BACHELOR THESIS

Tabea Rahel Bonus

The effect of stress on strategic decision-making using the Beauty Contest Game

Faculty of Behavioural Sciences
Department of Cognitive Psychology and Ergonomics

First Supervisor:	Dr. Matthijs Noordzij
Second Supervisor:	Erika van Lier, MSc
Study programme:	Psychology

Enschede, 24.06.2016

UNIVERSITY OF TWENTE.

Abstract

Background: The Beauty Contest Game (BCG) is a mathematic strategic decision-making game, which is commonly used to experimentally test strategic decision-making. The aim of this paper was to disentangle the different components underlying the decision-making process during BCG and to investigate the impact of stress, as there is striking evidence that stress impairs strategic decision-making (e.g. Leder, Häusser & Mojzisch, 2013; Leder, Häusser & Mojzisch, 2015). **Methods:** Stress has been induced using the Sing-a-Song Stress Test 18 minutes before the BCG (Brouwer & Hogervorst, 2014). The amount of skin conductance responses was used as physiological measurement of stress and a questionnaire was used in order to unravel the mathematical understanding and strategizing of the participants. **Results:** Stress has not shown to influence the outcome of the BCG significantly. Neither mathematical understanding nor the ability to strategize was influenced significantly by stress. Surprisingly, mathematical understanding tended to enhance under stress. However, this trend was not significant. **Conclusion:** The outcome of the BCG might not be unambiguously interpretable. Therefore, it is suggested to assess strategizing independently of the BCG's outcome. The trend of enhanced mathematical understanding under stress is in line with the findings of Pabst and colleagues (2013). According to their paper, stress could enhance decision-making when induced 18 minutes pre-task. Although their study did not include strategic decision-making, the present study might suggest that their findings could also hold true within the paradigm of the BCG.

Introduction

Often, decisions that have immense consequences must be carried out under stress. Many studies found that stress impairs decision-making (for recent reviews, see, e.g., Starcke & Brand, 2012; Starcke, Wolf, Markowitsch, & Brand, 2008; Pabst, Brand & Wolf, 2013). Moreover, stress has been shown to have a significant impact on several processes related to decision-making (Leder, Häusser, & Mojzisch, 2015; Cohen, 1980). For example, Cohen (1980) showed that stress exposure is related to a reduction of sensitivity towards others. Assessing the choices others might make (also called strategizing) is particularly salient for economic decisions, for example when trading stocks on the stock market (Leder, Häusser, & Mojzisch, 2015). Considering its importance, there are surprisingly few studies that investigate the effect of stress on strategic decision-making (for an exception, see Leder, Häusser, & Mojzisch, 2013; Leder et al., 2015). How exactly stress contributes to strategic decision-making is still not yet fully understood. The aim of this paper is to experimentally investigate the effects of stress on strategic decision-making.

A commonly used paradigm to shed light into strategic decision-making processes is the Beauty Contest Game (BCG). The BCG is based on Keynes idea that professional investment behavior essentially attempts to be one step ahead of average behavior (Duffy & Nagel, 1997). This idea was introduced in his work "The General Theory of Unemployment, Interest and Money". Keynes (1936) reasoned "professional investment may be likened to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole" (Keynes, 1936). Thus, the competitors must not pick the faces they find most attractive themselves. Rather, they need to think about the others competitors' choices and choose accordingly. Behavioral economists picked Keynes (1936) idea up in order to test it experimentally (Leder et al., 2013). They developed the so-called Beauty Contest Game. During the BCG the players have to pick a number within an interval of 0 and 100, instead of a face. The average of all chosen numbers is calculated and multiplied with p (with $0 < p < 1$). In most experimental studies, p is equal to 2/3. In any case, the value p is explained to all players before they make their own estimate. Depending on the number a player has chosen, conclusions about the players' level of reasoning can be made, with higher numbers indicating lower levels of reasoning, and lower numbers are associated with higher levels of reasoning (Ho, Camerer, & Weigelt, 1998; Sutter, 2005; for more information about the levels of reasoning, see next paragraph).

In her experimental study about levels of reasoning during the BCG, Nagel (1995) defines a hierarchical model of reasoning about beliefs of others. In the context of the BCG, a *zero-order belief* is considered a randomly applied strategy resulting in an arbitrarily chosen number. In this case, the participant does not hold any beliefs about the other participants. When a *first-order belief* is formed, the participant considers the other participants to choose a number at random and consequently picks the number that is most likely to win (e.g. 33,3 with a p of 2/3). A participant that assumes the other players to form a *first-order belief* and chooses his number accordingly, has formed a *second-order belief* (Nagel, 1995). As mentioned in the previous paragraph, higher numbers are associated to lower depths of reasoning and thus lower degrees of strategizing (Nagel, 1995).

