ERF 31ST Annual Conference
YOUTH
DEMOGRAPHIC DIVIDEND & MIGRATION IN MENA:
CHALLENGES & OPPORTUNITIES IN UNCERTAIN TIMES
2025

reshaping the future

Is ChatGPT Detrimental to Innovation? A Field Experiment among University Students

Mazen Hassan,
Engi Amin,
Sarah Mansour
and Zeyad Kelani

ECONOMIC RESEARCH FORUM
منتدى البحوث الاقتصادية

Is ChatGPT Detrimental to Innovation?
A Field Experiment among University Students

Mazen Hassan[1], Engi Amin[2], Sarah Mansour[3], Zeyad Kelani[4]

[1] Political Science Department, Faculty of Economics and Political Science, Cairo University, Giza, Egypt. Email: m.hassan@feps.edu.eg
[2] Socio-computing Department, Faculty of Economics and Political Science, Cairo University, Giza, Egypt. Email: engi.amin@feps.edu.eg
[3] Economics Department, Faculty of Economics and Political Science, Cairo University, Giza, Egypt. Email: sarah.mansour@feps.edu.eg
[4] Political Science Department, Faculty of Economics and Political Science, Cairo University, Giza, Egypt. Email: zelkelani@feps.edu.eg

Abstract

Purpose:

This paper investigates the potential collateral effects of A.I innovations, specifically ChatGPT, on three key variables: innovation, readiness to exert effort, and risk behavior

Methodology:

A field experiment was conducted involving nearly 100 senior university students at a public university in Egypt, at a time when ChatGPT had not yet been legally operational. Over a one-month period, participants submitted three graded essay assignments. The treatment group utilized ChatGPT to write the essays, while the control group completed the assignments without such assistance. After submission, both groups participated in a lab-based innovation game, a risk game, and a real effort task to measure their respective innovation, risk aversion, and effort exertion.

Findings:

The results revealed that students who used ChatGPT demonstrated significantly lower levels of innovation and risk aversion compared to the control group. Although the reduction in effort exerted by the ChatGPT group was not statistically significant, the overall trends suggest a potential decrease in effort related to the use of A.I. applications.

Originality:

This study is among the first to empirically test the impact of ChatGPT on innovation, effort, and risk behavior in a real-world academic setting. It provides preliminary evidence of the potential negative effects of A.I. applications on these variables, offering valuable insights for further research into the broader implications of A.I. on human behavior.

Keywords: Artificial Intelligence; ChatGPT; Field experiment; Innovation; Risk behavior; Effort task

Introduction

Even before the rise of ChatGPT and other artificial intelligence text generators (AITGs), there has been a global debate – that is not short of controversy – on the risks and benefits of automation and artificial intelligence (Pasquale, 2015), with some calls for regulation (Wachter and Mittelstadt, 2019), or at least a pause on developing the technology until further studies assess their impacts. The ascend of ChatGPT has certainly intensified this debate, with worries about the new technology from scientists, educators, and even students (see, for example, (Farhi *et al.*, 2023)). Indeed, recent survey evidence has shown that two-thirds of students use ChatGPT in their studies (von Garrel and Mayer, 2023). In this context, this paper seeks to test the effect of the continuous usage of ChatGPT on innovation, effort levels, and risk behavior among university students.

Our assumption is straightforward and intriguing: that repeated reliance on technology to come up with answers to questions and to solve problems (just as ChatGPT does) creates a social norm of dependence on technology to innovate and – if used over a sustained period of time – crowds out the innovation drive by humans. To explore this question, we conducted a pre-registered field experiment with nearly 100 senior university students at a public university in Egypt, where we tested the effect of using ChatGPT over a month in doing assignments on three dependent variables usually assumed by the literature to be A.I.-collaterals: innovation, risk behavior, and readiness to exert effort. The field experiment extended for around four weeks and involved participants submitting three graded essay assignments during that period. In the treatment (ChatGPT) group, students were asked to write the essays using ChatGPT, whereas in the control group, such an option was neither mentioned nor allowed (the experiment was fielded before ChatGPT was legally operable in Egypt). One week after all assignments were submitted,

the two groups were asked to take part in a lab experiment – without knowing what the treatment was – where they were asked to play an innovation and a risk game. Our findings are nuanced; the ChatGPT treatment group was significantly less innovative (measured by how frequently they changed the sales strategies at the 95% confidence level) and less risk averse (at the 90% confidence level). The treatment group also exerted less effort (measured by how frequently they recorded their strategies for reference over the rounds), although this result was not statistically significant.

This paper is divided into five sections that detail our theory and design. The following section outlines our theory, whereas section three presents the experimental design. Section four includes the findings, and section five concludes with a discussion of the possible limitations of the study and avenues for future research.

Theory

Technological advances have become a significant factor affecting human behavior and, consequently, the social norms generated from – and at a later stage, guiding – that behavior. The invention of personal computers and laptops has made working from home beyond office hours an expected behavior from employers. The rise of mobile phones has made people hardly available. The rise of social media has affected the attention span of its users (Firth *et al.*, 2020; Ophir *et al.*, 2009), giving rise to social comparison and making us less social (Fischetti, 2016). This is hardly surprising given the penetration by technology of almost every aspect of our everyday life coupled with our increasing reliance on them to make our lives easier – and, in parallel, also different.

The premise of this paper is that technologies have spill-over effects across domains and are capable of affecting behavior in areas and, at times, beyond those in which they are used

initially. For example, it is now almost a fact that smartphones make us less attentive to other tasks when we do not use smartphones (Altmann *et al.*, 2014). They also negatively affect our recall accuracy and behavioral control (Chen *et al.*, 2016; Tanil and Yong, 2020), making their effect survive long after we stop using them – although of course, such effects depend on duration and intensity of usage.

We contend that a similar spill-over effect happens with regard to one of the newest – but at the same time one of the fastest growing – technologies: ChatGPT. We argue that continuous usage of ChatGPT would also leave its imprint on socio-economic behavior in domains where ChatGPT is not mainly used. We focus on three behavior types central to economic growth: innovation, effort, and risk behavior.

Starting with innovation; it is indeed a value that is crucial to economic activity, productivity, and even social relations (e.g. Amabile, 2019; Dahl and Moreau, 2002). Innovation, however, is a mentally pressing task. It requires individuals to keep thinking about a specific task (and how to improve it) through a reiterative process of asking questions and trying to answer them. A central dimension of innovation, therefore, involves problem-solving (Amabile, 2019; Bieser, 2022; Sternberg and Lubart, 1999); whereby knowledge is produced via experimentation (Arrow, 1969; Weitzman, 1978) and where early failure is an integral part of the process (Manso, 2011). ChatGPT – and other AITGs – however, predominantly involve asking an algorithm a series of questions and waiting for the chatbot to come up with answers. If the task of answering the question is repeatedly outsourced, however, over time, the otherwise supposed innovator would not exert the same mental effort to answer the important questions and solve the pressing problems at hand, driving innovation (or at least human-driven innovation) down. Taking the mental exercise

out of problem-solving is likely to lead to counter-innovation attitudes. Our first hypothesis, therefore, is:

H1: *The continuous use of ChatGPT over a sustained period of time, would drive down innovative behavior.*

We do acknowledge that A.I. can also help innovation via different mechanisms. These might include freeing up the time of humans otherwise consumed in monotonous tasks (Baska, 2018) that are not structured to lead to innovation (although some studies do show that we might not be making use of this time in innovative tasks, just in more screen time instead (Ortiz-Ospina and Roser, 2023). Another mechanism could be A.I., which helps to process a large body of data to identify patterns that might unearth a problem and suggest a solution (see Bieser, 2022). We argue, however, that these mechanisms are (a) more about incremental innovation (solving problems with a focus on one variable to be optimized), rather than the type of ground-breaking innovation that involves a radical shake-up of the status quo for the better, and (b) that they largely talk about machine-driven innovation (either partly or entirely). In this paper, however, we are interested in studying human-driven innovation and particularly whether the potential mechanisms outlined above would – over time – crowd out the innovation drive by humans that has been so central to human progress (Bloom *et al.*, 2020).

A second socio-economic attitude that could be affected by AITGs is risk behavior. While a reasonable level of risk is required for economic activity to thrive, substantially low-risk aversion could lead to reckless behavior and endanger economic enterprises. We argue that the continuous usage of ChatGPT could increase risk tolerance via two mechanisms. First, there is the moral hazard context that this specific type of automation creates (Rowell and Connelly, 2012). The costless experimentation of asking a chatbot as many questions as one likes until one receives an

answer (that one likes) is likely to prime users with a sense of insurance (Winter, 2000) that allows for endless experimentation at virtually no or low cost, thereby decreasing their risk aversion. A second mechanism by which repeated usage of ChatGPT could affect risk behavior is what previous research has shown that when humans interact with machines and algorithms, they tend to apply a different set of values: mainly getting less emotional and less concerned with social rules of conduct (for a review see Chugunova and Sele, 2020). Such decreased pro-sociality has been shown in trust games (Schniter *et al.,* 2020), ultimatum games, as well as, in dictator and public goods games (Melo *et al.*, 2016). It is these social values, however, that often restrict reckless behavior and make an individual stop short of taking major risks. Our second hypothesis, therefore, reads as follows:

H2: *The continuous use of ChatGPT over a sustained period of time, would decrease risk aversion.*

Effort is usually a likely collateral of many technological advances. Because most technologies mainly aim at making a machine, software, or an algorithm replace some aspect of human effort (e.g., sending voice message instead of typing it, voice or face recognition instead of pressing buttons), increased reliance on technologies is likely to decrease our perception of the amount of effort required from us, leading to complacency (Wickens *et al.,* 2015). Relevant research has shown how technology has made users less active (Woessner *et al.,* 2021). Such effects are not confined to people with prior low skills: (Galletta *et al.*, 2005) has shown that using a spell-check program makes even individuals with high writing skills miss spelling errors later on. In health care domains, it was also shown that the capacity of professional staff to read mammograms went down (making them miss cancers) if they used computer-aided detection systems (Alberdi *et al.,* 2004). Our third hypothesis, therefore, reads as follows:

H3: *The continuous use of ChatGPT over a sustained period of time, would decrease the level of effort one exerts to complete a task.*

Material and Methods

We conducted a pre-registered field experiment in which 94 senior students from an Egyptian public university participated. We chose a university as our experimental context for two reasons. First, universities and students are usually expected to be primary engines for innovation by developing novel ideas, debating big questions, and having the time to experiment with multiple solutions. Indeed, most breakthroughs have usually originated from university campuses (Lawton-Smith, 2006). Second, the potential negative effect of ChatGPT on university students was among the first worries raised when the technology made its debut. Just as artificial intelligence text generators (AITGs) started to make headlines, many feared that students would use the chatbot to do assignments, write theses, and hence change – or harm – their learning trajectory (for a discussion, see (Lo, 2023; Qadir, 2023; Rudolph *et al.*, 2023; Tlili *et al.*, 2023). This worry deepened after several studies pointed to ChatGPT's reasonable performance in handling academic tasks, such as taking exams and generating academic abstracts (Friederichs *et al.*, 2023). A university campus, therefore, seemed to be an ideal venue for a field experiment on the probable effects of ChatGPT on innovation.

