REPLICATION STUDY

Outcome sequences and illusion of control - Part I: An online replication of Langer & Roth (1975)

This is an Accepted Manuscript of an article published by Taylor & Francis in International Gambling Studies on 03.10.2022 and can be found on https://doi.org/10.1080/14459795.2022.2133906. This article may not exactly represent the final published version. It is not the copy of record.

ABSTRACT

The illusion of control is an important feature of both problematic and nonproblematic gambling behavior. Crucially, this construct is incorporated in most cognitive models of problem gambling, and is also central in numerous approaches to gambling disorder treatment (e.g., psychological interventions using cognitive restructuring to mitigate the illusion of control). In the present preregistered study, we tried to replicate the illusion-of-control effect, as defined and investigated in the seminal work by Langer & Roth (1975), in an online context. Using the same trial procedure and a similar cover story as the original study, we presented three groups of healthy participants (N = 289 in total; crowdsourced sample) with three different sequences of wins and losses in a coin-tossing task. Consistent with the original study, we found that participants presented with more wins at the beginning of a sequence estimated their ability to predict the outcome of a coin toss higher than participants presented with more losses at the beginning, or those presented with a random sequence, although the effect sizes were small to medium (biggest Hedge's g_av = 0.49) compared to the original study which yielded larger effect sizes (biggest $\eta p^2 = 0.14$). Thus, we could replicate the findings of Langer & Roth (1975) in an online context, although the effect size was smaller than expected.

KEYWORDS

control beliefs, illusion of control, gambling, replication

1. Introduction

In the gambling literature, the illusion of control is an important construct, and refers to the ’expectancy of personal success probability inappropriately higher than the objective probability would warrant” (Langer, 1975, p. 311). It can be influenced by various factors or task characteristics, such as giving participants control over parts of the task, inserting skill parameters, or different sequences of wins and losses (for more details on the ’illusion of control’ in gambling, see Langer, 1975; Stefan & David, 2013). Illusion of control has downstream influences on different aspects of gambling behavior, and has been associated with the development of problematic gambling (Clark & Wohl, 2021). Crucially, the construct of illusion of control, in one form or another, is incorporated in most cognitive models of problem gambling (Blaszczynski & Nower, 2002; Sharpe, 2002), and is also central in numerous approaches to gambling disorder treatment (e.g., psychological interventions using cognitive restructuring to mitigate the illusion of control).

One of the most cited studies in the illusion of control literature is the study by Langer and Roth (1975), in which participants had to guess the outcome of 30 coin tosses. Unknown to the participants, the outcome of the game was predetermined, with an overall probability of winning of exactly .50. Importantly though, the sequence of wins and losses was manipulated, resulting in three groups: for the ’descending’ group, most of the *wins* occurred at the beginning of the sequence (so most losses occurred towards the end); for the ’ascending’ group, most of the *losses* occurred at the beginning (so most wins occurred towards the end); and for the ’random’ group, wins and losses were equally distributed throughout the sequence. After 30 tosses, participants were asked five questions about their ability to guess the outcome of the coin toss. Langer and Roth (1975) found that the descending group (i.e., more wins at the beginning) estimated their ability to guess the outcome of the coin toss higher than the ascending group (i.e., more losses at the beginning) and the random group.

There have been only few (more or less successful) replication attempts of this seminal study though (e.g. Coventry & Norman, 1998; Frank & Smith, 1989; Ladouceur & Mayrand, 1984). Although a meta-analysis reported a big effect size (average weighted D = .89), it was only based on 4 studies with a total sample size of 312 participants (Stefan & David, 2013). Moreover, the confidence interval around the effect size in this meta-analysis was quite wide, which suggests there is still much uncertainty about the size of the effect (considering also that the original study only found a medium effect size across the questions). Therefore, we tried to replicate the original study in an online context. Online testing allows to test bigger and (often) more diverse samples than university students (for further information see https://www.prolific.co; Palan & Schitter, 2018), which supports the robustness and generalizability of findings. Note that replications can either be direct ('close') or conceptual. In direct replications researchers follow the method of the original study as closely as possible when it comes to participants, materials, and procedure. This is usually highly challenging to achieve (especially for older studies, like the one by Langer and Roth, as e.g, the population – from which participants are recruited – might have changed in meaningful ways). Conceptual replications, which examine the same hypothesis or research question with (slightly) different participants or methodologies than the original study, are thus more common. Deviations from the original study can therefore be related to e.g., a different cultural context or a different participant recruitment procedure (for a further discussion see Wohl, Tabri, & Zelenski, 2019). Therefore, in this study, we aimed to conceptually replicate the original study of Langer and Roth (1975) in an online context.