However, Leder and colleagues (2015) proposed two other possible explanations for high numbers. Firstly, a given participant might not have understood the underlying mathematical mechanisms. Secondly, higher numbers could result from a player's belief that the competitors choose for high numbers (Leder et al., 2015). Consequently, high numbers are not unambiguous indicators for *zero-order beliefs*. Leder et al. (2015) were the first disentangling the effect of stress on the three proposed explanations (failure of strategizing, failure of understanding underlying mathematics and altered beliefs about others). In their study, they conducted the BCG and compared participants exposed to stress (they were attending an exam) and participants who were not exposed to stress. Half of the participants received a brief elicitation in which they were requested to actively think about which numbers other participants might have had chosen and to write it down. The number reflecting the expected average of other participants was expected to provide insight into reasoning processes. The results implied that stressed participants were neither able to understand the underlying mathematics, nor were they able to strategize.

However, the previous described study did not test mathematical understanding of the game directly. Rather, they interpreted the chosen number in relation to the expected number as an indicator for: a) understanding underlying mathematics, b) estimation of others' reasoning skills and c) ability to strategize. It is important to note that it is difficult to discern whether a high estimated average was due to *zero-order beliefs* (thus an inability to strategize) or due to the *first-order belief* that the other players will pick high numbers. For example, imagine two persons: person A and person B, both exposed to stress. Person A is convinced that his competitors will choose for high numbers because they are stressed. Thus, person A will estimate 2/3 of the average of chosen numbers of e.g. 90, and therefore chooses 60. Contrary, Person B fails to strategize under stress and chooses randomly for 90 during the brief elicitation and takes 2/3 of it. Although, both participants do not differ regarding their data, they clearly differ in depths of reasoning. Considering the lack of differentiation between strategizing and estimated beliefs, further research is needed to disentangle those two.

It is now a well-established phenomenon that stress impairs decision-making. However, how time could contribute to this phenomenon remains poorly understood. In their study, Pabst and colleagues (2013) indicated that temporal development of this phenomenon could be a major factor when it comes to decision-making processes. However, no studies considered the temporal development as a factor within the BCG. Previous studies focusing on the underlying mechanisms of the decision-making process induced stress immediately before the decision-task (Leder et al., 2013), or chose for participants currently under real life stressors (Leder et al., 2015). In order to contribute to the understanding of time development of the phenomenon within the BCG, the stressors are neither induced while-, nor immediately before the BCG, but about 18 minutes and 10 minutes prior to the BCG. The individual stress responses to the stressor are measured in this study, since people do not respond equally strong when exposed to the same stressor (Kudielka, Hellhammer, & Wüst. 2009). Even though, the effect of stress on strategic decision-making has been studied, yet few studies actually measured stress in this context. Therefore, measuring stress levels could provide further insight into how stress influences strategic decision-making.

This study focuses on monitoring electro dermal activity (EDA). EDA is part of the sympathetic nervous system (SNS), one of the human stress response systems which is activated under stress (Skoluda et al., 2015). Measuring EDA as an indicator for SNS activation is suitable, since it is the only autonomic psychophysiological index that is not contaminated by parasympathetic activity (Braithwaite, Watson, Robert, & Mickey, 2013; Brouwer & Hogervorst, 2014; Westenberg et al., 2009). Furthermore, studies using comparable methodologies have proven to elicit strong EDA responses (for more information, see Brouwer & Hogervorst, 2014).

As mentioned above, there is evidence that participants under stress choose for higher numbers. The first hypothesis is stated: The more the participants' skin conductance responses increase, the higher the number the participant chooses for during the BCG (1). In order to disentangle the three possible causes of high chosen numbers a questionnaire is used. After stress exposure and completion of the BCG, the questionnaire will test the participants' mathematical understanding of the BCG, their applied level of reasoning, as well as the participants' beliefs about others' choices. The following hypothesis will be tested: The more the participants' skin conductance responses increase, the less likely the participant is to understand the underlying mathematical mechanisms of the BCG (2). Based on the research described above, the third hypothesis is stated: The more the participants' skin conductance responses increase, the less likely the participant is to form *first-order*, *second-order* or even *higher-order beliefs* about other participants' beliefs (3).

