Simulating Persuasive Dialogues on Meat Reduction with Generative Agents

Georg Ahnert[1*†], **Elena Wurth**[1*], **Markus Strohmaier**[1,2,3], **Jutta Mata**[1,4]

[1]University of Mannheim
[2]GESIS - Leibniz Institute for the Social Sciences, Cologne
[3]Complexity Science Hub, Vienna
[4]Max Planck Institute for Human Development, Berlin

Abstract

Meat reduction benefits human and planetary health, but social norms keep meat central in shared meals. To date, the development of *communication strategies that promote meat reduction while minimizing social costs* has required the costly involvement of human participants at each stage of the process. We present work in progress on simulating multi-round dialogues on meat reduction between *Generative Agents* based on large language models (LLMs). We measure our main outcome using established psychological questionnaires based on the *Theory of Planned Behavior* and additionally investigate *Social Costs*. We find evidence that our preliminary simulations produce outcomes that are (i) consistent with theoretical expectations; and (ii) valid when compared to data from previous studies with human participants. Generative agent-based models are a promising tool for identifying novel communication strategies on meat reduction—tailored to highly specific participant groups—to then be tested in subsequent studies with human participants.

1 Introduction

Research Objectives Reducing meat consumption offers substantial benefits for human and planetary health, yet entrenched social norms continue to place meat at the center of shared meals (Godfray et al. 2018). Vegans, vegetarians, and flexitarians have the potential to challenge these norms and drive important social change (Judge et al. 2024). However, in mixed-diet social settings, they often self-silence to avoid social costs (e.g., Bolderdijk and Cornelissen 2022; Romo and Donovan-Kicken 2012). To date, the development of communication strategies that promote meat reduction while minimizing social costs has required human participants at each stage of the process (e.g., Carfora et al. 2019; Pabian et al. 2020). This is not only very costly—both financially for the researcher and in terms of participant effort—but also severely restricts the range of communication strategies and participant diversity that can be explored. Generative Agent simulations based on large language models (LLMs) have the potential to overcome these limitations (Bail 2024).

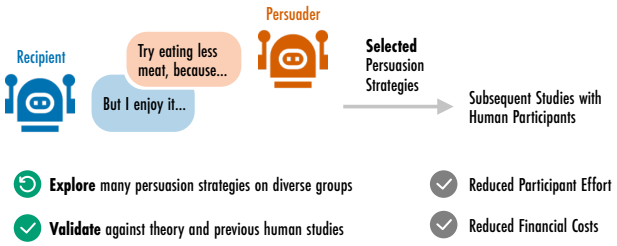


Figure 1: **Simulation of Persuasive Dialogues.** We use generative agents to develop and select persuasion strategies for meat reduction with minimal social costs, to be tested in subsequent studies with human participants.

Approach We present work in progress on simulating multi-round dialogues on meat reduction between *generative agents* based on LLMs, as shown in Figure 1. Related work on generative agent simulations already shows promising results (Park et al. 2022; Törnberg et al. 2023). Still, it remains an open question how well such simulations can inform subsequent studies with human participants, especially in our context of dialogues on meat reduction. This paper therefore addresses the following **Research Question**:

To what extent can generative agent simulations reliably reproduce persuasive human dialogues about meat reduction? To answer this, we validate the behavior of our agents both internally and against data from prior experiments with human participants. This validation is a necessary step before using these simulations to inform the design of future studies involving human subjects.

Results Initial experiments with the `Llama 3` family of models indicate that generative agents can effectively simulate persuasive communication, producing reliable and valid response patterns across key psychological constructs. However, issues such as uniformity in some constructs and the potential influence of instruction-tuning on strategies will require further investigation.

Generative agent-based models can be a promising tool for identifying novel communication strategies for meat reduction, tailored to highly specific participant groups, to then be tested in subsequent studies with human participants.

*These authors contributed equally.
†corresponding author: ahnert [at] uni-mannheim [dot] de

Proceedings of the First Workshop on Integrating NLP and Psychology to Study Social Interactions (NLPSI) @ICWSM '25, pages 1–10

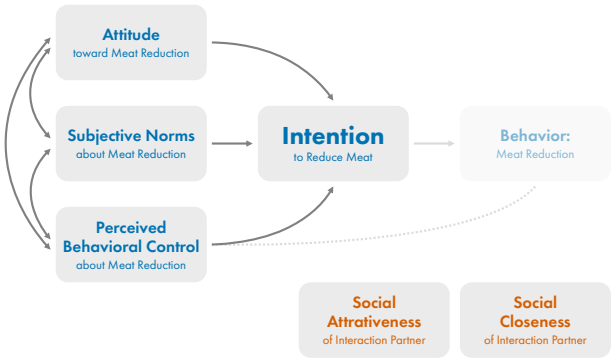


Figure 2: **Constructs Measured for the Recipient, Comprising Theory of Planned Behavior (TPB) and Social Costs.** Behavior itself (i.e., meat reduction) is unobservable to us in our simulation, but previous research has established that self-reported **Intention to Reduce Meat**—our main outcome—adequately predicts actual reduction (Berndsen and Van Der Pligt 2005; Saba and Di Natale 1998).

