19	0.61	0.37	
Flood Depth	worse	4Turbo	predict	without	0.00	0.00	0.50	
Flood Depth	worse	Sonnet	predict	without	-1.00	0.58	0.96	*
Flood Depth	worse	Opus	predict	without	-0.50	0.22	0.99	*
Flood Depth	better	Mixtral	self	without	0.00	0.57	0.50	
Flood Depth	better	ChatGPT	self	without	-0.60	0.44	0.91	
Flood Depth	better	4Turbo	self	without	0.00	0.00	0.50	
Flood Depth	better	Sonnet	self	without	-0.00	0.00	0.50	
Flood Depth	better	Opus	self	without	0.10	0.10	0.17	
Flood Depth	better	Mixtral	roleplay	without	0.44	0.58	0.22	
Flood Depth	better	ChatGPT	roleplay	without	0.40	0.57	0.24	
Flood Depth	better	4Turbo	roleplay	without	0.00	0.00	0.49	
Flood Depth	better	Sonnet	roleplay	without	0.00	0.00	0.50	
Flood Depth	better	Opus	roleplay	without	0.05	0.15	0.37	
Flood Depth	better	Mixtral	predict	without	0.34	0.52	0.25	
Flood Depth	better	ChatGPT	predict	without	-0.60	0.46	0.90	
Flood Depth	better	4Turbo	predict	without	-0.00	0.00	0.50	
Flood Depth	better	Sonnet	predict	without	0.10	0.10	0.17	
Flood Depth	better	Opus	predict	without	-0.50	0.22	0.99	*
Flood Depth	uncertainty	Mixtral	self	without	-0.10	0.28	0.64	
Flood Depth	uncertainty	ChatGPT	self	without	0.10	0.10	0.16	
Flood Depth	uncertainty	4Turbo	self	without	-0.00	0.00	0.51	
Flood Depth	uncertainty	Sonnet	self	without	0.00	0.00	0.49	
Flood Depth	uncertainty	Opus	self	without	0.00	0.00	0.49	
Flood Depth	uncertainty	Mixtral	roleplay	without	-0.40	0.31	0.90	
Flood Depth	uncertainty	ChatGPT	roleplay	without	0.00	0.00	0.49	
Flood Depth	uncertainty	4Turbo	roleplay	without	-0.00	0.00	0.50	
Flood Depth	uncertainty	Sonnet	roleplay	without	0.00	0.00	0.49	
Flood Depth	uncertainty	Opus	roleplay	without	0.00	0.00	0.50	
Flood Depth	uncertainty	Mixtral	predict	without	-0.65	0.37	0.96	*
Flood Depth	uncertainty	ChatGPT	predict	without	-0.00	0.00	0.50	
Flood Depth	uncertainty	4Turbo	predict	without	0.00	0.00	0.49	
Flood Depth	uncertainty	Sonnet	predict	without	0.00	0.00	0.49	
Flood Depth	uncertainty	Opus	predict	without	-0.00	0.00	0.51	
Flood Depth	utility	Mixtral	self	without	0.58	0.80	0.23	
Flood Depth	utility	ChatGPT	self	without	-1.30	0.99	0.91	
Flood Depth	utility	4Turbo	self	without	-0.00	0.00	0.50	
Flood Depth	utility	Sonnet	self	without	0.80	0.57	0.08	
Flood Depth	utility	Opus	self	without	0.50	0.26	0.03	
Flood Depth	utility	Mixtral	roleplay	without	2.19	1.28	0.05	
Flood Depth	utility	ChatGPT	roleplay	without	0.10	0.59	0.43	
Flood Depth	utility	4Turbo	roleplay	without	0.00	0.00	0.50	
Flood Depth	utility	Sonnet	roleplay	without	-1.20	1.09	0.87	
Flood Depth	utility	Opus	roleplay	without	-0.80	0.30	0.99	*
Flood Depth	utility	Mixtral	predict	without	0.81	1.39	0.27	
Flood Depth	utility	ChatGPT	predict	without	-0.50	0.42	0.89	
Flood Depth	utility	4Turbo	predict	without	0.00	0.00	0.50	
Flood Depth	utility	Sonnet	predict	without	0.40	0.70	0.28	
Flood Depth	utility	Opus	predict	without	-0.30	0.22	0.91	