Method

Participants

Fifty-five participants (17 female, 38 male, mean age: 24, SD= 4,4 years) took part in the experiment. The participants were students at the University of Twente or students at the Academy of Pop Music and MediaMusic. Students of the University of Twente participated in exchange for course credits in this study. All participants had chance to win a 25 Euro voucher upon condition that their number was closest to the target number. Informed consent was obtained from all participants. The Ethics Committee of the University of Twente approved that the study is in accordance with the standards listed in the faculties' Protocol about Ethics and Research.

Apparatus

The experiment was programmed with Python 2.7 and ran by Psychopy 1.8. The instructions were presented on a Windows 7 Laptop. EDA was measured with the skin conductance sensor- SA9309M by thought technology. The physiological data was recorded with ProComp Infiniti System w/ BioGraph Infiniti Software- T7500M by thought technology and processed with MATLAB.

Task

Stress was induced using the Sing-a-Song Stress Test (SSST). During the SSST, the participants sit in front of a screen that presents two consecutive neutral messages in time intervals of forty seconds. Then, the participants are informed that they have to sing a song later on, and that they should now start to think about a song they could sing. After the anticipation period of 30 seconds, the final message requests the participant to start singing and not to stop for 30 seconds. Used materials are based on the paper of Lars Nijboer (2015).

After completion of the SSST the participants are exposed to a noise stressor. The participants are instructed to relax and focus on their breathing. Beep sounds with 1000 Hz frequency, and 200ms duration each, are implemented in random time intervals. In total, the noise stress test last for 5 minutes. During the noise test, 26 beep sounds of 1000Hz occurred randomly for five minutes.

Finally, a one-shot Beauty Contest Game was conducted, since it is a promising way to test decision-making in an experimental setting. During the BCG the players pick a number within an interval of 0 and 100 [0, 100]. The average of all numbers chosen is then multiplied by 2/3. The player who's number is nearest to the target number wins (Grosskopf & Nagel, 2008; Ho, Camerer, & Weigelt, 1998). The rules described above appeared onscreen. After a countdown of 30 seconds the participants had to say the chosen number aloud.

Questionnaire

A questionnaire was used to assess the participants' level of reasoning, mathematical understanding and estimated beliefs. The questionnaire starts with an open-ended question, asking the participant to describe his decision-making process as accurate as possible. The second part of the questionnaire consists of a total of 8 questions. In question 1 the participants are asked to give their chosen number, question 3 asked for the estimated average. The questions 2-7 are contingency questions, most of them being dichotomous. Each of these questions describes one possible step of the decision-making process during BCG and requests the participant to state whether the step was part of their own decision-making process or not. The combination of both parts of the questionnaire gives insight into the individual decision-making process in order to judge whether a participant understood the mathematical mechanisms, whether the participant formed a *zero-order, first-order, second-order* or *higher- order belief* (for more information, see Appendix 2).

Procedure

At the start of individual sessions, the participants were asked to sit in front of the laptop. They were told that they are attached to appliances during the experiment and to move as little as possible, because movement could contaminate retrieved data. Also, they were informed that all instructions needed appear onscreen and that the experiment lasts approximately 45 minutes. Moreover, they were told that participation is voluntary, that they can stop the experiment whenever they want to and that the data is being processed anonymously. Then, the participants were asked to read and sign an informed consent.

Next, the researcher attached the appliances to the participants' fingers and wrists. The E4 wristband is attached to the participants' left wrist, the ECG to the right and left wrist, and the skin conductance sensors on the right ring finger and index finger. Additionally, the participants had to wear headphones during the experiment. After all appliances were attached, the researcher started all programs needed and entered the participants' gender and participation number. The researcher took a seat next to the participant and the experiment started. Instructions of the SSST, noise stressor and BCG appeared onscreen, respectively. Baseline and perceived stress were measured before and after each task. Furthermore, the participants were instructed to fill out a perceived stress scale (7- point Likert scale) before and after each task.

The researcher de-attached the appliances and the participants were given the first question of the BCG questionnaire. After completion, the participants received the remaining questions. Finally, the researcher debriefed the participants and thanked for participation.

Data analysis

Two researchers judge the participants' mathematical understanding and their level of strategizing (*zero-order, first-order, second-order* or *higher-order beliefs*) during the BCG based on the BCG questionnaire. Cohen's Kappa is calculated on mathematical beliefs and on levels of strategizing to test inter-rater reliability. In case no judgment is possible based on the open question, judgments of the closed ended questions complete the information needed. The open question of the BCG questionnaire has priority when it comes to conflicts, as the closed ended part might be more vulnerable to social desirability biases.

