Business Analytics - Group 10 report

Zeynep Guler, Francesco Iaccarino, Monika Kaczorowska, Ilia Koldyshev
Alfonso Gambardella, Claudia Frosi, Abhinav Pandey
Bocconi University

Introduction

Decision-making frameworks provide the foundation for navigating uncertainty and evaluating competing options in complex strategic environments. This paper explores the application of decision-making theories through a series of experiments and analyses, with a focus on how structured approaches refine strategies and improve outcomes.

The first part of the paper examines two case studies, Google and Meta, each facing pivotal strategic decisions. These cases illustrate how organizations use theories of value and experiments to navigate decision-making.

We conducted three analyses to test and expand decision-making theories. The first, using the data from case study simulations, examines whether testing lower-probability theories leads to greater success by fostering knowledge gains and better model refinement, aligning with insights on asymmetric information gain. The second, using data from the survey we created, explores how the framing of information and individual pre-existing trust levels influence adjustments in trust, revealing how biases and prior beliefs shape decision-making. The third analysis explores how external shocks and tailored information delivery influence participants' confidence in their theories, providing deeper insights into the dynamics of decision-making under uncertainty.

Lastly, we developed a taxonomy to classify attributes generated by Aristotle and proposed an approach to analyze patterns around how they are generated and what probabilities they are assigned.

Socrates - Play

1) Decision-Making on Google's Strategic Future

Eric Schmidt, the CEO of Google, evaluated two competing theoretical frameworks to determine the future direction of Google's strategy in the online video market.

The first, referred to as the "All-Media Ecosystem Theory," emphasized the creation of a comprehensive and integrated ecosystem encompassing video content, technological innovation, and platform synergies. This approach sought to ensure sustained, long-term growth by leveraging the interconnected nature of these elements.

In contrast, the "Synergy-Driven Accelerated Growth Theory" prioritized rapid expansion through the acquisition of complementary assets, utilizing YouTube's existing platform, audience base, and data analytics as the foundation for accelerated growth.

Theory 1: All-Media Ecosystem

Key Attributes and Initial Probabilities

The first key attribute of the All-Media Ecosystem Theory is "Technological Advancements," to which we assigned a high base probability of 80 percent. This reflects our assessment of the continuous and highly probable progression of innovation in the industry, driven by prevailing trends, substantial investments, and ongoing research. We determined that technological advancements are instrumental in driving growth in the online video sector, enhancing tools, platforms, and delivery methods. These enhancements increase the likelihood of online video growth by 20 percent. Additionally, technological progress plays a pivotal role in enabling integration by reducing technical barriers and providing essential tools, thereby raising the probability of successful integration by 60 percent.

The second key attribute, "Content Quality," has an assigned baseline probability of 50 percent. This value reflects an average level of user satisfaction with video content, which we believe hinges on the contributions of content creators and the platform's features. High-quality content is fundamental in attracting and retaining users, a dynamic that marginally increases the probability of online video growth by an additional 10 percent.

The "Online Video Growth" attribute is particularly significant, with a high base probability of 70 percent. We assigned this probability to capture YouTube's established dominance and the rapid adoption of video content by users. However, we also recognized that such rapid growth in the online video sector may provoke competitive responses from market rivals, increasing the likelihood of competitor actions by 40 percent. At the same time, the success of online video platforms directly contributes to achieving the broader goal of an integrated media ecosystem, boosting its probability by 45 percent.

We assigned the attribute of "Integration" a base probability of 20 percent, reflecting the challenges posed by technical and organizational complexities. While this is relatively low, we acknowledge that technological advancements and strategic alignment can significantly improve integration outcomes. Given the critical role integration plays in achieving a cohesive media ecosystem, we estimated that it increases the probability of realizing this goal by 45 percent.

The attribute of "Competitor Response" carries a very low base probability of 5 percent. This reflects our assessment that competitors are unlikely to significantly hinder Google's growth trajectory. However, we included this attribute to account for the minor negative impact competitors' actions could have on the development of an all-media ecosystem, reducing its likelihood by 10 percent.

Finally, the ultimate goal of the All-Media Ecosystem Theory is the establishment of a fully integrated media ecosystem. We assigned this state a base probability of 10 percent, reflecting its ambitious and complex nature and its low probability of success independent of previous attributes. This lower probability underscores the inherent difficulty in achieving a fully integrated media ecosystem and highlights the importance of interconnected attributes in driving success.

Analysis of Probabilities for the All-Media Ecosystem Theory

The All-Media Ecosystem Theory presents an Expected Probability of the Theory ($V\Theta$) at 82%, while the Subjective Confidence in the Theory (ω) we assigned is 60% to account for the balance between its promising potential and the inherent uncertainties in its implementation. Additionally, the Expected Probability if the Theory is False ($V\Theta\sim$) is set at 50%. Using the Bayesian framework outlined in Professor Gambardella's paper (Camuffo et al.), the updated probability is calculated using the formula:

$$V = \omega \cdot V\Theta + (1 - \omega) \cdot V\Theta\sim$$

Based on our assigned values, the Expected Probability of Success for the All-Media Ecosystem Theory is updated to approximately 69 percent.

Theory 2: Synergy-Driven Accelerated Growth Theory

Key Attributes and Initial Probabilities

The first critical attribute of the Synergy-Driven Accelerated Growth Theory is "Original Content Creation," to which we assigned a base probability of 30 percent. This reflects the inherent challenges of consistently generating engaging and compelling content. While original content plays a role in retaining users and attracting advertisers, its direct impact on platform expansion is relatively modest, contributing only a 5 percent probability boost to accelerated growth because original content serves as a supplementary, rather than primary, driver of growth.

The second attribute, "Search Integration," carries a base probability of 40 percent, reflecting Google's expertise in search algorithms and the technical feasibility of integrating YouTube's functionalities within its broader ecosystem. We determined that embedding YouTube into Google's search results streamlines user access to video content, directly increasing the platform's growth probability by 10 percent.

The third attribute, "Content Growth," is assigned a base probability of 50 percent. This reflects YouTube's established position as a leading content platform, supported by its extensive and diverse creator base. Although content growth's role in platform acceleration is indirect, it enhances synergies such as search integration and data analytics, offering a modest 5 percent increase in growth probability. Moreover, content growth significantly influences creators' motivation to produce original content, providing an additional 40 percent probability boost to the overall framework.

"Competitive Edge" is a pivotal attribute in this theory, with a high base probability of 80 percent. We based this probability on YouTube's strong market position, user base, and brand reputation, which enable effective data collection and utilization. This competitive advantage directly improves targeting and user insights, increasing data synergies by 40 percent. Furthermore, it amplifies discoverability and engagement, boosting the likelihood of successful search integration by another 40 percent.

The attribute of "Cultural Integration" has a base probability of 60 percent, reflecting Google's extensive experience in expanding to the world and fostering international collaborative team environments. We determined that cultural alignment reduces operational friction and decision-making delays, contributing a 10 percent boost to the platform's growth probability.

Finally, "User Data Synergies" is assigned a base probability of 50 percent, emphasizing the importance of data analytics in enhancing user engagement and optimizing monetization strategies. Leveraging these insights significantly accelerates platform growth, contributing a 20 percent probability boost.

The ultimate goal of the Synergy-Driven Accelerated Growth Theory is "Accelerated Growth," with a base probability of 5 percent. This low probability underscores the ambitious and interdependent nature of this objective, which relies on the combined success of all contributing attributes and is nearly impossible without the success of previously mentioned attributes.

Analysis of Probabilities for the Synergy-Driven Accelerated Growth Theory

The Synergy-Driven Accelerated Growth Theory presents an Expected Probability of the Theory (VΘ) at 72%, while the Subjective Confidence in the Theory (ω), which we assigned at 70%, reflects its well-supported assumptions, strategic alignment with YouTube's strengths, and tangible potential for accelerated platform expansion through synergies. Additionally, the Expected Probability if the Theory is False (VΘ~) is set at 50%.

Using the Bayesian framework previously articulated in Professor Gambardella's seminal paper, the Expected Probability of Success for the Synergy-Driven Accelerated Growth Theory is updated to approximately 65 percent.

Experiment Ran and Update of Probabilities

To determine which strategic theory to test, we conducted a detailed evaluation of the expected probabilities associated with both options. Based on initial calculations, the All-Media Ecosystem Theory, labeled as H1, exhibited a higher expected probability of success at 69 percent compared to the 65 percent attributed to the Synergy-Driven Accelerated Growth Theory, referred to as H2. Consequently, we prioritized the All-Media Ecosystem Theory for experimental validation.

The experiment subsequently provided additional data, leading to a significant update in the expected probability of H1. Following the experiment, the probability associated with the All-Media Ecosystem Theory increased from 69 percent to 81 percent. This marked improvement prompted us to reassess our causal beliefs regarding the factors influencing the theory's success. Specifically, we adjusted the Competitor Response impact parameter, reducing its negative contribution from -10 percent to -5 percent because the experiment's results strengthened the belief in the All-Media Ecosystem Theory, suggesting that advancements in video streaming and content recommendation algorithms could mitigate the negative impact of competitor responses.

Post-Experiment Calculations

To update the probabilities associated with H1, we utilized the Bayesian framework previously described. For the All-Media Ecosystem Theory, the updated probability (VΘ) of 84 percent reflects the revised probability after incorporating the experiment results into our model. The subjective confidence (ω), to which we assigned 60 percent, accounts for the balance between the theory's potential and inherent uncertainties. Additionally, the probability of success if the theory were false (VΘ~) was estimated at 50 percent. Using this framework, the final updated probability for the All-Media Ecosystem Theory was calculated to be approximately 70 percent. This update aligns with the experimental results, which provided a stronger basis for confidence in the theory's viability.

Conclusion and Decisions

The post-experiment analysis confirmed that the All-Media Ecosystem Theory was the most viable strategic option for Google in the online video market. We chose this theory because it offered the highest potential to achieve a fully integrated media ecosystem by leveraging advancements in technology, content innovation, and platform synergies. Unlike alternative strategies, this theory prioritized sustainable growth through integration and innovation, addressing market challenges and securing long-term competitive advantages. The post-experiment updated probabilities reinforced confidence in the feasibility of the theory, while the adjustment of parameters, such as reducing the negative impact of Competitor Response, reflected a refined understanding of how internal and external factors interact to shape the success of the theory.

2) Decision-Making on Meta's Metaverse Strategy

Meta's strategic approach to shaping the metaverse revolves around two prominent theories. The first, the "Immersive Social Network Theory," prioritizes the development of augmented reality (AR) and virtual reality (VR) devices to create an expansive network of avatars, thereby enhancing user interactions within immersive digital environments. Conversely, the "Creative Economy Theory" is centered on the establishment of a decentralized digital economy powered by blockchain and artificial intelligence (AI), allowing avatars and content creators to trade goods and services seamlessly.

Theory 1: Immersive Social Network

Key Attributes and Initial Probabilities

The cornerstone of the Immersive Social Network Theory is "Investment in AR/VR," to which we assigned a high base probability of 90 percent. This reflects Meta's steadfast commitment to AR/VR technologies as the foundation of its metaverse strategy. Such investment signals strong organizational focus and prioritization, which we determined significantly increases the likelihood of widespread AR/VR adoption, contributing a 40 percent boost to AR/VR usage.

"Competition" is another pivotal attribute, with the base probability assigned at 80 percent. This reflects our assessment of the high likelihood that competitors will also invest in AR/VR technologies to capture a share of the emerging metaverse market. Competitor investments are expected to positively influence overall market adoption, increasing the likelihood of AR/VR growth by 30 percent.

The attribute of "AR/VR Usage" is assigned a base probability of 30 percent. This reflects the conditional nature of widespread AR/VR adoption, which depends heavily on continued investments and competitive forces. Greater adoption is directly correlated with a 60 percent increase in active avatars within the metaverse, highlighting the centrality of AR/VR usage in shaping the digital ecosystem.

"AI Technology" plays a critical role in this theory, with a base probability of 30 percent. This probability reflects our evaluation of the potential for significant advancements in AI technologies to enhance avatar functionality and improve the user experience. These innovations are projected to increase avatar adoption by 25 percent, illustrating the importance of cutting-edge technology in driving engagement and participation within the metaverse.

The attribute of "Data Security" is assigned a moderate base probability of 65 percent. This reflects the reliance of the metaverse on user trust and the critical role of robust security measures in maintaining that trust. Strong data security frameworks are expected to build confidence among users, contributing to a 10 percent increase in avatar adoption.