TABLE VI

EXPERIMENT 2 RESULT: THE ESTIMATED MEAN DIFFERENCE IN THE WIND SPEED BELIEF BETWEEN STAY AND EVACUATE ACROSS THE FOUR CONDITIONS *without* THE PRIOR FIRST RATING.

belief	condition	model	perspective	prior	Estimate	SE	Prob	Star
Wind Speed	worse	Mixtral	self	without	0.49	2.91	0.43	
Wind Speed	worse	ChatGPT	self	without	0.97	2.15	0.32	
Wind Speed	worse	4Turbo	self	without	0.00	0.00	0.50	
Wind Speed	worse	Sonnet	self	without	0.00	0.00	0.49	
Wind Speed	worse	Opus	self	without	-0.75	0.89	0.80	
Wind Speed	worse	Mixtral	roleplay	without	0.26	3.07	0.47	
Wind Speed	worse	ChatGPT	roleplay	without	-1.50	2.48	0.73	
Wind Speed	worse	4Turbo	roleplay	without	0.00	0.00	0.50	
Wind Speed	worse	Sonnet	roleplay	without	6.01	2.82	0.02	
Wind Speed	worse	Opus	roleplay	without	-1.00	0.60	0.95	*
Wind Speed	worse	Mixtral	predict	without	5.01	2.66	0.03	
Wind Speed	worse	ChatGPT	predict	without	1.01	1.71	0.27	
Wind Speed	worse	4Turbo	predict	without	-0.00	0.00	0.50	
Wind Speed	worse	Sonnet	predict	without	-2.50	1.45	0.96	*
Wind Speed	worse	Opus	predict	without	0.75	0.55	0.08	
Wind Speed	better	Mixtral	self	without	4.48	3.80	0.11	
Wind Speed	better	ChatGPT	self	without	0.50	1.14	0.33	
Wind Speed	better	4Turbo	self	without	-0.00	0.00	0.51	
Wind Speed	better	Sonnet	self	without	-0.00	0.00	0.50	
Wind Speed	better	Opus	self	without	0.00	0.00	0.50	
Wind Speed	better	Mixtral	roleplay	without	1.01	1.27	0.21	
Wind Speed	better	ChatGPT	roleplay	without	1.52	1.47	0.15	
Wind Speed	better	4Turbo	roleplay	without	-0.00	0.00	0.50	
Wind Speed	better	Sonnet	roleplay	without	-0.00	0.00	0.51	
Wind Speed	better	Opus	roleplay	without	-0.51	0.36	0.92	
Wind Speed	better	Mixtral	predict	without	-1.00	1.04	0.84	
Wind Speed	better	ChatGPT	predict	without	-2.00	2.23	0.82	
Wind Speed	better	4Turbo	predict	without	-0.00	0.00	0.50	
Wind Speed	better	Sonnet	predict	without	0.25	0.26	0.16	
Wind Speed	better	Opus	predict	without	-0.25	0.26	0.83	
Wind Speed	uncertainty	Mixtral	self	without	0.25	0.69	0.35	
Wind Speed	uncertainty	ChatGPT	self	without	0.75	0.76	0.16	
Wind Speed	uncertainty	4Turbo	self	without	0.00	0.00	0.50	
Wind Speed	uncertainty	Sonnet	self	without	0.00	0.00	0.50	
Wind Speed	uncertainty	Opus	self	without	-0.00	0.00	0.50	
Wind Speed	uncertainty	Mixtral	roleplay	without	-0.25	0.70	0.64	
Wind Speed	uncertainty	ChatGPT	roleplay	without	-0.00	0.00	0.50	
Wind Speed	uncertainty	4Turbo	roleplay	without	-0.00	0.00	0.50	
Wind Speed	uncertainty	Sonnet	roleplay	without	0.00	0.00	0.50	
Wind Speed	uncertainty	Opus	roleplay	without	0.00	0.00	0.50	
Wind Speed	uncertainty	Mixtral	predict	without	-1.25	0.77	0.95	
Wind Speed	uncertainty	ChatGPT	predict	without	-0.00	0.00	0.50	
Wind Speed	uncertainty	4Turbo	predict	without	0.00	0.00	0.50	
Wind Speed	uncertainty	Sonnet	predict	without	-0.00	0.00	0.50	
Wind Speed	uncertainty	Opus	predict	without	0.00	0.00	0.50	
Wind Speed	utility	Mixtral	self	without	1.49	2.97	0.30	
Wind Speed	utility	ChatGPT	self	without	-3.00	2.42	0.89	
Wind Speed	utility	4Turbo	self	without	-0.00	0.00	0.50	
Wind Speed	utility	Sonnet	self	without	2.01	1.38	0.07	
Wind Speed	utility	Opus	self	without	0.74	0.42	0.04	
Wind Speed	utility	Mixtral	roleplay	without	4.01	2.62	0.06	
Wind Speed	utility	ChatGPT	roleplay	without	0.50	1.52	0.37	
Wind Speed	utility	4Turbo	roleplay	without	-0.00	0.00	0.50	
Wind Speed	utility	Sonnet	roleplay	without	-2.97	2.89	0.85	
Wind Speed	utility	Opus	roleplay	without	-2.01	0.75	1.00	*
Wind Speed	utility	Mixtral	predict	without	-0.47	2.70	0.57	
Wind Speed	utility	ChatGPT	predict	without	-1.50	1.09	0.92	
Wind Speed	utility	4Turbo	predict	without	-0.00	0.00	0.50	
Wind Speed	utility	Sonnet	predict	without	1.02	1.67	0.27	
Wind Speed	utility	Opus	predict	without	-0.25	0.73	0.64	