2. Method

All raw and processed data, code, materials and the preregistration can be found on https://osf.io/uk6t3 and https://osf.io/qm2a8. We also report (here and in our preregistration) how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

2.1. Sample size

To test if we could induce an illusion of control, we focused on the comparisons between the ascending and descending group for the five illusion of control questions (see below). A meta-analysis (Stefan & David, 2013) showed that sequences of outcomes have an average effect size of Cohen's D = .89 weighted across samples (therefore written in the capital letter D). Based on this effect size, we would need 32 participants per group (calculated with G*Power, Faul et al., 2007; two-tailed t-test with alpha = .01 and power = .8; the alpha level was adjusted for multiple comparisons, as we planned to compare five questions). Note that Langer and Roth (1975) did not find a difference between the groups on the distraction and practice question (for full details on these questions, see Procedure section). However, from a theoretical standpoint, a difference might be expected. Therefore, we assumed to find differences on all five questions with a highly powered sample.

We set a minimum sample size of 32 participants per group and a maximum sample size of 96 participants per group (which was triple the sample size indicated by the power analysis). To determine the exact sample size, after recruiting 32 participants per group, we used sequential Bayesian hypothesis testing (see Analysis section for further details) by adding 48 extra participants (16 per group) until either a decisive Bayes Factor (BF_{10} larger than 10 or smaller than 1/10) was obtained (Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2017), or we reached the maximum number of 288 participants (96 per group). As will be shown below, the eventual Bayes Factors for the five individual contrasts (ascending vs. descending) were not decisive, so we reached the maximum number of 288 participants (with one extra in the random condition).

2.2. *Participants*

In total 289 participants completed the entire online experiment and were included in the analyses (96 in the descending group, 96 in the ascending group, and 97 in the random group due to a mistake in the testing procedure. Across groups, there were 130 females, 157 males and 2 indicated non-binary; age M= 26.7 years, SD = 8.6 years; range = 18-67; see the Table A1 for age and gender information per group).

Participants were recruited via the crowd-sourcing platform Prolific where every eligible participant is notified as soon as a study is published online. The researcher can indicate the number of participants per study and participants can then sign up for it (as long as the requested number of participants has not been reached). Once signed up, participants have a certain amount of time to finish the study (56 minutes if the experiment is 15 minutes long) before they get timed-out (and free up a place again).

Only participants who spoke English and never participated in an experiment by the first author were allowed to participate. We did not have any exclusion criteria regarding gambling involvement as we wanted to keep the sample as general as possible (in comparison to the original study, which only tested undergraduate students). Participants gave their consent before starting the experiment and were randomly assigned to one of the three groups. The experiment took about 15 minutes complete, and participants were rewarded with £ 1.50 plus the £ 1 bonus they earned (£ 2.50 in total). The study was approved by the local research ethics committee of the faculty of Psychology and Educational Sciences at Ghent University.

In addition to the 289 participants who completed the experiment, an additional 7 participants finished the experiment and got paid but were excluded from data analyses in accordance with our preregistered exclusion criteria: three were replaced because they started the experiment again, possibly because they missed the completion code for Prolific or thought they could do better to win more money; two did not have data for the past and/or the future question (see section 2.4. below); one had more than 5 % missing trials; and one provided a completion code but no data were stored. One extra participant was rejected and therefore not paid (in accordance with the Prolific guidelines) as they never provided a completion code nor any data set. An additional 32 participants signed up for the experiment but never started or completed it. From our experience, this seems to be a common practice on Prolific to 'reserve' a spot in well-paid experiments.

In the OSF repository, we included a more detailed explanation of the Bayesian sampling and the replacement procedures. All data were acquired between 23rd February 2021 and 1st March 2021 (except for one of the replaced participants, who was tested in April 2021).