In order to analyze EDA, the amount of skin conductance responses during a baseline (before the SSST) and during the SSST phase where participants were preparing to sing a song (stressful situation where there are no motion artifacts from the actual singing) were taken. The two variables were transformed into z-scores in order to enable further analysis with difference scores. A new variable, which subtracted the baseline z-scores from z-scores during the SSST, was computed (Zdifference). The variable Zdifference is an indicator of the relative increase of skin conductance responses and was used as stress variable for further analysis.

The first hypothesis has been tested with a linear regression analysis, with the dependent variable chosen numbers and independent variable Zdifference. The second hypothesis has been tested using a binary logistic regression analysis. The dependent variable is mathematical understanding with value 0 meaning math has not been understood, and value 1 meaning math has been understood (0= no; 1= yes). The independent variable is still the increase of skin conductance responses. The third hypothesis has been tested using an ordinal regression analysis, with increase of skin conductance responses as independent and order belief as dependent variable.

Results

1.1 SCR and chosen numbers

Figure 1 shows the correlation between the increase of SCR and the chosen numbers. The Pearson Correlation is low negative with r= - .13. The linear regression model explains 1.6 % of the variance in chosen numbers (R^2= .016). Also, the main effect of increased SCR was not significant, $F(1,53)<1$, p= .17 > .05. The chosen number is predicted to be about 2.53 lower per increase of 6 SCR. A participant without an increase of SCR is predicted to choose for 40. The mean of all chosen numbers is 37.69.

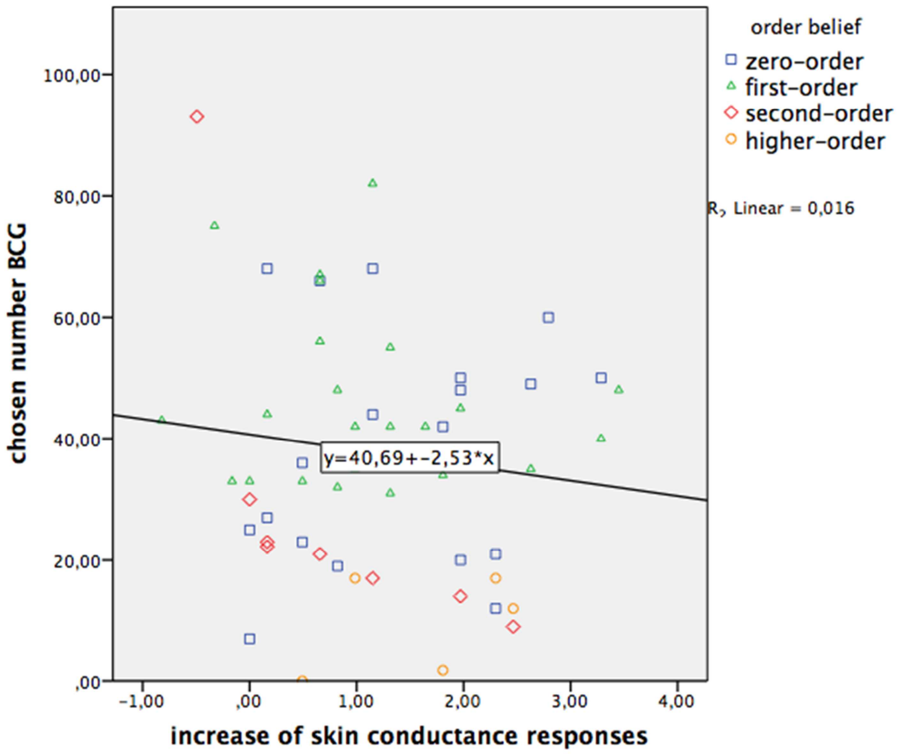

Figure 1 Correlation of the increase of skin conductance responses and the chosen numbers.

1.2 SCR and mathematical understanding

Cohen's Kappa was .74 for classifying mathematical understanding, which is a substantial agreement. The number of participants who understood the mathematical mechanisms (N=29) does not differ statistically significant from the number of participants who did not (N=26) with p= .69 > .05. A binary logistic regression analysis shows that an increase of SCR is not a significant predictor of mathematical understanding when using a significance level of .05 (p= .06> .05). Adding the amount of skin conductance responses to the model increased the ability to predict whether a subject understood the math significantly, with p= .047 < .05. Figure 2 shows that there is a trend towards an increase of SCR and understanding the math. However, this trend is only weak with ρ= .26.