The ultimate goal of the Immersive Social Network Theory is represented by the "Number of Avatars," which is assigned a base probability of 5 percent. Success in this area depends on the effective interplay of investments, competition, technological advancements, and data security attributes happening. This low baseline probability reflects the ambitious nature of achieving widespread avatar adoption without the previous attributes.

Analysis of Probabilities for the Immersive Social Network Theory

The Immersive Social Network Theory has an Expected Probability of Success (VΘ) of 73% based on our model. The Subjective Confidence (ω), assigned at 75%, represents a strong belief in the reliability of the theory's assumptions, given Meta's substantial investment in AR/VR and its alignment with the company's strategic priorities. The probability of success if the theory were false (VΘ~) was set at 18 percent, highlighting the low baseline likelihood of success without these attributes.

Using the Bayesian framework to integrate these inputs, the Expected Probability of Success for the Immersive Social Network Theory was calculated to be approximately 59 percent. This probability combines subjective confidence with alternative scenarios, offering a nuanced evaluation of the theory's feasibility and potential impact.

Theory 2: Creative Economy

Key Attributes and Initial Probabilities

The first of these is "Blockchain," for which we assigned a high base probability of 65 percent. This decision reflects the widespread acknowledgment of blockchain technology as a fundamental enabler of virtual economies. Blockchain facilitates secure and efficient transactions, which we determined would contribute a 10 percent increase in the overall metaverse GDP. Additionally, we concluded that blockchain's decentralized infrastructure significantly enhances user and creator engagement, adding a 15 percent boost to engagement levels to account for blockchain's central role in building trust and fostering activity in the digital economy in our model.

The second attribute, "AI Technologies," was assigned an even higher base probability of 80 percent. We based this on the strong confidence in ongoing advancements and widespread adoption of artificial intelligence within the metaverse. We reasoned that AI technologies would play a critical role in improving user experiences and optimizing creator tools, leading to a 25 percent increase in engagement among users and creators.

We assigned "Engagement" a relatively low baseline probability of 20 percent, reflecting the challenges inherent in achieving high levels of user and creator participation despite enabling technologies. However, we determined that engagement has an outsized impact on the metaverse economy, with increased participation positively influencing economic activity and contributing a 40 percent boost to the overall metaverse GDP.

Finally, we assigned "Metaverse GDP" a modest base probability of 5 percent. This decision reflects the ambitious and interconnected nature of achieving significant economic growth in the metaverse. We reasoned that this outcome is highly dependent on upstream factors, such as engagement, blockchain adoption, and AI advancements; therefore, without the success of those attributes, the success of "Metaverse GDP" is unlikely to happen.

Analysis of Probabilities for the Creative Economy Theory

The Creative Economy Theory initially presented an Expected Probability of Success (VΘ) of 58 percent, reflecting the potential of its foundational attributes in driving economic activity within the metaverse. The Subjective Confidence (ω), which we assigned at 61 percent, captures our balanced perspective on the theory, considering the promising synergies of blockchain and AI alongside the inherent challenges of engagement. The probability of success if the theory were false (VΘ~) was set at 44 percent, reflecting a moderately favorable baseline in the absence of full attribute alignment.

Using the Bayesian framework previously described, the Expected Probability of Success for the Creative Economy Theory was updated to approximately 53 percent.

Experimentation and Probability Updates

To determine which theory to test, we conducted an evaluation of their expected probabilities. Initially, the Immersive Social Network Theory (H1) appeared more favorable, with an Expected Probability of Success of 59 percent, compared to 53 percent for the Creative Economy Theory (H2). While H1 focused on enhancing user interactions through AR/VR devices, H2 offered a decentralized and scalable framework better suited to supporting a robust and adaptive metaverse economy. However, despite H2's lower initial probability, we ultimately chose to experiment with it. This decision was driven by the strategic importance of blockchain and AI technologies, which are central to the long-term economic sustainability of the metaverse.

Post-Experiment Calculations

Following the experiment, Meta updated the probabilities for the Creative Economy Theory (H2) using the Bayesian framework. The initial probability (VΘ) was at 58 percent, and the subjective confidence (ω) was revised upwards to 71 percent, reflecting greater confidence in the viability of H2 based on experimental insights. The probability of success if the theory were false (VΘ~) remained consistent at 44 percent. Using this framework, the updated Expected Probability of Success for H2 was calculated to be approximately 54 percent. This update incorporates the revised confidence level, highlighting the theory's strengthened position after experimentation and aligning with its potential to build a sustainable economic foundation for the metaverse.

Conclusion and Decisions

After analyzing the experimental results, we concluded that the Creative Economy Theory (H2) presented a more compelling path forward based on the updated probabilities provided by the experiment and the robustness models created. The experiment supported our model's success, which has blockchain and AI technologies as key enablers of economic scalability and long-term sustainability.

Socrates - Data

1) Background

We analyzed 388 observations where participants ran 1 or 2 experiments, excluding cases with an expected probability of success equal to 0. The observations came from students running different simulations on the Bocconi simulation platform named “Socrates”. The platform presented a historical scenario with key attributes and asked the participant to come up with probabilities of each attribute and their causal relationships, contributing to a final probability for each theory. After coming up with initial theories, the users had a chance to run an experiment to get additional information on their theories. Ultimately, participants were asked to decide which of these two theories was closest to reality. Based on that final decision, the user won the game if they chose the correct theory.

2) Theory

The research question: *Does running an experiment on the theory with a lower probability increase the probability of winning?*

Our theory: *Running an experiment on a theory with a lower probability increases the participant's chances of success.*

The reasoning behind our theory is that testing a lower-probability theory leads to greater knowledge gains for the individual. Theories with lower probabilities are associated with lower confidence and success expectations, which depend on the given attributes in the simulation and the participant’s probability assignments. Since participants cannot add additional attributes to the model, a lower final probability reflects both the perceived incompleteness of the given attributes and the participant’s confidence in their own model. This higher lack of robustness and confidence in lower-probability theories creates greater opportunities for learning and updating compared to higher-probability theories.

Gambardella emphasizes the asymmetric information gain of less probable theories: they provide higher upside potential while maintaining comparable downside risk, underlining that researchers can uncover significant deviations from established knowledge by exploring high-uncertainty theories. This aligns with our theory that *running an experiment on a theory with a lower probability increases the participant's chances of success* because of enhanced knowledge updates, which enables participants to better evaluate and refine their models, ultimately increasing their likelihood of selecting the correct theory and winning the simulation.

3) Model

$$\mathrm{Prob(win)} = \frac{\exp(\beta_0 + \beta_1 \mathrm{Less\_certain\_exp} + \gamma X)}{1 + \exp(\beta_0 + \beta_1 \mathrm{Less\_certain\_exp} + \gamma X)}$$

We ran a logistic regression with different sets of variables related to the probability of winning and to the decision to run an experiment on a lower probability theory.

A) Interpreting the variables in the model:

a. Dependent variable

- Win: a binary outcome (1 = the game is won, 0 = not won)
- Prob(win): probability of winning the game, probability of win=1.

b. Independent variable

Less_certain_exp: binary variable: 1 = if the player experimented on the theory with a lower expected probability; 0 = otherwise)

c. Controls

Grouped into three different categories for inclusion depending on both the relevance and the level of balance (See Appendix 2.3).

1. *Confidence and Expectations:* Theory 1 - Expected Probability, Theory 2 - Expected Probability represent the expected probabilities of both theories based on the model success and the confidence. Theory 1-Omega, and Theory 2 - Omega capture participants' confidence. The expected probability of theory 1 and Omega of theory 2 showed unbalances above the threshold. Controlling for these variables accounts for the influence of perceived likelihood and trust in the theories on decision-making and outcomes.
2. *Robustness and Effort:* Theory 1 - Density and Theory 2 - Density measure the interconnectedness and sophistication of the theories. Density of theory 2 showed above the threshold unbalances. These variables ensure that variations in the structural complexity of the models do not confound the relationship between Less_certain_exp and the probability of winning. Minutes_Spent reflects the time participants spent on each case, serving as a proxy for their engagement and effort. This control ensures that differences in participant commitment do not bias the results.
3. *Conditioning by the game played:* The difference in the game played may suggest different outcomes depending on the content and complexity of the game. Therefore games played more than 10 times were added as controls in the final model, as winning particular games could have been more probable than winning others.

B) Hypothesis:

We theorized that running an experiment on a theory with a lower probability increases the participant's chances of success. Therefore, we are theorizing upon β_1 being greater than 0, positively affecting the probability of winning.

Null Hypothesis (H_0): $\beta_1 = 0$, indicating that running an experiment on the less certain theory (Less_certain_exp = 1) does not affect the probability of success.

Alternative Hypothesis (H_1): $\beta_1 > 0$, indicating that running an experiment on the less certain theory (Less_certain_exp = 1) increases the probability of success.

If we reject the null, we can prove that $\beta_1 > 0$ with a one-sided t-test.

4) Analysis

A) Data Manipulation

I. *Row Filtering:*

Observations with an expected probability of success equal to 0 were removed. Observations with zero probability of success represent participants who were completely disengaged or uninterested in the experiment. Their zero probability suggests they fundamentally did not care about the outcome, leading them to exit or not participate meaningfully.

II. *New variable: Density*

In order to include information about attributes and links and avoid multicollinearity, we introduce a new variable computed from the number of attributes and number of links - density (Appendix 2.0).

B) Descriptive Analysis

The descriptive analysis provides insights into the dataset used to test whether experimenting on theories with lower probabilities increases participants' chances of success. Participants' expected probabilities for Theory 1 and Theory 2 both have a mean of 0.38, with a standard deviation of 0.20, suggesting moderate confidence levels but notable variability. These probabilities reflect participants' initial beliefs about success and are central to the hypothesis, as lower probabilities indicate less robust models with greater learning potential. Including these variables captures how initial expectations influence decision-making and outcomes.

The density of Theory 1 and Theory 2, with mean values of 0.26 and 0.25, respectively, measures the interconnectedness of attributes within the theories. The relatively low values suggest participants worked with simpler models overall, though the range (0.12 to 0.50) shows variation. Density is included to assess whether the complexity of a theory impacts both success and the decision to experiment, as more interconnected models may reflect greater robustness. Confidence in the theories, measured by Omega, has a mean of 0.60 for both theories. This variable represents participants' trust in their models and decisions. By including Omega, the analysis accounts for the aspect of confidence in shaping participants' choices. Participants spent an average of 34.52 minutes on each case, with a wide range (2 to 121 minutes). Time spent is included as a proxy for effort, controlling for its potential effect on outcomes. With 40% of participants choosing to experiment on the less certain theory, this variable directly tests whether such a decision improves success rates, isolating its effect on winning.

We performed a balance check on the covariates to ensure all are equally balanced (see Appendix 2.2 for full results). We also performed a VIF Check to confirm there is no multicollinearity (Appendix 2.3). We analyzed the correlation matrix (Appendix 2.4) and found that Theory 1 and Theory 2 were highly correlated across expected probabilities, densities, and omegas. This indicates that the two theories are structurally similar and do not represent distinct approaches. If the theories are closely aligned, their ability to represent contrasting levels of uncertainty diminishes, which could limit the potential to observe differences in success rates or uncover the transformative effects of testing unconventional ideas.

C) Regression Results

5 different models were constructed, adding variables step by step into regression as seen in Table 1 below:

1. Model 1: Baseline with only independent variable
2. Model 2: Model 1 + first group of controls
3. Model 3: Model 2 + second group of controls
4. Model 4: Model 3 + unbalanced case dummy variables
5. Model 5: Model 4 + all the rest of the selected case dummy variables

Table 1: Regression results for Socrates analysis

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
const	0.8864***	1.4007***	0.8913	0.8940	1.7124**
	(0.1441)	(0.4938)	(0.6421)	(0.6615)	(0.8193)
Less_certain_exp	-0.0237	-0.0254	-0.0515	-0.0321	-0.1450
	(0.2273)	(0.2314)	(0.2345)	(0.2413)	(0.2530)
Theory 1 - Expected Probability		-1.2167*	-1.2189*	-0.9221	-1.2169
		(0.6815)	(0.6957)	(0.7230)	(0.7729)
Theory 1 - omega		0.2822	0.3379	0.3334	0.4040
		(0.6654)	(0.6671)	(0.6842)	(0.7119)
Theory 2 - Expected Probability		0.7813	0.7444	0.6185	0.6983
		(0.7091)	(0.7158)	(0.7449)	(0.8005)
Theory 2 - omega		-0.8400	-0.8385	-0.7743	-1.0651
		(0.6099)	(0.6136)	(0.6350)	(0.6813)
Theory 1 - density			0.4059	0.2266	-0.6458
			(1.6005)	(1.6673)	(2.0110)
Theory 2 - density			1.0258	1.7950	2.0594
			(1.3462)	(1.4320)	(1.5951)
minutes_spent			0.0042	0.0015	0.0027
			(0.0051)	(0.0052)	(0.0059)
Luxottica				-0.9242***	-1.3544***
				(0.3433)	(0.4624)
Red Bull				-2.6682**	-2.9995**
				(1.1233)	(1.1701)
Lego					-1.2935**
					(0.6586)
Other case name covariates	NO	NO	NO	NO	YES
AIC	473.88	476.87	481.11	470.57	484.81
BIC	481.80	500.64	516.76	514.14	579.88

Standard errors in parentheses | * p<.1, ** p<.05, ***p<.01
Other case name covariates: Spotify, Booking.com, Netflix, Google, Apple, Amazon Web Services,Marvel, Porsche, OpenAI, Adobe, SpaceX, PayPal.