We designed our experiment to mimic educational institutions' most basic concerns about ChatGPT: whether students seeking assistance with the technology would be less innovative over time. Our participants were undergraduate students who had the same major (social sciences) and minor (social science computing) but were enrolled in two different courses: *Social Network Analysis* and *Data Mining*. We used the enrollment in either of these two courses as our unit of

randomization to assign participants to either the treatment or the control. We took this decision to avoid potential across-treatment contamination if the intervention was applied within the same course. In the next section, we present analyses showing that both groups (the treatment and the control) were not significantly different in crucial background variables. The field component of the experiment was administered as follows. Students in both courses (i.e., the treatment and the control) were asked to submit three graded essay assignments over one month during term time in April 2023. The difference between both groups was that members of the treatment group (enrolled in the *Social Network Analysis* course) were asked to write the essays using ChatGPT, whereas, for members of the control group (enrolled in the *Data Mining* course), the ChatGPT option was neither mentioned nor allowed.

The time in which we conducted the experiment enabled us to use a small window of opportunity when access to ChatGPT was still unavailable in Egypt. It was only in November 2023 that Egyptians could create an account and access ChatGPT with local cell phone numbers and Egyptian I.P. addresses. Before November 2023, Egyptians could access ChatGPT via a VPN and foreign phone number, but this step was largely burdensome and, hence, quite rare at that time. For the treatment group, we created ChatGPT accounts for them using foreign cell phone numbers. This context allowed for a field experiment that minimizes the risk of contamination – where participants in the control group might also use ChatGPT – and ensured more reliable results.

One week after all assignments were submitted, the two groups were asked to take part in a lab experiment where they were asked to play an innovation game (Ederer and Manso, 2013) and a risk game (Crosetto and Filippin, 2013). In the innovation game, participants had to innovate – over seven rounds – to increase the sales of a hypothetical lemonade stand by deciding on the values of six parameters (after having received advice from the previous manager):

a. location of the stand (school, business, stadium).

b. lemonade color (green, pink).

c. sugar content (scale from 0 to 10).

d. lemon concentration (scale from 0 to 10).

e. price (scale from 0 to 10).

f. an advertising message or slogan with a maximum of 20 words.

After each round, participants were shown the profit they made using the parameter values they had chosen and were asked whether they wanted to change or keep their parameter choices for the next round. Payoffs were calculated according to the optimal parameters set by (Ederer and Manso, 2013), and any deviation from the optimal parameter values (unknown to the participants) was penalized. We added a further innovation task to the design by [36] that asked participants to write an advertising message or slogan. This additional task was not included in the payoff structure but was used to assess innovation further, as shown below. After the conclusion of the sessions, four external annotators were tasked with rating the advertising messages on a creativity scale from 0 to 10. The annotators were unaware of the research question or the experimental design to ensure coding impartiality.

To measure effort levels, at the beginning of each session, participants were given a sheet of outlined paper to record their strategies for reference over the rounds. Recording previous choices should have been essential for supposedly efficient participants to avoid repeating their choices in future rounds (given the many decisions they had to make in each round). Filling in the cells was voluntary, and participants were not told that recording their decisions and profits would be measured. Instead, they were told that the sheet was for them *if* they wanted to track their choices

and profits. These sheets were collected after the session concluded. We used the number of cells filled out in the sheets as our measure of participants' effort levels.

To measure risk behavior, participants completed a dynamic version of (Crosetto and Filippin, 2013) "Bomb Risk Elicitation Task." The task is a game that presents participants with 100 closed "boxes," 99 of which contain a small monetary reward (the amount in each of these 99 boxes is the same). However, one of the 100 boxes contains a bomb that would wipe out all monetary gains achieved thus far if opened. Once the game starts, a box gets collected every second until the participant clicks "stop" or all 100 boxes are collected. The underlying assumption is that risk-tolerant individuals will collect more boxes while risk-averse individuals will collect fewer. See S1 for the experimental script.

Results

Table I shows a summary of the characteristics of the participants in the control and treatment groups. A total of 94 participants were initially recruited for the experiment. However, one participant was discarded due to missing information during payouts. The majority of the participants were female – which is common in social science departments in Egypt (e.g. (Amin *et al.*, 2018; Haas *et al.*, 2021; Hassan *et al.*, 2023)). The average age was 21, and around 92% were Muslims. Balance tests conducted to compare the demographic composition of the control and treatment group revealed no randomization failures (t-test for age and financial status, chi-square tests for gender, religion, and residence; failed to reject Ho, p-value > 0.05). The experimental script and screenshots can be found in the online supplementary material[5].

[5] Online supplementary material can be accessed at https://osf.io/2pr3j/?view_only=e62f25ff509f4292b0966d08555abb64

[Table I. Subject Demographics]

In the following, we investigate whether the use of ChatGPT for four weeks before the lab experiment influenced (i) the innovative behavior based on the lemonade stand game, (ii) the level of effort calculated from the sheets used to record strategies, and (iii) the risk behavior concluded from the bomb risk elicitation task. Data across sessions was compiled and cleaned. After performing an initial outlier analysis, two anomalous observations were detected in the dataset, and we decided to remove them based on the interquartile range method (1.5xIQR rule).

First: Innovative Behavior based on the Lemonade Stand Game

To measure the level of innovativeness from the lemonade stand game, we started by looking at whether there was a significant treatment effect on any of the three continuous decision parameters (sugar content, lemon concentration, and price) that participants had to decide on in each round. Since we measured innovative behavior by the subject-specific standard deviation, as studied by Ederer and Manso (Ederer and Manso, 2013), we focused only on the three continuous decision parameters (sugar content, lemon concentration, and price). Although this is a random first check – because theoretically, we do not expect one parameter to be more conducive to innovation than the rest – it still is an essential first step to look at the parameters separately. Figure 1 below shows a significant difference in innovative behavior (measured by the subject-specific standard deviation) in the decision regarding lemon concentration (Mann Whitney p-value=0.036). However, there is no significant difference regarding the sugar content or the price when comparing the treatment to the control (Mann Whitney p-value > 0.1).

[Figure 1. Average Level of Innovativeness for Control and ChatGPT Groups (Individual parameters)]

Next, we measure innovativeness by examining the propensity of participants to explore different values of the three parameters combined (sugar content, lemon concentration, and price), thus capturing the degree of novelty in the sales strategy. The innovation metric was calculated as the average, subject-specific standard deviation of strategy choices for the three continuous variables (sugar content, lemon content, and price). Figure 2 shows the average level of innovativeness among the control and ChatGPT treatment groups. The control group had a significantly higher level of innovativeness when compared to the ChatGPT group (Mann Whitney U test, p-value = 0.037), confirming our hypothesis H1.

[Figure 2. Average Aggregate Level of Innovativeness for Control and ChatGPT Groups]

Table II presents the results of regression analyses on two models with the level of innovativeness as the dependent variable after controlling for potential confounders. The coefficient for ChatGPT is negative and statistically significant across the two models (at the 95% confidence level in model 1 and the 90% confidence level in model 2), indicating that the level of innovation for participants in the ChatGPT treatment is *significantly lower* than the control group, after controlling for other variables. The negative values suggest that ChatGPT usage is associated with a decrease in innovation scores by approximately 0.6 to 0.72 standard deviation points, depending on the model. Additionally, living in an urban area seems to have a positive and significant effect on innovation in the second model. Moreover, household income is negatively associated with innovation and is statistically significant, but only at the 0.1 level, suggesting a moderate confidence level that having higher income may reduce innovation scores. This observation could be because participants on higher incomes might not have valued the need to innovate to increase their payoffs as those on lower incomes.

[Table II. Regression Analysis for Level of Innovativeness in Lemonade Stand Game]

The second innovation test we ran was coding the advertising messages that participants were asked to write in each round along an innovation scale from 1 to 10. This task culminated in 265 unique messages (as some participants chose not to change their messages at some rounds). Four coders coded the results after the experiment. Figure 3 below shows no significant difference in the creativity of these messages between the treatment and the control (Mann Whitney U test, p-value > 0.05), running against H1. We also ran other tests tracing innovation that produced insignificant results. Additional results can be found in the online supplementary material.[6]

[Figure 3. Creativity scores of advertising messages between Control and ChatGPT Groups]

Second: Risk Behavior from the Bomb Risk Elicitation Task

To calculate risk behaviors from the Bomb Risk Elicitation Task, the propensity for risk-taking among participants was calculated using the number of boxes they left unopened before they decided to press the *stop* button. A larger number of boxes opened reflects a higher inclination towards risk-taking, as the probability of encountering a bomb that would destroy all earnings increases. A participant was considered risk-loving if s/he opened more than 1/3 of the boxes and risk-averse otherwise. Figure 4 illustrates the proportion of risk lovers in the control group and the ChatGPT treatment. The proportion of risk lovers in the ChatGPT group was higher than that of the control group, as suggested by the hypothesis H2. Moreover, this observed difference is statistically significant at the 90% confidence level (Mann Whitney U test, p-value = 0.079).

[6] Online supplementary material can be accessed at https://osf.io/2pr3j/?view_only=e62f25ff509f4292b0966d08555abb64

[Figure 4. Proportion of Risk Lovers among Control and ChatGPT Groups]

Table III presents regression analyses investigating the likelihood of risk-loving behavior. Across both models, the coefficients for the ChatGPT treatment are positive (at the 90% confidence level), indicating that individuals in the ChatGPT treatment group *are more likely* to become risk lovers than those in the control group (which is in line with the hypothesis). Moreover, these results are statistically significant. In the second model, being a female is negative and statistically significant, indicating a lower likelihood of being a risk-loving person.

[Table III. Regression Analysis for Risk Loving Behavior in the Bomb Risk Elicitation Task]

Third: Level of Effort as Elicited from the Sheet for Recording Strategies

To assess the level of effort participants exerted (measured by their recording of the sales strategies employed at each of the seven rounds), we calculated the number of cells filled out by each participant at each period (min=0 and max=8). As mentioned above, filling in the cells was voluntary, and participants were not told that recording their decisions and profits would be measured. Figure 5 shows the sheet given to participant at the beginning of the session if they wanted to record their choices and profits over the rounds.

[Figure 5. Sheet for Eliciting Effort]

Figure 6 displays the average effort exerted by the control and ChatGPT groups. It shows that the average effort exerted by participants in the control group is slightly higher than that exerted by the ChatGPT treatment group, which runs in line with our hypothesis H3. However, this difference is statistically insignificant (Mann Whitney U test, p-value > 0.05).

[Figure 6. Average Level of Effort for Control and ChatGPT Groups]

Table IV presents the results of regression analyses on three models with the level of effort as the dependent variable. Across all models, the coefficient for the ChatGPT treatment is negative (as expected by the hypothesis), suggesting that participants using ChatGPT exerted slightly less effort than the control group. However, these results are not statistically significant in any of the models. The coefficient for the period is negative and is statistically significant in models 2 and 3 at the 0.1 significance level, implying that as the game progressed, participants tended to exert less effort, possibly due to fatigue or diminishing interest. In model 3, residing in an urban area is associated with a significantly higher level of effort, suggesting that urban participants were more engaged or able to exert more effort in the game.