2.3. Apparatus and stimuli

The experiment was programmed in jsPsych (version 6.0.5) and only ran on desktop computers and laptops in Chrome (de Leeuw, 2015). Langer and Roth (1975) told their participants that the study was about how subtle social cues could influence performance on certain tasks. To mimic this in an online setting, we created an avatar video using ’crazytalk’ software (Reallusion Inc., n.d.). The video was of a woman asking ’Heads or tails?’. The coin toss always started with an image of a quarter dollar coin from 1998 showing heads. The coin then did a flip of 6x360 degrees (6 circles) to land on heads again or a flip of 6.5x360 degrees (6.5 circles) to land on tails. In addition to this visual input, we added a coin toss sound. The coin-tossing animation lasted 500 ms.

2.4. Procedure

We used the original procedure by Langer and Roth (1975) but adapted it to an online experimental setting. The general procedure included welcoming the participant, asking for consent, and requesting information about age, gender (with the options ’male’, ’female’, ’non-binary’ and ’I don’t want to say’), nationality, and whether the participants gamble on a regular basis. We then explained the experimental procedure. As a cover story, participants were told that we were investigating subtle social cues in avatars.

After participants had read the instructions, the coin-tossing task started. The task consisted of 30 trials. On each trial, we presented a video in which the avatar asked the participants ’Heads or tails?’. This video lasted about 1500 ms. Then participants could guess the outcome (without any time restrictions) by pressing the left (heads) or the right arrow key (tails). After indicating their guess, participants saw a coin for 500 ms. Then an animation of a coin toss was displayed for 500 ms. After that, the outcome in points was shown below the coin for 1000 milliseconds (”+25” for a win, and ”-25” for a loss). The points were converted into real money at the end of the experiment. For every 100 points, participants would get 1£ extra and they started with a balance of 100 points. Importantly, wins and losses were equiprobable, but their sequence was predetermined according to the group participants were randomly assigned to. Therefore, the final balance for every participant was 100 points (so they received 1£ extra). The sequences of wins and losses were the same as in Langer and Roth (1975):

- WWWWLWWWLLLWWWLLWLLWWLLLWLLLLW in the descending condition;
- WLLLLWLLLWWLLWLLWWWLLLWWWLWWWW in the ascending condition;
- WLWLLWLWLWWLLWLLWLWWWLWLWLLWLW in the random condition.

If a trial was predetermined to be a win, participant would ’guess’ correctly regardless of whether they guessed ’heads’ or ’tails’. The exact trial procedure is depicted in Figure 1.

After 30 trials of coin tossing, participants answered the same five questions about the task as in Langer and Roth (1975) by using visual slider (Questions 1-3; note the original study used 11-point Likert scales) or text input (Question 4-5). These questions were used to determine the illusion of control:

- (a) ”How good do you think you are at predicting outcomes like these?” (item ’Prediction’; 100-point slider, 0 = *Very bad*, 100 = *Very good*)
- (b) ”How well do you think you would do on the task if you were distracted?” (item ’Distraction’; 100-point slider, 0 = *Very bad*, 100 = *Very well*)
- (c) ”How much do you think you would improve with practice?” (item ’Practice’; 100-point slider, 0 = *Not at all*, 100 = *Very much*)
- (d) ”How many correct predictions did you make on these 30 trials?” (item ’Past’; text input)
- (e) ”How many correct predictions would you make in the next 100 trials?” (item ’future’; text input)

After this, they filled out the English short version of UPPS-P impulsive behavior questionnaire (Cyders, Littlefield, Coffey, & Karyadi, 2014). This questionnaire was included as part of a bigger individual differences study and the data were not analyzed here. Participants also filled in the predictive control scale of the Gambling Related Cognition Scale (GRCS, Raylu & Oei, 2004) to explore whether the GRCS data correlate with the self-reported feeling of control (as assessed by our five questions). This exploratory comparison would allow us to see whether people who generally feel more in control of events also felt more in control in our coin-tossing task. The results can be found on OSF (https://osf.io/vkxpq). After the experiment we debriefed the participants by explaining the original purpose of the experiment.

[Figure 1 about here.]