The results showed no statistically significant relationship between experimenting with less certain theories and improved game outcomes. This finding led to not rejecting of null hypothesis (H_0), as there is no evidence to suggest that testing less certain ideas had any measurable impact on success. The lack of differentiation between the theories used in the experiment (shown in the correlation matrix Appendix 2.4) likely limited the ability to observe such an effect. The inability to reject the null hypothesis reflects not only the similarity between the tested theories but also raises questions about whether the platform is able to capture the dynamics of uncertainty and discovery reflected in the theory.

An interesting outcome of the analysis was the effect of case-specific variables, such as the inclusion of dummies for cases like Luxottica. The dummy variable for Luxottica revealed a significant negative coefficient (-1.3544***), suggesting that case-specific factors might play a critical role in shaping outcomes usually by having a negative effect on them. This finding shows that some cases could have been much harder for the participants than others because of the negative effect of such cases on the probability of success.

Experiment - Data

1) Theory and Survey Design

A) Theory

In our experiment, we wanted to explore information bias, specifically how people adjust their trust after being exposed to new information. We developed 2 central theories:

- *Theory 1: The framing of information about AI privacy—neutral, positive, or negative—affects an individual's change in trust in AI after being exposed to new information.*
- *Theory 2: Individuals with varying levels of pre-existing trust in AI adjust their trust differently in response to the same content of information they received.*

Theory 1 tests for information bias, which occurs when the framing of information alters the perceived importance or reliability of the content, independent of its factual accuracy. In the context of this study:

- Neutral Framing (Baseline): Messages presenting technical mechanisms without emotional tone serve as a baseline, expected to elicit the least biased reactions.
- Positive Framing: Messages emphasizing advancements and benefits may enhance trust by highlighting optimism and progress.
- Negative Framing: Messages focusing on risks and failures can reduce trust by invoking fear and skepticism.

Theory 2 looks at whether the information tone effect is mitigated in any way by the pre-treatment trust level. The interaction between information framing and pre-treatment trust highlights how prior beliefs can moderate the impact of framed information. Rather than assuming uniform effects across all individuals, Theory 2 accounts for variability in responses based on baseline trust levels.

B) Survey Design

We designed a survey in Qualtrics to collect the necessary information to test our theory. The survey collected information including: baseline levels of trust and confidence in their decision on trust, demographic questions, and other data points that will serve as covariates in our analysis, including AI usage frequency, AI knowledge rating, digital privacy concern, and tech adoption openness. After answering these questions, the respondents received a randomized treatment: fake news about AI privacy written in 3 framings- neutral, positive, and negative. After being shown this information, they were asked again to rate their trust in AI privacy and their confidence in this decision. (See Appendix 3.1 for the flow chart of the survey and the content of fake news used as treatment.) This pre-post design was chosen to capture the immediate impact of information framing on trust and confidence. By comparing baseline (pre-treatment) and follow-up (post-treatment) measures, we could isolate changes attributable specifically to the treatments.

2) Model

Theory 1 Model:
$\Delta Trust = \beta_0 + \beta_1 \cdot$ Positive Treatment + $\beta_2 \cdot$ Negative Treatment + $\beta_3 \cdot$ Medium Pre Trust + $\beta_4 \cdot$ High Pre Trust + $\beta_5 \cdot$ Controls + ϵ

Theory 2 Model:
$\Delta Trust = \beta_0 + \beta_1 \cdot$ Positive Treatment + $\beta_2 \cdot$ Negative Treatment + $\beta_3 \cdot$ Medium Pre Trust + $\beta_4 \cdot$ High Pre Trust + $\beta_5 \cdot$ Positive Treatment * Medium Pre Trust + $\beta_6 \cdot$ Positive Treatment * High Pre Trust + $\beta_7 \cdot$ Negative Treatment * Medium Pre Trust + $\beta_8 \cdot$ Negative Treatment * High Pre Trust + $\beta_9 \cdot$ Controls + ϵ

A) Interpreting Variables in the model:

a. Dependent Variable:

Change in Trust (ΔTrust)
This variable captures the difference between post-treatment trust and pre-treatment trust levels, as measured by participants' responses to the question, "How much do you trust AI with your private information?" on a scale from 0 (no trust at all) to 10 (complete trust). Positive values indicate an increase in the change of trust after exposure to the treatment, while negative values indicate a decrease in the change of trust.

b. Independent Variables:

1. *Treatment Segments* (Intervention)
   A categorical variable indicating the treatment group to which participants were assigned categorizing the three types of exogenous shock:
   - Positive: Exposure to a positively framed message highlighting advancements in AI privacy.
   - Negative: Exposure to a negatively framed message emphasizing concerns and failures in AI privacy.
   - Neutral: Exposure to a neutral message explaining mechanisms of AI privacy without emotional framing.

2. *Pre-Treatment Trust Segment (used as a control in model 1)*
   Similar to confidence, this variable categorizes participants based on their pre-treatment trust in AI privacy decisions using the same thresholds:
   - Low Trust (score: 0–3)
   - Medium Trust (score: 4–6)
   - High Trust (score: 7–10)

3. *Interaction Terms for Treatment and Pre-treatment Trust Segments (only Theory 2 Model):*
   The interaction terms between treatment and pre-treatment trust were included to explicitly test H2, which suspects that baseline trust moderates the effect of framing. The coefficients of these

terms(β_5, β_6, β_7, and β_8) allow us to assess whether individuals with high pre-treatment trust respond differently to positive or negative framing than those with low or medium trust.

c. Control Variables:

We used the following controls we extracted from our survey to account for individual differences influencing the trust change beyond the treatment effects.

1. *AI Usage Frequency* - A continuous variable capturing how frequently participants use AI, based on their responses to "How often do you use AI?" on a scale from 0 (never) to 10 (every day).
2. *AI Knowledge Rating* - A continuous variable measuring self-assessed knowledge of AI technologies, derived from responses to "How would you rate your knowledge about AI?" on a scale from 0 (not knowledgeable at all) to 10 (expert knowledge).
3. *Digital Privacy Concern* - A continuous variable reflecting participants' concerns about digital privacy, based on responses to "How concerned are you about your digital privacy?" on a scale from 0 (not concerned at all) to 10 (very concerned).
4. *Tech Adoption Openness* - A continuous variable measuring openness to adopting new technologies, based on responses to "How open are you to adopting new technology?" on a scale from 0 (not open at all) to 10 (very open).
5. *Pre-Treatment Confidence Segment* - A categorical variable dividing participants into three segments based on their pre-treatment confidence levels:
    a. Low Confidence (score: 0–3)
    b. Medium Confidence (score: 4–6)
    c. High Confidence (score: 7–10)

We performed ANOVA checks for each control variable to find any unbalanced ones. From the ANOVAs, we found that Tech Adoption Openness is unbalanced among treatment groups with a p-value of 0.034. To address this, we made sure to include this covariate in all of our models. (See Appendix 3.2 for visualizations of these checks.)

B) Hypotheses

Hypothesis 1:

$\Delta\text{Trust} = \beta_0 + \beta_1 \cdot \text{Positive Treatment} + \beta_2 \cdot \text{Negative Treatment} + \beta_3 \cdot \text{Medium Pre Trust} + \beta_4 \cdot \text{High Pre Trust} + \beta_5 \cdot \text{Controls} + \epsilon$

The tone of information about AI privacy is theorized to influence trust in different ways:

- Positive framing: Expected to increase trust by emphasizing advancements.
- Negative framing: Predicted to decrease trust by highlighting risks.
- Neutral framing: Anticipated to have minimal or balanced effects, serving as a baseline for comparison.

Null Hypothesis (H_0): $0 = \beta_1 = \beta_2$

Positive and negative framing of information does not affect the update of trust differently than the neutral tone (baseline).

Alternative Hypothesis (H_1): $\beta_1 \neq 0$ or $\beta_2 \neq 0$
Positive or negative framing affects trust differently compared to the neutral tone.

Hypothesis 2:
ΔTrust = $\beta_0 + \beta_1 \cdot$ Positive Treatment + $\beta_2 \cdot$ Negative Treatment + $\beta_3 \cdot$ Medium Pre Trust + $\beta_4 \cdot$ High Pre Trust + $\beta_5 \cdot$ Positive Treatment * Medium Pre Trust + $\beta_6 \cdot$ Positive Treatment * High Pre Trust + $\beta_7 \cdot$ Negative Treatment * Medium Pre Trust + $\beta_8 \cdot$ Negative Treatment * High Pre Trust + $\beta_9 \cdot$ Controls + ϵ

Pre-existing trust affects responses to treatment: Individuals with varying levels of pre-existing trust in AI adjust their trust differently in response to the framing of information. This implies that the same tone of information can have differing effects on individuals based on their initial levels of trust.

Null Hypothesis (H_0): $0 = \beta_5 = \beta_6 = \beta_7 = \beta_8$
The treatment effect of the same framing is consistent across all levels of pre-treatment trust.

Alternative Hypothesis (H_1): $\beta_5 \neq 0$ or $\beta_6 \neq 0$ or $\beta_7 \neq 0$ or $\beta_8 \neq 0$
The treatment effect of the same framing is heterogeneous - different for individuals of different pre-treatment trusts.

3) Analysis

A) Data set cleaning

We collected a total of 107 responses on our Qualtrics survey. The data set included demographic information, baseline questions, pre-treatment trust and confidence levels, post-treatment trust and confidence levels, and treatment assignment.
We began the analysis by cleaning the data. This included removing any responses with unanswered questions, as these responses can't be evaluated. The removed rows were evenly balanced among treatment groups. The data cleaning resulted in a total of 82 responses evenly balanced across treatment groups (27, 26, and 29 for neutral, negative, and positive, respectively).