[Table IV. Regression Analysis for Level of Effort Exerted in Lemonade Stand Game]

Putting together the details of the full picture, our results seem to be nuanced. Exposure to ChatGPT over a few weeks in doing assignments was shown to significantly decrease innovation, as measured by our lemonade stand game. Moreover, the use of ChatGPT positively affected risk-taking behavior compared to the control, and the effect was statistically significant. On the other hand, ChatGPT also had a negative impact on the level of effort exerted in a task, albeit not statistically significant. However, the fact that one of our two statistically significant findings is at the 90% confidence level highlights the need for further testing.

Discussion and Conclusion

The fast spread of ChatGPT has changed the nature of human-computer interaction, providing users with sophisticated conversational interfaces that respond intelligently to different prompts and queries. In this paper, we examined the effect of continuous reliance on such a technological breakthrough when doing assignments on undergraduate students' innovative behavior, risk attitudes, and readiness to exert effort. Our findings have shown significant results on innovation and risk but not on effort levels.

We would like to use this discussion section to discuss some caveats, possible limitations of our study, and venues for future research. On caveats, despite the paper's focus and its pre-registered hypothesis on possible negative effects of ChatGPT, it is still important to highlight that A.I. applications and AITGs will likely have significant advantages. For example, taking the social clues out of the decision-making produced less discriminatory decisions than human decisions (e.g. Hoffman *et al.*, 2018; Kleinberg *et al.*, 2018). A.I. also has considerable potential in dealing with a size of data that is too large for humans to handle effectively. Depending on usage techniques, even ChatGPT itself could contribute to innovation development if it succeeds – by providing a vast amount of valuable information – to foster a culture of curiosity and continuous exploration (see Romero-Rodríguez *et al.*, 2023) among its users. This study, therefore, does not deny the multiple advantages that A.I. and AITGs can generate.

On limitations, the findings of our study should be read in light of possible design restrictions. Firstly, there is a relatively small sample size to which we were restricted by the number of students already registered in the two courses we had access to. Secondly, the worry about the spill-over effect between treatment and control made us take course registration as the criterion for randomization, which – ideally – is not the best randomization strategy. Therefore,

further testing of our preliminary findings whilst improving on these two possible limitations is essential.

Finally, on venues for future research, although field experiments will always have the advantage of high ecological validity, testing the effect of AITGs on behavior could also benefit from the controlled environment of lab experiments. In such designs, spillover worries would be minimal. It is also important to note that we only tracked one dimension of innovation: innovation related to problem-solving. There are, however, many other dimensions of innovation. One such dimension, for example, is the active reflection that individuals engage with occasionally (Bieser, 2022). Such aspects of creativity and innovation require a free mindset that does not explicitly think about a specific problem but wanders around and accumulates ideas that could be used later (Dyer *et al.*, 2011).

References

Alberdi, E., Povyakalo, A., Strigini, L. and Ayton, P. (2004), "Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography", *Academic Radiology*, Vol. 11 No. 8, pp. 909–918, doi: https://doi.org/10.1016/j.acra.2004.05.012.

Altmann, E.M., Trafton, J.G. and Hambrick, D.Z. (2014), "Momentary interruptions can derail the train of thought.", *Journal of Experimental Psychology. General*, United States, Vol. 143 No. 1, pp. 215–226, doi: 10.1037/a0030986.

Amabile, T.M. (2019), "Creativity, Artificial Intelligence, and a World of Surprises", *Academy of Management Discoveries*, Academy of Management, Vol. 6 No. 3, pp. 351–354, doi: 10.5465/amd.2019.0075.

Amin, E., Abouelela, M. and Soliman, A. (2018), "The role of heterogeneity and the dynamics of voluntary contributions to public goods: An experimental and agent-based simulation analysis", *Journal of Artificial Societies and Social Simulation*, Jasss, Vol. 21 No. 1.

Arrow, K.J. (1969), "Classificatory Notes on the Production and Transmission of Technological Knowledge", *The American Economic Review*, American Economic Association, Vol. 59 No. 2, pp. 29–35.

Baska, M. (2018), "AI will give workers back two weeks a year, says research", *People Management*, 14 September.

Bieser, J. (2022), *Creative through AI – How Artificial Intelligence Can Support the*

Development of New Ideas, *Gottlieb Duttweiler Institute Research Paper No. Forthcoming*, doi: 10.59986/CCHA2271.

Bloom, N., Jones, C.I., Van Reenen, J. and Webb, M. (2020), "Are Ideas Getting Harder to Find?", *American Economic Review*, Vol. 110 No. 4, pp. 1104–1144, doi: 10.1257/aer.20180338.

Chen, J., Liang, Y., Mai, C., Zhong, X. and Qu, C. (2016), "General Deficit in Inhibitory Control of Excessive Smartphone Users: Evidence from an Event-Related Potential Study.", *Frontiers in Psychology*, Switzerland, Vol. 7, p. 511, doi: 10.3389/fpsyg.2016.00511.

Chugunova, M. and Sele, D. (2020), "We and It: An Interdisciplinary Review of the Experimental Evidence on Human-Machine Interaction", *Center for Law & Economics Working Paper Series*, Center for Law & Economics, ETH Zurich, doi: 10.3929/ethz-b-000442053.

Crosetto, P. and Filippin, A. (2013), "The 'bomb' risk elicitation task", *Journal of Risk and Uncertainty*, Springer, Vol. 47 No. 1, pp. 31–65.

Dahl, D.W. and Moreau, P. (2002), "The Influence and Value of Analogical Thinking during New Product Ideation", *Journal of Marketing Research*, SAGE Publications Inc, Vol. 39 No. 1, pp. 47–60, doi: 10.1509/jmkr.39.1.47.18930.

Dyer, J., Gregersen, H. and Christensen, C. (2011), *The Innovator's DNA: Mastering the Five Skills of Disruptive Innovators*, Harvard Business Review Press, Boston.

Ederer, F. and Manso, G. (2013), "Is Pay for Performance Detrimental to Innovation?", *Management Science*, INFORMS, Vol. 59 No. 7, pp. 1496–1513.

Farhi, F., Jeljeli, R., Aburezeq, I., Dweikat, F.F., Al-shami, S.A. and Slamene, R. (2023), "Analyzing the students' views, concerns, and perceived ethics about chat GPT usage", *Computers and Education: Artificial Intelligence*, Elsevier, p. 100180.

Firth, J.A., Torous, J. and Firth, J. (2020), "Exploring the Impact of Internet Use on Memory and Attention Processes.", *International Journal of Environmental Research and Public Health*, Switzerland, Vol. 17 No. 24, doi: 10.3390/ijerph17249481.

Fischetti, M. (2016), "Social Technologies Are Making Us Less Social", *Scientific American*, October.

Friederichs, H., Friederichs, W.J. and März, M. (2023), "ChatGPT in medical school: how successful is AI in progress testing?", *Medical Education Online*, United States, Vol. 28 No. 1, p. 2220920, doi: 10.1080/10872981.2023.2220920.

Galletta, D.F., Durcikova, A., Everard, A. and Jones, B.M. (2005), "Does spell-checking software need a warning label?", *Communications of the ACM*, Vol. 48 No. 7, pp. 82–86, doi: 10.1145/1070838.1070841.

von Garrel, J. and Mayer, J. (2023), "Artificial Intelligence in studies—use of ChatGPT and AI-based tools among students in Germany", *Humanities and Social Sciences Communications*, Vol. 10 No. 1, p. 799, doi: 10.1057/s41599-023-02304-7.

Haas, N., Hassan, M., Mansour, S. and Morton, R.B. (2021), "Polarizing information and support for reform", *Journal of Economic Behavior & Organization*, Vol. 185, pp. 883–901, doi: https://doi.org/10.1016/j.jebo.2020.10.013.

Hassan, M., Amin, E., Mansour, S. and Voigt, S. (2023), "Incentivizing cooperation against a norm of defection: Experimental Evidence from Egypt", *Journal of Behavioral and Experimental Economics*, Vol. 107, p. 102121, doi: https://doi.org/10.1016/j.socec.2023.102121.

Hoffman, M., Kahn, L.B. and Li, D. (2018), "Discretion in Hiring", *The Quarterly Journal of Economics*, Vol. 133 No. 2, pp. 765–800, doi: 10.1093/qje/qjx042.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. and Mullainathan, S. (2018), "Human Decisions and Machine Predictions", *The Quarterly Journal of Economics*, Vol. 133 No. 1, pp. 237–293, doi: 10.1093/qje/qjx032.

Lawton-Smith, H. (2006), *Universities, Innovation and the Economy*, Taylor & Francis.

Lo, C.K. (2023), "What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature", *Education Sciences*, doi: 10.3390/educsci13040410.

Manso, G. (2011), "Motivating Innovation", *The Journal of Finance*, John Wiley & Sons, Ltd, Vol. 66 No. 5, pp. 1823–1860, doi: https://doi.org/10.1111/j.1540-6261.2011.01688.x.

Melo, C. De, Marsella, S. and Gratch, J. (2016), "People Do Not Feel Guilty About Exploiting Machines", *ACM Transactions on Computer-Human Interaction*, Vol. 23 No. 2, pp. 1–17, doi: 10.1145/2890495.

Ophir, E., Nass, C. and Wagner, A.D. (2009), "Cognitive control in media multitaskers.", *Proceedings of the National Academy of Sciences of the United States of America*, United States, Vol. 106 No. 37, pp. 15583–15587, doi: 10.1073/pnas.0903620106.

Ortiz-Ospina, E. and Roser, M. (2023), "Loneliness and social connections", *Our World in Data*.

Pasquale, F. (2015), *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press.

Qadir, J. (2023), "Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education", *2023 IEEE Global Engineering Education Conference (EDUCON)*, IEEE, pp. 1–9.

Romero-Rodríguez, J.-M., Ramírez-Montoya, M.-S., Buenestado-Fernández, M. and Lara-Lara, F. (2023), "Use of ChatGPT at University as a Tool for Complex Thinking: Students' Perceived Usefulness", *Journal of New Approaches in Educational Research; Vol 12, No 2 (2023)*, doi: 10.7821/naer.2023.7.1458.

Rowell, D. and Connelly, L.B. (2012), "A History of the Term 'Moral Hazard'", *The Journal of Risk and Insurance*, [American Risk and Insurance Association, Wiley], Vol. 79 No. 4, pp. 1051–1075.

Rudolph, J., Tan, S. and Tan, S. (2023), "ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?", *Journal of Applied Learning & Teaching*, Vol. 6 No. 1,

doi: 10.37074/jalt.2023.6.1.9.

Schniter, E., Shields, T.W. and Sznycer, D. (2020), “Trust in humans and robots: Economically similar but emotionally different”, *Journal of Economic Psychology*, Vol. 78, p. 102253, doi: https://doi.org/10.1016/j.joep.2020.102253.