2 Background

Studies on Meat Reduction The Theory of Planned Behavior (TPB; Ajzen 1991) explains behavior through three core constructs: *Attitudes*, *Subjective Norms*, and *Perceived Behavioral Control*. These shape behavioral *Intentions*, which in turn predict actual behavior. On average, intentions account for 28% of variance in health behavior (Sheeran et al. 2005). In meat reduction, these variables reflect recipient traits as follows: attitudes map personal beliefs about meat, norms relate to perceived expectations of others, and control captures confidence in one's ability to change eating habits. From theory and previous research, all three constructs reliably predict the intention to eat less meat, which in turn predicts actual meat consumption (Berndsen and Van Der Pligt 2005; Saba and Di Natale 1998).

Social Simulations with Generative Agents A rapidly growing body of research investigates the feasibility of using large language models (LLMs) to model survey responses (e.g., Argyle et al. 2023; Bisbee et al. 2023; Ahnert et al. 2025), or the outcomes of experimental studies (e.g., Hewitt et al. 2024; Aher, Arriaga, and Kalai 2024). LLMs promise to facilitate novel exploratory studies of human behavior (Bail 2024), but recent research indicates that LLMs might not always accurately mimic human study participants (Tjuatja et al. 2024). It remains an open question how well socio-demographically prompted LLMs can represent human study participants (Sen et al. 2025)—in particular, LLMs were shown to "flatten" the reported attitudes of identity groups (Wang, Morgenstern, and Dickerson 2024).

Generative Agents use LLMs to simulate individuals and their interactions. Previous research applied generative agents, for example, to simulate communication on social media platforms (Park et al. 2022; Törnberg et al. 2023). Persuasive dialogues between generative agents have been found to favor communication strategies similar to humans (Vaccaro et al. 2025), but also to converge to the inherent biases of the underlying LLMs (Taubenfeld et al. 2024). We apply a similar generative agent setup to a novel context: persuasion for meat reduction with minimal social costs. This allows us to draw from the well-established literature in health psychology to statistically validate our generative agent-based model and to assess its potential for informing subsequent studies with human participants. We test a variety of open-weight LLMs, since previous research found that model size is an important predictor of negotiation success (Davidson et al. 2024).

3 Approach

Psychological Constructs Based on two current meta-analyses (Harguess, Crespo, and Hong 2020; Kwasny, Dobernig, and Riefler 2022), we identify several personal characteristics relevant to meat consumption and its reduction, including demographics, values, and personality traits, as shown in Table 1. Based on these central characteristics and to test the potential of our method, we develop two contrasting personas. According to the two meta-analyses mentioned above, these personas should represent opposite ends of the expected susceptibility to meat reduction appeals: a progressive, open-minded, younger woman with high education and income living in an urban setting (*Easy to Persuade Recipient*), and a conservative, older male with limited education and low income residing in a rural area (*Hard to Persuade Recipient*).

We define the Theory of Planned Behavior (TPB) as our theoretical framework and the corresponding variables as our outcomes for persuasion, as shown in Figure 2. Consequently, we use validated instruments to assess these constructs reliably. As part of our simulation, we extract survey responses from the agent to be persuaded (*Recipient*), with **Intention to Reduce Meat** as our main outcome. The questionnaire measures our 6 central constructs: *Attitudes*, *Subjective Norms*, and *Behavioral Control* of the Recipient; the Recipient's *Intention to Reduce Meat Consumption*; as well as the *Social Closeness* and the *Social Attractiveness*

Persona 1: Easy to Persuade Recipient	**Persona 2: Hard to Persuade Recipient**
demographics	**demographics**
female, younger, living in the city, high income	male, older, living on the countryside, low income
values	**values**
self-transcendence, openness to change, encouraging empathy towards animals	self-enhancement, conservation, discouraging empathy towards animals
personality traits	**personality traits**
open, conscientious	conservative, careless

Table 1: **Contrasting Recipient Personas.** For our initial experiments, we focus on two contrasting personas that provide the greatest potential for identifying existing effects, as they represent opposite ends of the expected susceptibility to meat reduction arguments.

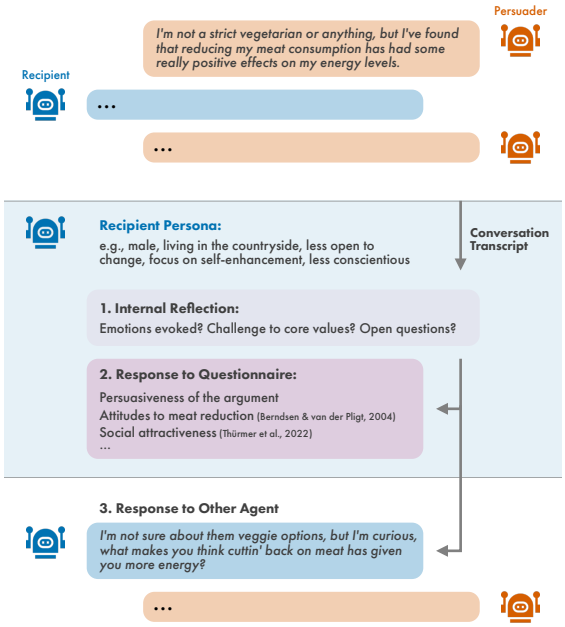


Figure 3: **Simulation Setup.** We simulate dialogues in which a Persuader agent aims to convince a Recipient agent to reduce their meat consumption. In each round of the simulation, agents first (1) perform internal reflection, and (2) answer a questionnaire, before they (3) generate a response to the other agent.

of the Persuader, as perceived by the Recipient. The first 4 constructs comprise the aspects of the TPB that we can measure in our simulation, while the latter 2 constructs measure Social Costs—see Appendix A for the full questionnaire. We expect individuals with favorable attitudes, strong normative support, and high perceived control to be more receptive to meat reduction messages.