B) Descriptive Analysis

Before running regressions, we observed the following change in trust among groups. Treatment-Neutral increase in change in trust of 0.667, Treatment-Negative decrease in change in trust of -0.346, Treatment-Positive increase in change in trust of 0.655. This shows us that, on average, both neutral and positive framing of information led users to trust AI privacy more, while negative tone, on average, decreased their trust. This makes sense based on the experiment design. (See Appendix 3.3 for more in-depth results and visualizations)

C) Regression Results

6 different models were constructed, adding variables step by step into regression, as seen in Table 2 below:

1. Model 1 = Baseline model: Treatment Segments' effect on Trust Change
2. Model 2 = Model 1 + Pre-Treatment Trust Segments

3. Model 3 = Model 2 + Tech Adoption Openness (unbalanced covariate)
4. Model 4 = Model 3 + remaining covariates
5. Model 5 = Model 3 + interaction effect between Treatment Randomizer and Pre-Treatment Trust Segment
6. Model 6 = Model 5 + all remaining covariates

Table 2: Regression results from survey analysis

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Intercept	0.6667	1.1770	0.8090	0.7475	1.0879	1.1648
	(0.3690)	(0.4302)**	(0.7440)	(0.8016)	(0.7574)	(0.8012)
Positive Treatment	-0.0115	-0.1704	-0.2322	-0.1215	0.0367	0.0364
	(0.5128)	(0.4927)	(0.5050)	(0.5029)	(0.8160)	(0.8003)
Negative Treatment	-1.0128	-0.8044	-0.8888	-0.8949	-2.1187	-2.5261
	(0.5268)	(0.5051)	(0.5258)	(0.5323)	(0.8723)*	(0.8667)**
Medium Trust		-0.3301	-0.3908	-0.4187	-0.5030	-0.6788
		(0.4581)	(0.4707)	(0.4862)	(0.8302)	(0.8193)
High Trust		-1.7464	-1.8230	-1.9669**	-2.6714	-3.2591
		(0.5548)**	(0.5712)***	(0.6117)	(0.9523)**	(0.9602)**
Tech Adoption Openness			0.0605	0.1501	0.0532	0.1935
			(0.0995)	(0.1280)	(0.1020)	(0.1305)
Positive Treatment: Medium Trust					-0.5629	-0.4767
					(1.1014)	(1.0922)
Positive Treatment: High Trust					-0.5278	-0.0622
					(1.5146)	(1.5454)
Negative Treatment: Medium Trust					1.3066	1.6746
					(1.1899)	(1.1605)
Negative Treatment: High Trust					2.5743	3.7729
					(1.2747)*	(1.3748)**
AI Usage				-0.2665*		-0.3124
				(0.1089)		(0.1064)**
AI Knowledge				0.1330		0.0906
				(0.1072)		(0.1083)
Privacy Concern				0.0454		0.0817
				(0.0896)		(0.0876)
Medium Confidence				0.4709		0.7253
				(0.6646)		(0.6517)
High Confidence				0.5282		0.2719
				(0.6924)		(0.6844)
R-squared	0.0584	0.1704	0.1744	0.2552	0.2444	0.3583
Adjusted R-squared	0.0345	0.1273	0.1201	0.1503	0.1500	0.2242

Theory 1:

Model 4 incorporates all covariates to evaluate the effects of positive and negative treatments on trust as posed in Theory 1. In this model, none of the treatments are statistically significant. Thus, the results do not provide evidence to reject the null hypothesis of Theory 1 (H0: $\beta1 = \beta2 = 0$). Consequently, there is no support in Model 4 for the claim that either positive or negative framing significantly influences trust in AI privacy compared to neutral framing. These findings suggest that, when accounting for additional covariates, the framing of information does not independently explain changes in trust. This means that the additional information (treatment) affected everyone in the same way regardless of the framing of information (treatment type), showing a homogeneous treatment effect.

However, the significant coefficient for High Trust (-1.7464, $p < 0.05$**)** is worth noting. This finding suggests that participants with high baseline trust levels experience a significant decrease in their change in trust after the treatment independent of the framing, compared to the baseline (low trust individuals shown neutral information). This aligns with theoretical considerations that individuals with high trust may have limited room for upward adjustments and are instead more likely to revise their trust downward when presented with new information.

Theory 2:

Model 6 extends the analysis by including interaction terms to explore how pre-treatment trust moderates the effects of treatments, as predicted in Theory 2. The coefficient for Negative Treatment (−2.5261, $p < 0.05$) is statistically significant, indicating that negative framing decreases trust for the baseline trust level (low trust). This supports the premise of Theory 1, which predicted that negatively framed information would reduce trust compared to the neutral tone. Thus, the negative treatment effect seen here provides partial evidence supporting Theory 1, even within the context of Theory 2.

Moreover, the interaction between Negative Treatment and High Pre - Treatment Trust (3.7729, $p < 0.05$) is also statistically significant. This finding suggests that individuals with high baseline trust experience a substantial positive adjustment to their change in trust when exposed to negative framing, proving against the overall homogeneous negative effect of the tone of the treatment. Therefore, we reject the null, H0, that the treatment effect of the same framings is consistent across all levels of pre-treatment trust. This means that depending on the pre-trust levels, the effect of the same treatment framing on the change of trust varies.

This variability underscores how individuals with different baseline trust interpret negative information differently. High-trust individuals may resist extreme negativity, perceiving it in a more positive light. This finding aligns with the confirmation bias framework, suggesting that these participants reinterpret overly negative messages that conflict with their existing beliefs.

Additionally, AI Usage (−0.3124, $p < 0.05$) is significant in Model 6. This finding indicates that frequent users of AI are more likely to experience a decrease in trust, potentially due to heightened awareness of AI-related risks or greater sensitivity to information about AI privacy concerns.

Aristotle/Plato - Data

1) Background:

The dataset comes from the Aristotle simulation, where individuals first created a theory using given information, optionally assisted by a general-purpose chatbot. In the second step, they were told to use a specialized chatbot, Aristotle; however, some did not. Participants were divided into four groups: 1) no treatment/no shock, 2) no treatment/shock, 3) treatment/no shock, and 4) treatment/shock. After each step, participants rated their theory's probability of success and their confidence in their assigned probability.

2) Theory:

Aristotle's impact on individuals' change in confidence in their theories is more favorable (greater confidence increase or smaller confidence decrease, or changing from decrease to increase) under shock conditions than under no-shock conditions.

In our theory, confidence is used as the measure of success because it captures a holistic view of both theory validity and trust in the process. Unlike the probability of success, which is constrained by limited information, confidence reflects participants' overall perception of the process. Even when the likelihood of success of the case worked on is low, individuals may still rate their confidence highly if they find the process comprehensive and effective. We are also theorizing about the change in confidence, not the final confidence, as we believe the magnitude and the sign of this update determine the success of Aristotle.

From our experience of running the simulation, we concluded that under no-shock conditions, Aristotle's impact is not acknowledged by individuals and, therefore, doesn't result in them being more confident in their theory. Without new information or updates, participants tend not to focus on Aristotle's benefits, such as providing concise, purposeful insights.

In contrast, the key idea is that under shock conditions, participants are more attentive due to the disruption caused by the shock. This heightened engagement clarifies Aristotle's value: it helps participants navigate uncertainty by offering focused, relevant updates instead of overwhelming them with excessive information. As a result, participants experience a significant boost in confidence.

Our theory hinges on the *update effect*—the difference in Aristotle's impact under shock and no-shock conditions—being positive. This means Aristotle's value is only fully realized in the presence of a shock, and that value is affecting confidence positively. While we do not propose a theory about Aristotle's effect without a shock (β1), we neither state it is negative nor positive because we believe the impact is not captured well by the individual due to lack of engagement; we argue that Aristotle's impact on confidence is significantly more favorable under shock conditions, which means depending on the magnitude and the sign of the effect of Aristotle under no shock, a less negative, more positive or even from negative to positive update on the confidence compared to Aristotle under no shock. This demonstrates that Aristotle's true utility lies in situations where participants face uncertainty, as shocks create the engagement necessary for its focused approach to be appreciated.

3) Model:

$$\Delta SD = \beta 0 + \beta 1 \cdot \text{Aristotle} + \beta 2 \cdot \text{Shock} + \beta 3 \cdot (\text{Aristotle} \times \text{Shock}) + \beta 4 \cdot \text{Controls} + \epsilon$$

A critical part involved cleaning the dataset by removing one participant with contradictory pre-experiment responses, indicating inattention or misunderstanding.

A) Interpreting Variables in the model:

a. Dependent Variable:

ΔSD (change in standard deviation) represents the change in confidence (SD is inversely related to confidence itself). These values are converted from confidence levels given by the test takers on the probability of success they have given to their own theory. A high standard deviation means low confidence in their own theory.

- Positive ΔSD: Participants standard deviation increased, and confidence decreased

- Negative ΔSD: Participants standard deviation decreased, and confidence increased

b. Independent Variables:

I. *Intercept:* The baseline change in confidence for participants who were not exposed to Aristotle and did not experience a shock, controlling for other factors.

II. *Aristotle:* A binary variable indicating exposure to Aristotle (1=exposed, 0=not exposed). β1 captures the effect of exposure on the change in confidence (ΔSD).

III. *Shock:* A binary variable indicating whether participants experienced a shock (1=shock occurred, 0=no shock). β2 reflects whether a sudden disruptive event impacts the change in confidence (ΔSD).

IV. *Aristotle × Shock:* A binary variable, taking the value of 1 only if both Aristotle and Shock are present. B3 captures the change in Aristotle's effect on confidence in the presence of a shock. Here are the meanings of B3 being positive or negative and its effect on the impact of Aristotle on the change in confidence. Here is the explanation of the coefficient β3's impact in two cases:

Case 1: β3 > 0 (Shocks worsen the impact of Aristotle on the confidence updates) (β1+β3>β1)

- If β1 < 0 (Aristotle alone improves confidence): Shocks weaken confidence improvement caused by Aristotle, making the effect of Aristotle on ΔSD present a smaller confidence increase (β1+β3> β1).
- If β1 > 0 (Aristotle alone worsens confidence): Shocks amplify the confidence decline caused by Aristotle, making the effect of Aristotle on ΔSD present a greater confidence decline (β1+β3>β1).
- If β1 = 0 (Aristotle alone doesn't affect the confidence): Only in the presence of a shock the effect of Aristotle on the confidence is captured, and it reduces the confidence improvement(0+β3>0).

Case 2: β3 < 0 (Shocks improve the impact of Aristotle in the confidence updates)(β1+β3<β1)

- If β1 < 0 (Aristotle alone improves confidence): Shocks enhance confidence improvement, making the effect of Aristotle on ΔSD present greater confidence increase (β1+β3< β1).
- If β1 > 0 (Aristotle alone worsens confidence): Shocks weaken confidence decline, making the effect of Aristotle on ΔSD present smaller confidence decrease(β1+β3<β1)
- If β1=0 (Aristotle alone doesn't affect the confidence): Only in the presence of a shock the effect of Aristotle on the confidence is captured, and it increases the confidence update(0+β3<0).

Note that in both cases, if the magnitude of B3 is bigger than B1, the effect of Aristotle can also change signs. In case 1, shocks still worsen the impact of Aristotle on the confidence updates as it made the effect of Aristotle on the change of standard deviation positive when it was negative under no shock. In case 2, shocks still improve the impact of Aristotle in the confidence updates as it made the effect of Aristotle on the change of standard deviation negative when it was positive under no shock.

c. Controls (B4):

To account for individual differences influencing ΔSD beyond Aristotle and shock effects, covariates were grouped by relativity and balance levels between the four groups (Appendix 4.2) and added step by step:

1. *Theory-Related Variables:*
    - V1, V2, Vdiff_nonabs: Represent participants' baseline, updated, and change in perceived probabilities of theory success. All exceeded the imbalance threshold, but only

V2 (updated probabilities) and Vdiff_nonabs (change in probabilities) were included to avoid multicollinearity.
    - SD1, SD2: Represent standard deviation levels in the two simulation steps, capturing variability in participants' confidence. Only SD_1 (first step) was included to avoid multicollinearity.
2. *Algorithmic Attitudes:*
    - Algorithmic Aversion & Algorithmic Liking: Reflect participants' biases for or against algorithms, which could influence confidence updates (ΔSD). Both exceeded the threshold, but only Algorithmic Liking was included due to its higher imbalance.
    - GPT Usage & GPT Familiarity: Represent participants' experience and familiarity with GenAI tools, which might affect their engagement with Aristotle and confidence updates. Only GPT_usage was included due to its higher imbalance.
3. *Baseline Knowledge Measures:*
    - Knowledge_Depth_1 & Knowledge_Depth_2: Reflect participants' initial understanding and exploration of ideas, which could influence how new information impacts confidence. Though within the imbalance threshold, both were included to test their impact on confidence.

To see the list of explanations of the variables used in the analysis in detail, refer to Appendix 4.1.

B) Hypothesis:

$$\Delta SD = \beta 0 + \beta 1 \cdot Aristotle + \beta 2 \cdot Shock + \beta 3 \cdot (Aristotle \times Shock) + \beta 4 \cdot Controls + \epsilon$$

Note that a high standard deviation means low confidence therefore ΔSD is negatively correlated with the change confidence.

Aristotle's impact on individuals' update of confidence is hypothesized to be more favorable (less negative or more positive) under shock conditions compared to the effect of Aristotle under no-shock conditions ($\beta 1+\beta 3<\beta 1$). This means the interaction term β3 is expected to be negative.

Null Hypothesis (H_0): $\beta 3=0$, indicating that the interaction effect is either nonexistent.
Aristotle's effect on the change in confidence is homogenous between no shock and shock conditions (B1+0=B1).

Alternative Hypothesis (H_1): $\beta 3<0$, indicating that the interaction effect is negative, therefore favorable.
Aristotle's effect on the confidence update under shock is more favorable (greater increase, smaller decrease) than the effect of Aristotle on the confidence update under no shock. Aristotle increases the improvement in confidence or reduces the decline in confidence under shock conditions ($\beta 1+\beta 3<\beta 1$).