Sternberg, R.J. and Lubart, T.I. (1999), “The concept of creativity: Prospects and paradigms”, *Handbook of Creativity*, Vol. 1 No. 3–15.

Tanil, C.T. and Yong, M.H. (2020), “Mobile phones: The effect of its presence on learning and memory”, *PLOS ONE*, Public Library of Science, Vol. 15 No. 8, p. e0219233.

Tlili, A., Shehata, B., Adarkwah, M.A., Bozkurt, A., Hickey, D.T., Huang, R. and Agyemang, B. (2023), “What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education”, *Smart Learning Environments*, Vol. 10 No. 1, p. 15, doi: 10.1186/s40561-023-00237-x.

Wachter, S. and Mittelstadt, B. (2019), “A right to reasonable inferences: re-thinking data protection law in the age of big data and AI”, *Columbia Business Law Review*, HeinOnline, p. 494.

Weitzman, M. (1978), *Optimal Search for the Best Alternative*, Vol. 78, Department of Energy.

Wickens, C.D., Clegg, B.A., Vieane, A.Z. and Sebok, A.L. (2015), “Complacency and Automation Bias in the Use of Imperfect Automation.”, *Human Factors*, United States, Vol. 57 No. 5, pp. 728–739, doi: 10.1177/0018720815581940.

Winter, R.A. (2000), “Optimal insurance under moral hazard”, *Handbook of Insurance*, Springer, pp. 155–183.

Woessner, M.N., Tacey, A., Levinger-Limor, A., Parker, A.G. and Levinger, I. (2021), “The evolution of technology and physical inactivity: the good, the bad, and the way forward”, *Frontiers in Public Health*, Frontiers, Vol. 9, p. 655491.

Figure 1. Average Level of Innovativeness for Control and ChatGPT Groups (Individual parameters)

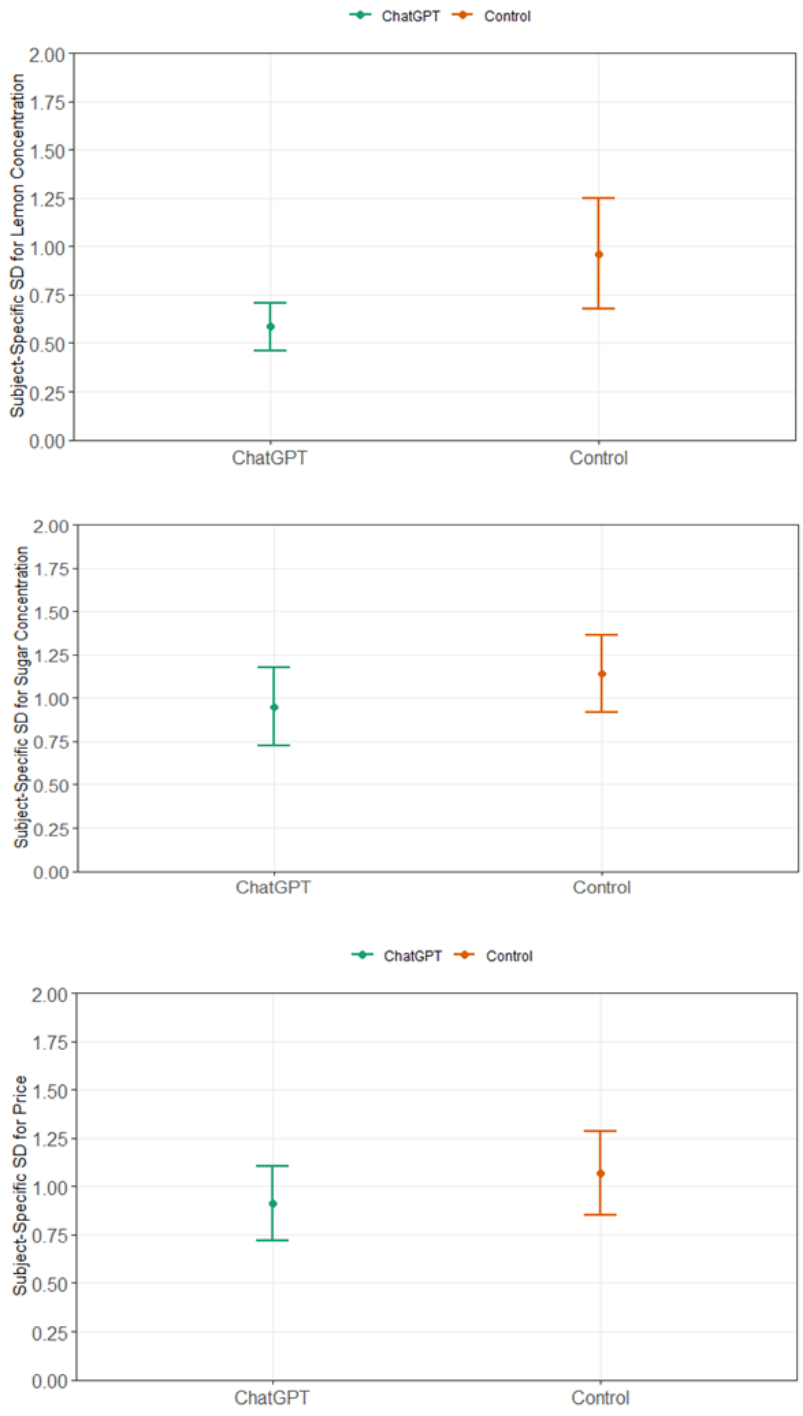


Figure 2. Average Aggregate Level of Innovativeness for Control and ChatGPT Groups

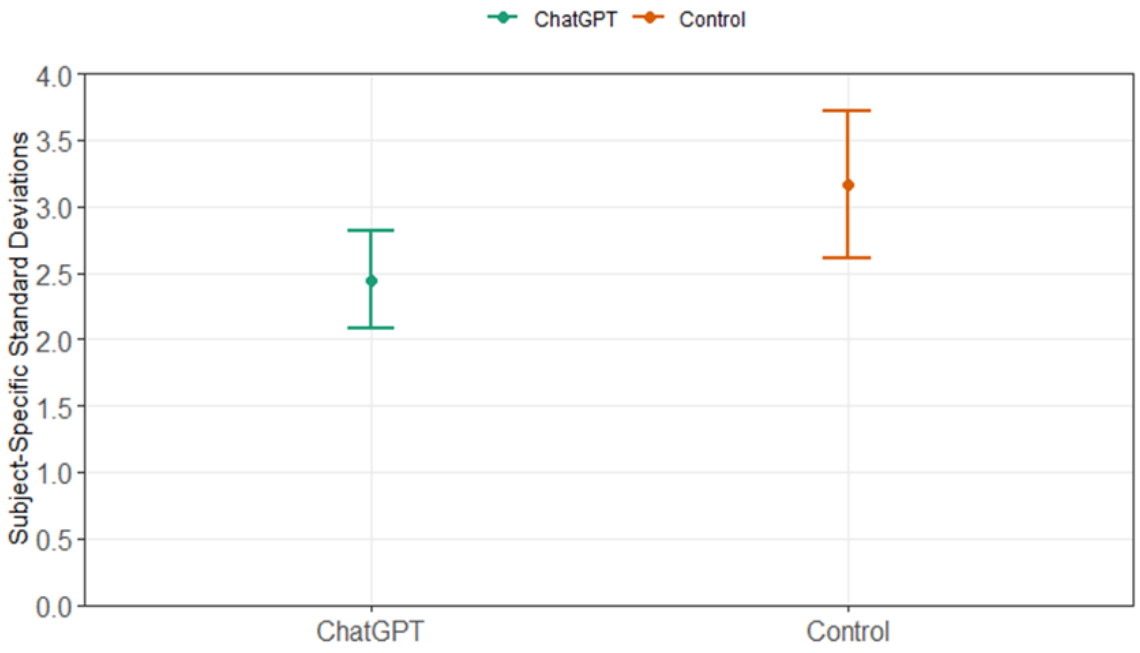


Figure 3. Creativity scores of advertising messages between Control and ChatGPT Groups