Simulation Setup We simulate meat reduction dialogues between two generative agents (*Persuader* and *Recipient*) as shown in Figure 3. Every conversation is started by the Persuader and consists of 5 *rounds*, in which both agents produce 1 message each. To produce a new message, an agent goes through the following steps: it **(1) performs internal reflection**; Recipient agents **(2) answer a questionnaire**; and finally, an agent **(3) generates a response** to the other agent. The Persuader has explicit instructions to persuade the other agent to reduce meat, while the Recipient is instructed to adopt one of the personas shown in Table 1. In our initial simulations, we do not prompt the Persuader to follow a specific persuasion strategy. We explain the simulation process in the following and publish all respective prompts as part of our code under MIT license.[1]

To perform **internal reflection**, all agents receive a transcript of the conversation so far and are instructed to reflect on *which emotions were evoked*, *which core values were challenged or supported*, and *which questions or uncertainties there are*. Only the Recipient agent, then, fills out the **questionnaire** described above. Finally, each agent **generates a response** to the other agent. This response takes into account only the agent's persona, the transcript of the conversation so far, and the agent's internal reflection. To ensure that the questionnaire does not affect the simulation, we remove it from the LLM's context.

For our initial analyzes, we simulate 200 dialogues for both contrasting Recipient *personas* that are shown in Table 1, i.e., 400 conversations in total, with 10 messages exchanged in each dialogue. We extract 10 independent questionnaire answers from the Recipient persona to improve robustness. We opt for *open-weight* LLMs to facilitate replication (Barrie, Palmer, and Spirling 2024), and run our simulation with 3 distinct LLMs to investigate the impact of model size: `Llama 3.3 70B`, `Llama 3.1 8B`, and `Llama 3.2 3B`. For the questionnaire, we employ *structured outputs* tailored to the respective answer options, to ensure that we obtain valid responses even from the smallest model. We run all simulations with Llama's default temperature of 0.6 and use `vllm` with *automatic prefix caching* to improve model throughput—the combined runtime for all simulations on two NVIDIA H100 in parallel was $\approx$ 70 hours.

Validation To assess whether generative agent simulations can meaningfully replicate persuasive human dialogues about meat reduction, we implement a two-step validation strategy. First, we perform an **internal validation** by evaluating whether the agents' responses exhibit psychometric properties comparable to those observed in human data. Using established psychological constructs and measurement instruments, we assess internal consistency, convergent and discriminant validity, and distributional plausibility. Second, we conduct an **external validation** by comparing the simulated results against empirical data from prior human studies with similar persuasion tasks (see Table 2). This dual approach enables us to examine both the internal coherence and the empirical realism of the simulated dia-

Construct	Original Study	*n*
Attitude & Intention to Reduce Meat	Pabian et al. (2020)	47
Subjective Norms & Behavioral Control	Wyker and Davison (2010)	204
Social Attractiveness	Thürmer, Stadler, and McCrea (2022)	260
Social Closeness	Monin, Sawyer, and Marquez (2008)	70
Threat of Freedom	Dillard and Shen (2005)	196
Meat Attachment	Graça, Calheiros, and Oliveira (2015)	318

Table 2: **Studies with Human Participants Used for External Validation.** We compare published means against the survey responses from our generative agent simulation.

[1] https://github.com/dess-mannheim/MeatlessAgents

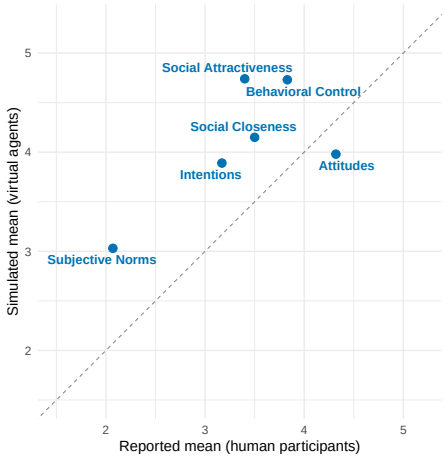


Figure 4: **Simulated Survey Responses Are on Average Close to Responses from Human Participants.** Mean values of our six central constructs based on human-reported data and simulated agent data. Each point represents one construct. The dashed diagonal line indicates perfect agreement between simulated and human means. All response scales range from 1 to 7.

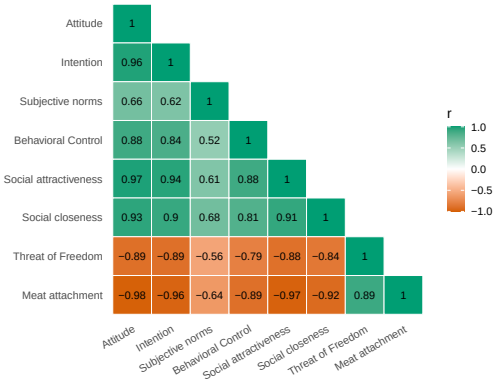


Figure 5: **Strong Pearson Correlations** between *Attitude*, *Intention*, *Behavioral Control*, *Social Attractiveness* and *Social Closeness* support **Convergent Validity**. Weaker correlations with theoretically more distinct constructs such as *Subjective Norms* support **Discriminant Validity**. *Threat of Freedom* and *Meat Attachment* were included to help assess if constructs form a distinct and coherent cluster.

logues. Crucially, this validation serves as a necessary foundation: only if our simulation proves **sufficiently reliable and valid** can it be used to **inform the design of future experiments**—with real human participants—under comparable conditions.