2.5. Analyses

We conducted all analyses in line with the original study and our preregistered data analytic plan. All data processing and most analyses were completed with R (version 4.0.2, R Core Team, 2018) using the packages reshape (version 0.8.8; Wickham, 2007a), reshape2 (version 1.4.4; Wickham, 2007b), ez (version 4.4-0; Lawrence, 2016), Hmisc (version 4.4-2; Harrell, 2020), Rmisc (version 1.5; Hope, 2013), doBy (version 4.6.8; Højsgaard and Halekoh, 2020), and tidyverse (version 1.3.0; Wickham et al., 2019). The Bayesian analyses were conducted with JASP (version 0.13.1.0, JASP team, 2022).

Our main interest was the comparison between the descending group and the ascending group. Therefore, we directly compared the five illusion of control questions between these two groups with two-tailed independent t-tests (Bayesian and frequentist). For completeness, we also report the remaining comparisons between the three groups and the one-way ANOVAs (frequentist and Bayesian), as in the original study, for each of the five illusion of control questions, with group (descending vs. ascending vs. random) as a between-subjects factor.

We report both p-values and Bayes factors for statistical inferences. Significance was determined at an alpha level .01, to correct for multiple tests (as we tested five comparisons between the ascending and the descending group). For the pairwise comparisons, Hedge's g_{av} is the reported effect size measure (Lakens, 2013). For Bayesian analyses, we report the Bayes Factor BF_{10}, which quantifies the evidence for the alternative hypothesis against the null hypothesis. We used the default priors in JASP.

For Bayesian t-tests the default prior width is 0.707 and for the Bayesian ANOVA the default prior width is .5. We tested the robustness of the Bayes Factors by varying the priors in JASP, which can be found on OSF (https://osf.io/qm2a8/ in the folder Part I/code/analyses/ JASP).

3. Results

[Figure 2 about here.]

The descriptive statistics are displayed in Figure 2. The outcomes of the one-way ANOVAs and all the t-tests appear in Tables 1 and 2 respectively. Here we focus only on the theoretically most interesting comparisons, namely the difference between the ascending and descending groups. These comparisons revealed some numerical trends: compared with the ascending group, participants in the descending group gave higher ratings for (a) their ability to guess the outcome of a coin toss with distraction, (b) their performance in the past 30 trials, and (c) their future performance in this task. For these three questions, the Bayesian analyses provided moderate ($BF_{10} > 3$) support for the alternative hypothesis (see Table 1 in Wagenmakers et al., 2018, for further information on the cut-off values). However, none of the five comparisons between the ascending and the descending group reached our pre-determined threshold for the Bayes Factor ($BF_{10} > 10$ or $BF_{10} < 1/10$) and only one comparison (the past performance question) was statistically significant (after correction for multiple comparisons).

[Table 1 about here.]

[Table 2 about here.]

As none of the illusion of control questions alone seems to exactly capture the illusion of control, we decided (post-hoc) to also calculate a sum score of all five items and compare this sum score between the groups. As shown in Table 2, the comparison between the ascending and the descending group with this sum score yielded moderate evidence for the alternative hypothesis (and the frequentist t-test was significant, even after correction for the extra test), tentatively indicating that the descending group estimated their task proficiency differently than the ascending group and the random group.

4. Discussion

In this online replication of Langer and Roth (1975), we were able to create an illusion of control as defined in their seminal work. Specifically, for the same three out of five questions which were significant in the original study, we found moderate evidence (BF > 3) for a difference between the ascending and descending groups in the expected direction. Furthermore, the (post-hoc) sum score of all illusion-of-control items revealed a medium-sized difference (g_{av} between .40 and .50) between the descending group and the two other groups (BFs > 5). Thus, we could replicate the original study, although the observed effect sizes (biggest $g_{av} = 0.49$, which reflects a small to medium effect) were smaller than the ones reported in a meta-analysis (averaged D = 0.89. which reflects a large effect size; Stefan & David, 2013) and in the original study (biggest $\eta p^2 = 0.14$, which also reflects a large effect size; Langer & Roth, 1975). Nevertheless, it is important to emphasize that our study generally replicates the findings from a seminal and frequently cited illusion of control study. Our study also shows that the illusion of control effect, as defined and operationalized by Langer and Roth (1975), is relatively robust with regard to small changes in the procedure, showing that outcome sequences can indeed create different estimations of the perceived ability to guess the outcome of a coin-toss. As mentioned in the introduction, the illusion of control construct is viewed as an important factor in gambling behavior and in the development of problematic gambling behavior from a cognitive perspective (Clark & Wohl, 2021). Therefore, it is important that these seminal studies on illusion of control are tested at a large scale, to establish their robustness and replicability (Heirene, 2020; LaPlante, 2019).