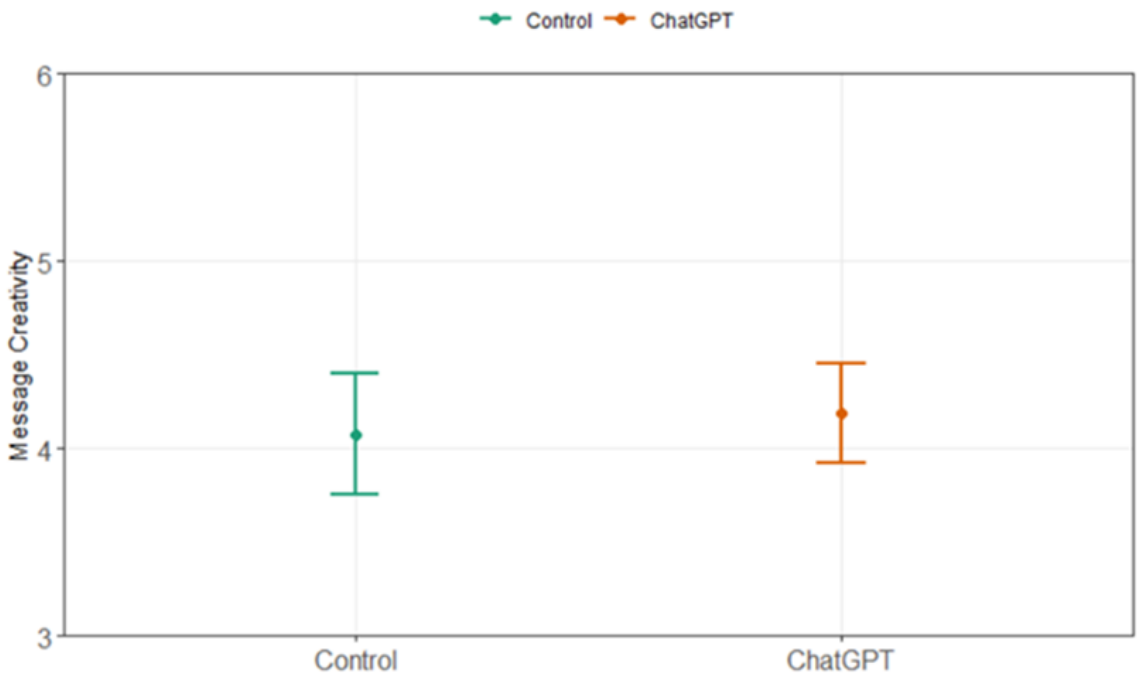


Figure 4. Proportion of Risk Lovers among Control and ChatGPT Groups

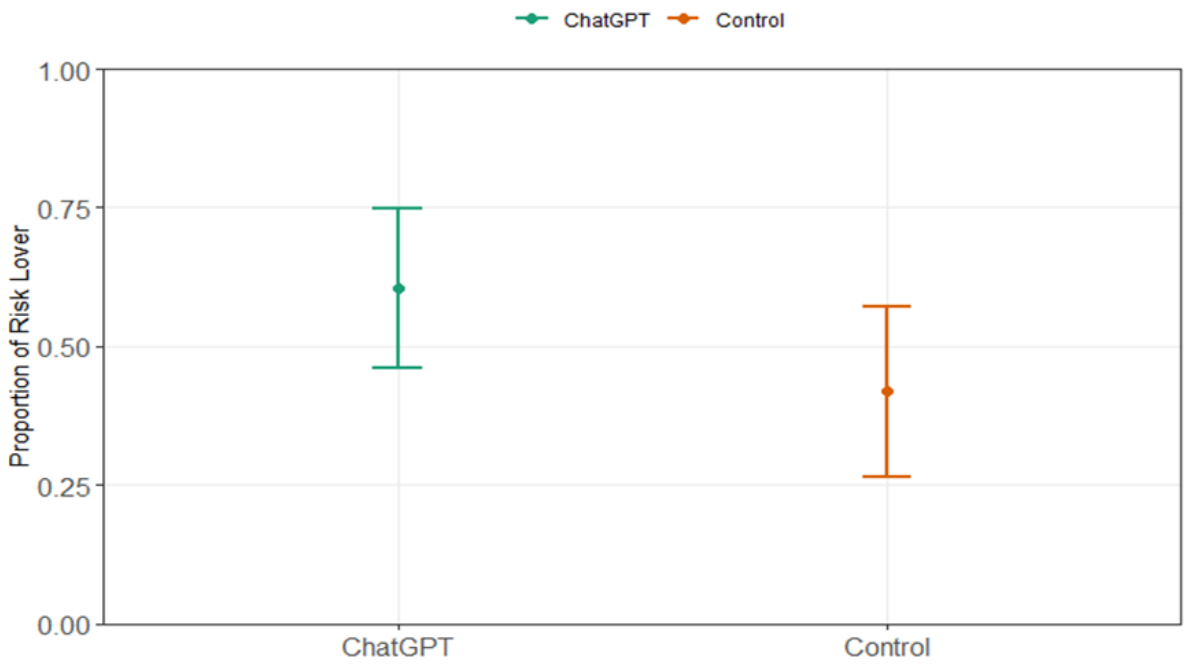


Figure 5. Sheet for Eliciting Effort

Record of Strategies/Results							
Round	Stand location	Sugar content	Lemon concentration	Lemonade color	Price	Advertising message	Profits
1							
2							
3							
4							
5							
6							
7							

Figure 6. Average Level of Effort for Control and ChatGPT Groups

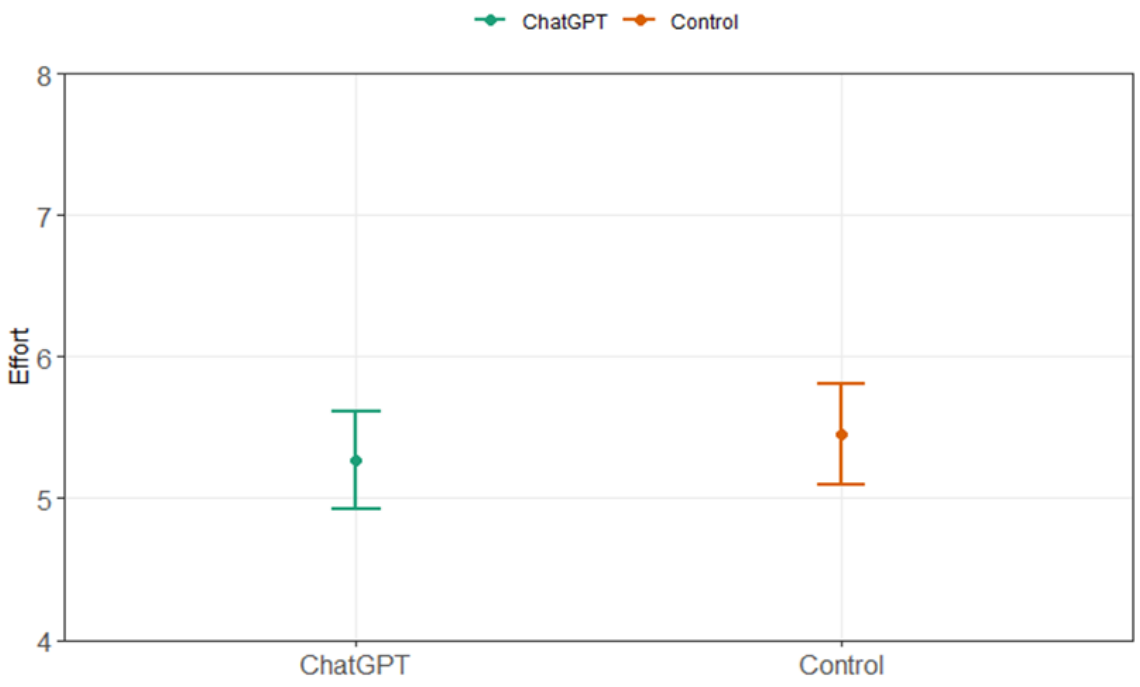


Table I. Subject Demographics

	Control	ChatGPT	Total
Gender			
Male (%)	25%	16%	20.5%
Female (%)	75%	84%	79.5%
Age (SD)	21.1 (1)	21 (0.74)	21 (0.9)
Household Income (SD)	2.42 (0.62)	2.54 (0.61)	2.5 (0.6)
Residence			
Urban (%)	100%	96%	97.8%
Rural (%)	0%	4%	2.2%
Religion			
Muslim (%)	86%	98%	92.5%
Christian (%)	14%	2%	7.5%
Observations	43	50	93

Table II. Regression Analysis for Level of Innovativeness in Lemonade Stand Game

	Dependent variable:	
	Exploration of sales strategies	
	(1)	(2)
ChatGPT	-0.719**	-0.604*
	(0.326)	(0.328)
Female		-0.320
		(0.462)
Age		-0.030
		(0.173)
Urban		0.673**
		(0.322)
Financials		-0.510*
		(0.273)
Constant	3.167***	4.081
	(0.271)	(3.529)
Observations	91	91

Notes: Results are clustered by Subject ID, *** p<0.01, ** p<0.05, * p<0.1

Table III. Regression Analysis for Risk Loving Behavior in the Bomb Risk Elicitation Task

	Dependent variable:	
	Risk Lover	
	(1)	(2)
ChatGPT	0.751*	0.818*
	(0.427)	(0.449)
Female		-1.135*
		(0.631)
Age		-0.002
		(0.252)
Urban		0.236
		(1.484)
Household Income		0.639
		(0.400)
Constant	-0.329	-0.608
	(0.309)	(5.386)
Observations	91	91
Akaike Inf. Crit.	126.909	130.316
Note:		*** p<0.01, ** p<0.05, * p<0.1

Table IV. Regression Analysis for Level of Effort Exerted in Lemonade Stand Game

	Dependent variable:		
	Effort		
	(1)	(2)	(3)
ChatGPT	-0.184	-0.184	-0.035
	(0.615)	(0.615)	(0.611)
Female			-0.206
			(0.852)
Age			-0.411
			(0.392)
Urban			4.285***
			(0.681)
Household Income			-0.187
			(0.546)
Period		-0.216***	-0.216***
		(0.041)	(0.041)
Constant	5.449***	6.313***	11.106
	(0.449)	(0.432)	(7.963)
Observations	637	637	637

Notes: Results are clustered by Subject ID, *** p<0.01, ** p<0.05, * p<0.1

Experimental Script and Screenshots

Screen 1: General Instructions

Thank you for participating in this scientific study that attempts to examine certain behaviors among individuals.

At the end of today's session, you will earn a sum of money, which will be determined based on your decisions during the session.

The profit you will achieve will be converted from points to Egyptian pounds according to this equation:

1 point = 1 Egyptian pound

Please pay attention to the instructions that are displayed:

- Participants in this study are not allowed to talk to each other until after the session has ended. Any violation of this rule may result in the cancellation of the study for the offending participant.
- If you have any inquiries, please raise your hand and the person in charge of the study will come to answer your questions.
- All data you provide will be treated confidentially and only used for scientific research.
- This study is divided into 3 sections:

  In the first and second sections, you will be asked to make some decisions, which you will enter on a computer.

In the third section, you will answer some identification questions.

- The other participants will not know your decisions and the amounts you earn; they are private to you and will be treated confidentially by the researchers.
- It is not allowed to use mobile phones during the study.

Do you agree to participate in this study?

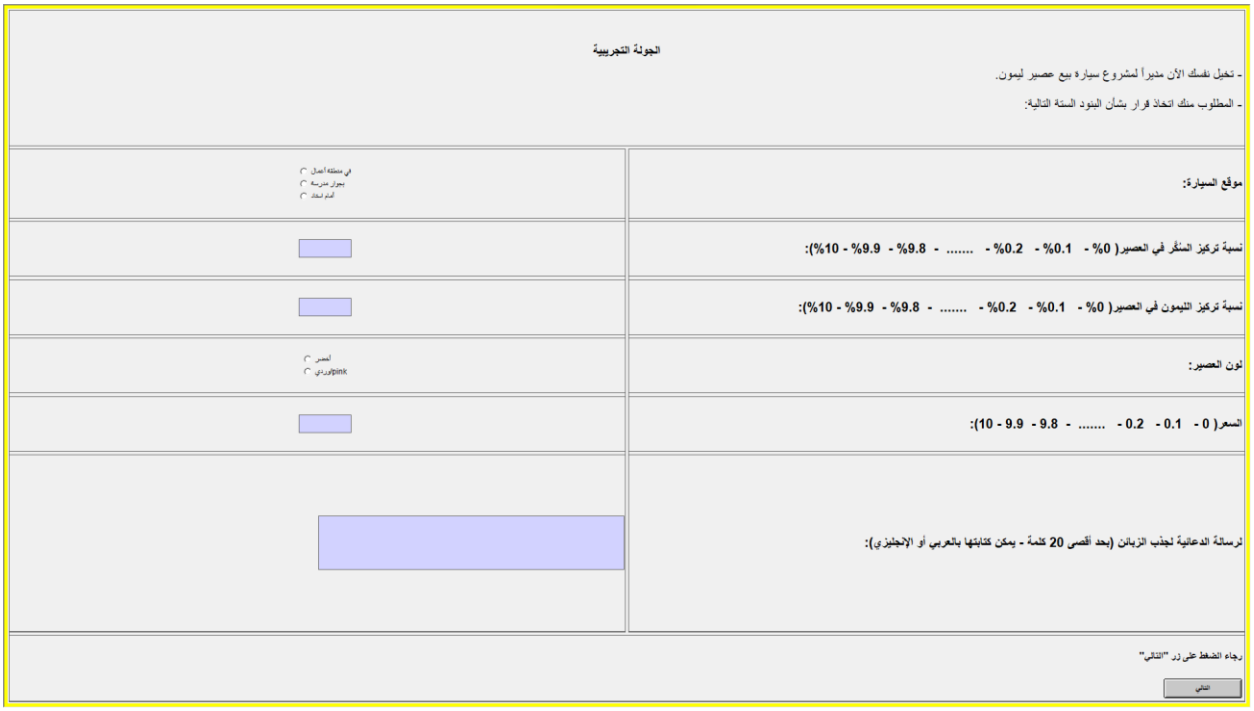
- Yes
- No
- I don't know
- I abstain from answering

Screen 2: Instructions for your role in the experiment

- In this study, you will be a person running a small business, specifically a mobile lemonade stand.
- There will be 7 rounds, during which you will have to decide how to manage this project.
- The six decisions you will make are: (1) choosing the location of the stand, (2) choosing the amount of sugar to be added to the lemonade, (3) choosing the concentration of lemon juice, (4) choosing the color of the lemonade, (5) setting the price of a cup of lemonade, and (6) crafting a promotional message to attract customers, with a maximum of 20 words.
- The decisions you make in any round will become the default choices for the following rounds unless you decide to change them in subsequent rounds.