4) Analysis

A) Descriptive Analysis:

The descriptive statistics (Appendix 4.3) highlight key characteristics of the dataset. SD_1 (mean = 4.73) and SD_2 (mean = 5.02) display similar means and variability, indicating stable confidence levels across simulation steps. SD_diff (mean = 0.29, range = -1 to 3) captures changes in confidence, where negative values reflect improved confidence and positive values indicate increased uncertainty. Theory-related variables, V_1, V_2, and V_diff_nonabs, show substantial ranges and variability, reflecting participants' differing evaluations of theory success, making these variables essential for controlling individual differences in theory assessments. Algorithmic Liking and GPT Usage exhibit moderate variation, reflecting differences in AI familiarity and attitudes. Knowledge Depth measures show consistent baseline understanding across participants, influencing how they process information.

The group-level summary (appendix 4.4) of ΔSD reveals important insights into the effects of Aristotle and shocks on confidence updates. Group 3 (Aristotle without shocks) has the highest mean ΔSD (0.43) and high variability (Std = 0.94), suggesting that Aristotle's concise updates, in isolation, result in greater confidence loss but inconsistent responses. Group 4 (Aristotle with shocks) mitigates the confidence loss seen in Group 3 (mean ΔSD = 0.31 vs . mean ΔSD = 0.43) but exhibits the highest variability (Std = 1.01), indicating diverse participant reactions to the combination of shocks and Aristotle. The decrease in standard deviation supports our theory that Aristotle with shock will lead to a smaller increase in standard deviation. However, these findings suggest that Aristotle's effectiveness is inconsistent due to the high standard deviation, particularly under uncertainty; therefore, adding covariates is justified to reduce this diverse effect.

B) Regression Results:

The regression results for ΔSD have been summarized with a stepwise addition of control variables across five models, as seen in Table 3 below:

1. Model 1: Baseline model without controls; includes only Aristotle, Shock, and their interaction.
2. Model 2: Adds highly unbalanced variables V_2 and V_diff_nonabs (avoiding multicollinearity by excluding V_1).
3. Model 3: Adds SD1 (avoiding multicollinearity by excluding SD2) to Model 2.
4. Model 4: Adds GPT_Usage and Algorithmic_liking to Model 3 (Worsens the significance and R-squared).
5. Model 5: Adds Knowledge_Depth_1 and Knowledge_Depth_2 to Model 3.

Table 3: Regression results from Aristotle/Plato analysis

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	0.2667	-0.0158	1.0356**	1.0074	-0.0026
	(0.2313)	(0.9717)	(0.0247)	(0.1474)	(0.9971)
Aristotle	0.1619	0.0367	-0.3365	-0.3630	-0.4170
	(0.6117)	(0.9054)	(0.2319)	(0.2040)	(0.1292)
Shock	-0.0902	-0.0487	-0.2089	-0.1674	-0.2081
	(0.7666)	(0.8688)	(0.4176)	(0.5287)	(0.4051)
Aristotle x Shock	-0.0259	0.2797	0.6350	0.6044	0.6971*
	(0.9528)	(0.5320)	(0.1111)	(0.1324)	(0.0717)
V_2		0.0009	0.0143**	0.0150**	0.0137**
		(0.8837)	(0.0207)	(0.0166)	(0.0233)
V_diff_nonabs		0.0335**	0.0234**	0.0231**	0.0239**
		(0.0115)	(0.0424)	(0.0487)	(0.0343)
SD_1			-0.3670***	-0.3880***	-0.3819***
			(0.0000)	(0.0000)	(0.0000)
GPT_Usage				-0.0344	
				(0.7168)	
Algorithmic_liking				0.0827	
				(0.2760)	
Knowledge_Depth_1					-0.1065
					(0.2390)
Knowledge_Depth_2					0.2958**
					(0.0162)
R-squared	0.0118	0.1303	0.3631	0.3789	0.4295
Adjusted R-Sq	-0.0393	0.0527	0.2936	0.2851	0.3434

p-values: *<0.01, **<0.05, *<0.10**

Looking at the fifth model, neither Aristotle alone (β1) nor shocks alone (β2) significantly influence the change in confidence, with only the interaction term (β3) being significantly different from zero. This indicates that the effect of using Aristotle is only captured in the presence of a shock, where the combination of Aristotle and a disruptive event becomes measurable because of the increased engagement with the chatbot.

The null hypothesis (β3=0) is not rejected even though B3 is significantly positive because we conducted a one-tailed t-test to test our hypothesis. The t-statistic for the coefficient Aristotle×Shock is approximately 9.72, but the one-sided p-value for testing if the coefficient is less than 0 is effectively 1, meaning there is no evidence to suggest that the coefficient is less than 0. This indicates that Aristotle under shock conditions does not improve the change in confidence compared to Aristotle under no-shock conditions. However, the positive and significant coefficient for β3 ($p<0.1$) suggests that Aristotle under shock conditions actually worsens confidence. Considering that the effect of Aristotle and Shock alone is not significant, therefore 0; under the presence of both shock and Aristotle, the change in standard deviation is 0.6971 points higher than the control scenario. Thus participants facing both Aristotle and Shock update their confidence more negatively. This result supports our reasoning that Aristotle's short and focused information doesn't have any effect on the confidence update in general scenarios, in the presence of a shock or uncertainty causes participants to decrease their confidence. In a state of heightened uncertainty, participants might find Aristotle's concise updates insufficient to address their doubts fully, which could lead to reduced confidence. Instead, broader, more generalized information offered by general-purpose chatbots might better meet their needs during shocks. This highlights a limitation of Aristotle's design, as its compact insights may not address the increased informational demands and reassurance required in shock scenarios.

Other findings align with the importance of controls and individual variability in shaping confidence updates. The consistent significance of V2 and Vdiff_nonabs highlights their critical role in capturing individual differences in participants' probability assessments of success. The strongly negative coefficient for SD1 underscores how initial confidence levels influence updates, with greater initial confidence leading to a negative impact on change in standard deviation therefore, a positive impact on the change of confidence. Interestingly, including GPT_Usage and Algorithmic_Liking worsened the model, as shown by reduced explanatory power, and these variables were excluded from the final model. The stepwise addition of controls is validated by increasing R-squared and Adjusted R-squared values, reaching 0.3434 in Model 5.

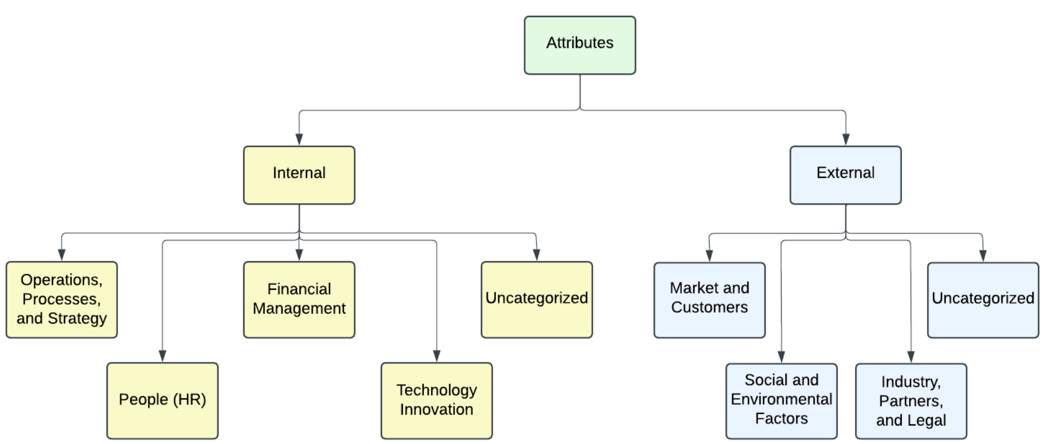
Overall, the findings didn't support our theory. Even though B3 is significantly positive, H0 is not rejected because we conducted a one-sided t-test for B3<0. Instead, we found that the impact of Aristotle on confidence updates under shock is actually less favorable than the impact of Aristotle on confidence updates under no shock. This suggests that during uncertainty, broader and more comprehensive information—like that from general-purpose chatbots—may better address participants' informational needs and foster confidence. Finally, the fact that B1 is not significant supports our idea of the importance of the engagement, indicating that the effect of using Aristotle is only captured in the presence of a shock, where the combination of Aristotle and a disruptive event becomes measurable because of the increased engagement with the chatbot.

Aristotle - Attributes

1) Background

As the final part of the assignment, we were given a dataset of theories created by Aristotle. This data set included information on the decision (company name, decision maker, decision problem) and the attributes that were generated by Aristotle for the theories of value. Each attribute had an associated attribute probability as well as insight into the causal links in that theory.

2) Attribute Taxonomy



The taxonomy developed organizes attributes into two primary categories: Internal and External. These categories further branch into subcategories based on areas:

A) Internal Attributes:

- Operations, Processes, and Strategy: Covers an organization's internal mechanisms and strategic decisions.
- Financial Management: Focuses on financial planning, funding, and budgeting.
- People (HR): Includes talent acquisition, employee engagement, and the development of organizational culture.
- Technology Innovation: Captures advancements and innovations (R&D) driven internally.
- Uncategorized: Attributes that do not fit into the above classifications.

B) External Attributes:

- Market and Customers: Addresses customer trends, market demands, and brand engagement.
- Social and Environmental Factors: Encompasses corporate social responsibility, environmental impacts, and societal expectations.
- Industry, Partners, and Legal: Involves external partnerships, regulatory compliance, and industry trends.
- Uncategorized: Attributes that do not fit into the above classifications.

3) Theories:

Theory on the frequency of attributes: We hypothesize that there is less publicly available information online about internal organizational processes than external ones. Consequently, Aristotle, when used to generate theory attributes, will produce fewer internal attributes than external ones.

Theory on assigned probabilities: We hypothesize a greater prevalence of negative information online about external attributes than internal ones. This asymmetry causes Aristotle to assign lower unconditional probabilities to external attributes than internal ones.

4) Proposed Plan to Measure Attribute Distribution:

1. *Categorization of Attributes:*
    - Utilize AI tools like ChatGPT and Aristotle to categorize each attribute from the dataset into the taxonomy categories we came up with.
        - Perform this step incrementally, analyzing a few rows of data at a time and giving feedback to ensure the tools correctly classify each attribute.
2. *Analysis of Attribute Frequency (Count)*:
    - Evaluate the count of attributes assigned to each attribute sub-category.
    - Compare to our hypothesis, the count of internal attributes is lower than that of external attributes.

    Note: To validate our reasoning that there is less information online about internal organizational processes than external ones, we could perform representative search queries for internal and external attributes. The number of search results for each category can be recorded and analyzed using statistical tests to determine significant differences.
3. *Analysis of Attribute Assigned Probability (Avg probability):*
    - Evaluate the average unconditional probability for each sub-category

- Compare it against our hypothesis, the probability of external attributes being lower than the probability of internal ones using a t-test.

Note: To validate our reasoning about the greater prevalence of negative information online on external attributes compared to internal ones, we could conduct sentiment analysis on the training data by sampling external and internal attributes, analyzing their sentiment, and measuring the proportion of negative content in each category.

Conclusion: Recommendations and Insights:

Our findings highlight key insights into the interplay between experimental design, decision-making frameworks, and the outcomes observed while also uncovering areas for refinement and enhancement.

Our analysis revealed that the probability of winning simulations has a significant random component, as the system selects the final comparison probabilities with variability. Upon inquiry with the teaching staff, it was clarified that the mechanism involves a degree of randomness rather than being entirely rational. This introduces variability that may overshadow the theoretical aspects we aim to test. This explains why our Socrates theory on the higher probability of winning based on running the lower probability experiment proved statistically insignificant, as winning the game is based on more than just making the correct decision, which is not incorporating the framework of experimenting on more surprising outcome is more beneficial to the individual. We believe that this framework could be added to the mechanism or the simulation for better connectedness with the theory.

Our analysis found statistically significant results on the effect of negative treatment for low (baseline) and high trust individuals partially confirming our theory. Post-experiment reflections revealed inconsistencies in the treatments, where both the information's tone and content varied. Such variability makes it difficult to isolate the specific effects of information framing. In future experiments, we would like to make the content consistent across treatments while only altering the tone. This would allow for a more precise analysis of how framing alone influences trust change, avoiding the confounding impact of content differences. Additionally, the analysis could have been more significant if the sample size had been larger. In future experiments, a larger sample size is recommended.