4 Results

To evaluate the quality of our simulation, we conducted several reliability and validity analyses. All results shown below refer to `Llama 3.3 70B`. Respective results for the smaller Llama models are provided in Appendix B. Note that with smaller Llama models, reliability and validity of the measures also decrease, but to a still acceptable level.

Descriptive Statistics & Distributions Appendix Figure 7 visualizes the distributions of the measures across all conversations and the two personas using boxplots. *Attitude* and *Intention* showed the highest median values and the highest variability in the sample. These wide distributions likely reflect the contrasting tendencies of the two personas, resulting in a broader spread of responses. *Behavioral Control*, *Social Attractiveness*, and *Social Closeness* revealed similarly high median values with more narrow spreads. This suggests more uniform perceptions across personas in terms of perceived control and social perception of the persuader. These constructs may have been modeled less distinctly between personas or are inherently perceived as less polarized. *Subjective Norms* exhibited both the lowest median and least variability, indicating that both personas shared similarly low perceptions of social expectations to reduce meat consumption. This uniformity could reflect limitations in the way social dynamics were depicted within the simulation framework.

To assess the internal consistency of the measured constructs, we calculated Cronbach's Alpha coefficients for each scale. The results indicated excellent reliability for most constructs (e.g. *Meat Attachment*, *Attitude*, *Intention*, *Social Attractiveness*, and *Social Closeness* $\alpha s \geq .98$). In contrast, *Subjective Norms* ($\alpha = .76$) and *Behavioral Control* ($\alpha = .70$) showed comparatively lower but still acceptable internal consistency. A detailed overview of the reliability coefficients, including comparisons with the values obtained from human samples, is provided in Appendix Table 3.

Comparing Simulated and Human Data To assess the similarity between simulated and human data, we compared the mean values of key constructs with those reported in previous studies using the same measures with human participants. As we did not have access to the original datasets, the comparisons were based on published mean values. Table 2 shows the original studies belonging to the respective constructs with their respective number of human participants. A scatterplot of the simulated versus reported means is shown in Figure 4. Most constructs produced higher mean values in the simulation compared to human data, with the exception of *Attitudes*, where the simulated mean was slightly lower. The largest discrepancy was observed for *Subjective Norms*, which were generally low in both cases, but especially in the human samples. In general, the points cluster closely around the identity line, indicating a high degree of similarity between the simulated and human-generated data, particularly in the relative positioning of the constructs. This supports the validity of the simulation in capturing psychologically plausible response patterns. However, these findings should be interpreted as preliminary evidence.

Construct Validity To further assess the construct validity of the simulated responses, we examined the intercorrelations between all measured variables, as visualized in Figure 5. The results reveal strong positive correlations between constructs that are theoretically linked, such as *Attitude*, *Intention*, *Social Attractiveness*, and *Social Closeness* (e.g., $r = .96$ between Attitude and Intention). This supports convergent validity. Correlations with less strongly related constructs, such as relations with *Subjective Norms*, are notably lower (e.g., $r = .52$ between Behavioral Control and Subjective Norms), supporting discriminant validity. These two constructs were deliberately included as control variables to assess whether the central constructs form a conceptually distinct cluster. However, some correlations between central variables are exceptionally high (approaching or exceeding $r = .95$), which is uncommon in empirical data and may reflect the polarized nature of the simulated personas. This could lead to overly homogeneous response patterns, limiting generalizability, and possibly inflating convergence artificially. Future simulations could benefit from a broader and more nuanced range of personas to capture more natural variance.

Persuasion Effectiveness Figure 6 shows the reported intention to reduce meat consumption—our main outcome—and each point represents the mean response of a simulated Recipient across the 3 respective questionnaire items and 10 repeated survey responses. As expected, the Hard to Persuade Recipient indicates a much smaller intention for meat reduction than the Easy to Persuade Participant. For `Llama 3.3 70B`, we also observe much more variance in the responses from the Hard to Persuade Recipient. Surprisingly, during the first rounds of the conversation, the simulated survey responses for the Hard to Persuade Recipient show a decreasing intention to reduce meat consumption. This could be interpreted as a backlash to the persuasion attempt, a pattern that is also observed with human participants (Shen, Sheer, and Li 2015). After approximately 3 conversation rounds, we observe an increasing mean intention for meat reduction, across both participants and all LLMs. This trend continues in round 5 for the Hard to Persuade Recipient and `Llama 3.3 70B`, which we take as a reason to extend future simulation to more rounds.

5 Conclusion

Our preliminary findings demonstrate that generative agents can engage in plausible discussions about meat reduction. However, limitations such as uniformity in subjective norm responses require further investigation. As next steps, we will extend the simulation to more rounds and run targeted simulations in which we instruct the Persuader to employ a specific persuasion strategy. Additionally, we will investigate more diverse personas.

Encouraging more people to reduce meat consumption remains a vital goal for both individual well-being and planetary health. Generative agent-based simulations offer great potential for large-scale exploration of persuasion strategies for meat reduction, which will then be tested in subsequent studies with human participants.