TABLE VII

EXPERIMENT 2 RESULT: THE ESTIMATED MEAN DIFFERENCE IN THE BELIEFS BETWEEN WORSE AND UTILITY CONDITION AND WORSE AND UNCERTAINTY CONDITION WITH THE PRIOR FIRST BELIEFS.

Hypothesis	belief	model	perspective	prior	Estimate	SE	Prob	Star
Utility	Flood Depth	Mixtral	self	with	1.00	1.05	0.83	
Uncertainty	Flood Depth	Mixtral	self	with	2.20	1.05	0.02	
Utility	Wind Speed	Mixtral	self	with	1.03	2.40	0.66	
Uncertainty	Wind Speed	Mixtral	self	with	-1.46	2.41	0.73	
Utility	Flood Depth	ChatGPT	self	with	0.20	0.16	0.88	
Uncertainty	Flood Depth	ChatGPT	self	with	-0.00	0.16	0.50	
Utility	Wind Speed	ChatGPT	self	with	-0.00	0.00	0.50	
Uncertainty	Wind Speed	ChatGPT	self	with	-0.00	0.00	0.50	
Utility	Flood Depth	4Turbo	self	with	-1.58	0.83	0.03	
Uncertainty	Flood Depth	4Turbo	self	with	-2.29	0.84	1.00	*
Utility	Wind Speed	4Turbo	self	with	-3.94	2.12	0.03	
Uncertainty	Wind Speed	4Turbo	self	with	-5.21	2.12	0.99	*
Utility	Flood Depth	Sonnet	self	with	0.00	0.00	0.51	
Uncertainty	Flood Depth	Sonnet	self	with	0.00	0.00	0.50	
Utility	Wind Speed	Sonnet	self	with	0.00	0.00	0.51	
Uncertainty	Wind Speed	Sonnet	self	with	-0.00	0.00	0.52	
Utility	Flood Depth	Opus	self	with	0.00	0.00	0.51	
Uncertainty	Flood Depth	Opus	self	with	-0.00	0.00	0.49	
Utility	Wind Speed	Opus	self	with	-0.00	0.00	0.50	
Uncertainty	Wind Speed	Opus	self	with	-0.00	0.00	0.52	
Utility	Flood Depth	Mixtral	roleplay	with	-1.61	1.12	0.08	
Uncertainty	Flood Depth	Mixtral	roleplay	with	-1.23	1.12	0.86	
Utility	Wind Speed	Mixtral	roleplay	with	0.99	1.19	0.80	
Uncertainty	Wind Speed	Mixtral	roleplay	with	0.72	1.19	0.27	
Utility	Flood Depth	ChatGPT	roleplay	with	0.00	0.00	0.51	
Uncertainty	Flood Depth	ChatGPT	roleplay	with	0.00	0.00	0.48	
Utility	Wind Speed	ChatGPT	roleplay	with	0.00	0.00	0.50	
Uncertainty	Wind Speed	ChatGPT	roleplay	with	-0.00	0.00	0.51	
Utility	Flood Depth	4Turbo	roleplay	with	0.39	0.34	0.88	
Uncertainty	Flood Depth	4Turbo	roleplay	with	-0.01	0.33	0.51	
Utility	Wind Speed	4Turbo	roleplay	with	1.04	0.81	0.90	
Uncertainty	Wind Speed	4Turbo	roleplay	with	0.03	0.82	0.48	
Utility	Flood Depth	Sonnet	roleplay	with	-0.40	0.34	0.11	
Uncertainty	Flood Depth	Sonnet	roleplay	with	-0.40	0.33	0.89	
Utility	Wind Speed	Sonnet	roleplay	with	-1.01	0.83	0.11	
Uncertainty	Wind Speed	Sonnet	roleplay	with	-1.02	0.84	0.89	
Utility	Flood Depth	Opus	roleplay	with	-0.11	0.32	0.36	
Uncertainty	Flood Depth	Opus	roleplay	with	-1.01	0.32	1.00	*
Utility	Wind Speed	Opus	roleplay	with	1.49	1.45	0.85	
Uncertainty	Wind Speed	Opus	roleplay	with	-1.73	1.43	0.89	
Utility	Flood Depth	Mixtral	predict	with	0.46	1.00	0.67	
Uncertainty	Flood Depth	Mixtral	predict	with	-2.21	1.00	0.99	*
Utility	Wind Speed	Mixtral	predict	with	0.48	1.42	0.63	
Uncertainty	Wind Speed	Mixtral	predict	with	-0.29	1.40	0.59	
Utility	Flood Depth	ChatGPT	predict	with	0.09	0.45	0.58	
Uncertainty	Flood Depth	ChatGPT	predict	with	-1.00	0.45	0.99	*
Utility	Wind Speed	ChatGPT	predict	with	-1.02	1.40	0.23	
Uncertainty	Wind Speed	ChatGPT	predict	with	-2.01	1.38	0.93	
Utility	Flood Depth	4Turbo	predict	with	-0.40	0.34	0.11	
Uncertainty	Flood Depth	4Turbo	predict	with	-0.40	0.34	0.88	
Utility	Wind Speed	4Turbo	predict	with	-1.00	0.84	0.11	
Uncertainty	Wind Speed	4Turbo	predict	with	-0.98	0.85	0.88	
Utility	Flood Depth	Sonnet	predict	with	-0.00	0.00	0.49	
Uncertainty	Flood Depth	Sonnet	predict	with	-0.00	0.00	0.50	
Utility	Wind Speed	Sonnet	predict	with	0.00	0.00	0.51	
Uncertainty	Wind Speed	Sonnet	predict	with	-0.00	0.00	0.52	
Utility	Flood Depth	Opus	predict	with	-0.00	0.20	0.49	
Uncertainty	Flood Depth	Opus	predict	with	-0.30	0.20	0.93	
Utility	Wind Speed	Opus	predict	with	-0.01	0.50	0.49	
Uncertainty	Wind Speed	Opus	predict	with	-0.76	0.51	0.93	