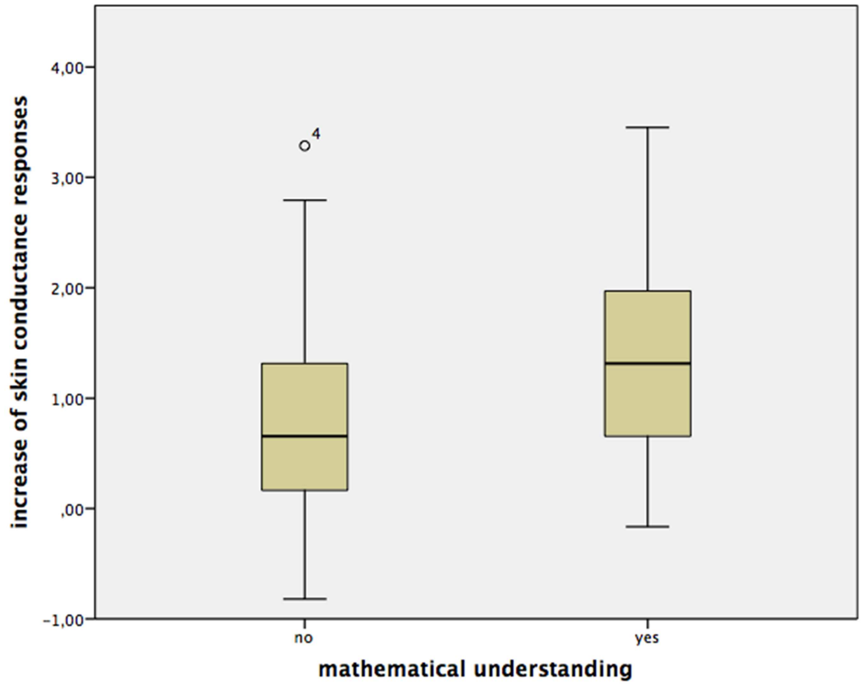

Figure 2 Correlation of the increase of skin conductance responses and mathematical understanding.

1.3 SCR and order beliefs

Table 1 provides the distribution of participants over the different order beliefs as well as the mean chosen number per order belief. Most of the participants formed zero-order or first-order beliefs. Few participants formed second-order or higher-order beliefs. Figure 3 illustrates the negative trend of increased SCR and order beliefs. Order-beliefs higher than second-order are an exception with the highest increase of SCR. But it is important to note that the sample size of higher-order beliefs is only 5. Because of its low sample size, the higher-order beliefs will be neglected for further analysis. The correlation between SCR and order beliefs in general is weakly negative with ρ= -.198. The Wald test with Zdifference as predictor is 2.038 with p= .15. The regression coefficient for increased SCR is not statistically different from zero in estimating the order-beliefs and the null hypothesis cannot be rejected. The increase of SCR is calculated to explain 4.8% of the variance in order beliefs. Cohen's Kappa was .92 for order beliefs, which is an almost perfect agreement.

		n	marginal percentage	mean chosen number BCG
order belief	zero-order	19	34.5%	38.68
	first-order	23	41.8%	46.13
	second-order	8	14.5%	28.65
	higher-order	5	9.1%	9.55
valid		55	100%	
missing		0		
total		55		37.69

Table 1 Distribution of participants over order-beliefs with mean BCG number per order belief.