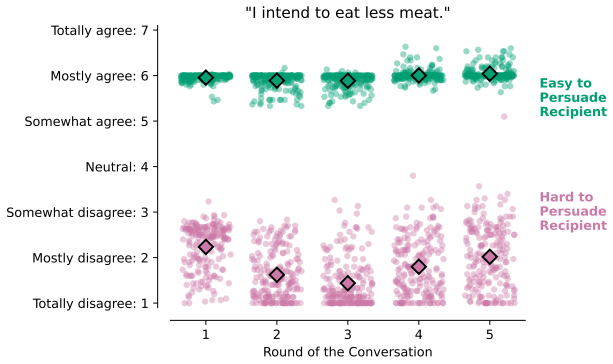


Figure 6: **Recipients' Intention to Reduce Meat.** Combined survey responses after each conversation round, where ◊ indicates the mean over all conversations per round. Following our expectations, the Easy to Persuade Recipient responds with a generally high intention to reduce meat. Over the first rounds of the conversation, the Hard to Persuade Recipient shows an initially decreasing intention to reduce meat, similar to human participants (Shen, Sheer, and Li 2015).

Ethical Considerations

From an ethical standpoint, our findings underscore both the potential and risks associated with using large language models for persuasive communication. While our study focuses on promoting reduced meat consumption—a goal aligned with individual and planetary health—the underlying persuasive strategies and technical setup are generalizable to a wider range of topics.

This generalizability is methodologically promising, but it also raises important ethical considerations. The same techniques used to encourage socially beneficial behavior change could potentially be repurposed to mislead or exert undue influence. However, effective persuasion still depends on a person's openness to change—people are unlikely to be swayed to adopt behaviors they fundamentally reject, especially those they view as ethically wrong. Moreover, we would expect persuasive dynamics in humans to be more complex and often less effective, given that conversational agents are inherently designed to be accommodating, assistive, and agreeable—traits that may amplify the success of persuasive strategies in artificial contexts compared to real human interactions. Nonetheless, without clear guidelines and safeguards, there remains a real risk of these technologies being used to promote misinformation, ideological agendas, or exploitative commercial practices.

We anticipate that applications of persuasive AI will continue to expand rapidly. Therefore, we believe it is essential to publish and critically examine what LLMs are already capable of in this space. Our goal is not only to inform future research, but to contribute to a proactive and transparent discourse about the ethical governance of persuasive AI technologies.

References

Aher, G.; Arriaga, R. I.; and Kalai, A. T. 2024. Using Large Language Models to Simulate Multiple Humans and Replicate Human Subject Studies.

Ahnert, G.; Pellert, M.; Garcia, D.; and Strohmaier, M. 2025. Extracting Affect Aggregates from Longitudinal Social Media Data with Temporal Adapters for Large Language Models. ArXiv:2409.17990 [cs].

Ajzen, I. 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2): 179–211.

Argyle, L. P.; Busby, E. C.; Fulda, N.; Gubler, J. R.; Rytting, C.; and Wingate, D. 2023. Out of One, Many: Using Language Models to Simulate Human Samples. *Political Analysis*, 31(3): 337–351.

Bail, C. A. 2024. Can Generative AI improve social science? *Proceedings of the National Academy of Sciences*, 121(21): e2314021121.

Barrie, C.; Palmer, A.; and Spirling, A. 2024. Replication for Language Models.

Berndsen, M.; and Van Der Pligt, J. 2005. Risks of meat: the relative impact of cognitive, affective and moral concerns. *Appetite*, 44(2): 195–205.

Bisbee, J.; Clinton, J.; Dorff, C.; Kenkel, B.; and Larson, J. 2023. Synthetic Replacements for Human Survey Data? The Perils of Large Language Models. preprint, SocArXiv.

Bolderdijk, J. W.; and Cornelissen, G. 2022. "How do you know someone's vegan?" They won't always tell you. An empirical test of the do-gooder's dilemma. *Appetite*, 168: 105719.

Carfora, V.; Catellani, P.; Caso, D.; and Conner, M. 2019. How to reduce red and processed meat consumption by daily text messages targeting environment or health benefits. *Journal of Environmental Psychology*, 65: 101319.

Davidson, T. R.; Veselovsky, V.; Josifoski, M.; Peyrard, M.; Bosselut, A.; Kosinski, M.; and West, R. 2024. Evaluating Language Model Agency through Negotiations. ArXiv:2401.04536 [cs].

Dillard, J. P.; and Shen, L. 2005. On the Nature of Reactance and its Role in Persuasive Health Communication. *Communication Monographs*, 72(2): 144–168.

Godfray, H. C. J.; Aveyard, P.; Garnett, T.; Hall, J. W.; Key, T. J.; Lorimer, J.; Pierrehumbert, R. T.; Scarborough, P.; Springmann, M.; and Jebb, S. A. 2018. Meat consumption, health, and the environment. *Science*, 361(6399): eaam5324.

Graça, J.; Calheiros, M. M.; and Oliveira, A. 2015. Attached to meat? (Un)Willingness and intentions to adopt a more plant-based diet. *Appetite*, 95: 113–125.

Harguess, J. M.; Crespo, N. C.; and Hong, M. Y. 2020. Strategies to reduce meat consumption: A systematic literature review of experimental studies. *Appetite*, 144: 104478.

Hewitt, L.; Ashokkumar, A.; Ghezae, I.; and Willer, R. 2024. Predicting Results of Social Science Experiments Using Large Language Models.