Although the results of our study were promising, some cautionary notes are warranted. First, we predicted and preregistered differences between the descending and ascending groups for all five questions. We did not observe this pattern, and none of the comparisons reached the crucial Bayes factor (i.e., BF > 10). We reasoned after collecting and analyzing the data that the sum score of all five questions might better capture the overall illusion of control. However, this analysis was not planned and also did not reveal a BF higher than our crucial BF of 10. These inconclusive results in the

Bayesian analyses can be connected to the fact that we assumed a medium effect size in our prior whereas the real effect seems to be smaller, which in turn might be because we were testing online. Compared to the original study, we had some procedural changes. In the original study, the experimenter flipped a coin in front of the participants, and participants had to call out the results while the coin was still in the air. Participants might have believed that they could predict the outcome based on some peripheral cues. In our online implementation, participants may be more likely to hold the belief that the outcome is randomly determined by the program, hence our manipulation overall has a smaller influence.

Additionally, it cannot be excluded that our results (and those of Langer & Roth in the original study) might reflect a primacy effect. However, if serial order would matter, we should also have observed a recency effect (e.g., Healey & Kahana, 2014), which was not the case (i.e., the ascending and random groups were similar). This might be related to the novelty of the stimuli at the beginning (e.g., Farrell & Lewandowsky, 2002). Nevertheless, it remains an empirical question what contributes to the effect observed by Langer & Roth (1975) and if the illusion of control as defined and operationalized by the authors actually reflects an overestimation of success probability.

In conclusion, we were able to conceptually replicate the finding that the outcome sequence can influence the perceived task performance on guessing the outcome of a coin-toss which was defined as an ’illusion of control’ by Langer and Roth (1975).

5. References

Blaszczynski, A., & Nower, L. (2002). A pathways model of problem and pathological gambling. *Addiction*, *97*(5), 487–499. Retrieved 2021-11-15, from https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1360-0443.2002.00015.x (eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1360-0443.2002.00015.x)

Clark, L., & Wohl, M. J. A. (2021). Langer's illusion of control and the cognitive model of disordered gambling. *Addiction*, *n/a*(n/a). Retrieved 2021-11-01, from https://onlinelibrary.wiley.com/doi/abs/10.1111/add.15649 (eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/add.15649)

Coventry, K. R., & Norman, A. C. (1998). Arousal, erroneous verbalizations and the illusion of control during a computer-generated gambling task. *British Journal of Psychology*, *89*(4), 629–645. Retrieved 2021-11-03, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2044-8295.1998.tb02707.x (eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2044-8295.1998.tb02707.x)

Cyders, M. A., Littlefield, A. K., Coffey, S., & Karyadi, K. A. (2014, September). Examination of a short English version of the UPPS-P Impulsive Behavior Scale. *Addictive Behaviors*, *39*(9), 1372–1376.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12. Retrieved 2019-07-10, from http://link.springer.com/10.3758/s13428-014-0458-y

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin and Review, 9,* 59–79.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39, 175*-191.

Frank, M. L., & Smith, C. (1989). Illusion of control and gambling in children. *Journal of Gambling Behavior*, *5*(2), 127–136. Retrieved 2021-11-03, from http://link.springer.com/10.1007/BF01019759

Harrell, F. E., Jr. (2020). Hmisc: Harrell miscellaneous [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=Hmisc (R package version 4.4-2)

Healey, M. K., & Kahana, M. J. (2014). Is memory search governed by universal principles or idiosyncratic strategies? *Journal of Experimental Psychology: General, 143,* 575–596.

Heirene, R. M. (2021). A call for replications of addiction research: which studies should we replicate and what constitutes a 'successful'replication?. *Addiction Research & Theory, 29*(2), 89-97.