The analysis of Plato/Aristotle simulations showed that using Aristotle (the specialized chatbot) under conditions of uncertainty yielded less favorable (at a 10% significance level) outcomes than stable conditions. While this does not imply a lack of capability, it highlights a misalignment between the tool's design and the information demands of uncertain scenarios. Compact and focused information, offered by Aristotle, may not be optimal when participants face shocks or uncertainty. One interesting idea to explore would be an enhanced version of Aristotle that dynamically adapts to the context—offering concise information in stable conditions but expanding its scope under uncertainty to provide the reassurance and breadth of detail users may need.

Incorporating these improvements and aligning experimental mechanisms more closely with theoretical frameworks will provide deeper insights and foster more effective applications of decision-making strategies in practice.

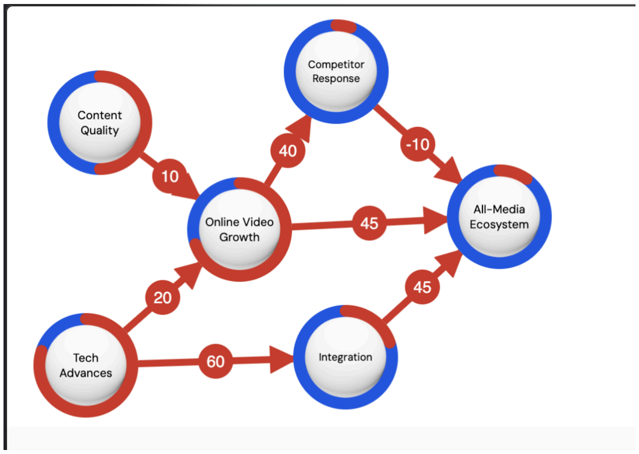
Works Cited

Camuffo, Arnaldo, Alfonso Gambardella, and Andrea Pignataro. "Theory-driven strategic management decisions." Strategy Science (2024). Accessed December 1st, 2024 .

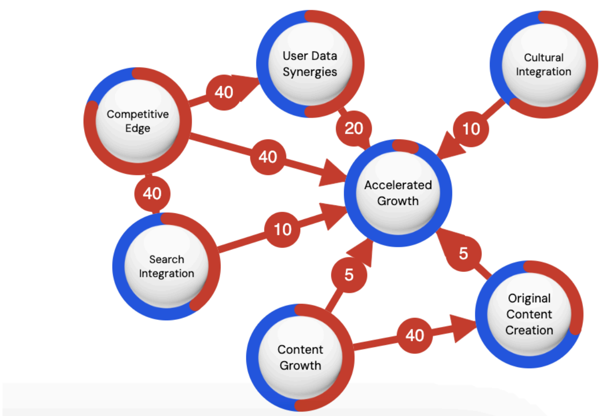
OpenAI. "ChatGPT: Conversational AI Model." Accessed December 1st, 2024 . Available at: https://openai.com/chatgpt.

Appendix

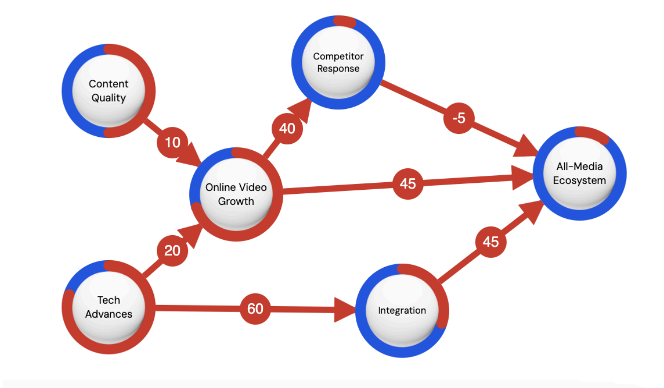
Appendix 1:



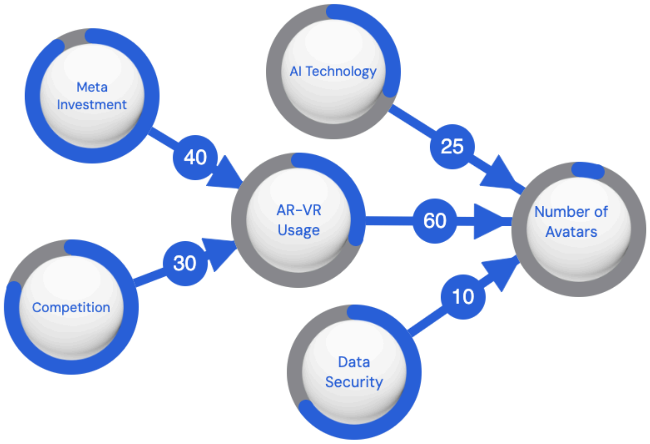
Dag Theory 1 Google Case



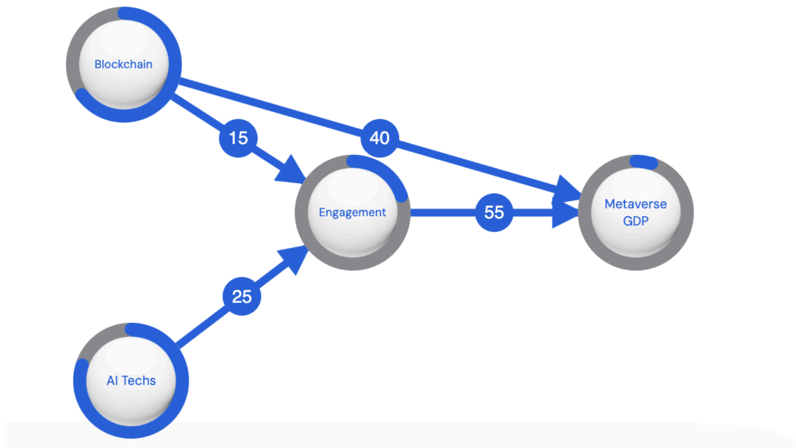
Dag Theory 2 Google Case



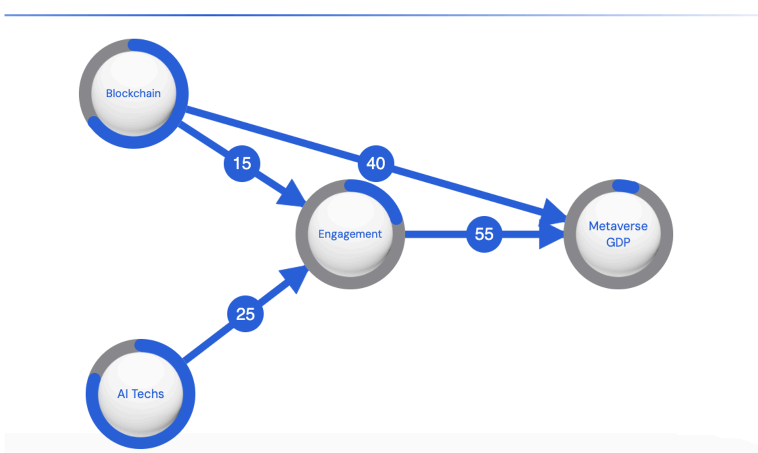
Dag Theory 1 Updated Google Case



Dag Theory 1 Meta Case



Dag Theory 2 Meta Case



Dag Theory 2 Updated Meta Case

Appendix 2:

Appendix 2.0: Density calculation

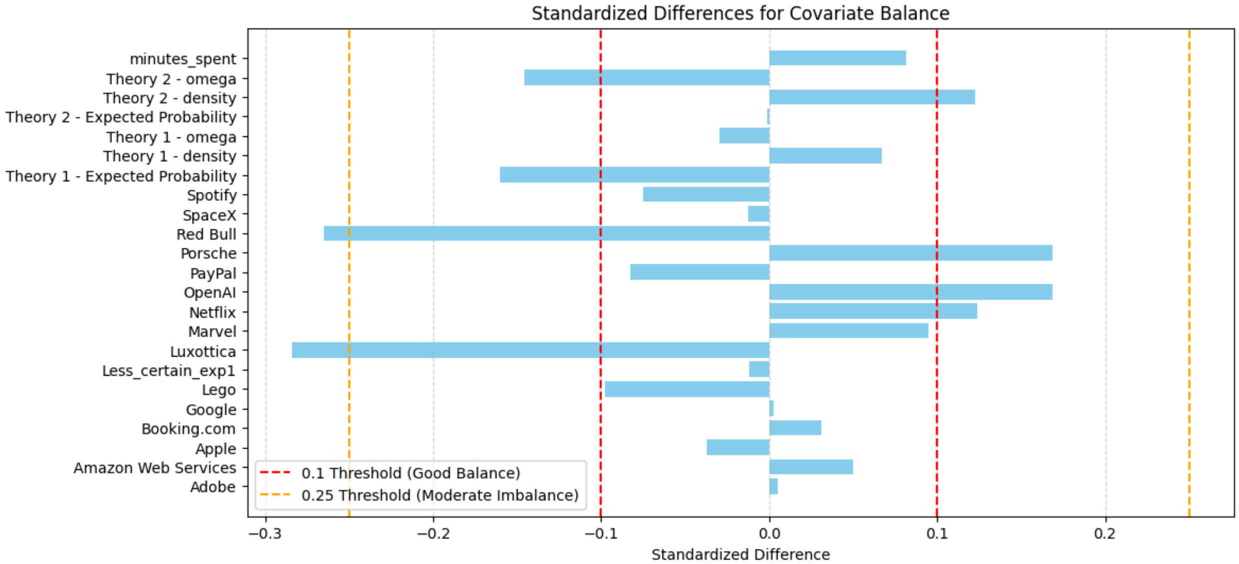
$$D = \frac{E}{V(V-1)}$$

V - number of attributes (PS: *V(V−1) is the maximum possible number of directed edges in a graph with V vertices (assuming no self-loops).*)
E - number of links
D=1: A fully connected directed graph (complete graph).
D≈0: A sparse graph with very few edges compared to the number of nodes.

Appendix 2.1: descriptive statistics

Statistic	const	Theory 1 - Expected Probability	Theory 1 - density	Theory 1 - omega	Theory 2 - Expected Probability	Theory 2 - density	Theory 2 - omega	minutes_spent	Less_certain_exp1
mean	1.00	0.38	0.26	0.60	0.38	0.25	0.60	34.52	0.40
std	0.00	0.20	0.09	0.19	0.20	0.11	0.21	23.30	0.49
min	1.00	0.00	0.12	0.01	0.00	0.12	0.00	2.00	0.00
25%	1.00	0.23	0.20	0.50	0.22	0.17	0.50	18.00	0.00
50%	1.00	0.35	0.25	0.60	0.38	0.20	0.60	29.00	0.00
75%	1.00	0.52	0.33	0.72	0.50	0.33	0.75	46.00	1.00
max	1.00	1.00	0.50	1.00	0.99	0.50	1.00	121.00	1.00