Judge, M.; Bouman, T.; Steg, L.; and Bolderdijk, J. W. 2024. Accelerating social tipping points in sustainable behaviors: Insights from a dynamic model of moralized social change. *One Earth*, 7(5): 759–770.

Kwasny, T.; Dobernig, K.; and Riefler, P. 2022. Towards reduced meat consumption: A systematic literature review of intervention effectiveness, 2001–2019. *Appetite*, 168: 105739.

Monin, B.; Sawyer, P. J.; and Marquez, M. J. 2008. The rejection of moral rebels: Resenting those who do the right thing. *Journal of Personality and Social Psychology*, 95(1): 76–93.

Pabian, S.; Hudders, L.; Poels, K.; Stoffelen, F.; and De Backer, C. J. S. 2020. Ninety Minutes to Reduce One's Intention to Eat Meat: A Preliminary Experimental Investigation on the Effect of Watching the Cowspiracy Documentary on Intention to Reduce Meat Consumption. *Frontiers in Communication*, 5: 69.

Park, J. S.; Popowski, L.; Cai, C.; Morris, M. R.; Liang, P.; and Bernstein, M. S. 2022. Social Simulacra: Creating Populated Prototypes for Social Computing Systems. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, 1–18. Bend OR USA: ACM. ISBN 978-1-4503-9320-1.

Romo, L. K.; and Donovan-Kicken, E. 2012. "Actually, I Don't Eat Meat": A Multiple-Goals Perspective of Communication About Vegetarianism. *Communication Studies*, 63(4): 405–420.

Saba, A.; and Di Natale, R. 1998. A study on the mediating role of intention in the impact of habit and attitude on meat consumption. *Food Quality and Preference*, 10(1): 69–77.

Sen, I.; Lutz, M.; Rogers, E.; Garcia, D.; and Strohmaier, M. 2025. Missing the Margins: A Systematic Literature Review on the Demographic Representativeness of LLMs.

Sheeran, P.; Milne, S.; Webb, T. L.; and Gollwitzer, P. M. 2005. Implementation Intentions and Health Behaviour. In Conner, M., ed., *Predicting Health Behaviour: Research and Practice with Social Cognition Models*, 276–323. New York: Open Univ. Pr. ISBN 978-0-335-21177-7.

Shen, F.; Sheer, V. C.; and Li, R. 2015. Impact of Narratives on Persuasion in Health Communication: A Meta-Analysis. *Journal of Advertising*, 44(2): 105–113.

Taubenfeld, A.; Dover, Y.; Reichart, R.; and Goldstein, A. 2024. Systematic Biases in LLM Simulations of Debates. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 251–267. Miami, Florida, USA: Association for Computational Linguistics.

Thürmer, J. L.; Stadler, J.; and McCrea, S. M. 2022. Intergroup Sensitivity and Promoting Sustainable Consumption: Meat Eaters Reject Vegans' Call for a Plant-Based Diet. *Sustainability*, 14(3): 1741.

Tjuatja, L.; Chen, V.; Wu, S. T.; Talwalkar, A.; and Neubig, G. 2024. Do LLMs exhibit human-like response biases? A case study in survey design. ArXiv:2311.04076 [cs].

Törnberg, P.; Valeeva, D.; Uitermark, J.; and Bail, C. 2023. Simulating Social Media Using Large Language

Models to Evaluate Alternative News Feed Algorithms. ArXiv:2310.05984 [cs].

Vaccaro, M.; Caoson, M.; Ju, H.; Aral, S.; and Curhan, J. R. 2025. Advancing AI Negotiations: New Theory and Evidence from a Large-Scale Autonomous Negotiations Competition. Version Number: 1.

Wang, A.; Morgenstern, J.; and Dickerson, J. P. 2024. Large language models cannot replace human participants because they cannot portray identity groups. ArXiv:2402.01908 [cs].

Wyker, B. A.; and Davison, K. K. 2010. Behavioral Change Theories Can Inform the Prediction of Young Adults' Adoption of a Plant-based Diet. *Journal of Nutrition Education and Behavior*, 42(3): 168–177.

A Questionnaires

In the following section, the questionnaires used in this study to assess all measured outcomes are described. An overview of the six key constructs is provided in Figure 2.

To assess the *persuasiveness* of the presented arguments, participants were asked to rate the item *"How persuasive did you find this argument?"* on a 7-point Likert scale.

Behavioral change was measured with the item *"To what extent did this argument make you consider changing your behavior?"*, also using a 7-point Likert scale.

Attitudes toward meat reduction were assessed using a four-item semantic differential scale based on Berndsen and Van Der Pligt (2005), rated on a 7-point scale:

- Pleasant – Unpleasant
- Useful – Useless
- Favorable – Unfavorable
- Good – Bad

Intentions to reduce meat consumption were captured with a three-item scale adapted from Graça, Calheiros, and Oliveira (2015), using a 7-point Likert scale:

- I intend to eat less white meat.
- I intend to eat less red meat.
- I intend to eat less processed meat.

Subjective norms were measured with four items adapted from Berndsen and Van Der Pligt (2005), each rated on a 7-point scale:

- How much do you feel your friends want you to reduce your meat consumption in the next year?
- How much do you feel your family want you to reduce your meat consumption in the next year?
- How much do you feel health experts want you to reduce your meat consumption in the next year?
- How much do you feel your colleagues want you to reduce your meat consumption in the next year?