- At the end of each round, you will learn about the profits you have made based on your choices.
- You will also receive some customer feedback, which may help you make choices in the following rounds.
- Before you start making decisions, you will be presented with a letter from the previous manager of the project explaining the strategy he used while managing the project, which was successful at a specific location. This strategy offers a detailed explanation of the previous manager's recipe, emphasizing that he did not try other recipes at other locations.
- To ensure you understand your role in the study, we will present you with a trial round before starting the actual rounds to familiarize you with the choices you will be asked to make while managing the lemonade project.
- Please click the 'Next' button to start the trial round.

Screen 3: Practice Round



• Imagine yourself now as the manager of a mobile lemonade stand project.
• You are required to make a decision regarding the following six items:

Car location:	○ In a business district ○ Next to a school ○ In front of a stadium

Percentage of sugar concentration in the lemonade (0% – 0.1% – 0.2% - ..... – 9.8% - 9.9% - 10%):	
Percentage of lemon concentration in the lemonade (0% – 0.1% – 0.2% - ..... – 9.8% - 9.9% - 10%):	
Juice color:	○ Green ○ Pink
Price (0 – 0.1– 0.2 - ..... – 9.8 - 9.9 - 10):	
Promotional message to attract customers (up to 20 words – can be written in Arabic or English):	

Screen 4: Results for Practice Round

لقد قمت باختيار ما يلي:

موقع السيارة:	في منطقة أعمال
نسبة تركيز السُكّر في العصير	10.0
نسبة تركيز الليمون في العصير	10.0
لون العصير:	وردي/pink
السعر	10.0
الرسالة الدعائية لجذب زبائن:	hbngfdz

بناء على خياراتك، فإن نتيجة الجولة التدريبية:

الأرباح المحققة في الجولة التدريبية: 12.5 نقطة

رأي العملاء في الليمون المقدم: أخبرك بعض عملائك بأن عصير الليمون به نسبة سكر كبيرة

من فضلك أدخل الكود الذي سيظهره المسؤول عن الدراسة لبدء الجولات الفعلية

استمرار

These were your choices:

Car location:	…
Percentage of sugar concentration in the lemonade (0% – 0.1% – 0.2% - ..... – 9.8% - 9.9% - 10%):	…

Percentage of lemon concentration in the lemonade (0% – 0.1% – 0.2% - ..... – 9.8% - 9.9% - 10%):	…
Juice color:	…
Price (0 – 0.1– 0.2 - ..... – 9.8 - 9.9 - 10):	…
Promotional message to attract customers (up to 20 words – can be written in Arabic or English):	…….

Based on your choices, the results of the practice round are as follows:
Total profits for the practice round: … points
Customers opinion regarding the lemonade: ……

Screen 5: Start of the experiment + Ex-Manager Letter

• Imagine yourself now as the manager of a mobile lemonade stand project.
• You are required to make decisions about: the location of the stand, the sugar concentration, the lemon concentration, the color of the juice, the price, and crafting a promotional message to attract customers.
• Before making your decisions on these items, please read the letter sent to you from the previous manager:

Dear New Manager,

I am providing you with the following instructions that might help manage the lemonade juice project. These guidelines are based on my previous experience running this project.

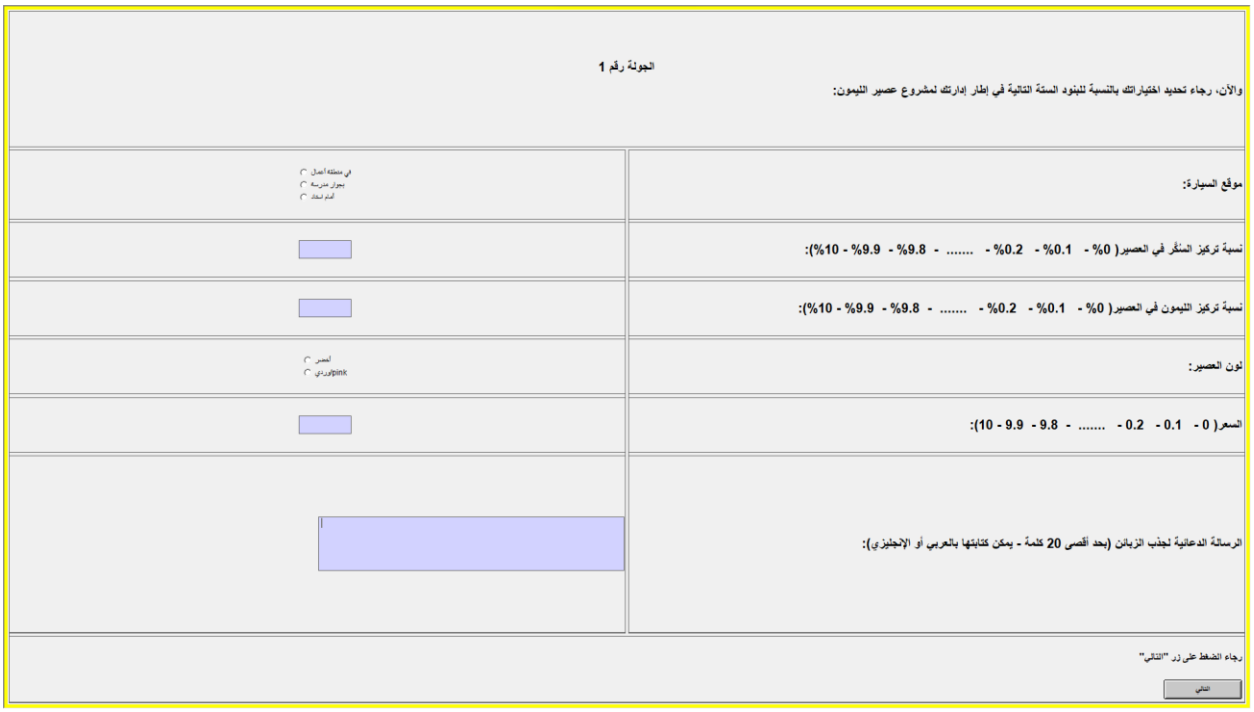
When I managed the lemonade juice project, I followed this recipe:
Location: Business district
Sugar concentration: 3%
Lemon concentration: 7%
Juice color: Green
Price: 8.2 pounds

I achieved an average profit of 85 pounds per round with this recipe.

I have tried some, but not all, options for (1) sugar percentage, (2) lemon percentage, (3) color, and (4) price. The options mentioned above are the best ones I tried that yielded profits.
I also did not try other locations, which may require very different strategies. Nor did I craft any promotional messages.

Best regards,
The Ex- Manager

Screen 6: Rounds of the game



Round 1

And now, please specify your choices for the following six items as part of your management of the lemonade stand project:

Car location:	○ In a business district ○ Next to a school ○ In front of a stadium
Percentage of sugar concentration in the lemonade (0% – 0.1% – 0.2% - ..... – 9.8% - 9.9% - 10%):	
Percentage of lemon concentration in the lemonade (0% – 0.1% – 0.2% - ..... – 9.8% - 9.9% - 10%):	
Juice color:	○ Green ○ Pink
Price (0 – 0.1– 0.2 - ..... – 9.8 - 9.9 - 10):	
Promotional message to attract customers (up to 20 words – can be written in Arabic or English):	

Screen 7: Results for each round

These were your choices:

Car location:	…
Percentage of sugar concentration in the lemonade (0% – 0.1% – 0.2% - ..... – 9.8% - 9.9% - 10%):	…
Percentage of lemon concentration in the lemonade (0% – 0.1% – 0.2% - ..... – 9.8% - 9.9% - 10%):	…
Juice color:	…
Price (0 – 0.1– 0.2 - ..... – 9.8 - 9.9 - 10):	…
Promotional message to attract customers (up to 20 words – can be written in Arabic or English):	…….

Based on your choices, the results of round 1 is as follows:
Total profits for the round 1: … points
Customers opinion regarding the lemonade: ……

Screen 8: Part 2 - Bomb Game Instructions

In this exercise, you will see a screen containing 100 boxes.

You will earn 1 Egyptian pound for each box that is opened. One box will be opened every second, starting from the corner at the top left of the screen. A counter will show you the number of boxes that have been opened, which equals your earnings. However, these profits are not guaranteed, as there is a time bomb in one of the boxes, which could destroy everything you have earned from the opened boxes. You will not know where this bomb is located until the exercise is completed. All you know is that it is in one of the boxes, randomly determined by the program with equal probability.

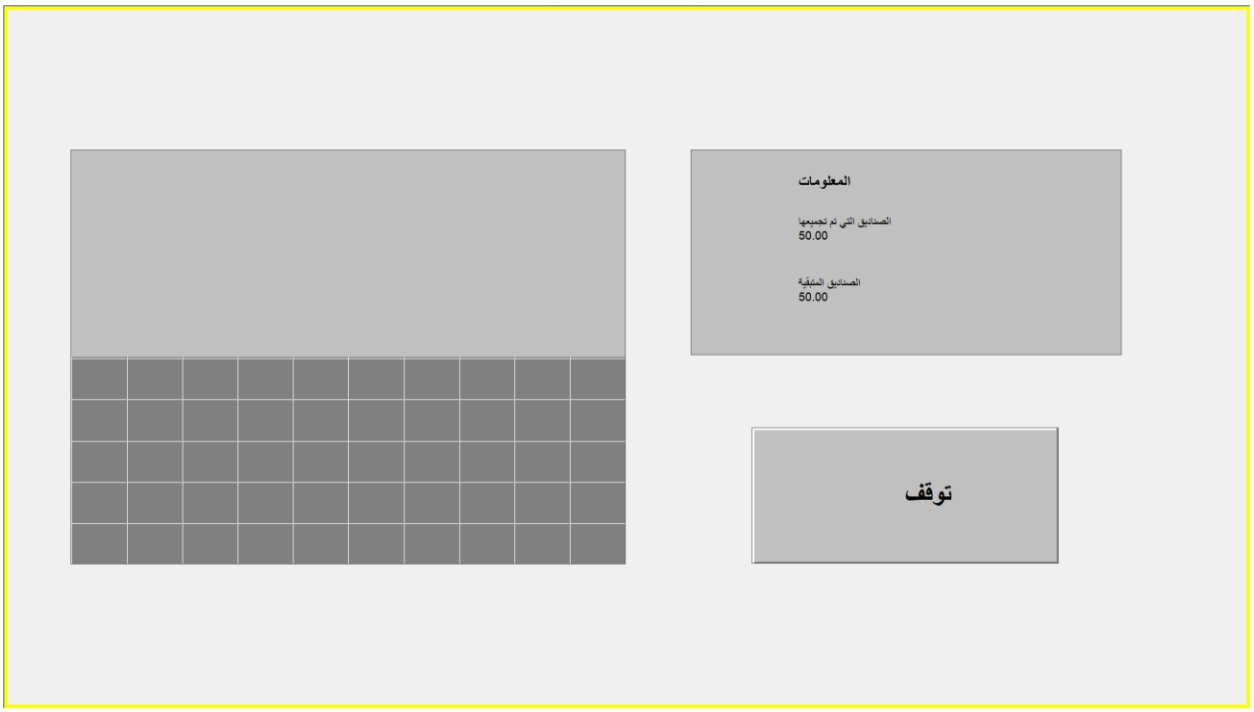
Your main task is to decide when to stop opening boxes and end the exercise. When you want to stop opening boxes, you can press the "Stop" button anytime.