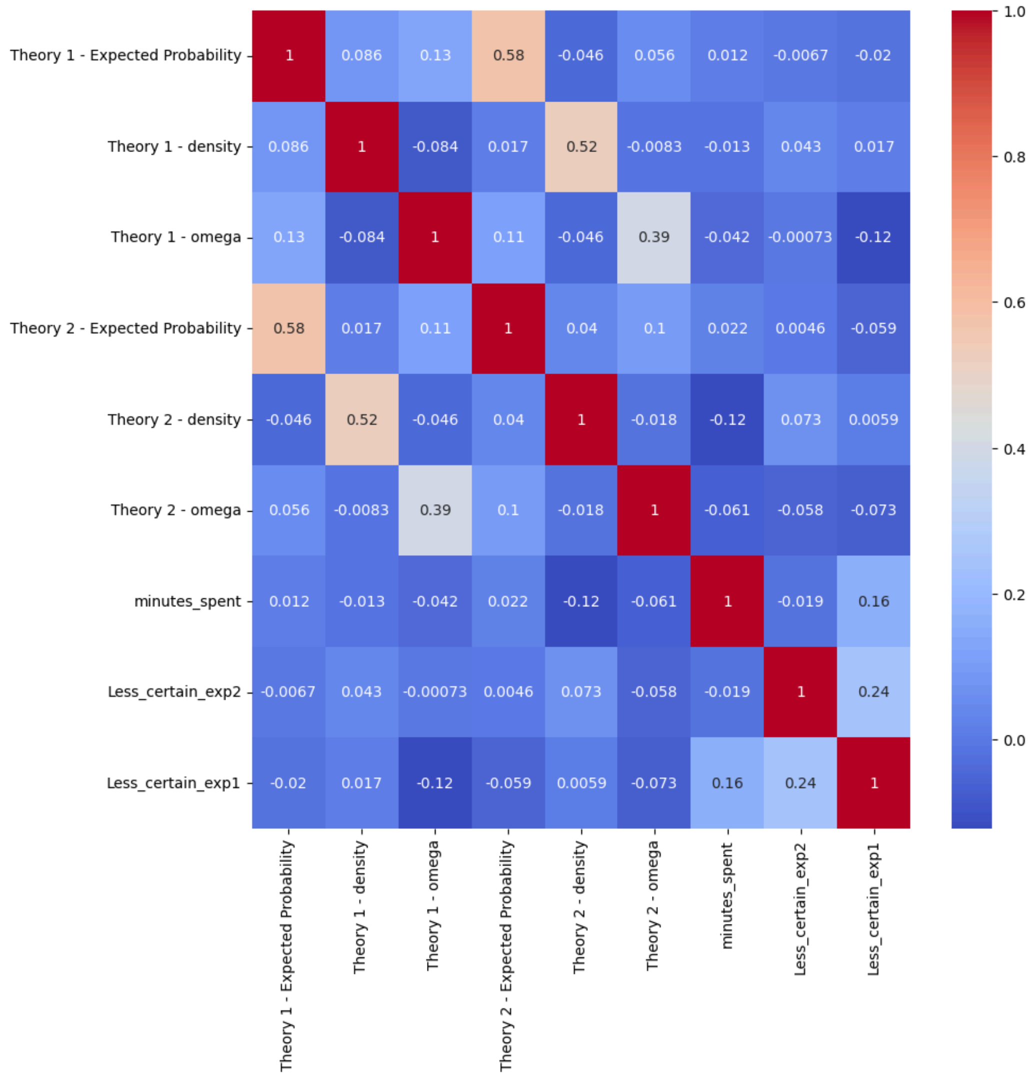
Appendix 2.2: Covariate Balance Check



Appendix 2.3: VIFs check for multicollinearity

	Variable	VIF
0	const	32.7016
1	Theory 1 - Expected Probability	1.56123
2	Theory 1 - density	1.42618
3	Theory 1 - omega	1.22498
4	Theory 2 - Expected Probability	1.53845
5	Theory 2 - density	1.44119
6	Theory 2 - omega	1.19716
7	minutes_spent	1.05505
9	Less_certain_exp	1.11106

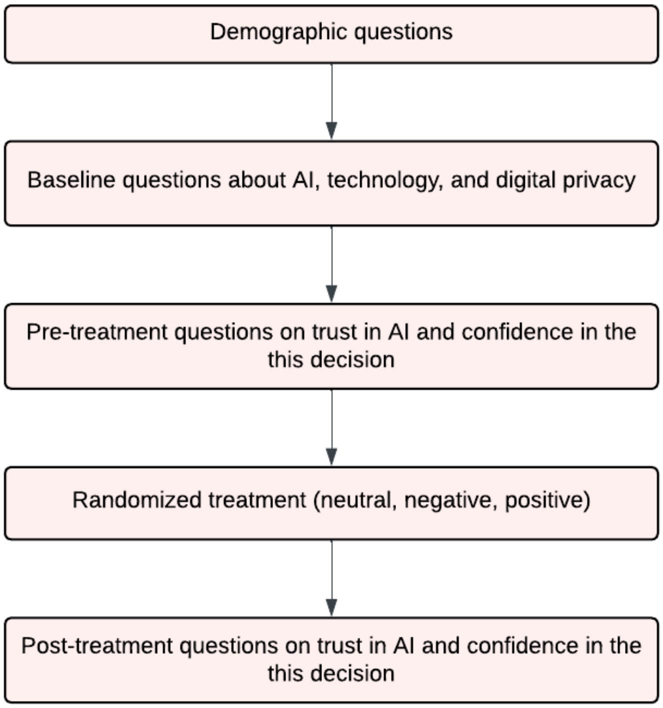
Appendix 2.4: Correlation matrix



Appendix 3:

Appendix 3.1 - Survey flow chart and content of fake news used as treatment

Survey Flow Chart



Treatment - Neutral

Technical Overview: Mechanisms of Data Privacy in AI Systems
Artificial intelligence systems utilize various data handling mechanisms aimed at managing user information. Techniques such as data anonymization remove identifiable elements from datasets, while federated learning enables data processing directly on user devices, reducing the need for centralized data collection. These methods are integrated to support the operational structure of AI systems.

AI frameworks also employ differential privacy, which introduces statistical noise to data outputs, making it difficult to trace back individual data points. These technical processes are part of the system's design, ensuring that data is processed within specified parameters without revealing specific user details. The application of these mechanisms reflects current approaches to managing data within AI-driven environments.

Treatment - Positive

Advances in AI Bring Unprecedented User Security and Renewed Public Trust
Recent developments in artificial intelligence have bolstered user security measures, reshaping the landscape of data privacy. Major tech firms are integrating end-to-end encryption, multi-layered security protocols, and user-centric data controls into their AI-driven platforms. These advancements not only meet global privacy standards but exceed them, offering users unmatched levels of safety and transparency.

Consumer trust is seeing a resurgence as more people feel empowered to control their personal data. Industry leaders emphasize that these technologies are designed with user security as a foundational element, highlighting a commitment to prioritizing individuals over profits. The result is a new era where AI is not just a tool for convenience but a trusted guardian of user privacy.

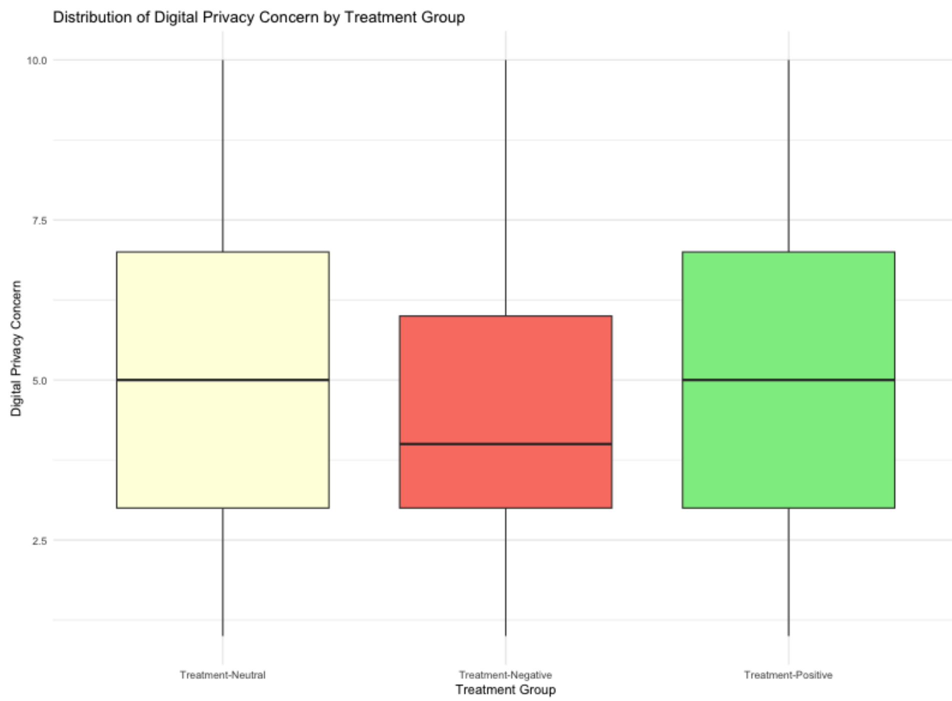
Treatment - Negative

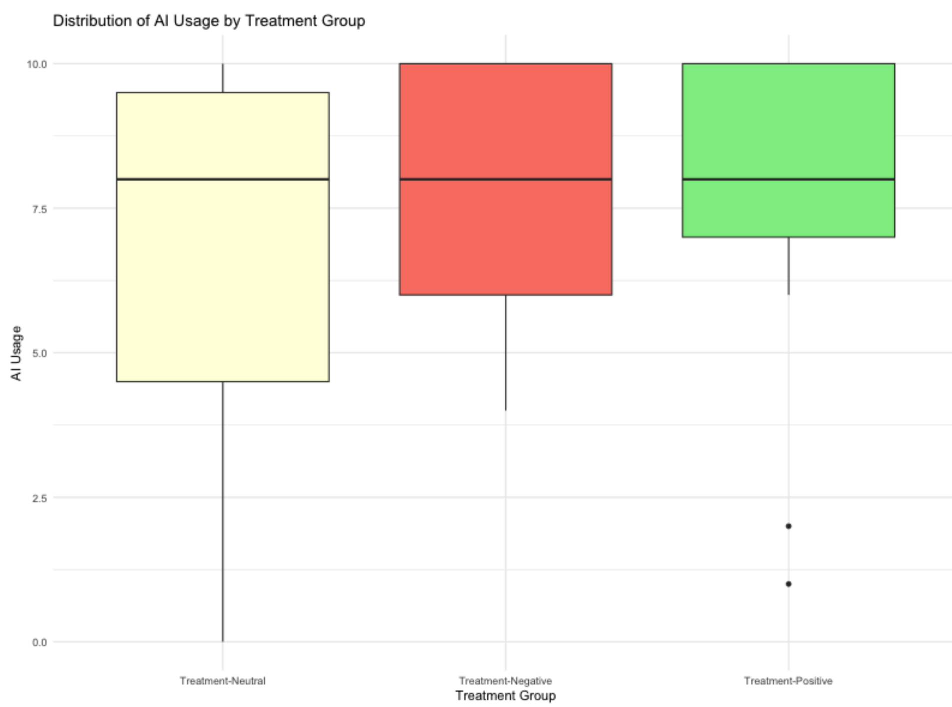
AI Privacy Failures Spark Growing Concerns Over User Security
Despite promises of better user security, recent incidents have highlighted the vulnerabilities in AI-driven systems. Data breaches and unauthorized data access reports continue to surface, exposing significant flaws in current AI security measures. Critics argue that while companies claim to prioritize user privacy, their practices often fall short of providing comprehensive protections.

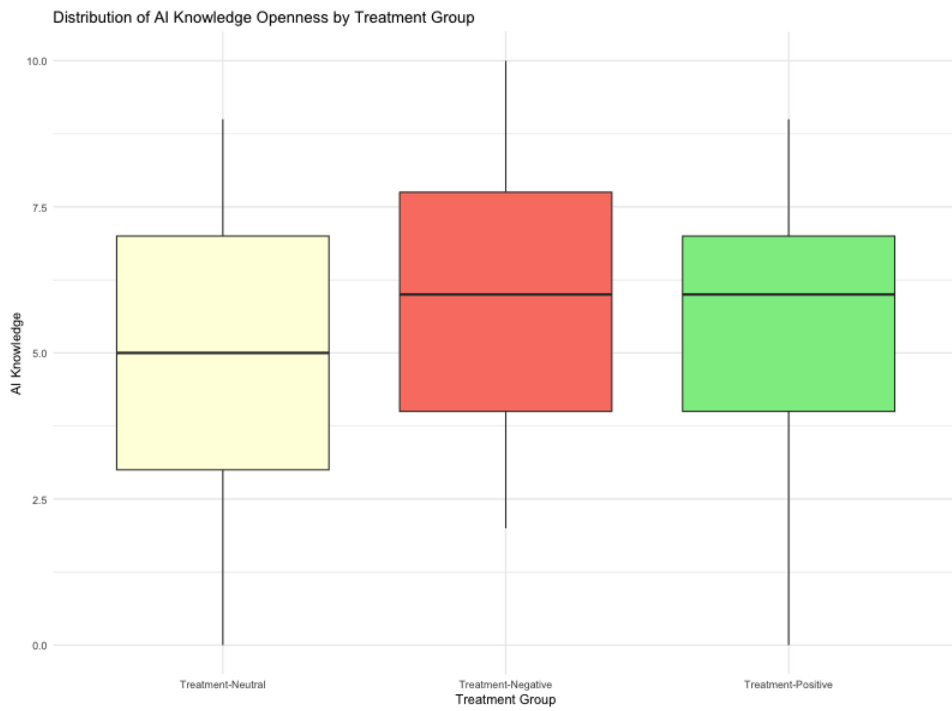
The erosion of public trust is becoming palpable as users express frustration over the lack of transparency and effective safeguards. Experts warn that without immediate improvements and stricter oversight, the situation could escalate, undermining confidence in not only individual companies but the broader use of artificial intelligence. Calls for enhanced regulations and accountability are growing louder, signaling a need for urgent industry-wide change.