Hope, R. M. (2013). Rmisc: Ryan miscellaneous [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=Rmisc (R package version 1.5)

Højsgaard, S., & Halekoh, U. (2020). doby: Groupwise statistics, lsmeans, linear contrasts, utilities [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=doBy (R package version 4.6.8)

JASP Team (2022). JASP (Version 0.16.2)[Computer software].

Ladouceur, R., & Mayrand, M. (1984). Evaluation of the "Illusion of control": Type of feedback, outcome sequence, and number of trials among regular and occasional gamblers. *The Journal of Psychology*, *117*(1), 37–46. Retrieved from https://doi.org/10.1080/00223980.1984.9923656 (Publisher: Routledge tex.eprint: https://doi.org/10.1080/00223980.1984.9923656)

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. Retrieved 2019-07-10, from https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00863/full

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, *32*(2), 311–328.

LaPlante, D. A. (2019). Replication is fundamental, but is it common? A call for scientific self-reflection and contemporary research practices in gambling-related research. *International Gambling Studies, 19*(3), 362-368.

Langer, E. J., & Roth, J. (1975, December). Heads I win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task. *Journal of Personality and Social Psychology*, *32*(6), 951–955.

Lawrence, M. A. (2016). ez: Easy analysis and visualization of factorial experiments [Computer software manual]. Retrieved from http://github.com/mike-lawrence/ez (R package version 4.4-0)

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

R Core Team. (2018). *R: A language and environment for statistical computing*. Retrieved from https://www.R-project.org/ (tex.organization: R Foundation for Statistical Computing)

Raylu, N., & Oei, T. P. S. (2004). The Gambling Related Cognitions Scale (GRCS): development, confirmatory factor validation and psychometric properties. *Addiction*, *99*(6), 757–769. Retrieved 2021-06-28, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1360-0443.2004.00753.x
(eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1360-0443.2004.00753.x)

Reallusion Inc. (n.d.). *Crazytalk*. Reallusion Inc. Retrieved from https://www.reallusion.com/crazytalk/

Sch¨onbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322–339.

Sharpe, L. (2002, February). A reformulated cognitive–behavioral model of problem gambling: A biopsychosocial perspective. *Clinical Psychology Review*, *22*(1), 1–25. Retrieved 2021-11-15, from
https://www.sciencedirect.com/science/article/pii/S0272735800000878

Stefan, S., & David, D. (2013). Recent developments in the experimental investigation of the illusion of control. A meta-analytic review. *Journal of Applied Social Psychology*, *43*(2), 377–386.

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... Morey, R. D. (2018, February). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76.

Wohl, M. J., Tabri, N., & Zelenski, J. M. (2019). The need for open science practices and well-conducted replications in the field of gambling studies. *International Gambling Studies, 19*(3), 369-376.

Wickham, H. (2007a). Reshaping data with the reshape package. *Journal of Statistical Software*, *21*(12). Retrieved from http://www.jstatsoft.org/v21/i12/paper

Wickham, H. (2007b). Reshaping data with the reshape package. *Journal of Statistical Software*, *21*(12), 1–20. Retrieved from http://www.jstatsoft.org/v21/i12/

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., Fran¸cois, R., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686.

Appendix A. Participant characteristics per group

[Table 3 about here.]

Table 1. Inferential statistics of group comparison on the score of each of the five questions.

Question	DFn	DFd	MSE	*F*	*p*	ges	BF_{10}
Prediction	2	286	403.81	2.58	.077	0.02	0.38
Distraction	2	286	493.36	6.11	**.002**	0.04	9.08
Practice	2	286	923.06	0.02	.980	0.00	0.04
Past	2	286	13.07	4.50	.011	0.03	2.16
Future	2	286	213.60	4.03	.018	0.03	1.42

Note: DFn = degrees of freedom in the numerator; DFd = degrees of freedom in the denominator; MSE = mean squared error; ges = Generalized Eta-Squared measure of effect size.