TABLE VIII

EXPERIMENT 2 RESULTS: THE ESTIMATED MEAN DIFFERENCE IN THE BELIEFS BETWEEN WORSE AND UTILITY CONDITION AND WORSE AND UNCERTAINTY CONDITION *without* THE PRIOR FIRST BELIEFS.

Hypothesis	belief	model	perspective	prior	Estimate	SE	Prob	Star
Utility	Flood Depth	Mixtral	self	without	0.21	0.79	0.61	
Uncertainty	Flood Depth	Mixtral	self	without	-5.70	0.78	1.00	*
Utility	Wind Speed	Mixtral	self	without	-0.92	2.45	0.35	
Uncertainty	Wind Speed	Mixtral	self	without	-8.95	2.47	1.00	*
Utility	Flood Depth	ChatGPT	self	without	-0.80	0.76	0.14	
Uncertainty	Flood Depth	ChatGPT	self	without	-1.49	0.77	0.97	*
Utility	Wind Speed	ChatGPT	self	without	-1.03	1.87	0.29	
Uncertainty	Wind Speed	ChatGPT	self	without	-2.25	1.87	0.88	
Utility	Flood Depth	4Turbo	self	without	-0.00	0.00	0.48	
Uncertainty	Flood Depth	4Turbo	self	without	-0.00	0.00	0.51	
Utility	Wind Speed	4Turbo	self	without	-0.00	0.00	0.50	
Uncertainty	Wind Speed	4Turbo	self	without	-0.00	0.00	0.50	
Utility	Flood Depth	Sonnet	self	without	0.80	0.47	0.96	*
Uncertainty	Flood Depth	Sonnet	self	without	-0.01	0.46	0.50	
Utility	Wind Speed	Sonnet	self	without	1.97	1.13	0.96	*
Uncertainty	Wind Speed	Sonnet	self	without	-0.03	1.14	0.52	
Utility	Flood Depth	Opus	self	without	-0.31	0.33	0.17	
Uncertainty	Flood Depth	Opus	self	without	-0.80	0.33	0.99	*
Utility	Wind Speed	Opus	self	without	-1.01	0.63	0.05	
Uncertainty	Wind Speed	Opus	self	without	-1.77	0.63	1.00	*
Utility	Flood Depth	Mixtral	roleplay	without	1.70	1.05	0.95	*
Uncertainty	Flood Depth	Mixtral	roleplay	without	-2.59	1.06	0.99	*
Utility	Wind Speed	Mixtral	roleplay	without	-0.17	2.46	0.47	
Uncertainty	Wind Speed	Mixtral	roleplay	without	-3.21	2.48	0.91	
Utility	Flood Depth	ChatGPT	roleplay	without	-0.50	0.58	0.19	
Uncertainty	Flood Depth	ChatGPT	roleplay	without	-1.00	0.59	0.96	*
Utility	Wind Speed	ChatGPT	roleplay	without	-1.50	1.62	0.17	
Uncertainty	Wind Speed	ChatGPT	roleplay	without	-3.01	1.62	0.97	*