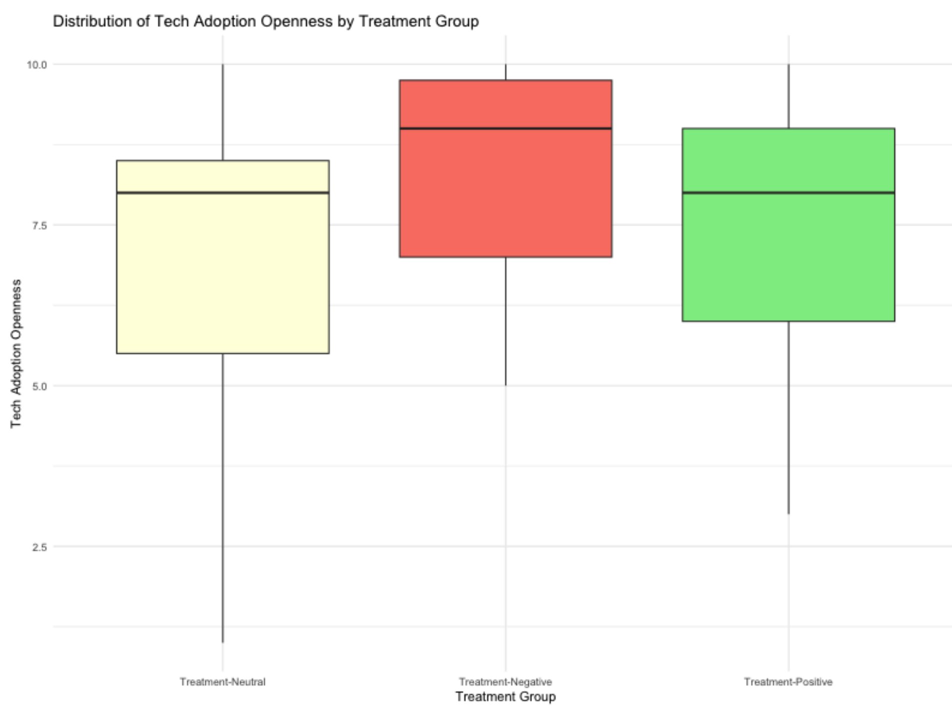
Appendix 3.2 - ANOVA tests on covariates

	Variable	P_Value
1	AI_Usage_Frequency	0.23025272
2	AI_Knowledge_Rating	0.27266420
3	Digital_Privacy_Concern	0.45949040
4	Tech_Adoption_Openess	0.03409479


Distribution of Digital Privacy Concern by Treatment Group
Digital Privacy Concern
10.0
7.5
5.0
2.5
Treatment-Neutral
Treatment-Negative
Treatment-Positive
Treatment Group


Distribution of AI Usage by Treatment Group
AI Usage
10.0
7.5
5.0
2.5
0.0
Treatment-Neutral
Treatment-Negative
Treatment-Positive
Treatment Group



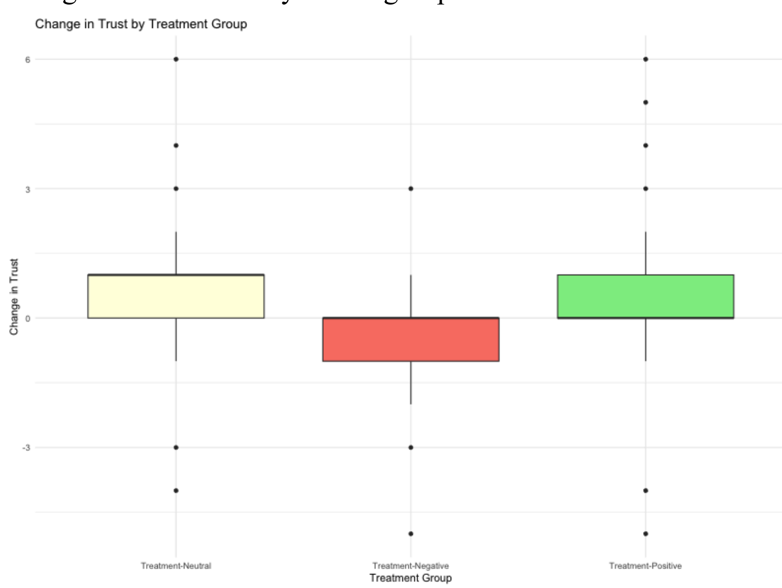


Appendix 3.3 - Summary statistics

Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Gender	82						
... Female	31	38%					
... Male	46	56%					
... Non-Binary	3	4%					
... Other	1	1%					
... Prefer not to say	1	1%					
AI_Usage_Frequency	82	7.4	2.5	0	6	10	10
AI_Knowledge_Rating	82	5.4	2.3	0	4	7	10
Digital_Privacy_Concern	82	4.9	2.5	1	3	7	10
Tech_Adoption_Openess	82	7.6	2.2	1	6	9	10
Trust_AI_Privacy_Pre	82	4.3	2.3	0	3	6	9
Decision_Confidence_Pre	82	6	2.4	0	5	7.8	10
Trust_AI_Privacy_Post	82	4.6	2.3	0	3	6	9
Decision_Confidence_Post	82	6.3	2.3	0	5	8	10
Treatment_Randomizer	82						
... Treatment-Neutral	27	33%					
... Treatment-Negative	26	32%					
... Treatment-Positive	29	35%					
pre_trust_segment	82						
... low	30	37%					
... medium	34	41%					
... high	18	22%					
pre_confidence_segment	82						
... low	13	16%					
... medium	33	40%					
... high	36	44%					
Trust_Change	82	0.34	2	-5	0	1	6
Confidence_Change	82	0.32	1.7	-6	0	1	3

Change in trust summary across groups:



	Treatment_Randomizer	Mean_Trust_Change	Median_Trust_Change	SD_Trust_Change	Min_Trust_Change	Max_Trust_Change	Count
	<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>
1	Treatment-Neutral	0.667	1	1.92	-4	6	27
2	Treatment-Negative	-0.346	0	1.52	-5	3	26
3	Treatment-Positive	0.655	0	2.21	-5	6	29

Appendix 3.4 - AIC of models and GVIF check

	df	AIC
model1	4	344.4085
model2	6	338.0207
model3	7	339.6235
model4	11	340.3597
model5	16	336.9678

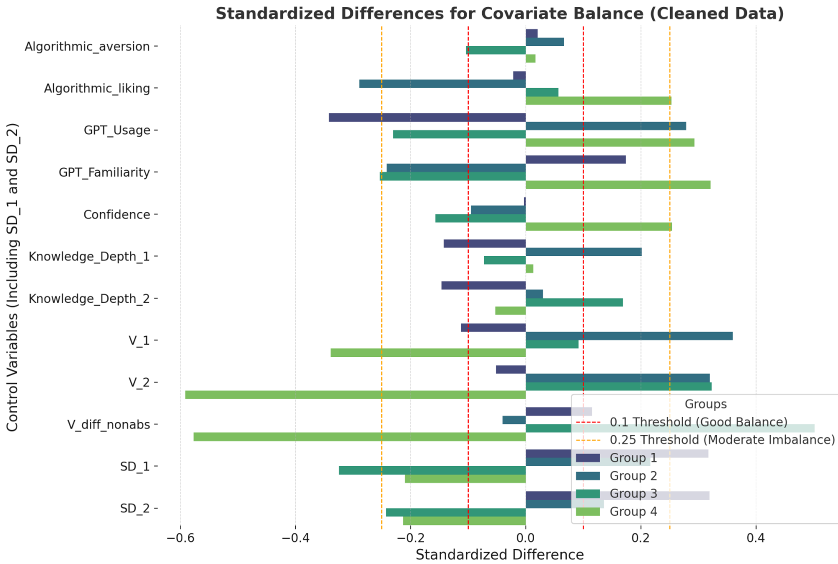
	GVIF	Df	GVIF^(1/(2*Df))	Interacts With
Treatment_Randomizer	2.680015	8	1.063552	pre_trust_segment
pre_trust_segment	2.680015	8	1.063552	Treatment_Randomizer
Tech_Adoption_Openess	2.269288	1	1.506416	--
AI_Usage_Frequency	2.017341	1	1.420331	--
AI_Knowledge_Rating	1.755257	1	1.324861	--
Digital_Privacy_Concern	1.353515	1	1.163407	--
pre_confidence_segment	1.869976	2	1.169389	--

Appendix 4:

Appendix 4.1: Explanations of the variables used in the model

- **SD_1:** Standard deviation for initial theory(i.e. confidence),
- **SD_2:** Standard deviation for updated theory (i.e. confidence)
- **ΔSD:** Change in standard deviation, SD_2 minus SD_1 (inversely related with change in confidence)
- **Aristotle:** Binary variable indicating if the participant was exposed to Aristotle (1: Yes, 0: No).
- **Shock:** Binary variable indicating if the participant experienced a shock (1: Yes, 0: No).
- **Algorithmic_aversion:** Score reflecting participants' discomfort with algorithms.
- **Algorithmic_liking:** Score reflecting participants' liking or trust of algorithms.
- **GPT_Usage:** Score indicating how frequently participants use GenAI-based tools.
- **GPT_Familiarity:** Score indicating participants' familiarity with GenAI-based tools.
- **Confidence:** Pre-treatment measure of participants' confidence
- **Knowledge_Depth_1:** Pre-treatment measure of how deeply participants understand or articulate their initial theories.
- **Knowledge_Depth_2:** Pre-treatment measure of how much participants found the knowledge useful
- **V_1:** Initial probability of success for the theory.
- **V_2:** The updated probability of success for the theory.
- **V_diff_nonabs:** Directional change in theory value ((V_2 - V_1)).

Appendix 4.2: Covariate Balance check



Appendix 4.3: Descriptive Statistics

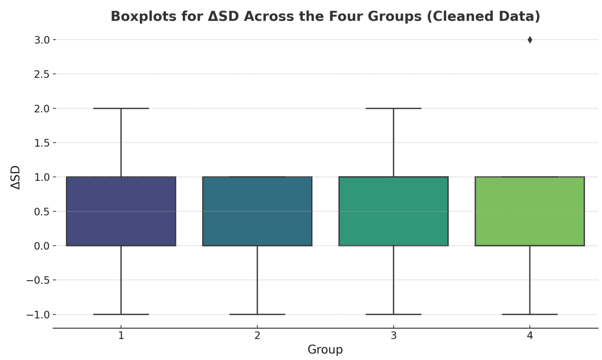
Overall Descriptive Statistics for Cleaned Data

Variable	Mean	Median	Std	Min	Max	Range
SD_diff	0.29	0.00	0.84	-1.00	3.00	4.00
SD_1	4.73	5.00	1.32	1.00	7.00	6.00
SD_2	5.02	5.00	1.22	1.00	7.00	6.00
Algorithmic_liking	3.42	3.00	1.27	2.00	6.00	4.00
Algorithmic_aversion	5.05	5.00	1.11	2.00	7.00	5.00
Confidence	4.81	5.00	1.02	2.00	7.00	5.00
V_diff_nonabs	5.44	5.00	9.18	-15.00	31.00	46.00
V_1	62.55	65.50	18.34	0.00	90.00	90.00
V_2	67.98	70.00	19.48	0.00	97.00	97.00
GPT_Usage	5.97	6.00	1.02	2.00	7.00	5.00
GPT_Familiarity	5.21	5.00	1.10	2.00	7.00	5.00
Knowledge_Depth_1	5.31	5.00	1.15	3.00	7.00	4.00
Knowledge_Depth_2	5.85	6.00	0.85	3.00	7.00	4.00

Appendix 4.4: Group-level Summary and Box-plot

Group-Level Summary for ΔSD

Group	Mean	Median	Std	Min	Max
Group 1	0.27	0.00	0.80	-1.00	2.00
Group 2	0.18	0.00	0.64	-1.00	1.00
Group 3	0.43	1.00	0.94	-1.00	2.00
Group 4	0.31	0.00	1.01	-1.00	3.00



Appendix 4.5: Multicollinearity check for the final significant model

VIF Analysis for Model 5

Variable	VIF
Intercept	65.75770272773633
Aristotle	1.0936569214019347
Shock	1.1431460620739862
V_2	1.5482392988321092
V_diff_nonabs	1.3353987894094907
SD_1	1.3894636206640978
Knowledge_Depth_1	1.404019560413277
Knowledge_Depth_2	1.3394093989072011