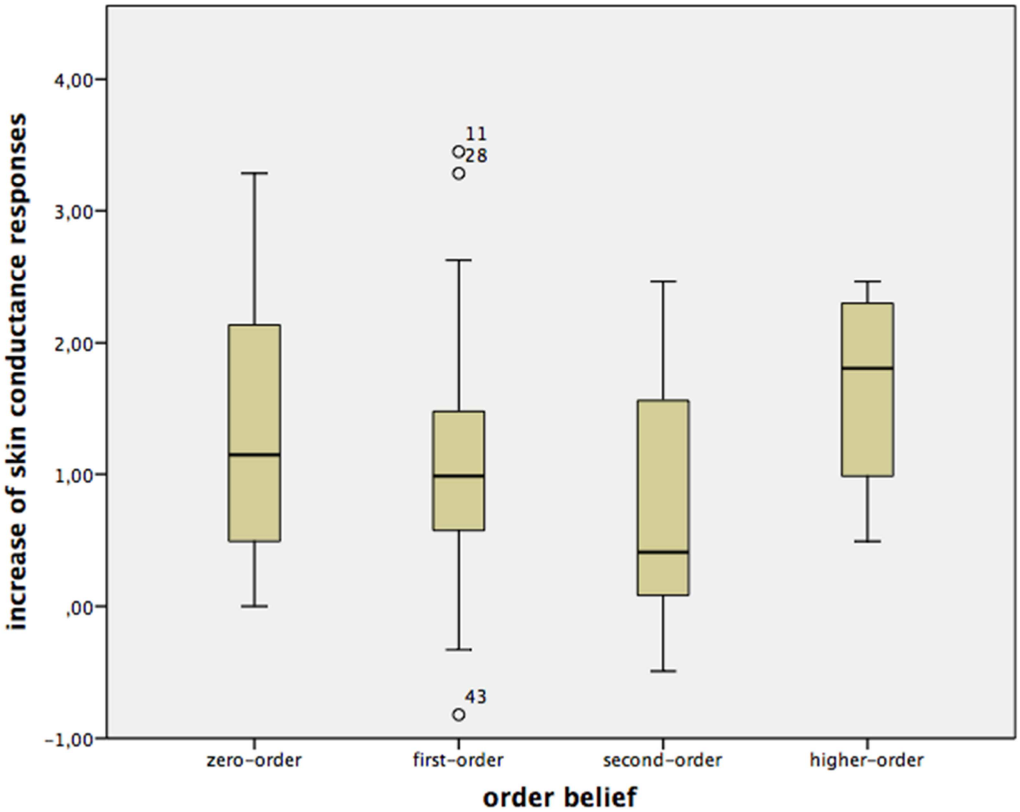

Figure 3 Increase of skin conductance responses per order belief.

Discussion

This paper sought to improve the overall understanding of the effect of stress on strategic decision-making. This was done by using the Beauty Contest Game and checking for a) mathematical understanding, b) beliefs about others' beliefs, and c) strategizing. Firstly, no significant effect of stress on the numbers that have been chosen during the BCG has been found. One interesting finding is that the correlation of increased SCR and chosen numbers was not only weak but also negative. This finding was unexpected and suggests that there is a trend towards choosing for lower numbers when stressed. Moreover, participants with no increase of SCR were predicted to choose for an even higher number than the overall mean. This outcome is contrary to previous studies that associated stress with choosing for higher numbers (Nagel, 1995). This finding will be discussed later in more detail. Secondly, it has been tested whether participants with increased SCR are less likely to understand the underlying mathematical mechanisms of the Beauty Contest Game. This experiment did not detect an impairment of mathematical understanding due to stress. Again, a weak trend contradictory to the hypothesis was found. Thirdly, it has been tested whether stress influences the order beliefs that are formed by the participants. A weak negative correlation has been found. However, participants who formed higher-order beliefs (those are order beliefs higher than second-order) showed the highest increase of SCR. Even though they were neglected in the analysis because of the small sample size, stress has not been found to be a significant predictor for order beliefs. Taken together, these findings do not support previous research investigating strategic decision-making by use of the Beauty Contest Game.

These rather surprising results could be explained by several factors. As Table 1 shows, the mean chosen numbers for zero-order beliefs were in fact lower than those chosen by participants performing first-order beliefs. An explanation was revealed when evaluating the questionnaire. Participants performing zero-order beliefs were found to base their reasoning on numbers they are familiar with. For example, two participants with zero-order beliefs chose for 42 and argued that this "is the answer to everything" (Adams, 1979). Some participants chose for their favorite number (e.g. 7), which was often fairly low. Most often, participants with zero-order beliefs chose for their age. Since the mean age of this study was 24, the tendency to choose for the own age could have decreased the mean of chosen numbers severely. These results might contribute to the fact that an increased amount SCR was not found to be a predictor for high chosen numbers. Importantly, this could suggest that chosen numbers during the BCG are no reliable indicators for different levels of strategizing on their own. Rather, the applied level of strategizing might need to be tested independently from the chosen number.

Although the results differ from some published studies investigating the Beauty Contest Game (Leder et al., 2013; Leder et al., 2013; Nagel, 1995), they are consistent with those focusing on the time dimension (Pabst et al., 2013). In their study, Pabst and colleagues (2013) found that stress impaired decision-making when the decision task was conducted 28 minutes after stress exposure. However, when the decision-task was conducted 5 or 18 minutes after stress exposure, stress did not impair decision-making. Even more interestingly, a trend of the 5 minutes and 18 minutes groups to perform better than the control group has been observed. In our study, the Beauty Contest Game has been carried out about 18 minutes after the Sing-a-Song Stress Test. Therefore the lack of significant findings could be due to the time between stress exposure and the Beauty Contest Game. Furthermore, the time between stressor and decision task could also explain the observed trends of being more likely to understand the mathematical mechanisms and choosing lower numbers with increased SCR. These results could suggest that the findings of Pabst and colleagues (2013) are not limited to decision-making in the context of risky behavior but could also hold true when applied within the paradigm of the Beauty Contest Game. However, the findings need to be interpreted with caution because no trend was significant and the correlations were weak.