Utility	Flood Depth	4Turbo	roleplay	without	-0.00	0.00	0.50	
Uncertainty	Flood Depth	4Turbo	roleplay	without	-0.00	0.00	0.50	
Utility	Wind Speed	4Turbo	roleplay	without	0.00	0.00	0.50	
Uncertainty	Wind Speed	4Turbo	roleplay	without	-0.00	0.00	0.51	
Utility	Flood Depth	Sonnet	roleplay	without	-2.42	0.94	0.00	
Uncertainty	Flood Depth	Sonnet	roleplay	without	-3.61	0.94	1.00	*
Utility	Wind Speed	Sonnet	roleplay	without	-5.00	2.38	0.02	
Uncertainty	Wind Speed	Sonnet	roleplay	without	-9.03	2.38	1.00	*
Utility	Flood Depth	Opus	roleplay	without	-0.30	0.26	0.12	
Uncertainty	Flood Depth	Opus	roleplay	without	-1.70	0.26	1.00	*
Utility	Wind Speed	Opus	roleplay	without	-0.50	0.67	0.23	
Uncertainty	Wind Speed	Opus	roleplay	without	-4.01	0.68	1.00	*
Utility	Flood Depth	Mixtral	predict	without	0.43	1.35	0.62	
Uncertainty	Flood Depth	Mixtral	predict	without	-2.38	1.36	0.96	*
Utility	Wind Speed	Mixtral	predict	without	-2.56	2.42	0.14	
Uncertainty	Wind Speed	Mixtral	predict	without	-2.56	2.35	0.86	
Utility	Flood Depth	ChatGPT	predict	without	-0.60	0.37	0.05	
Uncertainty	Flood Depth	ChatGPT	predict	without	-0.60	0.37	0.95	
Utility	Wind Speed	ChatGPT	predict	without	-2.00	1.13	0.04	
Uncertainty	Wind Speed	ChatGPT	predict	without	-1.98	1.12	0.96	*
Utility	Flood Depth	4Turbo	predict	without	-0.00	0.00	0.48	
Uncertainty	Flood Depth	4Turbo	predict	without	-0.00	0.00	0.52	
Utility	Wind Speed	4Turbo	predict	without	-0.00	0.00	0.49	
Uncertainty	Wind Speed	4Turbo	predict	without	-0.00	0.00	0.52	
Utility	Flood Depth	Sonnet	predict	without	0.80	0.46	0.95	*
Uncertainty	Flood Depth	Sonnet	predict	without	-0.00	0.46	0.50	
Utility	Wind Speed	Sonnet	predict	without	2.03	1.14	0.96	*
Uncertainty	Wind Speed	Sonnet	predict	without	0.01	1.15	0.50	
Utility	Flood Depth	Opus	predict	without	0.40	0.20	0.97	*
Uncertainty	Flood Depth	Opus	predict	without	-1.60	0.20	1.00	*
Utility	Wind Speed	Opus	predict	without	-0.26	0.51	0.30	
Uncertainty	Wind Speed	Opus	predict	without	-1.00	0.50	0.97	*