Perceived behavioral control was assessed via two items based on Wyker and Davison (2010), using a 7-point Likert scale:

- How much personal control do you feel you have about reducing your meat consumption in the next year?
- To what extent do you see yourself as being capable of reducing your meat consumption in the next year?

Attachment to meat was measured using the 20-item Meat Attachment Questionnaire (MAQ) developed by Graça, Calheiros, and Oliveira (2015), on a 5-point Likert scale. Items include:

- To eat meat is one of the good pleasures in life.
- Meat is irreplaceable in my diet.
- According to our position in the food chain, we have the right to eat meat.
- I feel bad when I think of eating meat.
- I love meals with meat.
- To eat meat is disrespectful towards life and the environment.
- To eat meat is an unquestionable right of every person.
- Meat consumption is crucial to my balance.
- A full meal is a meal with meat.
- I'm a big fan of meat.
- If I couldn't eat meat I would feel weak.
- If I was forced to stop eating meat I would feel sad.
- By eating meat I'm reminded of the death and suffering of animals.
- Eating meat is a natural and undisputable practice.
- I don't picture myself without eating meat regularly.
- Meat sickens me.
- I would feel fine with a meatless diet.
- Meat consumption is a natural act of one's affirmation as a human being.
- A good steak is without comparison.
- Meat reminds me of diseases.

The *perceived threat to freedom* was assessed using four items from Dillard and Shen (2005), rated on a 7-point Likert scale:

- The person tried to make a decision about my own diet for me.
- The person tried to influence me and my diet.
- The person threatened my freedom to decide about my own diet.
- The person tried to pressure me regarding my diet.

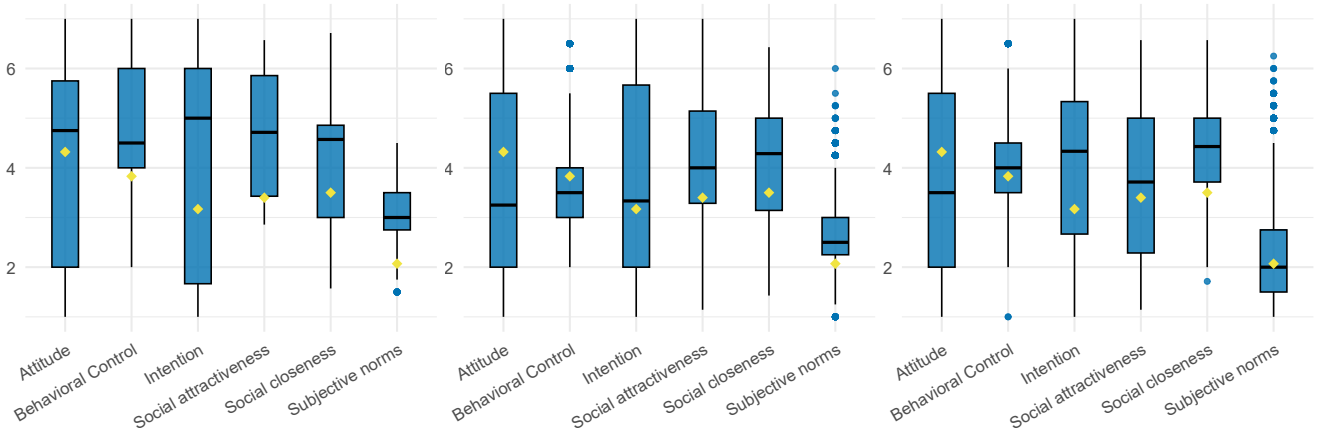
Social attractiveness was measured with seven items adapted from Thürmer, Stadler, and McCrea (2022), using a 7-point Likert scale:

- To what extent do you think your interaction partner is intelligent?
- ...trustworthy?
- ...friendly?
- ...open?
- ...likeable?
- ...respectable?
- ...interesting?

Social closeness was measured via seven items based on Monin, Sawyer, and Marquez (2008), rated on a 7-point Likert scale:

- I would like to get to know the other person better.
- I would like to have lunch with the other person sometime.
- I would like to work together with the other person on a task.
- I would like to have a conversation with the other person about food.
- I felt emotionally close to the other person.
- I wanted to have a conversation with the other person.
- I felt like I made friends with the other person.