Table 2. Pairwise comparison between the three groups on every item of the questionnaire and on the sum score.

	diff	Lower CI	Upper CI	df	t	p	BF_{10}	g_{av}
Prediction								
ascending vs. descending	-6.01	-11.80	-.22	188.85	-2.05	.042	1.10	.29
ascending vs. random	-.68	-6.47	5.11	189.69	-.23	.817	.16	.03
descending vs. random	5.33	-.23	10.89	191	1.89	.060	.82	.27
Distraction ascending vs. descending	-8.08	-14.27	-1.90	189.55	-2.58	.011	3.37	.37
ascending vs. random	2.65	-3.64	8.95	189.87	.83	.407	.22	.12
descending vs. random	10.74	4.29	17.18	190.85	3.29	**.001**	22.08	.47
Practice								
ascending vs. descending	.26	-8.40	8.92	189.90	.06	.953	.16	.01
ascending vs. random	.84	-7.83	9.51	190.87	.19	.849	.16	.03
descending vs. random	.58	-7.99	9.15	191	.13	.894	.16	.02
Past								
ascending vs. descending	-1.48	-2.54	-.41	181.03	-2.74	**.007**	5.01	.39
ascending vs. random	-.30	-1.24	.65	190.99	-.62	.534	.19	.09
descending vs. random	1.18	.11	2.25	182.48	2.18	.031	1.42	.31
Future								
ascending vs. descending	-5.67	-10.02	-1.31	190	-2.57	.011	3.29	.37
ascending vs. random	-1.16	-5.22	2.90	186.22	-.56	.574	.18	.08
descending vs. random	4.51	.46	8.56	186.44	2.19	.029	1.47	.32
Sum score								
ascending vs descending	-20.98	-35.72	-6.24	177.61	-2.81	**.006**	5.96	.40
ascending vs random	1.36	-13.64	16.35	181.76	0.18	0.86	.159	.03
random vs descending	-22.34	-35.28	-9.40	190.68	-3.40	**.001**	31.50	.49

Note: diff = difference; CI = confidence interval (95%); BF_{10} = Bayes Factor 10; g_{av} = Hedge's average g

Table A1. Participant characteristics per group

group	N	female	male	non-binary	age	sd	regular gambler	no regular gambler
ascending	96	48	46	2	27.75	9.34	21	75
descending	96	40	56	0	26.00	8.34	14	82
random	97	42	55	0	26.38	7.91	30	67

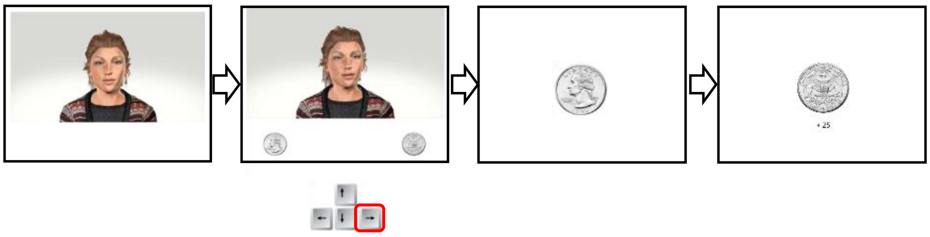

Figure 1. Trial procedure.

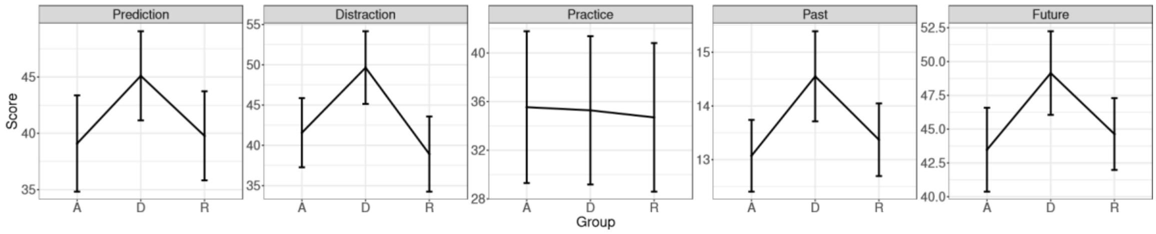

Figure 2. The estimated ability to guess the outcome of a coin-toss displayed by item and group. A = Ascending; D = Descending; R = Random. The error bars reflect the 95% confidence intervals.