After the boxes have been opened:

1. If it happens that one of the opened boxes contains the bomb, then you will lose everything you have collected and will not win anything.
2. If it happens that the bomb is not in any of the opened boxes, you will win an amount equal to the number of boxes that have been opened.

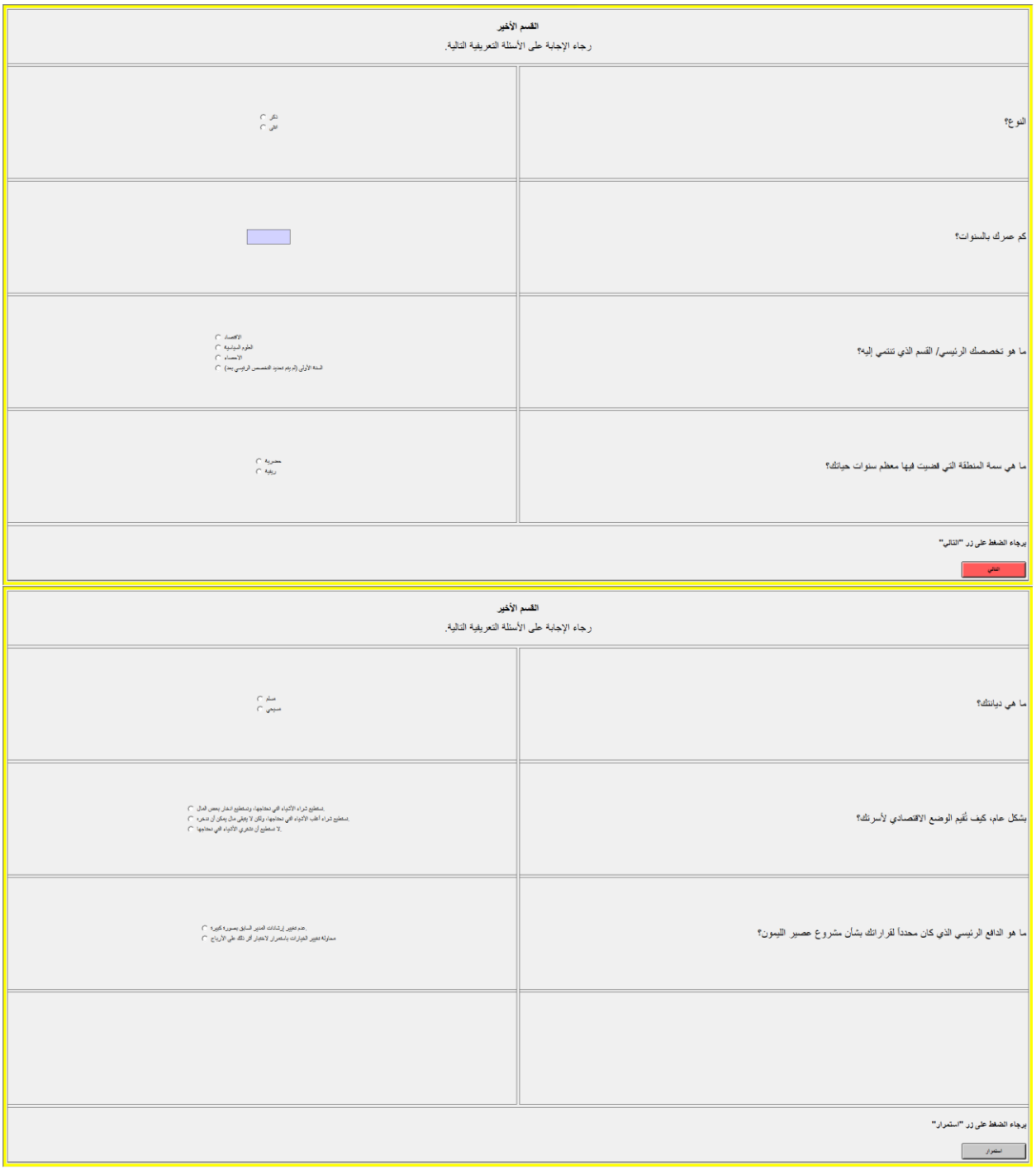
We will now do a trial round, after which the actual exercise will begin.

Screen 9: Bomb Game



Information
Number of boxes opened: …
Number of boxes remaining: …

Screen 10: The Final Section - Post-experimental Questionnaire



Please answer the following questions.

Gender?

- Male
- Female

How old are you in years?
Answer: ... years

What is the major/main department to which you belong?

- Economics
- Political Science
- Statistics
- First Year (Major not yet determined)

What is the characteristic of the area where you spent most years of your life?

- Urban
- Rural

What is your religion?

- Muslim
- Christian

In general, how do you rate your family's economic situation?

- We can buy the things we need and save some money.
- We can buy most of the things we need, but no money is left to save.
- We cannot buy the things we need.

What was the main motivator for your decisions regarding the lemonade stand project?

- Not making significant changes to the previous manager's instructions
- Constantly trying to change options to test their impact on profits

S2 Appendix
Other measures of Innovation

Regarding other measures of a subject's innovative behaviour, we tried the following:

a) ***Profits earned after each round,*** as a proxy of how close subjects came to finding the optimal strategy (Ederer and Manso, 2013) which turned out to be weakly significant at the 90% confidence level (Mann Whitney U-test, p-value = 0.097). See Fig A.1 and table A.1.

Fig A.1. Average Profit across control and ChatGPT groups

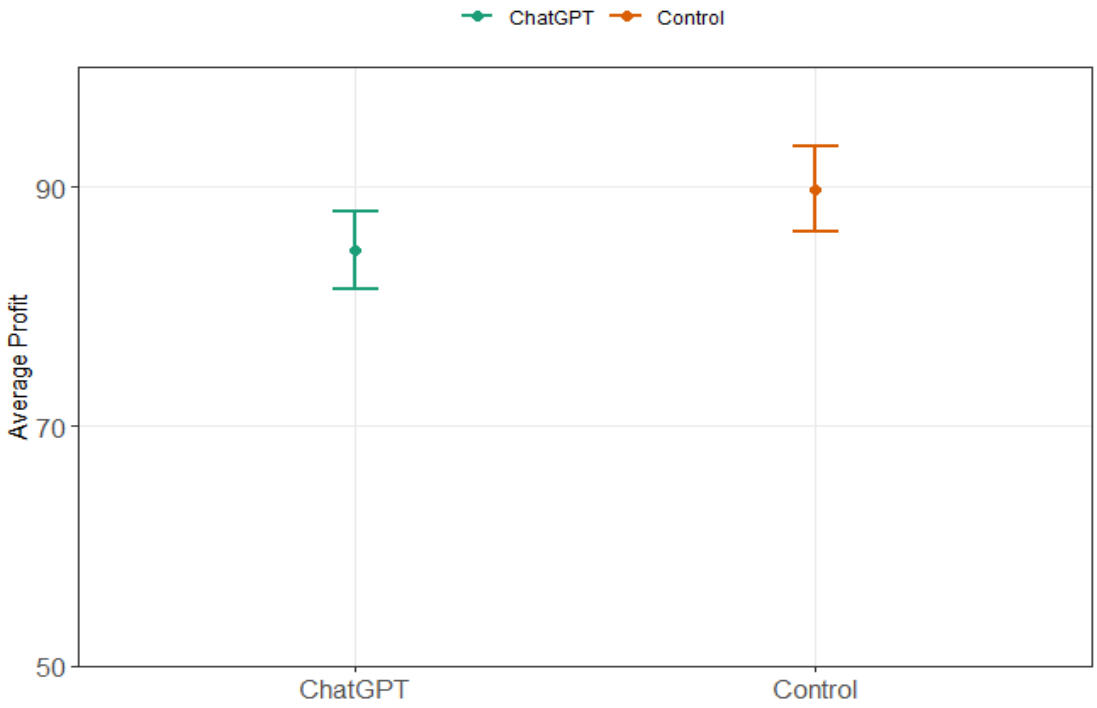


Table A.1. Regression Analysis for Profits Earned in Lemonade Stand Game

	Dependent variable:		
	Profit		
	(1)	(2)	(3)
ChatGPT	-5.103	-5.103	-4.540
	(4.951)	(4.951)	(5.073)
Female			5.420
			(6.334)
Age			0.837
			(3.096)
Urban			15.878***
			(5.263)
Household Income			-1.948
			(4.502)
Period		3.805***	3.805***
		(0.574)	(0.574)
Constant	89.799***	74.580***	39.792
	(3.532)	(3.386)	(61.427)
Observations	637	637	637

Notes:

1) **** p<0.01, ** p<0.05, * p<0.1*
2) *First column: presents results for fixed effects.*
3) *Second and third column: Results are clustered by subject ID*
4) *Robust standard errors clustered by participants: reported in parentheses to account for correlation between observations within the cluster (i.e., for same subject)*

b) ***Sum of changes in sales strategy,*** as a proxy of the subject's willingness to innovate. This variable was calculated as follows: sum of changes = Change_Location + Change_Color + absolute(Change_LemonConcentration) + absolute(Change_SugarConcentration) + absolute(Change_Price). It turned out to be weakly significant at the 90% confidence level (Mann Whitney p-value = 0.067). See Fig A.2 and Table A.2.