Another possible explanation is based on the Yerkes-Dodson law (Yerkes & Dodson, 1908). According to the Yerkes-Dodson law, performance on difficult tasks increases as mental and/ or physiological arousal increases, but only up to the point where an optimal level of arousal is reached. Further increase of arousal results in decreased performance. This could suggest that either the optimal level of arousal has not been exceeded in this study, or the optimal level of arousal has been exceeded but not while conducting the Beauty Contest Game, both explanations would ultimately result in enhanced performance. This would imply that the stressor was either not stressful enough to impair strategic decision-making, or the participants stress level decreased over time to the point where it enhanced decision-making. Since the physiological stress level of an individual varies over time (e.g. through coping strategies), it seems that the Yerkes-Dodson law supports the idea of time being a factor that could make the difference between enhanced or impaired decision-making under stress.

Overall, the present findings suggest that time dimension is a major factor when investigating the impact of stress on decision-making in the context of the Beauty Contest Game. Preferably, the stressor should be induced 28 minutes before the Beauty Contest Game, at least if impairments due to stress are of interest. Another issue that emerges from these findings is that the number that has been chosen does not provide enough information to make a conclusion about the levels of reasoning. Rather, more information about individual decision-making processes should be derived in order to judge whether a participant formed zero-order, first-order, second-order or even higher-order beliefs. A standardized questionnaire could enhance further research by unraveling the underlying variables that contribute to the decision-making process within the paradigm of the Beauty Contest Game. Furthermore, these results raise intriguing questions regarding the nature and extent of the influences of stress on strategic decision-making. Previous studies focused on how stress impairs strategic decision-making. None did address the question whether stress could also enhance strategic decision-making under specific temporal circumstances, making it an important issue for future research.

The experimental work presented here provides the first investigation into how physiological stress responses influence mathematical understanding, order beliefs and chosen numbers during the Beauty Contest Game. The findings reported in this study contribute to our understanding of stress and strategic decision-making and provide a basis for further research.

References

Adams, D. (1979). *The Hitchhiker's Guide to the Galaxy*. New York: Harmony Books.

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology, 49*, 1017-1034.

Brouwer, A. M., & Hogervorst, M. A. (2014). A new paradigm to induce mental stress: the Sing-a-Song Stress Test (SSST). *Frontiers in Neuroscience, 8*(224), 141-148.

Cohen, S. (1980). Aftereffects of stress on human performance and social behavior: a review of research and theory. *Psychological Bulletin, 88(*1), 82–108.

Duffy, J., & Nagel, R. (1997). On the robustness of behaviour in experimental 'beauty contest' games. *The Economic Journal, 107*(445), 1684-1700.

Grosskopf, B., & Nagel, R. (2008). The two-person beauty contest. *Games and Economic Behavior, 62*(1), 93–99.

Ho, T. H., Camerer, C., & Weigelt, K. (1998). Iterated Dominance and Iterated Best Response in Experimental "p-Beauty Contests". *American Economic Review, 88*(4), 947–969.

Keynes, M.J. (1936). The General Theory of Unemployment, Interest and Money. London: Macmillan.

Kudielka, B. M., Hellhammer, D. H., & Wüst, S. (2009). Why do we respond so differently? Reviewing determinants of human salivary cortisol responses to challenge. *Psychoneuroendocrinology, 34*(1), 2–18.

Leder, J., Häusser, J. A., & Mojzisch, A. (2013). Stress and strategic decision-making in the beauty contest game. *Psychoneuroendocrinology, 38*(9), 1503–1511.

Leder, J., Häusser, J. A., & Mojzisch, A. (2015). Exploring the underpinnings of impaired strategic decision-making under stress. *Journal of Economic Psychology*, *49*, 133–140.

Nagel, R. (1995). Unraveling in Guessing Games: An experimental Study.pdf. *The American Economic Review*, *85*(5), 1313–1326.

Nijboer, L. (2015). Measuring stress sensitivity: Personality tests, heart rate and sweat for the singing of a song. 53.