B Additional Results

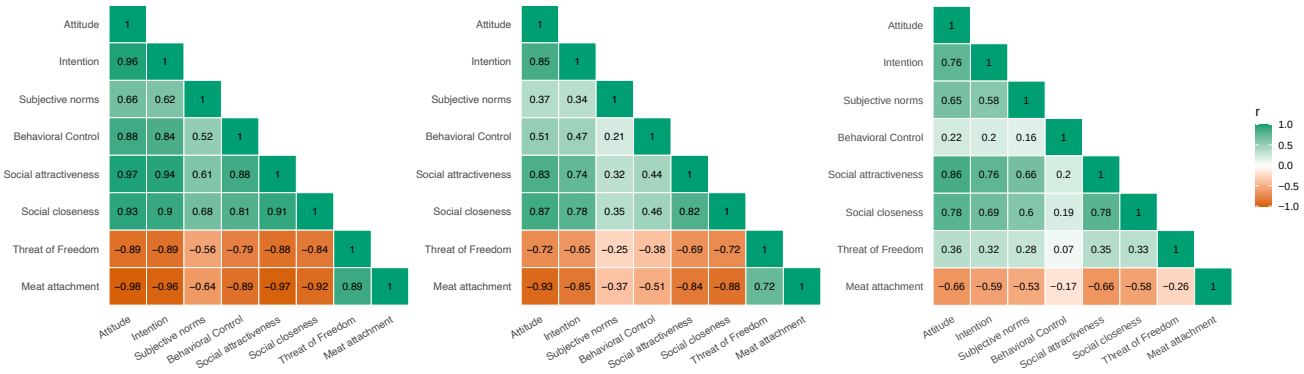


(a) **Llama3.3 70B** (b) **Llama 3.1 8B** (c) **Llama 3.2 3B**

Figure 7: **Smaller `Llama` models show less consistent construct response distributions.** The distribution of model-generated responses is shown across `Llama models (70B, 8B, 3B)` for key psychological constructs. Yellow diamonds indicate the mean from human samples for each construct. Compared to the 70B model, smaller models exhibit reduced variability and more outliers in several constructs (e.g., *Behavioral Control*, *Subjective Norms*), suggesting a loss of distributional alignment with human responses as model size decreases.

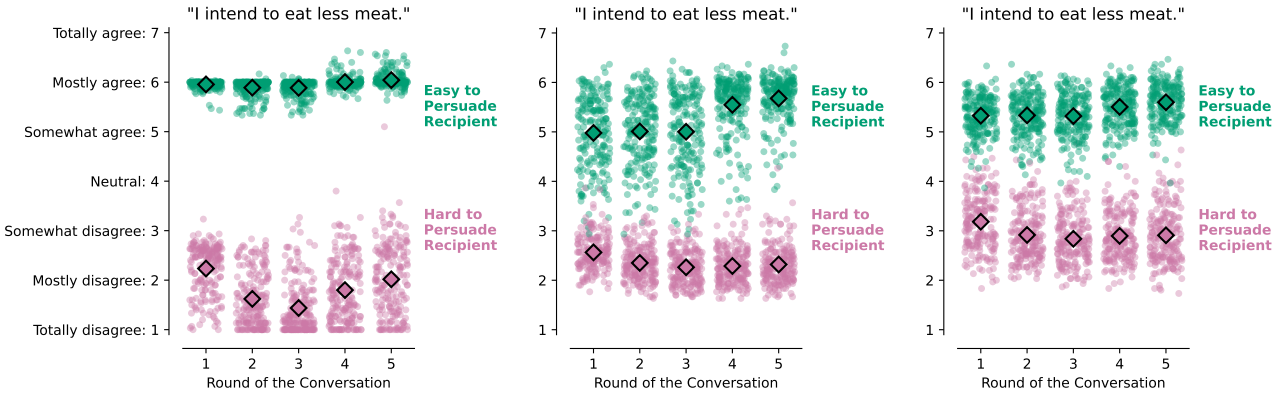
Scale	#items	Llama 3.3 70B	Llama 3.1 8B	Llama 3.2 3B	Human Samples	Original Study
Theory of Planned Behavior						
Attitude	4	.99	.97	.92	.86	Pabian et al. (2020)
Intention	3	.99	.93	.88	.85	Pabian et al. (2020)
Subjective Norms	2	.76	.59	.78	.75	Wyker and Davison (2010)
Behavioral Control	2	.70	.66	.68	.70	Wyker and Davison (2010)
Social Costs						
Social Attractiveness	7	.98	.93	.96	.91	Thürmer, Stadler, and McCrea (2022)
Social Closeness	7	.98	.94	.82	.91	Monin, Sawyer, and Marquez (2008)
Additional Constructs						
Threat of Freedom	4	.93	.87	.60	.84	Dillard and Shen (2005)
Meat Attachment	20	.99	.98	.71	.92	Graça, Calheiros, and Oliveira (2015)

Table 3: **Reliability of psychological constructs decreases with smaller Llama models.** Cronbach's Alpha values for each scale are compared across `Llama models (3.3 70B, 3.1 8B, 3.2 3B)` and human participant data to assess internal consistency. While larger models consistently show high reliability ($\alpha \geq .70$), several scales fall below the acceptable threshold in smaller models (shown in red)—most notably *Behavioral Control*, and *Threat of Freedom*. This suggests a degradation in the coherent representation of multi-item constructs as model size decreases.



(a) **Llama 3.3 70B** (b) **Llama 3.1 8B** (c) **Llama 3.2 3B**

Figure 8: **Correlation strength weakens with smaller Llama models, while overall construct patterns remain stable.** To assess construct validity, we examine Pearson correlations among psychological constructs across three Llama models of varying sizes. While general correlation patterns—including the two added constructs *Threat of Freedom* and *Meat Attachment*—remain largely stable, the strength of correlations systematically decreases as model size reduces. This suggests a loss of representational consistency in smaller models.



(a) **Llama 3.3 70B** (b) **Llama 3.1 8B** (c) **Llama 3.2 3B**

Figure 9: **Recipients' Response to: "I Intend to Eat Less Meat"**, obtained from simulations run with `Llama 3.3 70B`, `Llama 3.1 8B` and `Llama 3.2 3B`. Combined survey responses after each conversation round, where ◊ indicates the mean over all conversations per round. Compared to `Llama 3.3 70B`, we observe much greater variance in survey responses for simulations performed with smaller models. The intention of the Hard to Persuade Recipient stays more stable over time for the smaller models, while the intention of the Easy to Persuade Recipient shows a stronger increase in round 4, especially for `Llama 3.1 8B`.