Figure A.2: Sum of changes in sales strategy across control and ChatGPT groups

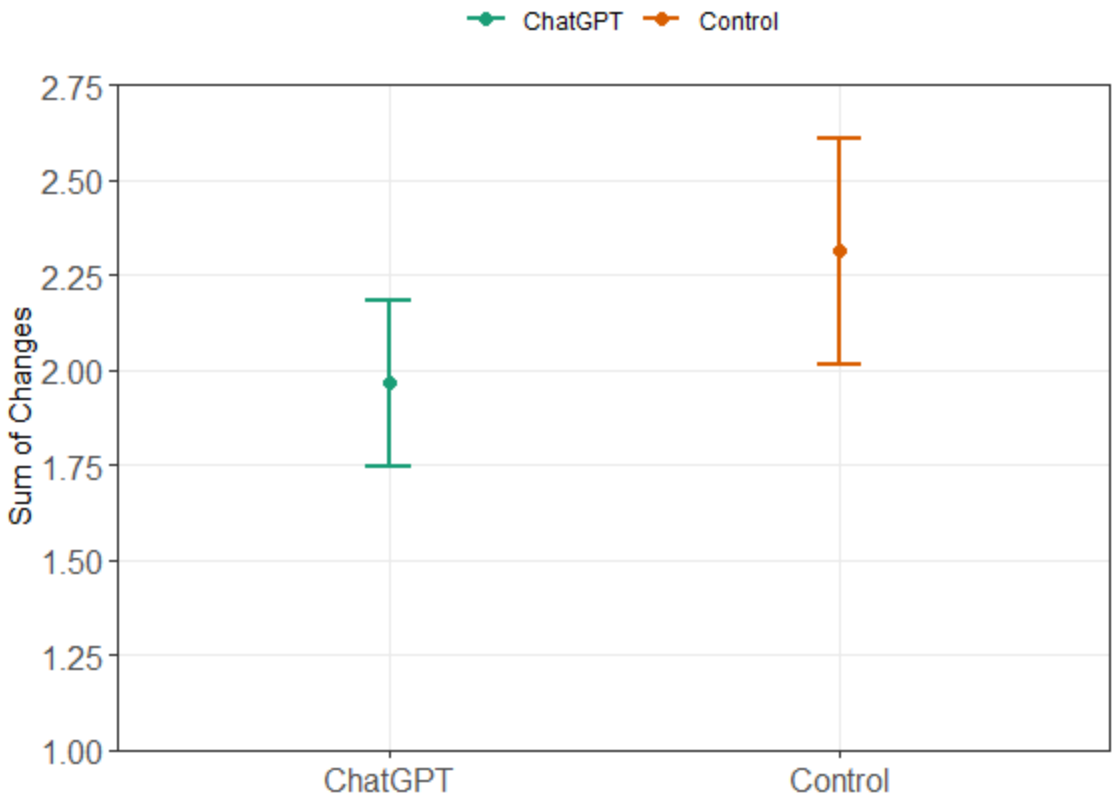


Table A.2. Regression Analysis for Sum of Changes in sales strategy

	Dependent variable:		
	Sum of Changes		
	(1)	(2)	(3)
ChatGPT	-0.349	-0.349	-0.340
	(0.253)	(0.253)	(0.252)
Female			-0.560
			(0.384)
Age			-0.250**
			(0.115)
Urban			-0.236
			(0.489)
Household Income			-0.096
			(0.237)
Period		-0.156***	-0.156***
		(0.055)	(0.055)
Constant	2.312***	3.015***	9.071***
	(0.199)	(0.314)	(2.522)
Observations	546	546	546
Note:			*** p<0.01, ** p<0.05, * p<0.1

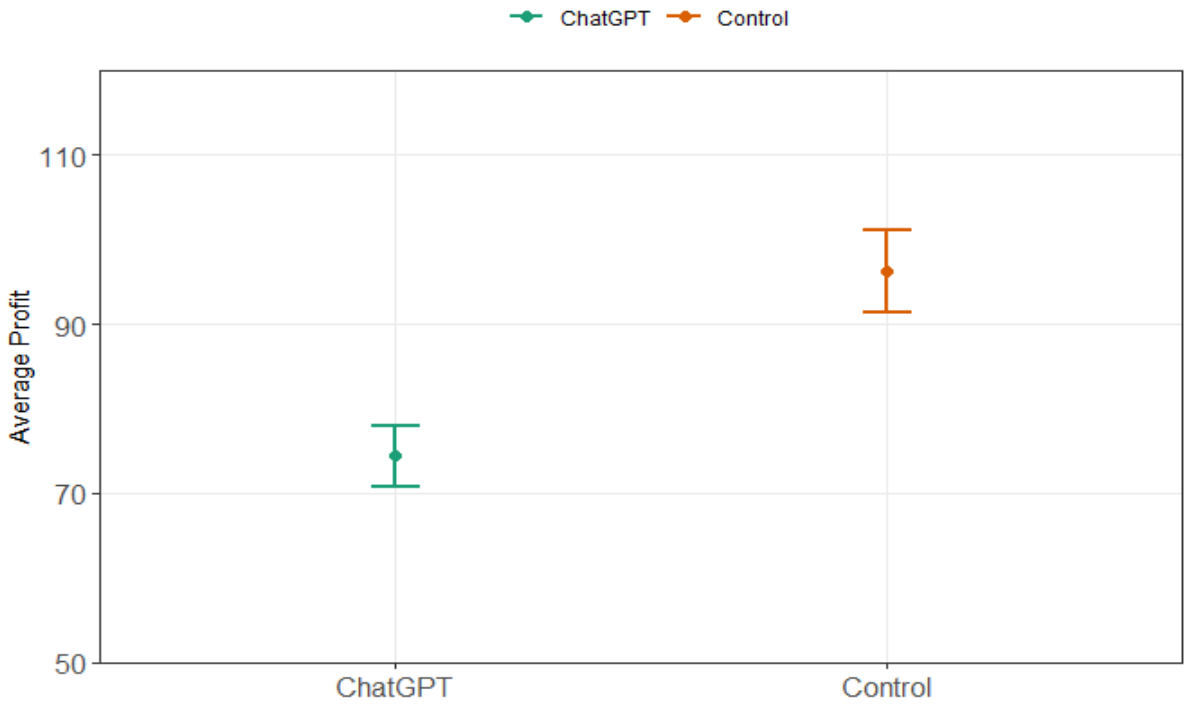
To better understand what drives this result, we classified subjects into risk averse versus risk lovers based on their behaviour in the risk game.[1] A subject is considered risk-averse if s/he opened less than 1/3 of boxes and non-risk averse otherwise (Risk-averse = 1 if the number of opened boxes < 33, Risk-averse = 0 otherwise).

[1] Two-proportions z-test, p-value =0.18

Table A.2: Percentage of Risk Averse across control and ChatGPT

	Control	ChatGPT
Risk Averse	18 (42%)	29 (58%)
Non-Risk-averse	25 (58%)	21 (42%)
Total	43 (100%)	50 (100%)

Figure A.2: For Risk Averse: Average Profit across control and ChatGPT groups[2]



[2] Mann Whitney U test, p-value < 0.05

Table A.3: For Risk Averse Subjects: Regression Analysis for Profits in Lemonade Stand Game

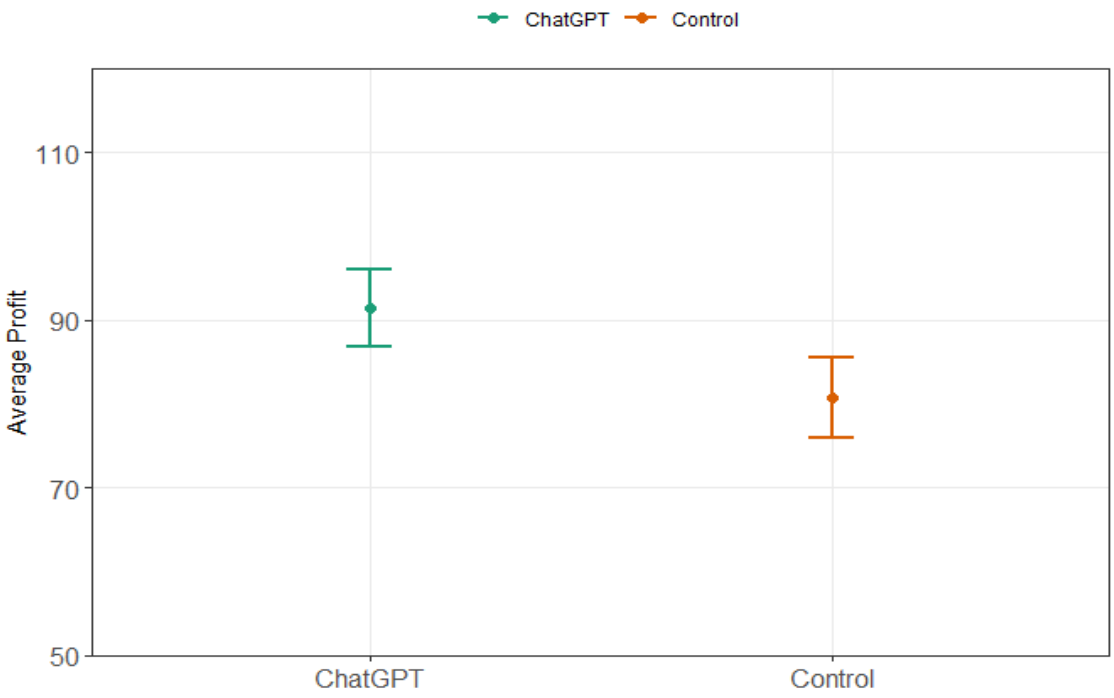
	Dependent variable:		
	Profit		
	(1)	(2)	(3)
ChatGPT	-21.896***	-21.896***	-19.481***
	(5.553)	(5.553)	(5.479)
Female			9.959
			(7.762)
Age			7.569
			(4.806)
Urban			4.419
			(5.624)
Household Income			-0.220
			(4.177)
Period		3.957***	3.957***
		(0.750)	(0.750)
Constant	96.291***	80.464***	-91.559
	(4.651)	(4.238)	(96.192)
Observations	637	637	637

Notes:

1) **** p<0.01, ** p<0.05, * p<0.1*
2) *First column: presents results for fixed effects.*
3) *Second and third column: Results are clustered by subject ID*

The above finding shows that the negative impact of ChatGPT, in terms of innovation, is manifested for risk averse subjects rather than risk loving subjects. In other words, subjects who are risk averse are the ones the most negatively affected by ChatGPT; being less innovative in the sense that they did not explore many strategies and thus were not close to finding the optimal strategy resulting in significantly lower profits. The opposite happens for risk lovers (check figure A.3 and table A.4).

Figure A.3: For Risk Lovers: Average Profit across control and ChatGPT groups[3]



[3] Mann Whitney U test, p-value < 0.05

Table A.4: For Risk Loving Subjects: Regression Analysis for Profits in Lemonade Stand Game

	Dependent variable:		
	Profit		
	(1)	(2)	(3)
ChatGPT	10.661	10.661	13.087*
	(6.905)	(6.905)	(6.796)
Female			-2.144
			(12.155)
Age			-5.786**
			(2.798)
Urban			29.121***
			(9.438)
Household Income			-8.386
			(8.652)
Period		3.663***	3.663***
		(0.870)	(0.870)
Constant	80.783***	66.133***	172.266***
	(4.726)	(4.963)	(59.036)
Observations	637	637	637

Notes:

1) **** p<0.01, ** p<0.05, * p<0.1*
2) *First column: presents results for fixed effects.*
3) *Second and third column: Results are clustered by subject ID*