Pabst, S., Brand, M., & Wolf, O. T. (2013). Stress and decision making: A few minutes make all the difference. *Behavioural Brain Research*, *250*, 39–45.

Skoluda, N., Strahler, J., Schlotz, W., Niederberger, L., Marques, S., Fischer, S., … Nater, U. M. (2015). Intra-individual psychological and physiological responses to acute laboratory stressors of different intensity. *Psychoneuroendocrinology*, *51*, 227–236.

Starcke, K., & Brand, M. (2012). Decision making under stress: A selective review. *Neuroscience and Biobehavioral Reviews*, *36*(4), 1228–1248.

Starcke, K., Wolf, O. T., Markowitsch, H. J., & Brand, M. (2008). Anticipatory Stress Influences Decision Making Under Explicit Risk Conditions. *Behavioral Neuroscience*, *122*(6), 1352–1360.

Sutter, M. (2005). Are four heads better than two? An experimental beauty-contest game with teams of different size. *Economics Letters*, *88*(1), 41-46.

Westenberg, M. P., Bokhorst, C. L., Miers, A. C., Sumter, S. R., Kallen, V. L., Pelt, J., & Blöte, A. W. (2009). A prepared speech in front of a pre - recorded audience : Subjective , physiological , and neuroendocrine responses to the Leiden Public Speaking Task. *Biological Psychology, 82*(2), 116-124.

Yerkes, R. M., & Dodson, J.D. (1908). The relation of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18,* 459-482.

Appendix 1 Beauty Contest Game – Instructions.

First page:

Each participant of this study will write down a number between zero (0) and one hundred (100). Zero and one hundred are also possible. We will calculate the average, which is the mean, of all numbers picked. Then we will multiply the mean with 2/3. The resulting number will be the target number. To win the game, you should pick a number that is as close as possible to this target number.

Second page:

The participant whose picked number is closest to the target number, 2/3 of the mean, will win the game and receives a 20€ voucher. Please say the number you've chosen out loud when the countdown has expired.

Appendix 2 Beauty Contest Game – Questionnaire.

Beauty Contest Game - Questionnaire

Identity Code:

Gender:

1. How did you come to your decision? Describe the process as accurate as possible.

Appendix 2 Beauty Contest Game – Questionnaire (continued).

Beauty Contest Game - Questionnaire

Identity Code:

Gender:

1. Which number did you pick?

☐

2. Did you think about the average of the others picks?

☐ Yes ☐ No (continue with question 8)

3. What do you think is the average of the others picks?

☐

4. How did you get to this number?

☐ I just guessed it

☐ I thought about the choices others might have made

☐ Others:

5. Did you take 2/3 of this number?

☐ Yes ☐ No (continue with question 8)

Appendix 2 Beauty Contest Game – Questionnaire (continued).

6. When estimating the average, did you take into account that the other competitors might have also estimated an average of all picks and took 2/3 of this average?

☐ Yes ☐ No (continue with question 8)

7. Did you take 2/3 of the number that you think would be the average after others already took 2/3?

☐ Yes ☐ No

8. If you would do this exercise again, would you change the number you picked?

☐ Yes ☐ No

If yes: Why?

☐ I did not rationally estimate or calculate my number

☐ I changed my mind about the others choices

☐ I got the calculation wrong

☐ Others:

Appendix 3 Beauty Contest Game – Coding.

Open question:

2 raters fill in this table: yes/ no/ no judgment possible

participant	mathematical understanding	altered beliefs	zero-order beliefs	first-order beliefs	second-order beliefs

Inter-rater reliability (Cohen's kappa) is calculated afterwards

If the answer does not allow a judgment of the given categories, no judgment is made. Instead, judgment is based on the closed ended questionnaire. If judgments of open vs. closed judgments do not match, the open question is treated true.

Closed ended Questionnaire:

Mathematical understanding:

If chosen number is 2/3 of estimated average math has been understood
If this is not the case but within a threshold of +/- 10: yes on question 5 also means that math has been understood
No on question 5 means math has not been understood properly

Altered beliefs:

If the chosen number and the estimated average are both high, and math has been understood, the high chosen number is due to altered beliefs

Strategizing:

Question 2
Yes: go on to question 4
No: zero-order belief

Question 4
B: first- or second order belief, go on to question 6
A: zero-order belief

Question 6
Yes: second-order
No: first-order

participant	mathematical understanding (1, 3, 5)*	altered beliefs (3)	zero-order beliefs (2)	first-order beliefs (2, 4)	second-order beliefs (2, 4, 6)

*Numbers in breaks indicate the questions that are relevant to code this item.